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

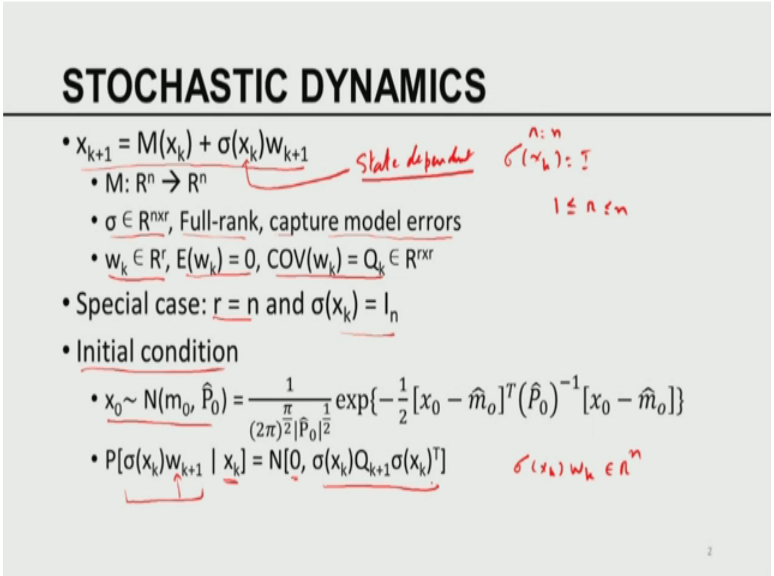
Lecture - 36
Nonlinear Filtering

So far we were concerned with the model that is linear, but stochastic observations are also linear functions of the state, the model noise, observation noise, the initial condition are all normally distributed this is a classic LQG problem, and in this case we had a complete solution for the filtering problem. So, in this sense Kalman's solved 1 of the fundamental data assimilation problem of assimilating data into an imperfect model where the imperfections are captured by stochastic model noise.

Now, we are going to be talking about extensions of Kalman's ideas to assimilating data; in this case the data may be non-linear function of the state, the model itself may evolve according to a non-linear map. So, we are going to be concerned with the extension to non-linear stochastic models, the stochasticity comes from our assumption relating to assuming that the model noise is again white, and the observations are again corrupted by observation noise which is Gaussian. We are again going to fall back on the Gaussian assumption for the both the model noise and the observation noise, initial conditions are also going to be random we will assume that to be also Gaussian as in the previous case the only the primary difference is that because the model is non-linear, the forecast loses the Gaussian any property right at the first step.

So, we have to contend with non-Gaussian processes are arising out of the non-linear systems, and that presents lots of challenges in the data assimilation process, and we are going to provide how to approach the filtering problem. In this case we will not be first talking about first moment second moment the mean and covariance, we will be talking about probabilistic characterization of the forecast the probabilistic characterization of the analysis, we will try to give an evolution of the forecast probability density analysis probability dense density, these are in general infinite dimensional problems because we are trying to talk about an evolution of the density function in the model space. And all the associated mathematical problems challenges, computational problems and challenges that is what we are going to see first.

(Refer Slide Time: 03:18)



So, let us consider a stochastic model; a non-linear stochastic model we are also trying to generalize the forcing Wk plus 1 is a model noise vector, we are going to assume the model noise vector is all dimensional. Sigma X is the coefficient that multiplies the model noise sigma X is the state dependent matrix functions matrix R functions the matrix sigma is n by R we assume it is full rank we assume it captures the model errors. So, if I assume sigma Xk is equal to identity, if I assume R is equal to n and sigma Xk is the identity then the observation that the model error is simply a sequence of state independent Gaussian random of variables, here this is the state dependent noise. So, so this is state dependent.

So, we are assuming that the model is driven by in general a state dependent in principle it could be a general state dependent noise process. R is a variable, R can be in principle less than n R refers to the they the degrees of freedom that the noise has in terms of in terms of it is ability to affect the evolution of the state, when R is equal to 1 there is only one scalar noise that affects all the components of the state vector, when R is equal to n there are n different noise components that can affect all the components of the state vector depending on the structure of the make the state dependent matrix sigma x.

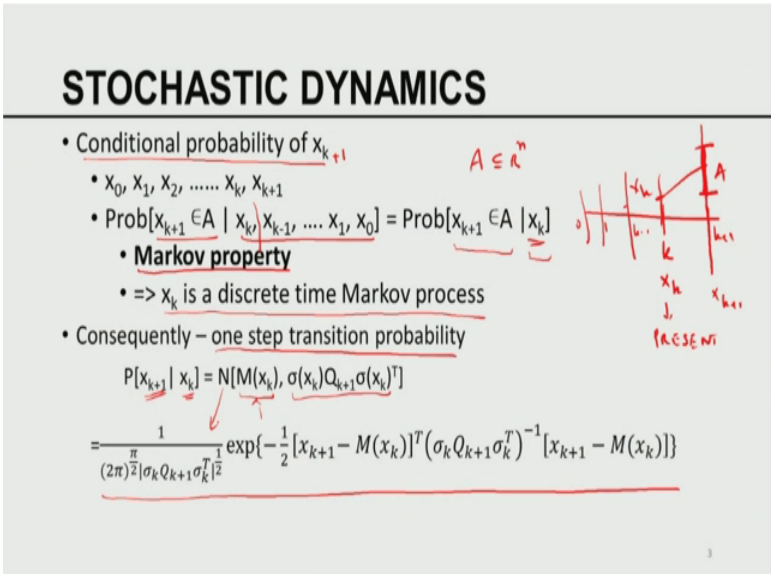
So, there are you can see from the set up by appropriately choosing the sigma matrix by appropriate choosing the value R, one can simulate quite a variety of assumptions one can realize quite a variety of assumptions relating to the nature and type of model noise into the system.

So, Wk is mean 0 Wk has a covariance Qk. Qk is a matrix of size R by R, the special cases are 2 special cases are R is n and the and sigma Xk is equal to I n is the identity matrix. Then origin and sigma Xk is equal to I n what is that we do? There is no state dependent noise the noise is independent of the state the noise becomes a pure Gaussian white noise

The initial conditions are random again it is a multivariate normal distribution with the mean M naught and covariance P naught hat. So, given Xk Wk is random. So, I can compute the distribution of sigma Xk, Wk plus 1 the probability density of this vector. So, please realize this is an n vector sigma Xk times w k is a n vector and. So, it is its mean 0 it is a covariance is given by this expression sigma Q k plus 1 sigma transpose.

So, we have talked about the choice of the model error or model noise we are also talked about the choice of initial conditions.

(Refer Slide Time: 07:36)



Now we are going to be talking about the properties of the evolution of the state of the system, then minimally the forecast, when there is no observation that is what is called analysis of stochastic dynamics.

We are interested in the conditional probability density of Xk, we are interested in the evolution of conditional probability of I should say Xk plus 1 given in the past. So, what is the probability that Xk plus 1 will find itself in a state in a set A. So, A is supposed to be in this case a subset of Rn please realize in some cases we used a as for the subsidy, in some cases use ef for matrices. So, the occasion will tell you what that symbol means. So, in this case a is a set. So, what is that we are trying to talk about? Given the past trajectory model starting from X naught to Xk plus 1, we are interested in trying to find out what is the probability that Xk plus 1 will belong to a set A, in here Xk is the present. So, we can think of this Xk, k Xk time k Xk is the present this is k plus 1, this is Xk plus 1.

So, we were trying we are trying to ask ourselves the following question let me draw the picture a little bit differently. This is k plus 1 this is Xk plus 1. So, given the state Xk what is the probability Xk plus 1 will be a subset of the set A. So, this is the set A at time k plus 1.

Now, I am given the state of the system at time k minus 1, I am given the state of the system as time 1, I am given the state of the system at time 0 even though I am given the complete history from 0 to k, if this probability depends only on X k, but not on the past. So, k is the present. So, given the present the future probability the future evolution of the states is the independent of the is independent the past; that means, the state of the system from 0 to k minus 1 does not play a role once Xk is given in determining, what fk plus 1 is going to be that kind of property is called Markov property. So, what does the Markov property say? Given the present past is in consequential to consider the future. So, the probability that at time k plus 1 due to belong to the set a depends only on the current state Xk and not in the past.

So, the model equation given in the previous slide in fact, represents a Markov process as it is evident from the relation if I am given Xk, I do not have to know anything, M of Xk can be computed sigma Xk can be computed, Wk k plus one can be generated, it is Wk plus 1 is the one that brings randomness into the in deciding what Xk plus plus 1 is going to be. So, given Xk the value of Xk plus 1 does not depend on anything before Xk it depends on Xk and the noise that comes into the system after the time k that is why the model is said to be a discrete time Markov model and the process generated by this model is called a discrete time Markov process.

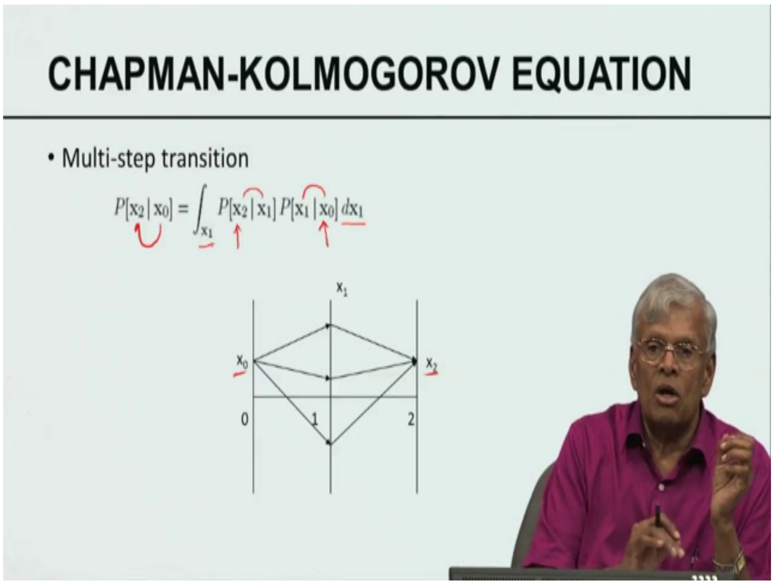
The notion of being Markov is very fundamental; it is a stochastic generalization of the deterministic principle what is the principle of deterministic determinism? If I have a differential equation if I know the state of the system at time k in principle that is enough to be able to compute the state of the system at time k plus 1, because there is the differential equation tells you the rate at which the system evolves starting from time k.

Therefore Markov property can in many ways thought of a simple extension of the fundamental properties of deterministic dynamical system, in this case we are considerable concerned with the discrete time evolution a continuous time evolution Markov process theory still exists, but that theory is little bit more technical, not to reduce the amount of mathematical technicality that one needs to know, we can find our attention to the analysis of non-linear difference equation, in which are driven by state which could be depend depending on state dependent noise vector, in the evolution together disc describing your discrete time Markov process.

Therefore we are interested in what is called the one step transition probability, given Xk what is the probability of Xk plus 1, this is the one step conditional probability is a one-step conditional transition probability. So, if I am given Xk, M of Xk plus 1 depends on M f Xk plus the noise term. Given Xk sigma Xk is given expected value of Wk plus 1 is 0 therefore, the mean is essentially the deterministic part M of Xk, the covariance of the state is given by sigma Q k plus 1 sigma transpose.

