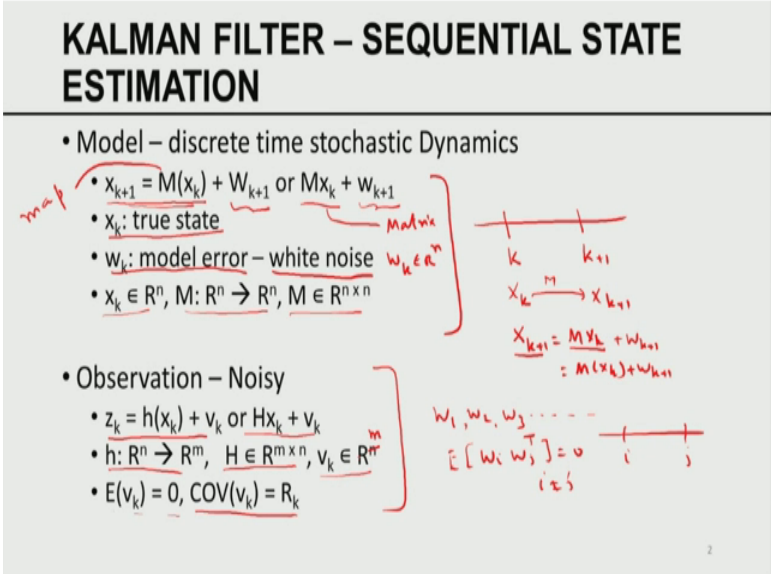
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture – 32
Linear Stochastic Dynamics- Kalman Filter

In this module 8 we are going to be talking about prediction in the context of stochastic models and observations character by noise everything is stochastic. Again here we are having a stochastic model which could be linear non-linear, we have a noisy observation which could be a linear function of the state or non-linear function of the state, again 1 can consider 4 different possibilities model being linear non-linear, the observation being a linear function of the state or non-linear function of the state.

I am going to start with the simplest possible case where the dynamic model is discrete time linear model, the observations are linear functions of time, the noise the model is not perfect. The imperfections in the model are compensated by some random input, we will talk about the properties of the randomized the input to if the there is meant to compensate for the deficiencies in the model. The whole aspect of data assimilation assimilating noisy observation into stochastic dynamic models is what we are after the data assimilation algorithm ultimately when we derive it is called the Kalman filter equation.

(Refer Slide Time: 01:50)



So, the name filter in Kalman filter has a very special connotation, we will talk about what is filtering what is a prediction and so on a bit later, but we will start with some of the basic descriptions of the model of the observation. The Kalman filter also refers to sequential state estimation you may recall from our discussions with static deterministic estimation case, we can have an offline or an online estimation techniques.

Online estimation techniques are called are also called sequential state estimation; in the sequential state estimation things keep moving forward in the 4 D var on the other hand the adjoin takes you back 4 D var methods are in general offline te or all offline techniques. So, this is an alternate to 2 4 D var where there is no going back everything keeps going forward the estimation the inverse problem everything is solved sequentially much like we had a recursive, linear least squares in the context of static deterministic problems.

So, the model is a discrete time stochastic or random model, this is the general description of the non-linear discrete time model we have already seen, M is the model map, M is from Rn to Rn, x case the state. This is the example of a non-linear map this is ab ob or this is an example of a linear system the addition of noise is new here, Wk plus 1 Wk plus 1. So, let us talk about the timing diagram and this is little bit a little bit of a notation, this is time k this is time k plus 1 at time k a I have known Xk I would like to know the state at time k plus 1, the model map maps Xk plus 1 to M. So, M is called 1 step transition map if it is a non-linear function, it is called 1 step state transition matrix. If it is a matrix, this is the matrix in the linear case, this is a map in the non-linear case we use the same symbol m, the juncture will tell whether it is the matrix or a map. The Wk plus 1 is the noise, Wk plus 1 is the noise that occurs after time k before time k plus 1; that means, I know Xk I would like to be able to compute Xk plus 1 if there is no noise Xk plus 1 would have been M of M times Xk or M of Xk depending on the models we linear non-linear.

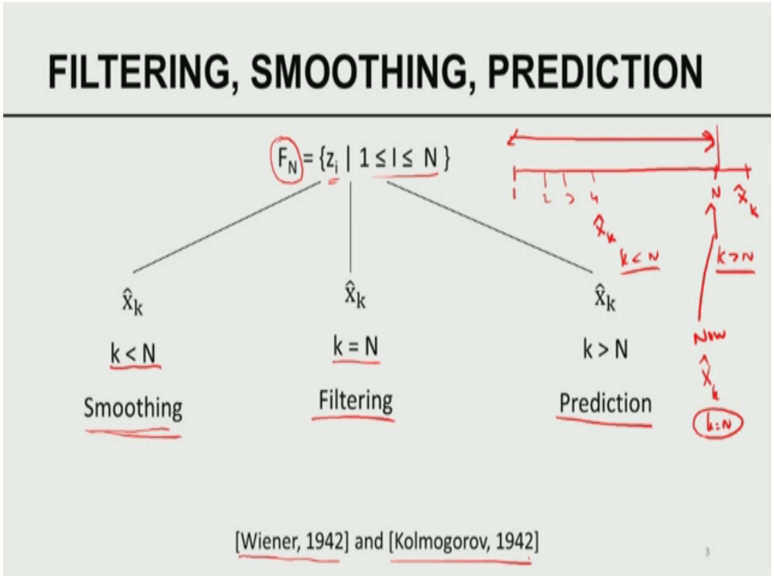
So, Xk plus 1 is a sum of what a deterministic model would have given you plus a noise that comes after time k. So, to emphasize the importance of the noise coming after time k I am going to denote it as Wk plus 1. So, Wk plus 1 is the noise that affects the evolution of the system, given Xk the noise occurs after time k before time k plus 1. So, Xk plus 1 the value of the state at time k plus 1 is the sum of the deterministic part plus the random part, that is the interpretation for making it Wk plus 1, in some textbooks you will see Wk plus 1 is called W k, really does not matter we thought in order to make it very clear that it is the noise affecting the system after the state k is known it provides a less room for confusion.

So, Wk what is that, it is represents a compensation for the model error. Come it is also a random process. One of the simplest is the random processes one can think of is white noise, what is white noise? W 1, w 2, w 3 this is the sequence of noise that affects the system. We say w is the white noise if there is no temporal correlation what is it mean? Expected value of w i w j transpose is equal to 0. In other words if i and j are 2 distinct moments in time, if I am considered the noise wi and wj they are temporally uncorrelated for all i not equal to j. So, that is what is called white. So, is an un is a 6 is a sequence of uncorrelated noise that affects the evolution of the system you may.

So, Xk is the true state I do not know that. So, I am trying to oh add a noise to make up for the deficiency and M could be a matrix or M could be a map now you got all the things associated with the model. With respect to the observation again nothing changes essentially, the same this is the non-linear function of the observation the linear function of the observation I am trying to do both of them simultaneously because we have gained a lot of expertise in trying to handle linear non-linear observation models and other things.