So, given Xk, Xk plus 1 is Gaussian and it has a density function whose expression is little bit complex and, but for explicit analysis I am giving the that the distribution in this particular form and that essentially tells you how the system evolves. So, once I know the initial distribution, once I know how they distribute how the system goes from time k to k plus 1, I should in principle be able to pull the system forward in time.

(Refer Slide Time: 14:31)



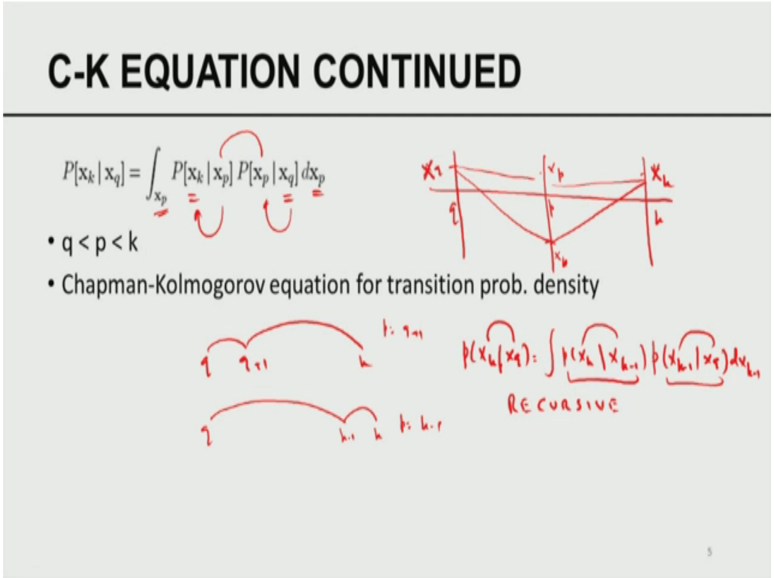
So, we are now going to talk about how knowing one step transition probability, we can compute multi-step transition probabilities, let us consider transition from time 0 to time 1 to time 2, let us assume I am in state X 0 to start with, I would like to be able to find out what is the probability that X 2 will be at the position shown in the figure.

So, in order to go from X naught to X 2, I had to go through an intermediary stage, the intermediary stage is the value of the state at time one. So, from X naught I can go to any 1 point in the one dimension in the X space in this case for simplicity I am trying to show um the state space as a vertical line as if it was a 1 dimensional, but the same thing applies to multi-dimensional, we are simply representing the multi-dimensional space by a vertical line. So, X 1 refers to the state at time 1. So, go from X naught is fixed, X 2 is fixed, to go from X naught to X 2, I had to go through some X 1 in the medium in the intermediary.

So, with this the probability of going from X naught at time 0 to X 2 at time 1 at time 2, is given by continued conditional probability of X 2 given X naught. So, that is what the conditional probability given X to X naught. So, I would like to be able to argue now, I started from X naught, I want to go to X 2, X 1 can take any intermediary values. So, I am going to go one step from X naught to X 1 and from the chose on X 1 I go to X 2, this X one can be any point in the space in the state space. So, I am going to have to multiply this conditional probability P of X 2 given X 1, times P of X 1 given X naught and integrated with respect to X 1 dx 1.

This sum total of the product of the conditional probabilities will give you a 2 step conditional probability is called a multi-step conditional probability. This relation of trying to find how a multi an expression for the multi-step transition based on one step transition, has come to be called chapman Kolmogorov equation.

(Refer Slide Time: 17:01)



It is a very basic equation I can now extend this to a general case instead of 0 to 2, I can now think of the following.

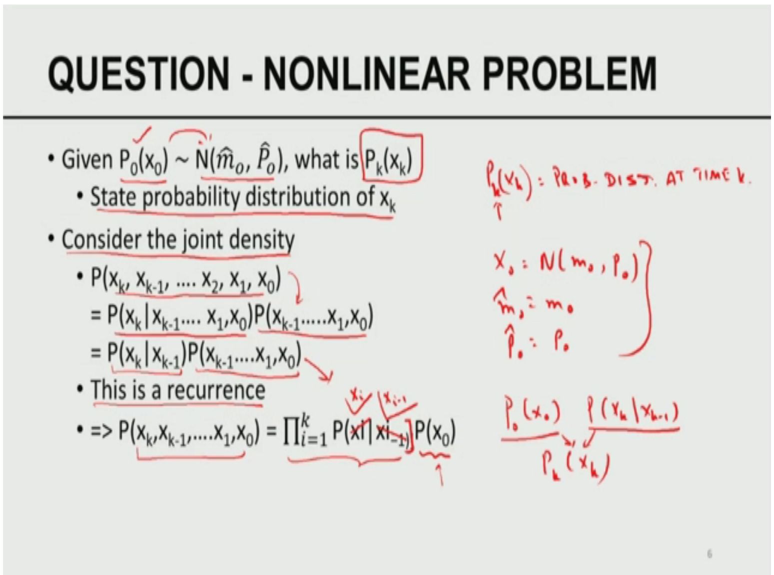
Let us assume I am in state q I am sorry I am in state X q at time q. So, q is some instant in time, I want to be able to go to k, I would be able to go to Xk at time k. If I want to go from q to k I am go I will have to go through some intermediate stage you. So, you can think of that intermediate stage to be you can think of that intermediate stage to be from q to p. So, P is the time this is X p. So, you go from X q to xp from xp to Xk, you go from xq to X P and xp to X k. So, xp can take any value in here therefore, q is fixed, k is fixed q the P, X P is our variable I am going to integrate with respect to X P, I can once I integrate this I get the transition probability from step q to step k.

So, what is this this is this is the transition from q to P this is the transition from P to k. So, this is a combination of 2 multi-step multi step transition probabilities. So, by I i I can. So, I can break this P to be many things. So, I can go from q to q plus 1 q plus and then q plus 1 to k that is a possibility in this case P is equal to P is equal to q plus 1 I can split it like q to k minus 1 to k.

So, I can. So, in this case P is equal to q plus 1, in this case P is equal to k minus 1 and. So, I can reduce the multi-step transition by a sequence of one step transitions. So, I would like to rewrite this. So, X i would like to be able to rewrite this equation recursively in this way P Xk given X q is equal to summation, P Xk given Xk minus 1 times P Xk minus 1 given times xq times d Xk minus 1. So, q to p is related to k minus 1 to k and q to k minus 1. So, this is one step transition this is the multi-step transition I can convert this .

So, this is a recursive relation using this recursive relation, I can compose multi-step transition probability involving any number of steps. So, this general this equation is called chapman Kolmogorov equation for probability density functions, multi-step transition probability. So, given a Markov process, a Markov process is uniquely defined by the initial condition or the initial distribution and the one step trait transition mechanism. If the one step trait transition mechanism is specified I can create multi-step transition mechanisms probability values using chapman Kolmogorov equations.

(Refer Slide Time: 20:56)



So, that is the conclusion so far now, what is the statement of the non-linear problem I am given an initial condition P naught X naught please understand. Pk Xk is the probability density of the state Xk at time k. In general this probability density will depend on k therefore; the subscript for P refers to the time varying density of the state Xk as the state evolves according to the dynamical system.

So, what is the question given P naught X naught this is the initial state distribution, which is given by X naught m naught hat and P naught hat please remember, that X naught is equal to m naught P naught. So, I am trying to define my initial analysis to be m naught, my initial analysis covariance to be P naught. So, with this initialization this represents the initial distribution, given the initial distribution of the state I would like to be able to find the distribution of the state at time k, this is what is called the probability distribution at time k. So, I am I want to remind the reader now there are several probability density functions we are involved in, one is the initial probability density another 1 is the transition probability from k minus 1 to k another 1 is Pk xk.

Now, in all this we know this transition density from the model, we are given this initial condition from external specification that is equivalent to specifying the initial conditions for the dynamical system. So, given these 2 our job is to be able to compute the state. The state probability density function at time k. So, this is called the state probability distribution of Xk ,I would like to be able to compute this quantity now we are going to look at means by which we can arrive at this evolution of the state probability density functions in time.

So, it will be take there once more, given the model, given the forcing, given the initial condition, the model defines the one sub transition the initial condition randomness is given. So, you can see there are 2 sources of randomness, one coming from the choice of initial condition another coming from the one step state transition, these 2 together decide the state probability density function. The state probability density function is called is the Pk Xk our ultimate goal is to be able to find out how the states of the model are distributed and in any given time Pk what is P k Xk for any k.

Given this I would like to be able to start from what is called the joint density. If I have 2 random variables, I can consider I am I am that that always exists in an appropriately chosen probability of a space, there is a joint density, then I can consider the marginal densities I am assuming that we are all familiar with the notion of marginal densities conditional densities joint densities all basic fundamental concepts.

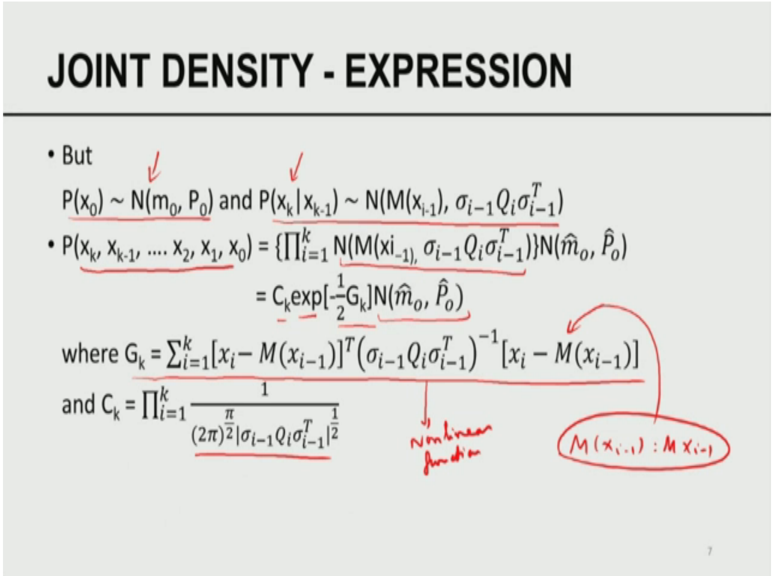
So, consider a joint density of the state from X naught to Xk. Using a simple conditional probability, I can express the condition at the joint density as the product of another joint density and the conditional density. So, this is P of Xk given Xk minus 1 through X naught times P of Xk through Xk minus 1 through X naught. So, conditioned on the knowledge from X naught to Xk minus 1 I can split this into a product of these 2.

But I have already assumed the process is Markov. So, knowing Xk minus 1, I do not have to know how I got to Xk minus 1 the past is of no consequence is deciding the future if the present is known. So, Xk minus 1 is the present. So, the this conditional probability reduces to your one step trait transition probability of the Markov process, we are considering.

So, this is the joint density from state 0 to state k minus 1. So, you can now see the joint density from 0 of the state from 0 to k mine k can be broken down into the product of joint density from 0 to k minus 1, and once the transition from k minus 1 to k, this recurrence relation and this I can apply this recurrence relation to this term on the right hand side. If I apply this continuously now I can you can readily see the joint density is expressible as the product of the conditional densities; that is a typo, this is X sub i, this is X i minus 1.

So, the product of the conditional densities times the initial density. So, this is the initial density, this is the conditional density. So, I can you can readily see the joint density is the product of the one state transition densities and the initial density. One sub tastes a straight transition density is given by the mark the model, this is the initial density. So, I am expressing the conditional density as a product of everything that I know.

(Refer Slide Time: 27:23)



So, I am now going to be looking for expressions for the joint density; a little bit more characterization. So, we now know P of X naught is normal, one step trance state transition probability which is the conditional probability that is also normal, we have already argued about the normality of the one step state transition probability for the model equations. In the previous step we have expressed the joint density as the product of the conditional densities and the initial densities. Therefore, I can express the joint density as the product of the normal densities and another normal density.

The product of k normal density is referring to the k step transition from 0 to k and the initial density. If I substitute the expressions for each of these normal densities and simplify I get a constant times exponential of minus Gk; minus 1 half of Gk times the initial density, where Gk has an expression which is the sum of Xk minus 1 minus M of X i minus 1 transpose the inverse of the covariance matrix of the one step transition times X i minus M of X i minus 1.

So, that is a quadratic form, this quadratic form is a non-linear is non-linear it is much more than quadratic because M in general is a non-linear function, this will become an actual quadratic form only when the model is linear, when the model is not linear. So, what do you mean by saying model is linear M of X i minus 1 is equal to M of X i minus 1 that is the linear case, in this case it is a quadratic form if not this is not a quadratic form. So, in principle this 1 is not a quadric form, it is it is more complex than a quadratic function is the non-linear function. Much of the difficulty in computing the joint density arises from the this complex nature of the non-linearity that enters into the description that in the description of the joint density.

So, the Ck is given by this constant. So, given this now we have computed the joint density at least mathematically in in the form given by Gk, in the form that is given by the exponent Gk the expression for the exponent Gk is a complex non-linear function.

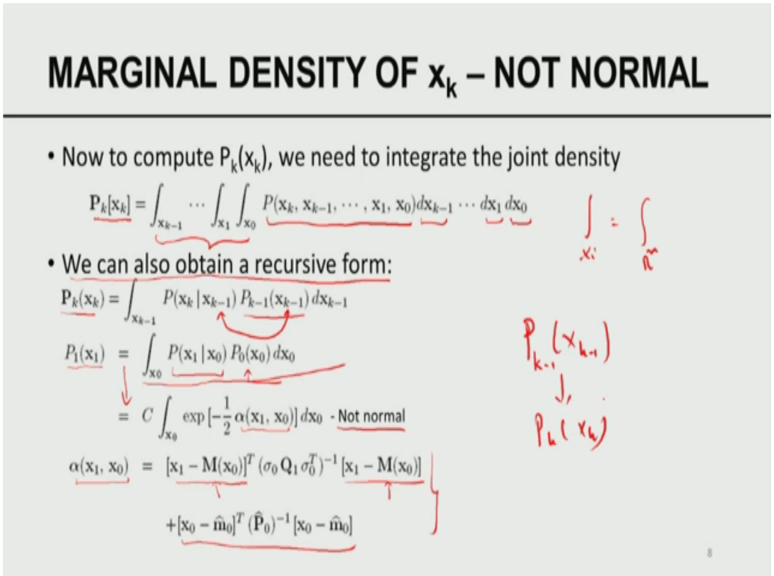
If you recall we are in the middle of discussion of non-linear filtering, in the case of non-linear filtering because of the non-linearity we cannot simply be content with first moment and second moment. The complete solution is derived is given by the entire probability density function for the forecast for analysis. Once you know the probability density function then we can compute any number of moments first moment second moment etcetera. This is largely because of the fact that there are non-linearity in the system even though the initial condition maybe in Gaussian distributed, maybe the one step transition probabilities of the non-linear system that defines the Markov process which are also Gaussian.

In spite of the fact if you want to be able to compute the state distribution at any given time, that is highly a non-linear function and it is far from being normal. We are trying to get a handle on on this important quantity namely P k, Xk you may recall from the previous page. What you speak Xk is, Pk Xk is the probability density function of the state at time k, there are 2 case 1 for the state X sub k and on the state itself is changing another there is a subscript for P, P of k of Xk P of k refers to the density it is the probability density of the state Xk at time k, the subscript k associated with P tells that the probability density function is changing in time. Not only the state is changing in time the probability density is also changing in time.

It is this quantity which is of interest to us and we are trying to express the joint densities to start with, I am trying to go over some of the things we have already done. So, I am trying to compute the joint density of the state from 0 from time 0 to time k, using this recurrence we just saw it can be expressed as a product of conditional densities and the initial density.

Initial density is normal condition density is normal. So, the joint density is given by constant Ck times exponential minus 1 half times Gk normal with mean M naught hat and P naught hat, the crux of the expression relates to analyzing what is contained in Gk G k is given by this complicated expression which is the exponent and the large that the most of the difficulty arises from the fact M is not a linear function, in case M of X i is equal to X i minus 1 is equal to M times X i minus 1 the exponent becomes a simple quadratic function because M is in general not necessarily a linear function this exponent is in general more non-linear than quadratic functions. So, in general they are quadratic functions and this is largely the major difficulty in trying to quantify the state distribution. The distribution of Xk at time k which is P k X k, but at least theoretically one can compute the joint density of the state from time 0 to time k.

(Refer Slide Time: 34:03)



Once you have the joint density from time 0 to time k, I am still interested in finding not the joint density, but the state density at time k, pk, xk you know that P, kx, k is the marginal distribution of the joint density. So, this is the joint density, if I integrate the joint density over all the variables other than X k. So, this is integration is from X naught X 1 through xk minus 1. So, there are k iterated integrals each of these iterate iterated integrals are integrals over Rn, because Rn is the state space. So, when I say integral over X i that is equivalent to integral over Rn. So, integral over X i integral over R n. So, this is the repeated integration in the in n dimensional space.

Please recall we are not trying to do the actual integration we are trying to develop the theory. So, the theory can go anywhere, but ultimately we are interested in trying to compute the probability density function for the state Xk at time k. We can obtain now your recursive form for this P k Xk why this Pk Xk can can be expressed in a recursive form? Pk Xk is the probability density of the state Xk at time k. So, if I now Pk minus 1 Xk minus 1 that is the probability density of the state k minus 1 at time k minus 1.

Then from Xk, I can go to Xk minus 1 by the one step transition probability rules. So, given in this I can compute the transition density from k minus 1 to xk. So, this integration is over Xk minus 1. So, this is this essentially follows from basic probability theory basic probability state arguments.

In particular when k is 1, P 1 X 1 that is in the probability density of the state at time 1 is equal to the initial state distribution that is the initial condition, this is the one step transition probability, the initial density is Gaussian, the conditional density is Gaussian, but the conditional density is a function of the model map, model map is highly non-linear therefore, P 1 X 1 can in principle be expressed by this integral, but it is far from being Gaussian. That is the primary difference between the linear of the non-linear filter much of the difficulty associated with the non-linear filtering, at least in one part comes from this inability to preserve normality under non-linear transformation.

So, now let us give a little bit more life to this P naught X naught is normal, P of X 1 given X naught is normal, but if I substitute all the normal expressions for this, this integral becomes equal to this well I am now going to talk about the exponent the exponent is alpha times X 1 X naught, that alpha times alpha of X 1 X naught is given by this function that is the exponent that describes the product density. The product of the conditional density and the initial density, you can readily see this M is a model map this product that tries to make it a non-linear function. So, I cannot simply rewrite it a Gaussian and this alpha of X 1 X naught has also another term that comes from the initial condition.

So, the initial condition current contribution is quadratic, but the contribution from one step transition is not, it is a combination of these 2 terms makes P 1 X 1 far from being Gaussian. That is not that is that is the real rub, then it comes to non-linear filtering as we move from linear to non-linear maps as you linear as you go from linear to non-linear models.

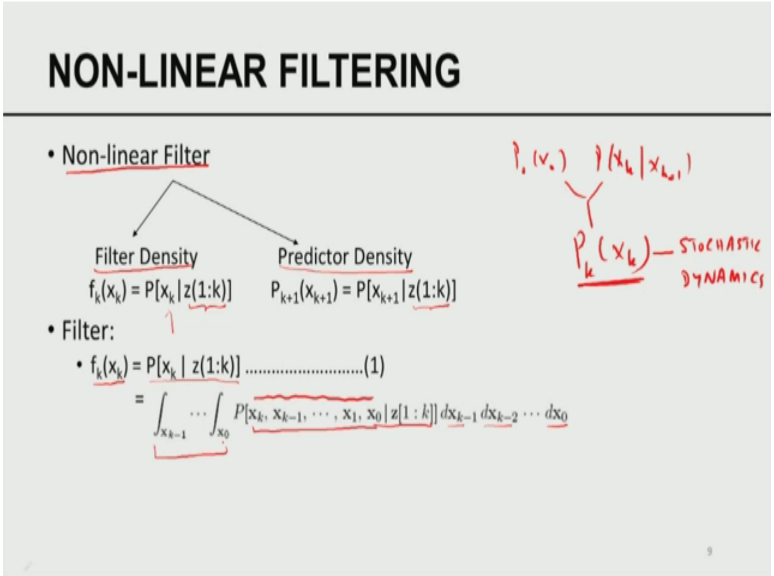
So, if I can. So, by this we have seen that P 1 X 1 is not normal therefore, P k Xk is not normal, the non-normality continues to dominate the show because the state distributions are not normal it is not enough to compute the mean and the variance I need to be able to compute the entire distribution. So, the non-linear filter seeks to up update not the mean and the covariance as their linear Kalman filter dead let us go back; what is that thing we did in the Kalman filter? We updated the forecast mean we update the forecast covariance, we updated the analysis mean we updated analysis covariance because everything is normal by knowing the mean and the variance I know the entire distribution that is not the case and that is largely the difficulty.

So, Pk Xk in principle one can compute the only way to be able to compute these things as numerically of course, there are still very many good numerical integration packages one can utilize to be able to compute this, but what is that we are seeking? We are seeking something a sequential algorithm, what is the sequential algorithm like in Kalman filter, I am going from state time k to at time k plus 1. So, knowing the analysis and it is covariance at time k, I would like to be able to compute the forecast and this covariance or time k plus 1 observation comes, I am now going to recompute the analysis and the covariance at the next time interval that is the sequential nature, and that is what we are looking for.

So, we would we would like to go from time k to time k plus 1, we would like to be able to update P k minus 1, Xk minus 1 to P k Xk if I can do that that is what the sequential algorithm is all about. Please recall each of these are functions each of these are continuous, I am assuming the density functions they are continuous these are functions continuous functions defined over the n dimensional space, and the integral of this must be 1 these functions have to be positive.