H is a map, h is a matrix covariance of Vk 0 covariance of Vk s Rk, I am sorry covariance of Vk is Rk mean of Vk 0, and Vk is again I should have said this is Vk is Rm sorry Rm. So, Wk is the model noise w k belongs to Rn, Vk is the observation or that belongs to Rm, hope the description of the model and the observations are clear.

(Refer Slide Time: 08:15)



Now, I am going to talk about the technical definition of the word filtering, smoothing and prediction. This definition is due to weiner, this definition is also due to Kolmogorov these kinds of definitions have been introduced in the literature since early 1940s.

Wiener 1942, Kolmogorov 1942 as I mentioned when I was doing optimal interpolation wiener and Kolmogorov independently were thinking of the same problem. Wiener was working in frequency domain that because he was an expert in Fourier analysis, Kolmogorov another hand was working in time domain. So, except for this difference in the domain of interest for analysis, they essentially uncovered the same set of results. So, let us give you a technical of what filtering is in Watson the word filtering is used in Kalman filters. In general it colloquially filter means something that stops from certain things going are going out for example, if I have a radio, the radio has as a tuner, the tuner essentially filters out all the signals that are that does not belong to a particular spectrum.

So, we can call a low pass filter, a high pass filter, bandpass filter, we will talk about coffee filters. So, we know what in Watson filter filtering in an ordinary sense is used; now technically filtering has a slightly different connotation. So, let us talk about that now; suppose I have observations from in the interval 1 to N. The observations are zi let us assume the observations are coming in discrete instances in time 2 3 4 k all the way up to N. Let zi I running from 1 to N be the collection of observation, let us call the collection of observation as f of capital F of N, N is a subscript N denotes the number of observations that we have, N also denotes the last instant we have the observation.

So, given a set of observations Z 1 to Z n, if I want to be able to make a prediction about the system let me do one thing I do not want to put a k here, that is right fine. If I want to be able to talk about the state of any estimate or the state of the system at time k, k greater than N, please understand this is the time interval over which I have the observation, I want to make an estimate of a state of system at a time k beyond N, that problem is called the prediction problem as you are rightly know as you rightly know.

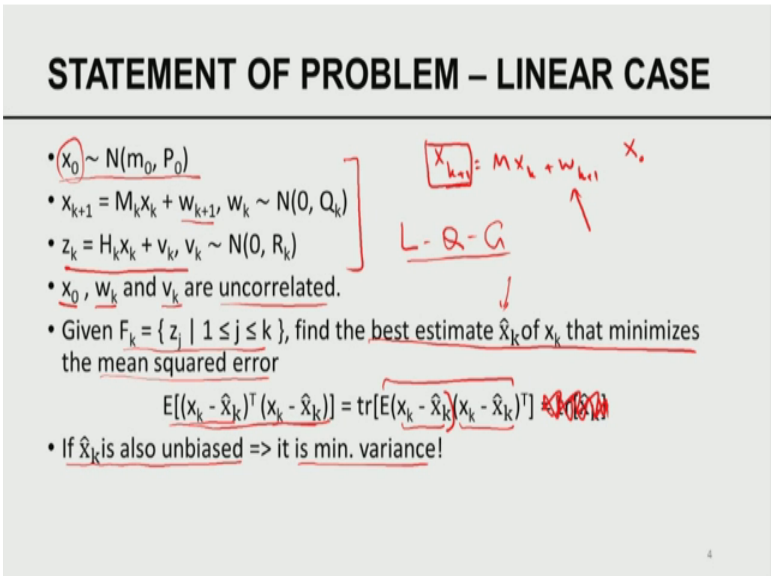
So, x bar k is the estimate of a state of a system at a time in future, how do I tell the time in future k is greater than N. So, you can think of N as today now. If k greater than N means it is the future. So, given all the information up to now trying to estimate the state of a system at a time in future, that is the prediction problem, that is the definition of prediction or forecasting. Knowing what I know today what will be the price of an IBM stock tomorrow that is the prediction. Knowing what I know today what will be the temperature distribution in early spring in North America that is a prediction problem. So on the other hand suppose I want I have known all the information from 1 to N I want to go back, I want to be able to evaluate the state of a system Xk, for some k less than N; that means, I have the benefit of information from 1 to N, and still trying to go back to estimating a quantity at a time k, k less than N.

So, k less than N means past, k greater than N means future. Estimating a quantity the past when k less than N that cause smoothing, because I have the benefit of the entire observation 1 to N, I am interested in trying to find an estimate at a time k in between 1 and N. I can I am allowed to exploit all the observations and that problem is called smoothing problem. So, prediction problem smoothing problem, if I want to make an estimate at the time k is equal to N now. So, what is the idea here I have been given a bunch of observation from 1 to N I would like to be able to get a state of the system at time k k is equal to N that is called the filtering problem.

So, filtering problem is an estimation problem, where I use all the information up to the time N and at that time I would like to be able to get the best estimate that is the filtering problem. Smoothing problem is given a set of observation from 1 to N, I would like to be able to estimate a state at the time in the past, prediction problem is trying to estimate a state given the set of observations at a time in the future. So, what is given to you, what you want to estimate, what is the relation of the time index at which you want the estimate to be? Depending on the relation of the amount of observation available and the relative relation of the time index k with respect to N, N is the last time at which the last observation is available; we have 3 problems smoothing, filtering, prediction.

So, filtering smoothing prediction are 3 classes of problems, there is a classic definition widely accepted. This is due to the pioneering work of wiener and these classifications are known since early 1940s.

(Refer Slide Time: 14:43)



So, what is the problem of Kalman filtering, that is what we are going to talk about we are going to assume everything is linear everything linear means what. So, let us talk about this now. I also want to take one more moment to talk about the problems that the problems relate to the models. So, we talked about the model being stochastic, let us consider the linear model Xk plus 1 is equal to M times Xk plus Wk plus 1 Wk plus 1 is a white noise therefore, Mk Xk plus 1 is also random, this is a model every model needs to have an initial condition x naught. I am going to assume even the initial condition is random the initial condition is picked from a realization from a prior you can think of there is a prior distribution, with an nor which is normally distributed M naught as the mean P naught as the covariance.

So, initial condition is random. So, if there is no noise, if there is no if the initial condition random the solution is random. If there if the initial condition is deterministic and if there is a noise affecting the model evolution, then the model solution is random. In here I am considering thus 2 sources of randomness that affects the state of the system, one is the randomness in the initial condition another is the randomness in the noise. So, the noise that affects the system, that noise that forces the system the initial condition that is random, the observation are noisy.

So, if I did not have noise analysis of a dynamical system with the random initial condition and random forcing, that is called analysis or stochastic dynamical system. So, I need to be able to first understand analyzing the properties of stochastic dynamical systems, how I can characterize the evolution of the state or what are the probabilistic properties, how they characterize the probability properties of the state, that is the first task.