Now, we can see the I am talking about non I am talking about positive function functions that are non-negative and whose integral has to be 1, and they are differently n dimensional space, and when n is large 100, 100,0, 10,000 million you can see the associated difficulties and trying to keep the non-negativity of this function, when you are trying to do the numerical computation these are some of the challenges one free one will find themselves in when you are trying to convert these things into numerical algorithms.

(Refer Slide Time: 41:18)



So, what is that we have accomplished? We have simply analyzed the model forecast with no observation. Starting from the initial distribution, knowing the one step transition probability of the underlying Markov process defined by the model, I am now at least theoretically be able to explain R X express Pk Xk. So, we need we have started from we P naught X naught ,we had access to Pk Xk minus 1 from the if I combining these now I have an expression for P k Xk no data is involved it is simply model analysis.

This idea of trying to explain the evolution of Pk Xk that is part of the stochastic dynamics that is the stochastic dynamics part. So, the complete information that one can hope to give in the case of stochastic dynamics is the probability is the evolution of the probability density of function.

Now, let us bring in the filter, filter means observation. So, when I am going to be developing expression for the non-linear filter I have to have now 2 kinds of densities, one is the predictor density another is the filter density what is the predictor density? It is the density of the state at time k plus 1 given all the observations. So, that embeds the model as well as the observation. So, we are going to call this as the predictor density.

The filter density is going to be f of k f Xk, which is which is given all the observations from 1 to k, I would like to get the best estimate of the state at time k. Please recall the Kolmogorov v in our definition given all the information up to time k estimating the state of the system at time k is called the filter problem, given all the information up from time

1 to k trying to know the state of the system at time k plus 1, there is a prediction problem here instead of simply predicting the mean of the covariance we have to predict the entire distribution itself. So, this is the predictor density this is the filter density.

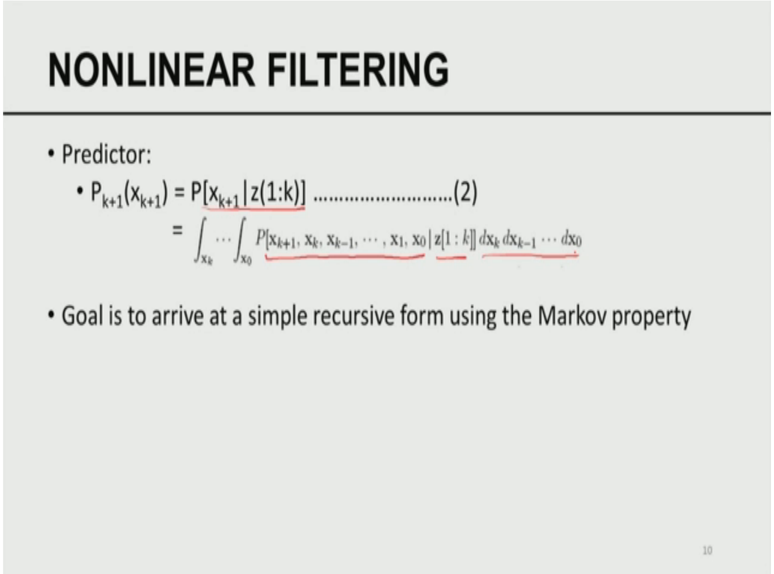
We have some idea of the state density evolution, now I would like to talk about the structure of the filter density. So, f of k Xk the filter density also changes in time. So, f of sub k of Xk, Xk changes in time f of k also changes in time much like the state density distribution changed.

So, by definition this is equal to the probability density of Xk given Z 1 to k, the observations from time 1 to time k. This can be written the join this can be written as the integral of the conditional density of the state of the system from X naught to Xk, given Z 1 to k, in the integration is from k minus 1 Xk minus 2 and X naught, and that that is the repeated integration in here.

That comes essentially from basic definition of the density functions, I want you to understand these are all mathematical possibilities, we want to know that it is first mathematically feasible to describe what I want leaving the computational problem after the feasibility studies have been completed

So, this conditional density of the state at time k given the observation from 1 to k, is essentially the marginal of the joint conditional density integrated over X 0 to Xk minus 1. I think that should be pretty clear from basic probability argument.

(Refer Slide Time: 45:41).



The predictor density now is Pk plus 1, I am giving again expressions for this filter density and predictor density, we saw the filter density in equation 1 we are now giving the predicted density in equation 2, predictor density is likewise you have the look at this now the joint predictor density condition on Z 1 to Z of 1 to k, integrated over 0 to k.

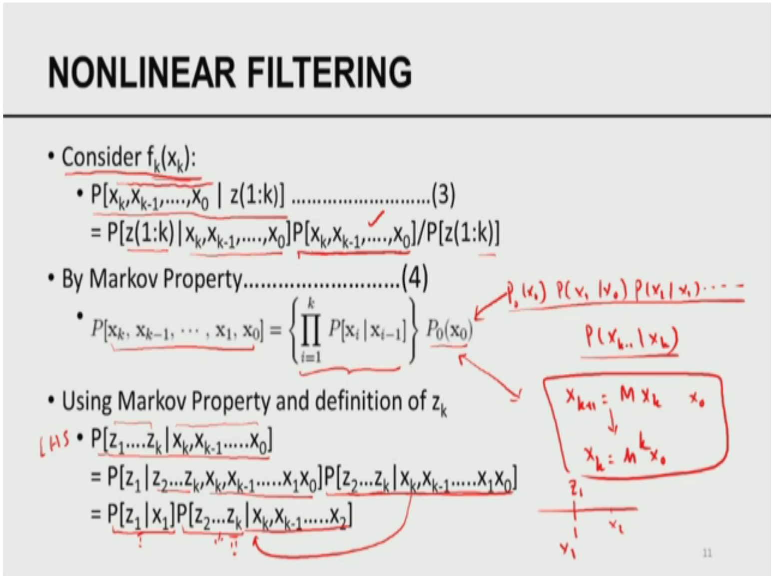
So, that is the expression for the predictor density. These 2 densities in principle make sense, but these are all not in the recursive form. So, what is our goal? To say I have a non-linear filter is to be able to rewrite these 2 equations 1 and 2 in a recursive form.

So, we arrive at a simple recursive form and who is going to provide the key to the recursive form, the Markov property the underlying Markov property of the stochastic model please understand that is the key. So, what is the Markov model essentially tells you if I know the state at time k and if I know what comes after time k, knowing what comes after time k I should be able to precisely probabilistically predict what the state will be. We do not know the exact value, but we would be able to tell the distribution of the state of the system at the next time.

The Markov property depends critically on the one step state transition probability, and we are going to exploit that property to be able to write equation 1 and equation 2 in a simple beautiful recursive form. One is that is accomplished at least in principle, we would have solved the non-linear filtering problem to tell how to make the prediction of the density of the predictor state, then given the predictor density and the distribution of the observation, I am going to get the filter density the filter density represents the analysis the predictor density represent the forecast step.

So, you can see essentially all the ingredients of the Kalman filter are alive and well, instead of computing the vectors and matrices which are the first and second moment we are going to have to update the entire functions over the n dimensional space. That is the key to understanding non-linear non-linear filter equations.

(Refer Slide Time: 48:18)



So, with this as the background now I am going to talk about manipulating the expressions for f k of Xk is what I want now we will start with the basic statement. Given the observation from 1 to k the conditioned on that I have a probability density over the state from X naught to Xk. So, that is the conditional probability distribution given the observation after time k, a you can you can see the following I am Xk is involved, I am also interested in the trajectory of the system starting from X X 0 to Xk, I am conditioned everything on all the observations up to including time k, Z 1 colon k Z 1 colon k represents all the information that are obtained from time 1 to time k. Of course, inherent in here is the model information to why going from X naught X 1 is given by the model. So, that is model information that is observation information. So, this is the goblet go mix of both the model and the observations.

Now, using Bayes rule, this conditional probability can be written as Z 1 k conditioned on the state 0 to k times the probability of the trajectory going from 0 to k divided by P the probability of observing the observations from 1 to k that essentially follows from simple Bayes rule.

So, we have already applied Bayes rule as in 3, now I am going to express this probability which is the probability of the trajectory starting from X naught Xk, we have already seen using Markov property, the joint probability of the trajectory starting from X naught to X 1 is given by P naught of X naught. So, we can talk about this now. So, P naught of X naught then P of X 1 given X naught, then P of X 2 given X 1 all the way up to P of Xk minus 1 times Xk.

So, what does it tell you? This is the probability of observing a particular trajectory of the system starting from X naught. That it is this is encapsulated in here, it is simply the product of the transition probabilities that define the path times the initial distribution.

So, what is this? This is a stochastic analog of simply recursing the set of equations. I will give you a quick analog, in a linear system if I have Xk plus 1 is equal to M times Xk X naught is given from here we would know Xk is equal to M to the power k X naught.

So, I should be able to relate the system at time k to the time X naught, through the case step transition probability matrix, the kth power of it, and that is what happens in the linear system in the non-linear system this cannot be done, but you can think of this to be an analog of what happens in the case of a linear system. So, for those of us who are who would like to have a link that is how you need to look at this. Given the initial condition given the state transition map, how does the state transition map and the initial condition together define the trajectory. So, this is the probability of observing the entire trajectory from X naught to X 1 to X I, Xk minus 1 to Xk.

Now, using the Markov property. So, this is one of the terms in the Bayes rule, it is the second term in the numerator on the right hand side of 3. Now I am going to consider the first term in the numerator of 3 on the right hand side of 3, and that is what is given by this. Z 1 colon k what is mean with what I have all the observations from Z 1 to Z k, I am conditioning it on the trajectory, now look at this now. I have already computed the probability of observing the trajectory. So, conditioning this is known. So, given the probability that a particular trajectory is observed, I can now conditioned on that particular trajectory, I can then compute the probability of the observation conditioned on the trajectory, that is the whole idea that is a very simple idea.

Again we are going to apply the manipulation of conditional probability again and again. So, this can be written as probability of Z 1 given Z 2, Z k X 0 to Xk times probability of Z 2 to Z k and the same trajectory.

Now, let us look at this now, Z 1 it is it is the observation at time 1. X naught X 1 these are the state at system at time 0 and time 1. X 1 depends on X naught and Z 1 depends on X 1, Z 1 does not depends on X 2 what is X 2, X 2 the state of the system at time 2; what is Z 1? Z 1 is the observation at time 1, the observation at time 1 does not depend on the future state I hope that it is very clear.