The second task is suppose I give you on the top of it observations, how do I bring in the observation in addition to the stochastic and model analysis, to be able to combine the model and the observation to get the an analysis. So, the model solution is now going to play the role of background, observations are going to still play the role that it has played all along. So, the model forecast playing the role of a background provides the prior, the observation is going to provide you the new information we are going to combine them.

So, you can think of Kalman filter again within the Bayesian framework. So, there are 3 sources of randomness; initial condition is random, model forcing is random, the observation noise is random we are going to assume that all the 3 noises are uncorrelated. So, what is the basic idea? Given a set of observation fk from time 1 to k, fine us to estimate of x hat head of Xk look at this now. I have given k observations from 1 to k, I am I want the best estimate x hat k you can see from the previous slide this is the filtering problem. So, x hat k is the filtered estimate, what is the characterization of filtered estimate, that minimizes the mean square error again a least squares, the magic of least squares comes again and again and again is inseparable.

So, what is the idea here? Xk is unknown; Xk hat is the estimate that is the error in the estimate. So, Xk minus Xk hat transpose times Xk minus k hat, that is the covariance of the error I am sorry I should not say the covariance of the error. This is inner product of the 2 anomalies I am sorry this is the sum of the variances of all components of the forecast of the filtered estimate and that is essentially given by the trace of this matrix a bracket is missing. So, this quantity is inner product there is a scalar this scalar is equivalent to trace of the covariance matrix please realize this is the covariance that is the trace. So, that is equal to trace of I should say this I do not think this is correct I am sorry.

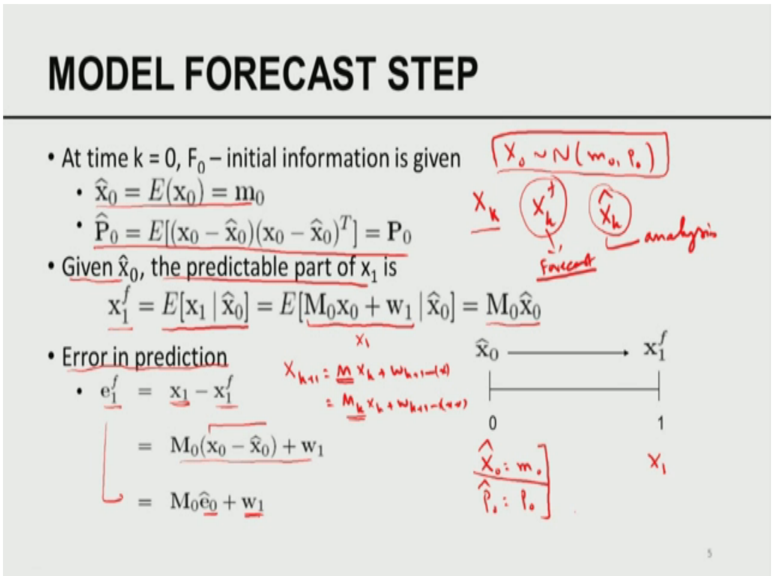
So, the that trace is enough. So, now, we have stated the problem, I want an estimate Xk hat of Xk such that it minimizes this mean square error, that is the statement of the problem; because I have given all the information up to k, if I am because I am interested in estimating the state Xk this is also called filtered estimate or the estimator that estimates Xk hat is called the filter equation that is where the notion of filter comes in.

If I also can show that this filtered estimate is unbiased, we have already seen minimum square error is equal to minimizing the variance if the bias is 0. We have we have seen this relation when we talked about the Bayesian setup therefore, it is very prudent to analyze and arrange things such that the estimate not only minimizes the mean square error, it is also unbiased. These 2 combined together will give you the minimum variance estimate. Now please understand, we are not now talking about linear minimum variance estimate, we the we are not bringing linearity right now we are simply say I want to have a minimum variance estimate, linearity refers to the structure of the estimator.

So, that is the I want the best that is exactly the whole idea here. I also want to bring out 1 more this problem is called linear quadratic Gaussian LQG, is that is the lethal combination. Linearity the model of the observation, quadratic mixture of the objective function to be minimized and the Gaussian nature of the noise involved. Kalman first showed this LQG combination is the lethal combination lethal in what sense we can get absolutely beautiful results is the 1 of the very few cases we have absolutely beautiful results.

So, one can ask yourself a question well seldom in life is linear why are you backing on linearity the problem is well non-linear problems are hard to solve anyway I cannot solve them I can only approximate them. So, we mathematically is interesting to ask you question which problems are solvable in closed form and what are the properties of the solution at least I want to enjoy the moment. So, the moment of enjoyment occurs when you deal with LQG problem. So, in the literature on control theory actually Kalman was the control theorist, Kalman introduced this within the context of control theoritic arguments therefore, within the con within the context of control theory, LQG theory is the very famous very popular very fundamental theory and this is an instantiation of the beauties of LQG.

(Refer Slide Time: 22:40)



So, before I talk about. So, where are we? We described our model, the model has 2 sources of randomness initial condition and the first thing observation has another source of randomness which is observation noise, given a bunch of observation, given a bunch of given the evolution of the dynamical state, I would like to be able to estimate. So, given the observation also given the model information, I would like to combine everything whatever you can do, you do give me the best estimate in the sense of minimizing the mean square error which in addition if I add the concept of unbiasness, also gives you the minimum variance estimate that is the problem we set up to solve, and that problem when everything is linear is called LQG.

So, before we I would like to separate the thing in 2 phases; first is model analysis model forecast analysis or model forecast step. Let us take the baby step 0 to 1, once I understand what goes from 0 to 1 then I can go from k to k plus 1. So, let us consider the transition from 0 to 1. Please recall my initial conditions are random I have assumed the initial conditions comes from a normal distribution with the mean M naught and the covariance P naught. Now I would like to be able to separate several quantities of interest to us, Xk is the pure is the true state, Xk f is the forecast is the forecast estimate of the true state Xk hat is the analysis. So, this is the analysis this is the forecast. So, these are the 2 quantities we will go back and forth Xk is the state of the system. So, Xk is the state of the system this is forecast there is analysis.

So, initially at time 0, I do not have any observation, the only information I have about the initial state is that it is normally distributed therefore, what is my initial analysis? My initial analysis is the mean of the initial distribution. So, what is the initial analysis covariance? Pk hat that is equal to P naught. Please understand analysis supposed to represent the best information I have. So, initially the analysis contains only information to right from the initial condition. If you give me a Gaussian random variable as an initial condition, what is the best estimate of the random variable is the mean.

What is the best estimate of the covariance is the covariance underline distribution. So, I am now going to postulate the initial analysis x naught hat is the M naught, initial analysis covariance P hat naught is P naught that is exactly this statement as well as this statement. Once I have initialized the analysis and its covariance, I want to be able to generate the forecast.

So, I would like to be able to now use them. So, again there is no observation now only model. Knowing what I know at time 0 if a feed this information to the model, if the model gives me an output how do I generate the forecast from the model output? Knowing that the model output is random, that is the question this is what is called stochastic dynamical system analysis. So, given x naught the predict I want to be able to compute the prediction of the state x 1.

So, x 1 is the state of the system, x 1 is going to be a random. So, I would when I am trying to talk about predicting a random phenomenon, I can only I need to think about decomposing the random process into 1 of 2 things there is a deterministic or a predictable part. There is an unpredictable part we can only control the predictable part un control unpredictable part we have no control. For example, the noise, but very nature is not predictable. Therefore, what is that I am going to define now again I want to bring out one more fact.

From the when we did the mean square error estimate, what is the theorem we have proven. Within the Bayesian within the mean square error analysis, the best estimate is the conditional mean we have already shown that. Therefore, I am going to want to create x 1 f please understand my notation, what is x 1 f ? X 1 f is the part of the forecast of the state at time 1 that I have control over that is a predictable part, and that is equal to conditional expectation of x 1 given x naught hat. What is x naught hat? X naught hat is the information about the initial state, who is going to create x 1 model is going to pull the x naught hat into x 1. X 1 is the true state of the model; the true state of the model according to the model equation is the M naught x naught plus w 1. Now I am going to talk about one more little thing, you can write the model equation like this Xk plus 1 is equal to M of Xk plus Wk plus 1 I can also write the equation to be M k of Xk plus Wk plus 1.

In star. So, this is double star these are all important things that is why I am trying to spend a little time. Both are linear, but in here M does not vary in time here M varies in time. If the algebra the mathematics of it is not much different between M varying in time M un invariant in time. So, I am assuming and I am sticking it k to m. So, what is double star mean? I have a linear time varying model. If I can allows the linear time varying model, that time invariant model you simply take the k out of m.

So, without loss of generality I can assume the model is yes its time varying therefore, M 0. So, if I am using my model to be time varying model x 1 is equal to M naught x naught plus w 1 that is the model equation, please go back to my model equation earlier. A that is

Xk plus 1 is equal to M of Xk plus Wk plus 1, for simplicity to get started I assumed M is a constant now I am sticking M sub k. M sub k essentially refers to the fact that model could also be varying in time, the algebra is no different if I can get a free ride why not I would like to get the maximum benefit out of it.

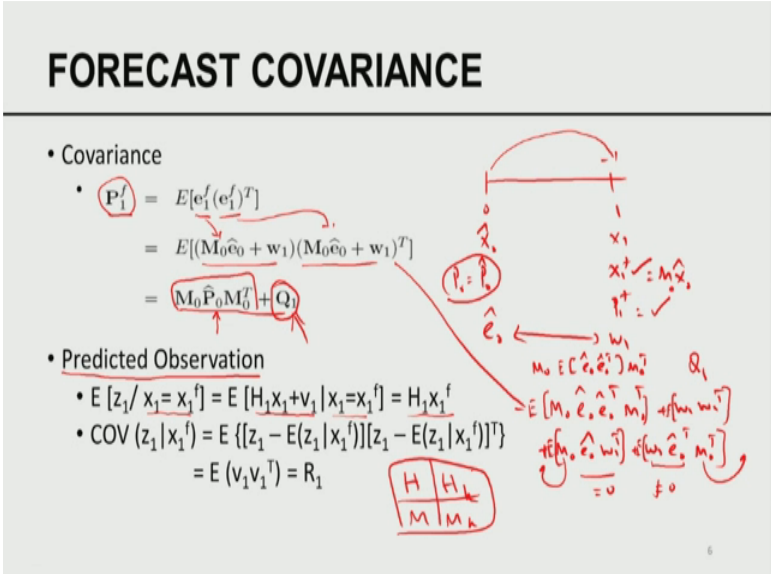
So, if I use the model, this is what the model will tell you as your x 1 is. I am sorry this is what the model will tell you r as your x 1 is, but x 1 is random what is the estimate of x 1 conditional x; what is the best estimate of their true state conditional expectation. So, conditional expectation of x 1 given x naught hat that is equal to M naught of x naught hat because look at this now conditional expectation of. So, conditional expectation of a linear operator, conditional expectation of a sum of the conditional expectations, the second conditional expectation is expectation of w 1 with respect to x naught hat, that is 0, because w 1 is white it does not depend on anybody else.

So, M naught I will I already know at time 0 x naught hat I already know coming from time 0. So, given x naught, the best forecast I can make at time 1 is M naught x naught hat. So, now, this error is going to this prediction is going to be in error. So, x 1 is the actual state, x 1 f is the predicted state the difference is called the error in prediction e 1 f, superscript f always refers to forecast or predicted quantities hat always refers to analysis quantities.

So, I am now going to get an expression for e 1 f that is equal to. So, x 1 is equal to M naught x naught plus w 1, x 1 f is equal to M naught x naught hat. So, if I substitute and simplify I get this, but by definition this is equal to e naught. So, now, I get a recurrence relation for the evolution of the forecast error, now look at this is the beautiful expression. The forecast error at time 1 is M naught times the analysis error at time 0 plus w 1. The analysis error in the previous step is going to dictate the forecast error in the next step that is how the analysis or the forecasts are related. Now I would like to come by. So, analysis is filtering, analysis the given time forecast is the predictions.

So, we had talked about smoothing prediction and filtering. So, in this process I already have filtering and prediction part of 2. So, you can think of in our hat is the filtered estimate, even f is called the forecast estimate that is a predicted estimate errors and. So, on I hope this is clear this is where the rubber (Refer Time: 32:27) road we need to combine several things in here.

(Refer Slide Time: 32:33)



So, what is P 1? P 1 is the analysis I am sorry the forecast covariance at time 1. So, let us lets go back in here, this is 0, this is 1, I had x naught hat I had P naught hat is equal to P naught, I have x 1, I have x 1 f, I have to have P 1 f. I already know this this is equal to M 0 of x naught hat, now I need to compute what this one is; if you if you understand this step, the step of going from k to plus 1 will become trivial. So, Pk plus P I am sorry P 1 f f is equal to e of P 1 f e 1 of f times e 1 of f transpose, e 1 of from the previous page is this expression that is that expression. There are 2 terms in each if you multiply there are going to be 4 terms we already know the error in the analysis the previous step on w 1.

So, the error in the analysis e naught 0 in this step and w 1 these are uncorrelated therefore, of the 4 terms the 2 of the terms will die because of this uncorrelated nature as well as w 1 is mean 0, I am left with only 2 quantities the 2 quantities are related by this. So, for because we are doing this for the first time let me try to write this down. What is this? This is equal to M naught, e naught hat, e naught hat transpose, M naught transpose plus w 1, w 1 transpose plus M naught, e naught hat, w 1 transpose plus w 1 times e naught hat I am sorry I made a mistake one second and you get w 1 times e naught hat transpose M naught hat. You can readily see the multiplication of these 4 terms leads to these 4 quantities; I am going to think of expectation of the whole. Expectation of the whole is equal to expectation of the individual quantities; I am now going to distribute that e to every term. So, that is equal to e of this, plus e of this plus, plus e of this term plus e of this term now M naught being a constant it comes out this is the previous analysis this is the feature noise these are uncorrelated.