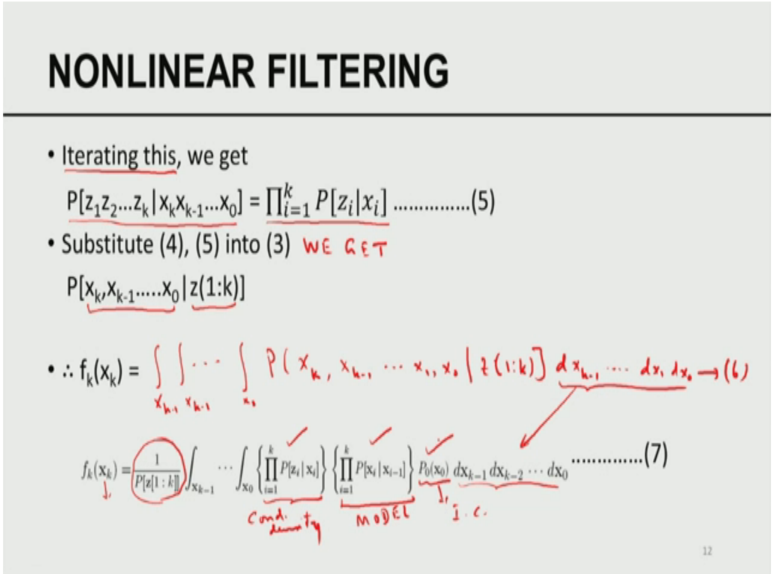
So, in view of the non-dependence of the observation Z 1 on states beyond X 1; that means, Z 1 does not depend on X 2, 2 Xk if it does not depend on X 2 Xk the conditioning has no value, I can drop that conditioning out of consideration. So, I can rewrite this term as P of Z 1 conditioned on X 1 times P of then the rest of this the rest of the term comes in here right now.

So, now we can see I am trying to bring a recursive structure into the system, and this recursive structure is again a consequence of Markov property. So, let me let me say it once more. If I have a time one state X 1, Z 1 is an observation that comes at time one, X 2 is a future state it makes sense to think about the system does not have the anticipatory power in other words my todays observation of temperature is not going to take into account tomorrows temperature, todays observation is measured on today and perhaps some of the past even that has taken place.

So, same consideration because of such simple arguments, you can readily see the conditioning even though I have conditioned this on Z 2 to Z k, X naught to Xk todays observation does not depend on tomorrows observation, todays observation does not depend on tomorrow state therefore, the dependence of Z 1 and Z 2 to Z k can be dropped dependence of Z 1 and state from X 2 to Xk can be dropped, therefore, Z 1 can at best depend on X 1. So, the first time simplifies as follows that again comes from the mass Markov property as we observed.

Now, look at this now; this is the left hand side, this is one of the terms in the right hand side, the this is the first term and this is the second term. tThe second term and the left hand side are exactly the same except that the left hand side is from 1 to k the right hand side from 2 to k. So, what does it mean? I have expresses a recursive structure, this recursive structure now can be adapted to the second term in the right hand side. I can further apply the recursive structure.

(Refer Slide Time: 56:45)



So, if I apply this recursive structure continuously open it up, I have that is called iterating you. So, iterating this we get this density is equivalent to product of probability of Zi versus X i, i is equal to 1 to k.

So, we can now I have 5, I have 4, I can substitute 5 and 4 look at this now what is 4? 4 relates the second term on the numerator on the right hand side of the Bayes rule what is a 5? 5 relates to the expansion in the property of the first term on the right hand side of the Bayes rule. So, I substitute 4 and 5 in 3 simplify the trajectory given the observation

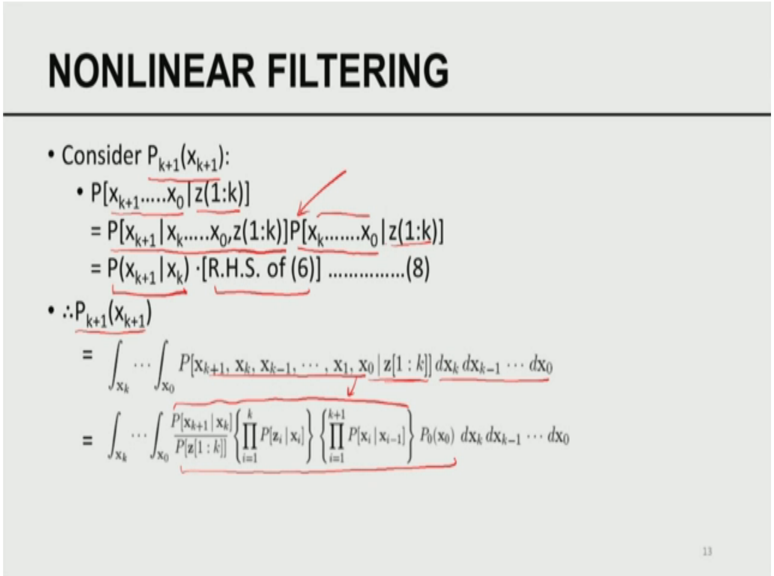
The filter density is simply the module density of this conditional density integrated with respect X naught 3 Xk minus 1. So, that is what this 1 is. So, this is integrated X minus 1 X minus 2 and X naught. So, the entire expression the for the filter density in full form is given by this look at this now. I am multiplying 1 over the probability of observing the k first k observation, 1 over probability of Z of 1 colon k, that comes from the denominator of the Bayes rule that was given in equation 3 the right hand side of the equation 3.

The numerator of the Bayes rule I have utilized the recursive property and broken down into several factors. So, one of them relates to one step transition probabilities now look at this now the structure is absolutely beautiful, this relates to the model transition probability, which is given by the Markov process. This is given by the conditional density of the observation, this is the initial density, initial condition if you want to call it and this is integration with respect to their the k time variable, that is what comes from here.

So, this is a this expression is the complex expression, it is, but it is easy to understand. So, the filter density what is the filter density? The density of the state at time k given all the observation is equal to 1 over the probability of observing all the observations times the integral the k fold integral along the path from 0 to k minus 1. And the integrand is the product of the initial density, the model one step transition probability and the conditional density the observation given the state. So, nothing could be more beautiful than this we know the conditional density, we know the state transition distribution, we also know the initial conditions what is the only thing we need to do? We need to have it all multiply them and that gives you an expression for the filter density. So, it is not that we cannot compute the filter density, it is simply that computation of this difficult conceptually disposable. So, this is one part of the solution of the non-linear filtering problem.

Now, I would like to come to the forecast density, please understand filter density is the analysis. So, we have done the analysis part.

(Refer Slide Time: 60:28)



Now, I would like to be able to do the forecast part. In Kalman filter equation, the forecast depends on the previous analysis the analysis depends of the previous forecast, it is this interdependency the forecast analysis that makes the sequential method a very beautiful and extremely effective.

So, now, let us try to compute the forecast density, what is them forecast density? Is the probability of given observation 1 to k, I would like to be able to explain not only what happened up to k, but also beyond k given Z, 1 to k what happened up to k that is filtered that is done. Now I want to know what is happening beyond, that is why this is called filtered density, again I am now talking like a broken record this conditional probability can be written again broken down by applying the conditional probability rule again to gain the recursive form.

So, this is equal to probability of Xk plus 1 having observed the stage from X naught to Xk and having observed the observation 1 to k times, the probability of being able to observe the state from 0 to k given the observations at 1 to k observation from 1 to k. I hope that transition is clear, it is simply a very simple probabilistic rule of trying to express the conditional density as a product of 2 other related conditional densities, it is a very simple mechanism.

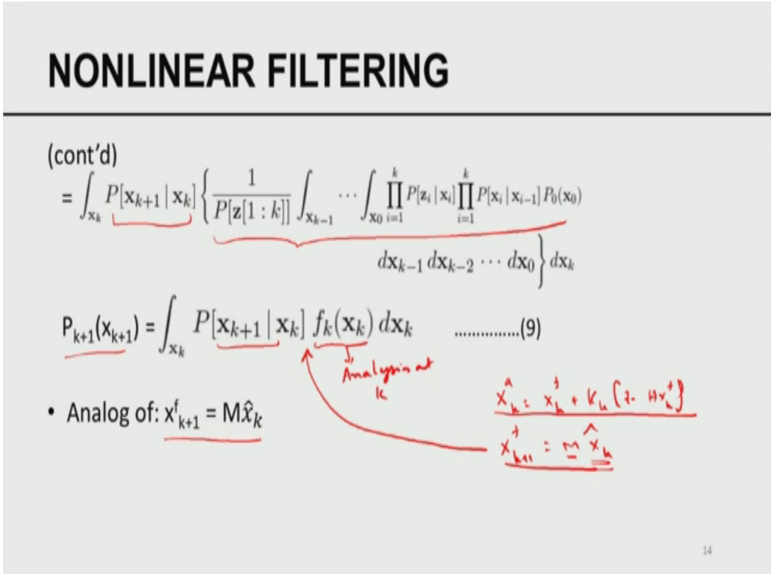
Now, let us come to the first term of this product term what is given? I am interested in the conditional density of Xk plus 1 different the entire trajectory from X naught Xk, and the observation from 1 to k, but the process case Xk is markov. So, once the Markov process Xk plus 1 depends only on Xk and nothing else nothing else matters therefore, the first factor depends the conditioning depends only on X k. So, the first factor reduces to probability of k plus 1 given Xk, that comes from the previous discussion that we have already had therefore, the second factor is the right hand side of 6. So, let us go back to the right hand side of 6.

The right hand side of 6 is the filter density. So, let us look at this now, I would like to a spend a minute on that. So, the probability density of Xk plus 1 given the entire trajectory is 0 through k and the observation times, probability given observation 1 to k and this now what is this part that is the filter density by 6. That is the right hand side of 6, therefore, by identifying this to be the filter density now I can express what I want. So, that is 8 is a beautiful expression for the forecast density.

So, the forecast density now can be written using the right hand side of 6 let us go back. The right hand side of 6 is given by this expression which is the k fold multiple integral. So, I am going to copy that multiple integral in here therefore, the forecast density is simply integral of the entire path given the observation and the integration is from 0 to that. And again I can decompose this from the previous argument to this integral, please understand we have already done that decomposition. Therefore, Pk plus 1 of Xk plus 1 is mouthful you can see it will probably take a 5 minutes to write this slide, I am trying to spend less than a minute on this you understand that, but all I am trying to do is nothing new it is simply manipulation of conditional probabilities that is all what it is.

So, except for the complications in these size of the expressions, the ideas are extremely simple. Therefore, by substituting this quantity from the previous slide I get this part that is integrand that is integrand hopefully that is clear to all of us.

(Refer Slide Time: 65:29)

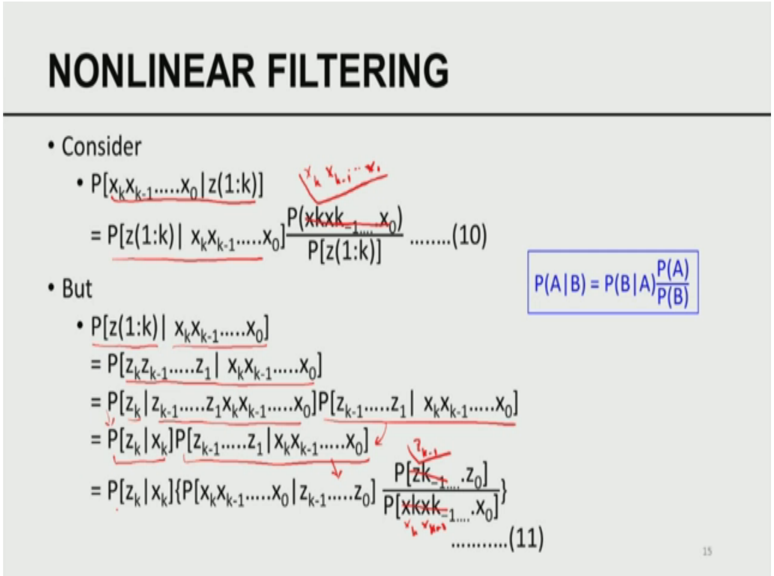


Now, I can rewrite that integrand into the product of conditional density, once of transition probability yeah. I can express this as a one-step transition probability times this quantity and that quantity is essentially fk Xk, and times one step 1 step transition a look at this this is beautiful what is that we have said? Analysis k is equal to forecast at time k plus the Kalman gain Vk minus Hf Xk f that is the a data assimilation step and to be able to get Xk plus 1 f is equal to this is M times Xk hat, that is the forecast equation this is the analysis equation we saw the embodiment of analysis equation previously with respect to the filter density this is the analog of this.

Look at this now the forecast of time k plus 1 in the linear case is model times the analysis time k. Now let us look at this here, f of k what is this this is an analysis time k why this color analysis time k this is not as vector that gives analysis this is the analysis density that is the filter density.

So, model operates on here model operates on analysis, here model operates on analysis density, what is the model one step type transition matrix therefore, if I multiply the analysis density with the 1 straight state transition probability and integrated over Xk I get this. So, I have already said this this is the analog of the forecast at time k plus 1 is equal to model times the for the analysis at time k Malay system time k.

(Refer Slide Time: 67:40)



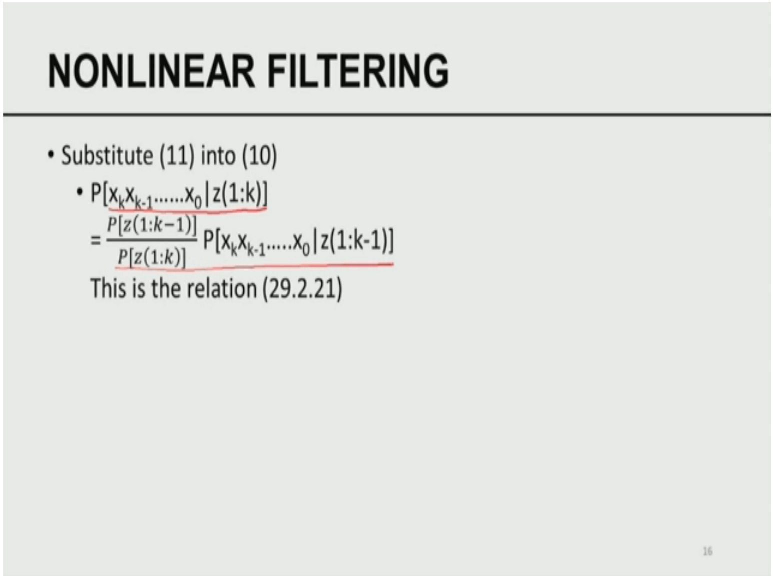
So, we have now considered both the equations.

So, let us l try to think of the non-linear filter is little bit more simplification is needed. If you consider this conditional probability, that can be written like this again by applying simple Bayes rule, this is a simple Bayes rule I am sorry this one I should have said Xk Xk minus 1, the k and X we are at the same level I am sorry for that this is product X naught. So, that is the numerator part of it.

So, this essentially for us from the Bayes rule, which is given in the right hand side, but Z the probability the observation given the state, again given by in long form given by this we already know probab the probability of observing k given the entire thing, depends can be written as a product of these 2 conditional densities, but this one depends only on the state at time k. So, that becomes the conditional probability distribution of the observation given Xk, and then we get the second quantity from here.

So, I get the recursive form, by again I can rewrite this in this form by the Bayes rule this must be Z k minus 1 sorry Z k minus 1, this again must be Xk , Xk minus 1 Xk minus 1 I applying the same Bayes rule here 11. So, you can readily see how I am able to compute the probability of observing the set of all observation given the trajectory in this particular form as given by 11, I am as given by 11 the substitute. So, now, look at this now.

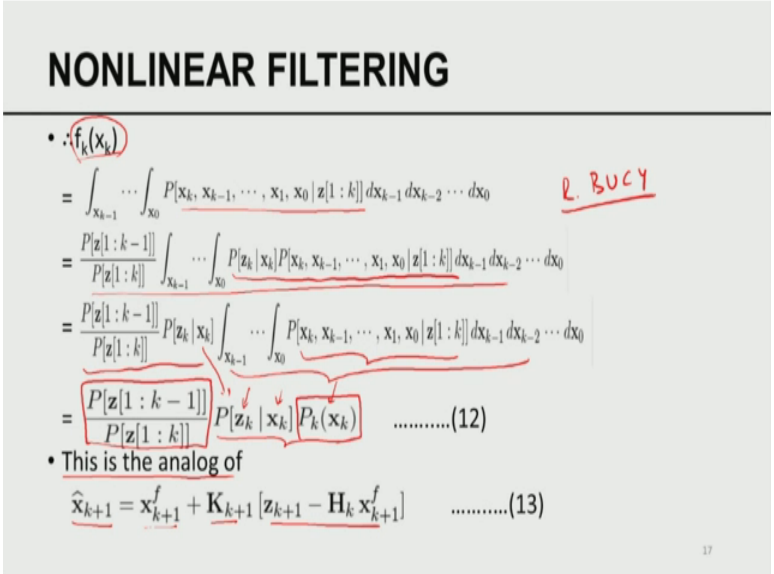
(Refer Slide Time: 69:43)



Substitute this 11 into 10, you then you essentially get what you want. So, this is the forecast sorry this is the filter density which can be written like this, which can be again written from substituting 11 into 10 we get this form, and this is exactly the relation given an in our book. I am sorry this is exactly the relation given in our book chapter 29 equation 21 and section 2. So, with this we have it readily seen the recurrence relation relating to f depends on P and P depends on f.

Let me go back to and say equation 9, the predictive density depends on the filter density in and the right hand side of 7 also depends can be rewritten as the predicted density. So, these 2 equation together gives you the expression for the non-linear filter in in in general term.

(Refer Slide Time: 71:10)



You can now I am going to further simplify f of k, Xk let me let me go back. You can see there are lots of things to be done this is the equation 7.

So, I am going to substitute 8, 9 and 10 and 11 to get back to the forecast, I am sorry the filter density coming back the filter density. So, the filter density again from fundamental principles is given by integral of this, and that can be written by using Bayes rule by this. Then in the previous slide in equation M 10 and 11, we have broken this down by applying Bayes rule that gave rise to the factor like this, which comes out of the integral and this is what I have within the integral I integrate this, I get an expression which is probability of Z k given Xk times probability of P k X k.

The probability of Pk Xk, this probability of Pk Xk comes from here. So, the probability this integral in the in it is entirety gives raise to Pk Xk, now look at this this is the analog of the analysis. The analysis of time k plus 1 is equal to forecast at time k plus on Kalman filter times this, what is the analog coming in here? This is the filter density, this filter density depends on the forecast density I am sorry this is the filter density depends on the forecast density and the observation. So, this is the density the observations and that gives you the recursive form, and this is analog of the an analysis stuff. This is analog of the analysis stuff yes this is easily said than done, I hope you are able to keep track of the all the major you are able to keep track of all the major issues in here, and that is the relation that relates to the forecast with the filter the forecast of the filter.

I am I hope it is it is clear. So, substituting again 11 in 10 you can see this relation is again given by the 29. 2.1 in our book and that is the important relation that relates the particular density with the filter density.

So, in summary what is that we have accomplished? Yes I know I may have some of you might think that I may have gone a little fast, but again this is an advanced course, in this course we are we will not be able to hang carry you and show you every little step, but we have shown all the basic major steps going from one step to another step is largely part of the exercise, I hope you will be able to pursue, but it this will provide you a good big picture modulo some of the algebra, I hope with this you are able to see the relation. So, 13 I am sorry 12, 12relates the forecast to the filter, 9 relates the filter to the forecast.

So, 9 is the model forecast step, 12 is the data assimilation step, you can readily see the data assimilation step let us spend 1 or 2 more minutes on this the filter density at time came is the predictor density at time k. If Xk is known I can conditioned on Xk, I can conditioned on Xk, then this is the conditional density of observation and once I have conditional density observation the this ratio comes is again as a multiplying factor which is essentially meant to induce that the density is that the condition for the densities observed what is the condition for the density? The integrals must we want.

So, you can think of this as simply a multiplying constant and 1 of those this is the ratio of the probabilities of observing the observations, yes the program the expressions look little complex, but the basic idea of going from forecast to analysis, analysis is the forecaster analysis in the function space must be clear. It is this iteration in the function space which is an infinite dimensional space, which makes this iterative scheme impractical.

Except for in very simple cases, what are the simple cases linear Gaussian chronic is one case where I can implement this because it reduces to updating the mean of the covariances. In the literature they have identified a few handful of other cases combinations of non-linear systems and associated noise, where they could explicitly express these integrals in closed form.

So, other than these elementary cases, these equations in general are not easy to compute and hence the difficulty of non-linear filtering. I want to reemphasized, it is not that we do not know how to do non-linear filtering; this has been done way back in the midsixties. Our derivation depends on the development in bucys buck our development depends on the book by Bucy I did not spell it correctly sorry our Richard Bucy. The original paper was be Kalman the second paper was Kalman and Bucy Kalman originally derived the filter in discrete time, at the same time bucy was also deriving the Kalman filter in continuous time, when bucy submitted the paper Kalman's first paper was already under we had has been accepted.

So, the reviewers asked both of them to get together and publish a common paper. So, first is Kalman second is Kalman Bucy and Bucy has been working in non-linear filtering ever since nineteen late fifties early sixties, and the derivation that we had given here is adapted from Bucy's papers and in a monograph he wrote. So, this essentially meant to provide you the idea that filtering problem what is filtering problem? The data the sequential data assimilation problem in a non-linear system is solved theoretically, but not computationally that is the story.

So, far we had concentrated on deriving the filter equation and the predictor equation on the function space, these are infinite dimensional in nature computationally extremely demanding. The next question is even if I spent lot of effort to be able to get the entire distribution, from the forecaster perspective what kind of forecast product I have to develop from these probability distributions. If you think about it for a moment more often than not, we are used to interpreting the mean we have a reasonably good interpretation of the variance, I do not know what would mean to a public consumption if I say the third moment is this the fourth moment is this.