So, this product is equal to 0, M naught comes out this the next noise this is the previous error. So, this term is 0, this term is essentially the noise covariance Q, Q 1, this is essentially M naught E of e naught, e naught transpose M naught transpose and what is that, that is P naught. Therefore, we get the expression M naught P naught M naught transpose plus Q 1. There are 2 things what they have noticed here the cove the the predicted coherence the at time 1 consists of 2 parts for example, one part if the initial covariance magnified by the model, say this is the initial covariance magnified with model, the second one is the covariance that is introduced by the model noise solid at are you all with me, please back then part you have to have an here this is an moment.

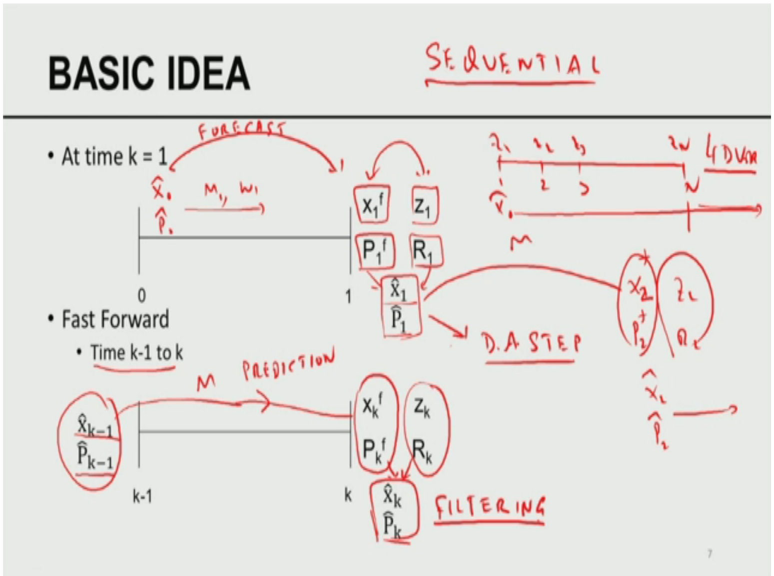
So, if a model is stochastic dynamics, if the randomness are coming from 2 directions Q 1 is the uncertainty in the prediction coming out of the model noise, the first term is an uncertainty that comes from the initial condition noise, initial uncertain the initial condition. Therefore, the prediction has 2 sources of randomness from the initial condition and the forcing, and those 2 together I did trivially contribute to the total covariance of the prediction, why this is added to?

Because we are assuming everything are uncorrelated. If there is correlation between e naught and w 1, then there will be other terms that is coming to cover, this equation the correlation will also try to increase the value. So, why do we assume things are uncorrelated, because I would like to have a plane simple elegant formulation and that is nothing can be simpler, nothing can be more beautiful than this formulation I hope you got. I hope you got the idea of going from step 1 as going from step 0 to step 1. Therefore, I now know the forecast; I now know the forecast covariance. One singular predicate once I have forecast I can create mischief.

What is the mischief? A forecast is the best estimate I can have of the state at time 1, I have the forward operator now I can use the forward operator and the predictor state to create what is called model predicted observation. So, that is what it is. What is the expected value of Z 1 given x 1 is equal to xf. Z 1 is already given to you from Z 1 is given by mother nature, but I am interested in the conditional of Z 1 expected value of Z 1 with respect to x 1 is equal to xf, that is what this is may now condition of the knowledge I would like to be able to get Z 1 from the model is H 1 x 1 plus V 1 x 1 is equal to x 1 f this is the gain conditional expectation is the linear operator, this the conditional expectation of V 1 given x 1 is equal to x 1 is 0.

Therefore the model predicted observation is H 1 x 1 f, again I am assuming the model operator the linear operator can be changing in time. So, am I considering H or H of k am I considering M or M of k. It turns out the arithmetic the algebra replacing by H by H, k M by M k are no different from keeping them time invariant. So, without loss of generality we will pull the time index all through that the idea. So, now, let us look at, this now this is the model predicted observation. So, Z 1 is the actual observation this is the model predicted observation. So, that is the error in the predicted value of the observation given the model state forecast. So, the product of the 2 is going to give you V 1 inverse V 1 that is equal to r 1 as it should be r 1 is the observational covariance error.

(Refer Slide Time: 39:56)



So, that is a check on what we are trying to do. So, the basic this is the basic idea, I have already illustrated. So, you gave me x naught, you gave me x naught hat and P naught hat I there you use my model.

I use my w 1 or model M 1 w 1 I created my forecast, I created my forecast covariance. As I do that somebody is giving you observation and somebody is giving you the observational covariance. So, I have 2 pieces of information in here, I would like to combine these 2 pieces of information to create an analysis and an analyst covariance at time 1. So, combining these 2 to get this is called the filtering. Going from here to here is called prediction. All we have done is to finish the predictive part we have not done yet the combination part. So, this step is called the data assimilation step D A step. This part is called the forecast step. So, now, you can imagine I started with x naught hat P naught hat I made a forecast, then I got the observation I created the new analysis x 1 hat x 2 hat P, P 1 hat then I am going to use the model to to get x 1 I am sorry x 2 f, P 2 f.

Then I am going to get Z 2 R 2, from these 2 I am going to get Z 2 hat P 2 hat and the system continues a that is a cycle that is a sequential process. So, where is the data assimilation step comes in? The data assimilation step comes in after the forecasters made, the forecasts plays the role of the background observation is a new information I am combining them you can see the Bayesian point of view. And that is repeated it is because of this we call it sequential. I think it is better to remind ourselves what do we do in 4 D var? In 4 D var we first decide a time horizon N we get observation Z 1, Z 2, Z 3 Z N we get all the observations, then you tried to fit all of them to be able to decide the best initial condition. Once you bet they decide the best initial condition, then we run the model forward anything beyond N is called a forecast that is what we do in 4 D var.

In sequential we never look back we keep only going forward. Sequential data assimilation is exactly what is being practiced in all forecast centers of the world, these days. So, sequential is in other words I know what I know I simply want to update given the new information to continue new to get the new analysis. So, if I know how to go from 0 to 1, that is essentially the same step to go from kn k minus 1 to k, k to k plus 1 or k minus 1 to k. So, what is the general step? Now that I have described the process of going from 0 to 1 very clearly, I can take lebordy with some other details in going from k minus 1 to k. So, I have been given the analysis and the analysis covariance at time k minus 1 from these 2 using the model M, I am going to create the forecast in the forecast covariance these 2 are generated using the model equations observations come.

I do the data assimilation part; it is here the data assimilation is done. So, forecast prediction data assimilations filtering. So, prediction filtering are continued to sequentially, this combination of prediction filtering somehow call Kalman filtering. So, what is Kalman filtering? Kalman filtering is the with the process that underlie assimilating data into a linear model when the observations are linear. So, Kalman filter essentially assumes LQG; LQG Kalman is in fact, a master of the LQG world, he essentially here he has he has done many wonderful things in control theory Kalman filtering is only one aspect of his multifaceted contribution to control theory, but it is this Kalman filtering that is applicable in the geophysical domain.