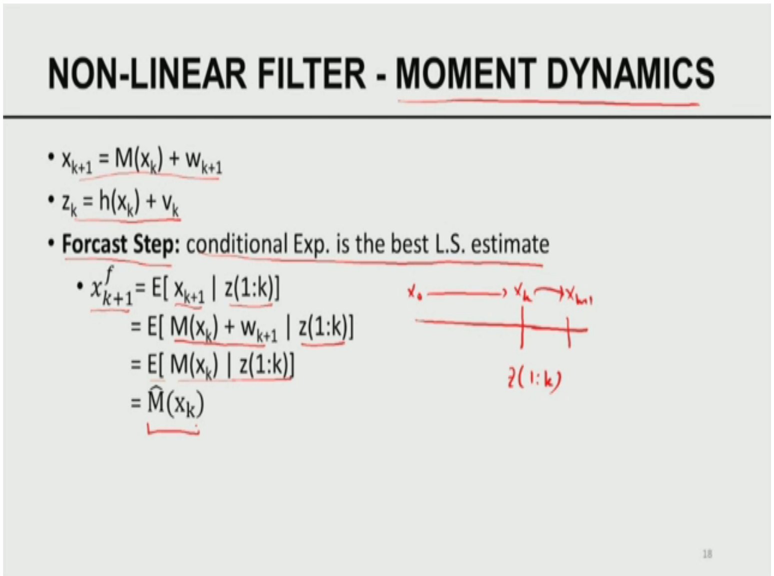
So, third moment relates to skewness of the distribution, fourth moment is called kurtosis. Skewness of the distribution essentially tells you the mean and the mold may be different or the whole thing can be tilted to one way or the other the kurtosis if the kurtosis was large what does that tell you? The tails are thick the kurtosis for the normal distribution is 3 is step 4 the standard normal distribution is 3, what is meant by saying the kurtosis is larger? The kurtosis larger statistically me implies that the probability mass for very large values of the state variable are larger. If the probability match for very large values of a state variables large means what? High impact events could occur with a larger probability that is what kurtosis larger kurtosis means that is called tail of the probability distributions.

So, if you are trying to develop a forecast product generally we can only process the first moment second moment, third moment I am not sure how we use it in our interpretation of events that could occur. Fourth moment if you say kurtosis is more than 3, I am not sure it is very easy to interpret the likelihood of rare events happening.

So, kurtosis means kurtosis is large means rare events can occur much more frequently than smaller kurtosis that is what it means. So, in principle larger kurtosis means the potential for extremely rare events to happen with a higher frequency, that is all what it means. So, looking from our ability as well as the usage of statistical quantity to interpret random phenomena in nature, we generally settle down our being a settle down on being able to predict the first moment which is the mean the second moment or the second centered moment which is the variance.

So, from that perspective while we have in principle derived expressions for the update of the filter equations or the predictor equation in the n dimensional space, more often than not we are interested simply in moment dynamics. What is moment dynamics, how the mean update themselves are evolve how the variance evolves. Please go back the Kalman filtering is essentially dynamics of the first 2 moments mean of the covariance, and that fits everything we generally know how to do in statistics and how to interpret in statistics.

(Refer Slide Time: 82:37)



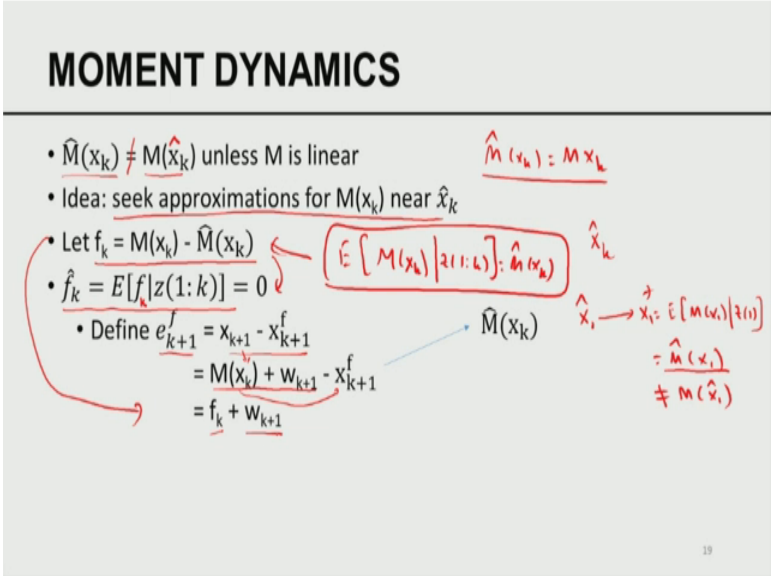
So, given these we are now going to look at how to derive the moment dynamics from the dynamics of probability density functions that is what we are after now. So, consider your non-linear dynamics given by this, consider non-linear observation given by this the forecast step what is the forecast step? The conditional expectation of the best linear estimate; I hope it is clear from our discussion of the statistics, conditional expectation is the best estimate.

The conditional expectation is the best linear estimate and we are now going to be looking at what is the best way to be able to compute this conditional expectation of the forecast.

So, the forecast at k plus 1 is equal to expected value of X of k plus 1 given Z at 1 to k. Z at Z from 1 to k means what? I have I have been given all the observations from Z 1 to Z k, but and I am I also no Xk, I also know X naught I know every state I would like to be able to predict Xk plus 1. So, that is. So, given all the information Xk as well as Zk, I would like to be able to predict Xk plus 1, and we have already seen in the derivation of the Kalman filter equation forecast the best estimate for the forecast is the conditional expectation of the state given all the observation.

This expectation is taken with respect to the predictor density, but Xk plus 1 from the model is given by this conditioned on that. Conditioned on observation 1 to k the expectation value of Wk plus 1 is 0 therefore, it reduces to the conditional distribution of the non-linear value of X of k given X 1 to k. This conditional density I am now going to call as M hat of Xk. M hat of Xk please understand evaluating this conditional expectation is not easy, but such a conditional expectation exists, I am going to call it M hat Xk what is the M hat Xk? Into the average of the value of the state passed through the non-linear map given all the observation, the expectation is with respect to the predicted density.

(Refer Slide Time: 85:11)



So, it is it you can really see M hat of Xk is equal to M hat of Xk is not equal to M of Xk unless M is linear, only in the case of linear linearity M of M hat of Xk which is equal to M of Xk that is the linear case.

So, in general the expectation that we got in the previous step the condition expectation in that we regard in the previous step is not equal to M of xk. So, what is the idea? I am trying to seek approximations to the conditional moment. So, what is the basic idea here? I have analysis X hat k which is. So, let us pretend I have an analysis X hat k I am going to be I am approximating M hat of Xk around X hat k.

So, let us look at this what is what is the idea here? Suppose I have X naught hat initial condition I know the analysis I am going to be able to make a prediction I would like to be able to make a prediction X 1 f and what is that this is equal to e of mf X 1 given Z 1, and that is equal to M hat of X 1 then general this is not equal to M of X hat.

So, I should have put yeah thing in here therefore, it is very difficult to compute M hat of X 1 because is a is a conditional is an integral relating to the conditional expectation. So, I can only approximate. So, what is that we are going to approximate M hat, the small neighborhood around X hat of k. So, what is X hat of k? X hat of k is an approximate analysis known at time k, I am going to approximate my forecast around that.

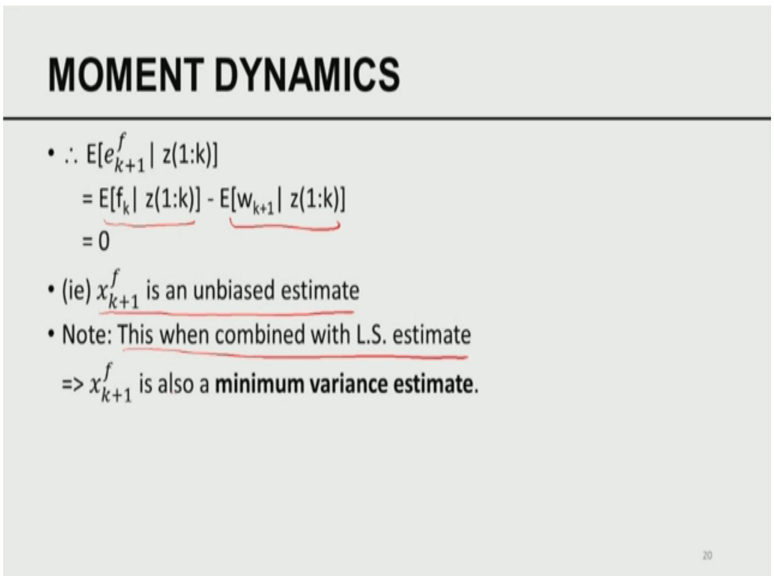
So, we seek an approximation of M of Xk near Xk hat. So, f of k is equal to M f Xk minus. So, what is that, that is the that is the error mf Xk is the actual value mf X is the expected value. So, it can think of that as an anomaly. F k hat is the conditional expectation you can see the hat opposes the conditional expectation. So, this is going to be this is fk. So, a conditional expectation of a k fk f of k given the all the observation that is 0, that you can readily see from the definition of f of k, because if you took the condition expectations of M of Xk given Z one colon k is equal to M hat of Xk please remember that is the definition. And that immediately when applied to this immediately implies this, then apply to this immediately implies that therefore, the error the anomaly has expected conditional expected value 0.

I am now going to define the forecast error which is given by this, the forecast error is equal to Xk plus 1 is given by this, that is a forecast from our definition f of k this is the first term that if we combine these 2 that is f of k therefore, if I use my definition of f of k, ek plus 1 is equal to f is equal to f of k plus w k plus 1 that is the forecast error.

So, you can write readily see how the approximation start building up. I have f of k I have e of k plus 1 that is called a forecast to error, the forecast to error has 2 terms 1 due to f of k another due to Wk plus 1. This is the forecast error term which is very similar to what we have in the linear case.

(Refer Slide Time: 89:46)

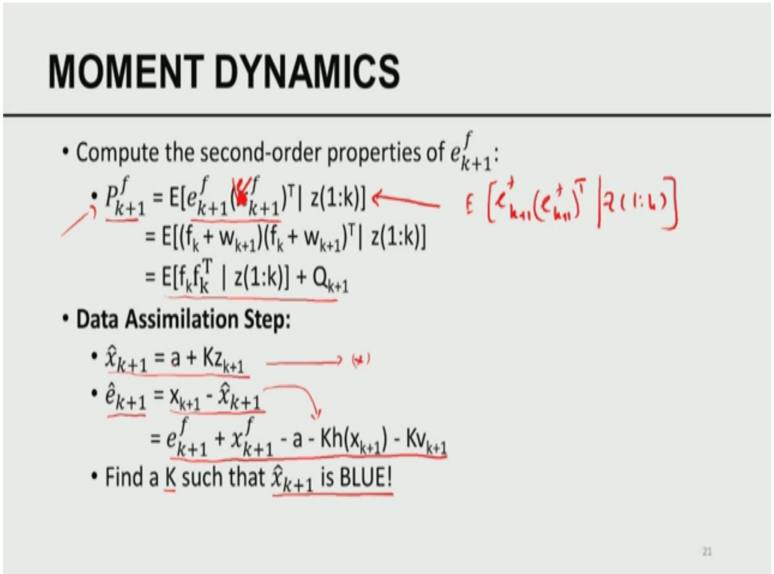
.



So, if I consider the X conditional expected value of e k plus 1 of f given this, I can now substitute ek plus 1which gives rest these 2 terms; the first term is 0 because of the definition of f of k, the second term is 0 because of the definition of the noise therefore, e k plus 1 f is unbiased. So, therefore, Xk plus 1 is an unbiased estimate, this when combined with the least square estimate it becomes the minimum variance estimation.

Therefore Xk plus 1 f is also a minimum variance estimation in our based on the basic statistical information we have created.