Because in geo physical domain since we are concerned with natural occurring system, is there is no way to control the natural (Refer Time: 45:01) you cannot control a hurricane, you cannot control ah an earthquake, you cannot control the blowing out of a top of a mountain a volcano. Natural ecosystem we can only observe we can only predict, but engineering occurring systems you can analyze you can predict you can design you can control. So, that is the fundamental difference between engineering approach to engineering problem and scientific approach to natural I mean natural occurring systems. So, geologists geophysicists and atmospheric science people are also interested in same kind of problems that engineers are interested, the only thing is engineers have an added advantage of being able to control whereas, in sciences you simply have to be able to predict.

So, for example, if there is going to be hurricane we cannot change the hurry motion of a hurricane. Well engineers have sometimes suggested, when at the time when the hurricane forms wanted to just drop a bomb and dissipate. It this is a typically in engineering idea. If you talk to an atmospheric scientist or anybody else they will simple laughs at it and go they would not even care to ah think of answering that.

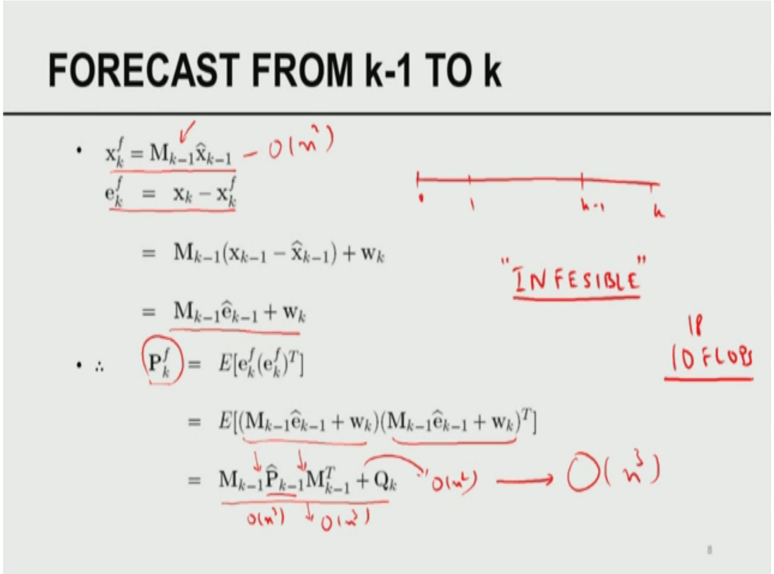
So, engineers mind is always, if you know that there is a danger why do not we control it and prevent the danger from occurring that is the engineering that is why engineers are design, but it is very difficult to be able to control natural occurring systems. But that is why much of what Kalman has done is unknown other than the Kalman filtering, because it is the Kalman filtering is the only one aspect of the engineering solution that he has developed is applicable to the data assimilation setup. In fact, Kalman did not call it data assimilation. In fact, the ultimate paradigm in data assimilation was stated by Kalman is embodied in Kalman filtering.

What is that? I have a stochastic dynamic model, I have observations I would like to be able to sequentially keep updating, and creating newer analysis from previous forecast and an observation. So, in my view in 1960-61 when Kalman published his paper called Kalman filtering that is forerunner of all the data assimilation systems known to mankind. For example, the 4 D var came out only in the mid-eighties, Kalman filter was applied to meteorological problems only in the early eighties.

So, Kalman filter as a solution to a data assimilation problem for earlier than many of the folks in geophysical world had imagined dreamed that is what I would like you to think about why? In the context of weather forecasting in 1960, what is the what was the kind of tool they were using? They were still using successive approximation there was no 4 D var there was no 3 D var, all these things came much later. Even the Kalman filtering idea for it to see through the engineering scientific literature it took well over twenty years. So, in my view Kalman filter is one of the earliest of the complete solution to the data assimilation problem in the context of linear stochastic models and linear of observations there are functions of the linear function of the observation corrupted by noise. I hope the sequential aspect of the idea is clear now.

So, what is that I am now going to do, I am simply going to run through the mill I have been given x k, Xk hat k minus 1 P hat k minus 1.

(Refer Slide Time: 48:46)



I am going to run through the model. So, my forecast at time k till the analysis this is the analysis till the analysis through the model, you get the forecast, then you get the forecast error. If you have the forecast error, this is the expression for the forecast error then this is the forecast covariance which is the product of these 2 terms, because the cross term vanish this is the expression for the forecast covariance. Look at this now no data only model. So, I go from 0 to 1, I go from k minus 1 to k. So, let us look at the computational aspects of this now. To be able to generate the forecast I have to run the model, running the model is essentially matrix vector multiplication that is cheap. So, this is going to be o N square, but let us compute let us let use see what happens in the update of the forecast error covariance assume. Pk minus 1 is given P Pk minus 1 is a N by N matrix, M is a N by N matrix.

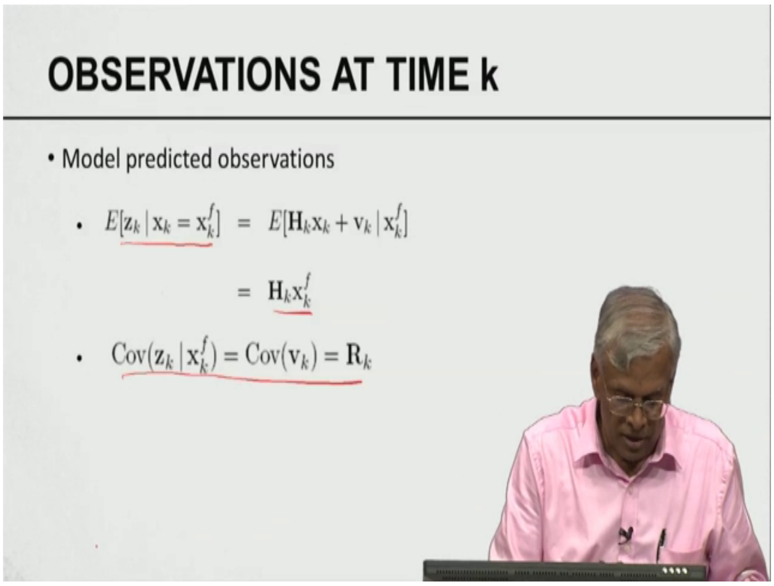
So, I have to do one matrix multiplication, another matrix multiplication each matrix multiplication is going to cost me o of n cube, this is going to cost me o of n cube. Then I have to add 2 matrices that is going to cost me o of n square. So, which is the most expensive part in Kalman filter equation it is not the forecast it is. Updating the forecast error covariance updating the forecast error covariance is of the complexities of the order of N cube, you please remember that we did. If you want to multiply 2 matrices of each size 1 million, and teraflop machine you took about 12 and half 13 days, we have already examined that. So, this multiplication will take 13 days, this multiplication will take thirteen days and this is probably several hours. We are talking about a month's time and there is only to do the forecast error covariance. So, what does it mean? For large systems where the state of the system is in the order of 10 to the power of 6 or higher.

While I know what to do in Kalman filters? Its impractical to get it done in practice because of course, of dimensionality such problems in computing field is called infeasible. It is not that I do not know how to solve I do not know how to solve it is simply that with the kind of environment computing environment I have, I cannot finish this what is the added outcome of this? This promotes an idea telling computer folks you folks, you need to build me a larger machines sorry what is the need it was megaflop machines then it was a teraflop machines, then it became petaflop machines, peta flop machines are 10 to the power of 15 there are very few petaflop machines around the world, now they are talking about exascale machines, where the flop rating is 10 to the power of 18 flops

Japan china I am sure India also has joined this race America. So, almost all the wealthy countries in the world the government related in the wealthy countries in the world they are putting enormous resources, in the development of faster and faster computers because it is the availability to the faster and faster computers, that are going to be useful in making major technological innovations in the future. So, weather forecasting in this sense is one of the hardest computational problems, not because we do not know how to solve, but because we do not have powerful enough computers to support us to be able to perform large computations in a short time.

So, that is a aside story that coming out of this, then fed us to need deal with larger computers essentially come from these kinds of arguments.

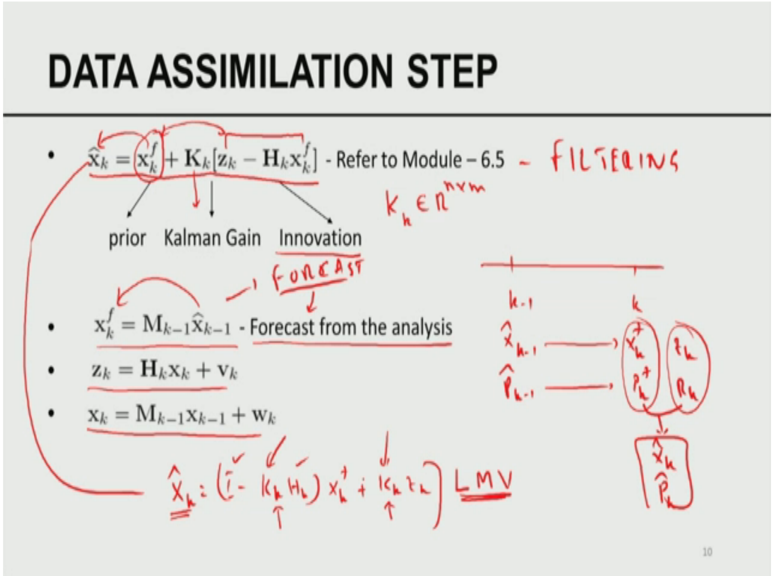
(Refer Slide Time: 53:01)



Again I am going to quickly run through some of the things, which we have already talked about from going from 0 to 1. So, the conditional expectation of Zk given the forecast is this the I can talk about the condition the covariance of Zk given x k, now please understand everything is conditional. Why I am conditioning everything on the amount of information that is available, what are the information available one coming from the model another coming from the observation.

So, everything is a conditional analysis. So, what is the data assimilation step?

(Refer Slide Time: 53:39)



Now I would like to be able to come back to the data assimilation step you go back, I am sorry I can I can I can read here. So, what is the data assimilation step? I want to cut through the mess and give you. So, this is k minus 1, this is k I had Xk minus 1, I had P k minus 1, I have xkf, I have P kf, I have Z k, I have Rk, I am going to combine the 2 to get Xk hat, and Pk hat. So, this is the analysis that is a forecast that the observation.

So, what is that I am now trying to do? Please go back we already have a lot we know a lot of things we already have done the static Kalman filter. Now do you see this is the static Kalman filter look at this now? I have I have a forecast Xk f and its and its covariance, I have observation and its covariance. So, what is it I can do I want to be able to combine them.

I can combine them in the Bayesian way, I can combine them the linear minimum variance way I can combine them as a 3 D var way, we have seen all of them almost all these earlier experience leads to the fact the analysis of time K is equal to the forecast at time k plus zk minus Hk times Xk f, what is H k times skf? That is the model predicted observation what is zk is the actual model observation.

Other than my place if I have model I should be able to predict everything you should never be afraid of prediction, the prediction may come true prediction may be a bust, if the difference is large prediction is the bust. Even though the prediction is the bust I am learning something. So, Z minus Hk s xkf is the innovation, is the new information that the observation brought to 4, that I did not I could not have known at the time I am I when I made the prediction. So, I am going to combine the forecast with innovation, the coefficient of linear combination is a matrix k. K is called the Kalman gain matrix, k is the rectangular matrix of size r n by m. So, what is that we have done we have made the best forecast I hope you understand that.

So, Xk f is the best forecast available that is the best background information that is available to me now I need to estimate. So, this is the estimator this is structure of the estimator what is that is the linear structure? This this can be written as why this is the linear structure Xk hat is equal to I minus K k Hk x kf plus K k, Zk do you remember this looks like L times Xk f plus k times Z k. So, L is a matrix, k is a matrix, in this particular case k is k k k l is this. So, why am I bringing this this is the linear. So, the estimators are linear structures, I would like this estimate to be unbiased, I would like this estimate to be minimum variance. So, I am going to fall back on linear minimum variance estimation a that exactly what Kalman did and I do not know how to do too much because I have already covered linear minimum variance estimation.

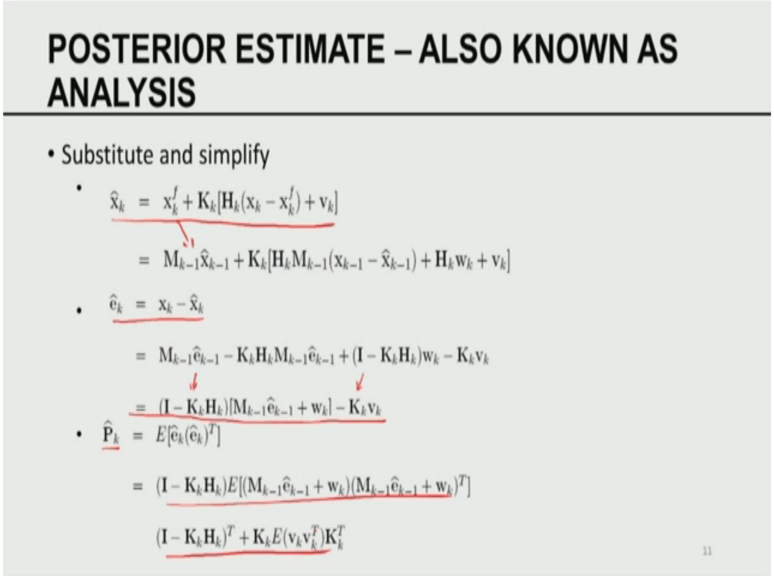
So, what is that? I have done I have essentially prepared all the concoctions needed, I simply need to mix them it is like a fast food chain in McDonalds every order is met within 5 minutes, how did they do that? They anticipate and a certain amount of sale, they tried to prepare all the ingredients the ingredients are already stored whenever an order comes, they simply need to assemble they already ingredients to make the product. So, every product they can assemble a in a short time, that is why this is called fast food chain. And that is the approach I have taken here I have prepared all the concoctions that are needed to do the Kalman filtering. So, what are the various things we did? We have talked about given 2 pieces of information one called background another called observation.

How to mix them, how many different ways in which we have we have looked at them Bayesian way, we had to looked at linear minimum variance way we are talking about 3 d var way. So, I have all kinds of concoctions ready and now I am facing the problem, I can simply call any one of these framework and be able to solve the problem. So, I now know from the theory we have already talked about how to determine Kk, we already have the formula from gauss to Kalman one of the modules we have seen. So, this gives raise to I am simply trying to remind ourselves this is the forecast volume this is the observation this is the actual model.

So, 4 acts true state forecast observations and the Kalman filter equations, and also I would like you to understand the forecast the analysis feeds in to the forecast and the forecast feeds into analysis look at that now. The previous analysis provides the next forecast, the current forecast and the new observation decides the current analysis that is filtering. So, this is the filtering step, I hope you enjoyed this is the forecast step oh I already here I have it here. So, the whole question is my analysis is the linear function in the forecast and observation, the linear function of the forecast and observation I would like to be able to I one more.

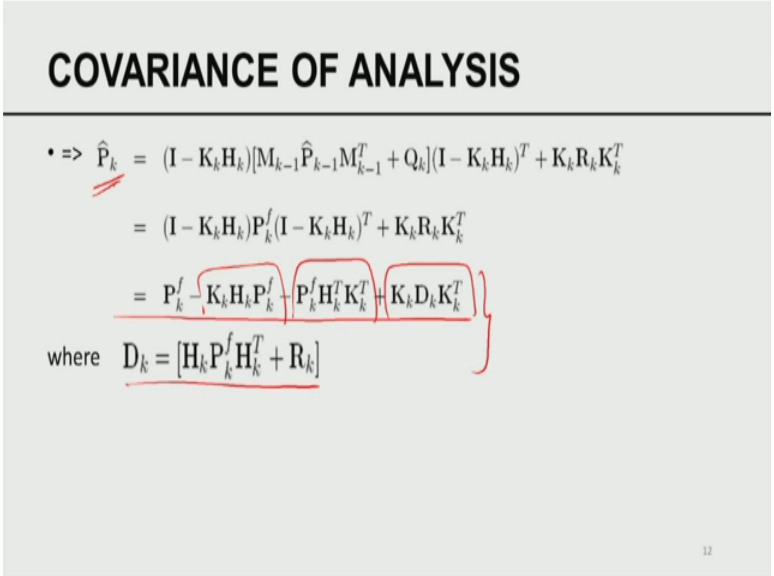
So, I will erase this part, you already know this part therefore, what is involved in here? I already know i, I already know h k. So, the only thing I need to determine is Kk, linear structure is involved I only need to be able to compute the covariance of Xk hat, the covariance of Xk hat is going to be a function of K k, I am going to minimize the trace of the covariance with respect to the elements of Kk a that is the minimum variance estimation, I have already talked about the methodology for doing this in the previous class.

(Refer Slide Time: 60:37)



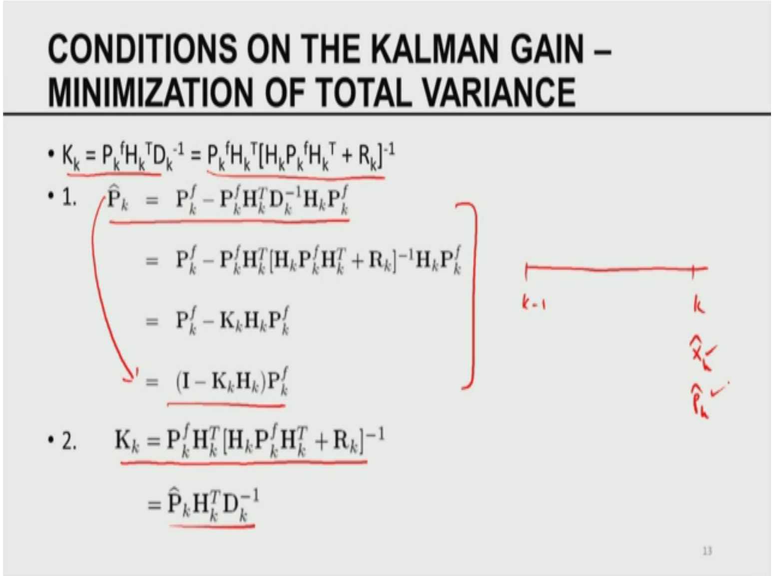
So, I have earned the right to quote the results, but again even instead of simply quoting, I am trying to probe you through the various steps I am not let me quickly tell you the various steps I am not. So, this is the analysis structure, I already know from the previous step how the variables in time k are related to variables in time k minus 1. So, I am relating this to I am sorry, I am jumping I am relating these 2 variables in time k plus 1 from k to k minus 1, k to k minus 1 I am having the error, I substitute the error in here I get the structure of the analysis error. So, this is the structure of analysis error, now please understand the structure of the analysis error involves the Kalman gain, I have not determined Kalman gain I only talked about what is that therefore, the analysis covariance is given by this structure and I have to be able to come ok.

(Refer Slide Time: 61:47)



So, if you do these I am simplifying sorry sorry. So, I hope you got this structures again as a yes is easy for me to do this because look at this now ek is the sum of 3 terms 1, 2 and 3. So, there are going to be 6 terms when you multiply in pk. If you carefully analyze these terms and simplify you get your Pk to be this expression this is the ultimate expression Dk comes in here. Please recall this is exactly the expression we have done in one of the earlier modules, this is the term this is quadratic in k, this is linear in k this is linear in k, I need to be able to minimize that trace of Pk with respect to Kk, Kk has nm elements.

(Refer Slide Time: 62:43)



The conditions for the Kalman gain minimization of the total variance we have already done, and that analysis I am now quoting the optimal Kalman gain is given by this.

So, now I have computed the optimal gain. Once I compute the optimal gain I wanted to go back the optimal gain depends on kk. So, I know the value of k k. So, if I substitute it the optimal gain value in here, I get the minimum expression for the minimum value of the analysis covariance. The minimum expression for the minimum value of the analysis covariance is given by this, which can also be re written like this. Yes there is ton of algebra to be done, I am assuming you will do the algebra and I not only do the algebra, but enjoy the lessons coming out of this algebra is a very educative algebra as any other algebra is.

So, the expression for the Kalman gain is given by this, the expression for the optimal covariance analysis covariance is given by this. So, I have completed the Kalman filter equations. So, what is that now we have said? Let me do it once more in here, I know some of you might feel that I have gone a little faster, but there is nothing I have done here is new I have already built everything in here. So, if I am going from k minus 1 to k I have to I know how to compute x hat and Pk hat. So, that is of the Kalman filter equations. We summarize this Kalman filter equation in a tabular form little later, before that we are going to provide several comments relating to the structure of the Kalman gain.