(Refer Slide Time: 90:30)



So, compute the second order approximation properties of ek plus 1 f, second order properties relates to the covariance structure. So, the covariance structure Pk plus 1 that is equal to this must be this must be e.

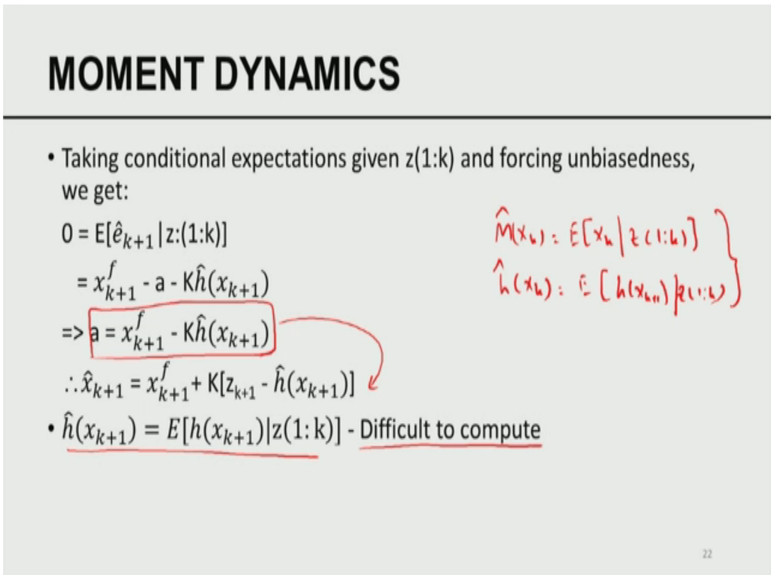
So, this must be ek plus 1, f times e k plus 1 f transpose given Z 1 to k expected value that is the expression. Ek plus 1 is the f of k plus Wk plus 1, f of k plus Wk plus 1 transpose if you multiply both of them f of k and Wk plus 1 are uncorrelated because f of k depends only one up to time k, Wk plus 1 is what happens after time k. So, this reduces to this equation. So, you can think of Pk plus 1 that is the second moment you can think of. So, let us look at what is the we have accomplished.

We have a forecast look at this now we have a forecast in we have a forecast in page 18 you have the forecast covariance approximation in page, I am sorry in the at this stage it is not approximation we have a computed everything reasonably exactly, but we are going to later see fk is not easy to handle because fk has M bar M bar has to be approximated therefore, , but at least in principle this is the forecast covariance in other words, I am trying to derive the general expressions assuming everything is possible without worrying about computational issues right now.

So, the data simulation step now can again be given by Xk plus 1, I am trying to do what I have did in the case of a linear case I am going to do the derivation from the scratch ground up. So, if I have I am going to make my analysis depends on a plus k times zk plus 1, you may remember in 1 of the earlier discussions of statistical estimation, this is the structure of the linear estimation a is the vector k is the matrix. So, this is the an. So, you can express the analysis of the linear function of the observation where a and k have to be determined to make the analysis, um unbiased and also analysis of minimum variance.

So, with that in mind I have e bar, I am sorry e hat of k plus 1 is equal to the basic definition in here which is given by this equation which is given by this equation. So, this is the equation for Xk plus1. So, my job is to be able to find a k such that this is a blue X hat k plus 1 is a blue. So, if I did that sorry.

(Refer Slide Time: 99:38)



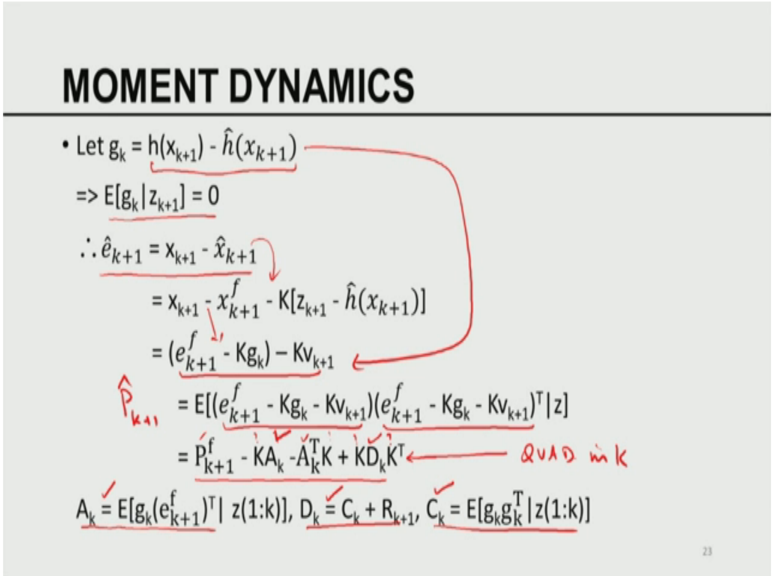
I am going to take the conditional expectation on both sides forcing unbiasness that gives rise to the value of a.

Now, please understand I have k hat h hat h hat essentially comes from the fact that I have been given the expression for the update. So, this is the expression for the update. So, I have a computed according to this relation sorry I have a computed according to this relation therefore, if I substitute at this a in this expression which is in this expression sorry in this expression in page in in page 21. So, this is the expression I am going to compute I have I have already computed a I am going to substitute a in here in star if I did that, I am going to get a structure for a given by this. This been substituted in the previous expression gave the structure you can see this is the typical update from Kalman.

So, what is h bar? H bar is given by the conditional expectation, please understand this is as difficult to compute as M bar Xk, h bar Xk this is equal to expected value of Xk let us go back I want to be able to remind you where it was. So, look at this in page eighteen is given by the conditional expectation of the conditional expectation of the conditional expectation of M of Xk given observations one through k and that is exactly what we are going to sorry oh did I that is right this is this is conditional expectation given Z 1 colon k likewise for h, h of Xk bar h of Xk bar in this case is equal to e of e of h of Xk plus 1 given Z 1 colon k 1 colon k and these are the 2 difficult quantities to compute.

So, even though these are difficult quantities to compute, but we know such quantity exists mathematically. So, we have derived the underlying expressions following the derivations of the linear Kalman filter, assuming all the calm of complicated integrals can be evaluated.

(Refer Slide Time: 96:31)



Now I am going to derive the moment dynamics let Gk be the difference hk, h of Xk plus 1 minus h of Xk plus 1 hat we already know from this definition if I take the conditional expectations of both sides that is 0 therefore, this the analysis error is given by Xk plus 1 minus X hat of k plus 1 I already know the structure of the X hat k plus 1 using the Kalman filter equation.

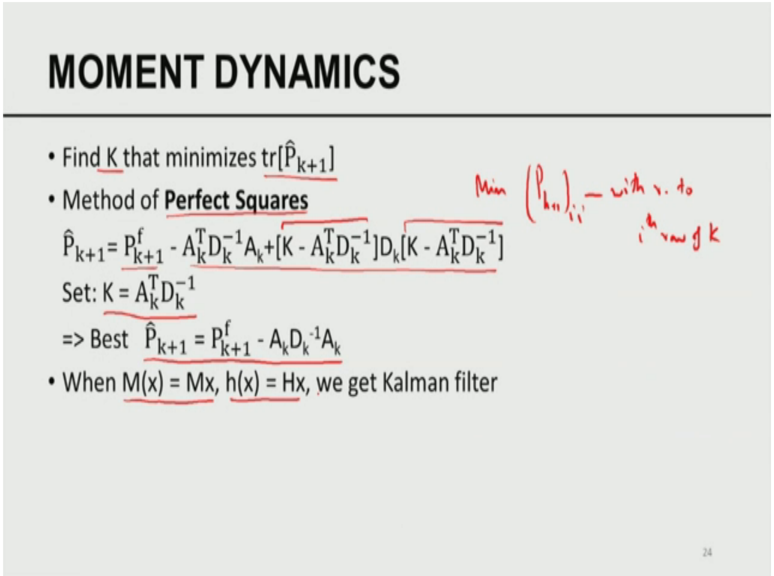
So, this can be rewritten using the definition of Gk, this can be rewritten as this when combined with this equation. So, ek plus 1 look at this now the analysis error is equal to forecast error minus correction Gk is a random quantity Vk is a random quantity Gk and Vk are random quantity I want to be able to compute the analysis covariance in this case this must be I think the left hand side must be P k plus 1 hat. Pk plus 1 hat is equal to if this is the analysis error expression this times it is transpose conditioned on Z is going to be Pk plus 1 hat. If I multiply these 2 after I do lot of algebra I get this expression you can readily see in this expression I have Ak is a matrix Ak is a matrix given by gk ek plus 1 f conditional expectation on Z 1 to k Dk is given by Ck plus Rk plus 1 and Ck is given by gk gk transpose again conditioned on that.

So, there is a lot of lot of notations in here I can compute I have an expression for Ak I have an expression for Dk I have an expression for Ck. So, if I look at this expression I know I know Dk I know Ak I know Pk plus 1 the only thing I do not know is k I am sorry I know Ak, but I do not know k sorry. So, my job is to be able to find k such that.

So, this this d is known I want to be able to find k such that. So, Ak is known Dk is known pkf is known k is not known.

So, I am going back to the old homework how do I make the trace of Pk plus 1 minimum to the appropriate choice of k, that is the minimization problem that we have already involved in this expression is a quadratic in k. So, this gives rise to a quadratic minimization problem. So, giving this quadratic minimization problem I can I need to find k go back. I need to find k such that it minimizes a trace of P hat k plus 1 which was given in the previous page.

(Refer Slide Time: 99:30)



So, by method of perfecting the squares again I am going back to the exercise as I have done earlier in the context of Goss to Kalman I am doing exactly the same thing the mathematics are absolutely similar. So, I am trying to rewrite the equation for Pk plus 1 hat this is the method of perfecting the perfect square, the method of perfecting the square if I want to be able to now look at the expression on the right hand side this is another method for minimizing.

In the earlier case what did we do we computed we minimized the i th term the i th term in other words we minimized Pk plus 1 i with respect to a given row of k, minimize with respect to the i th row in here I am in demonstrating another basic principle we are simply trying to express the previous quadratic expression as the product of these 2 this is called method of perfecting the square. So, this is 1 term this is the another term d is known.

Now look at the structure now Pkf is the sum of 3 terms, first term is known is that is independent of k second term is known that is independent of k, the third term is known it depends on k. So, if some term does not depend on k I cannot choose k to be able to change it. So, the only way to be able to affect k is to make those terms that are dependent on k 0 therefore, by picking k is equal to a k transpose Dk inverse I can make this quadratic term to vanish in that case my best or optimal covariance is given by this expression where Ak is the matrix already known Dk is the matrix that is already known.

I would like to emphasize the following fact that this derivation that we had given is the generalization of the linear filter derivation if M. M of X is equal to M times X is h of X is equal to h times X, this derivation I had given in the past 5 6 slides reduces to the derivation of the Kalman filter, the moment dynamics for the Kalman filter. So, by camouflaging the difficulty in computing certain conditional expectation and we are able to derive the moment dynamics the first moment dynamics and second moment dynamics the mean and the covariance forecast mean forecast covariance analysis mean analysis covariance.

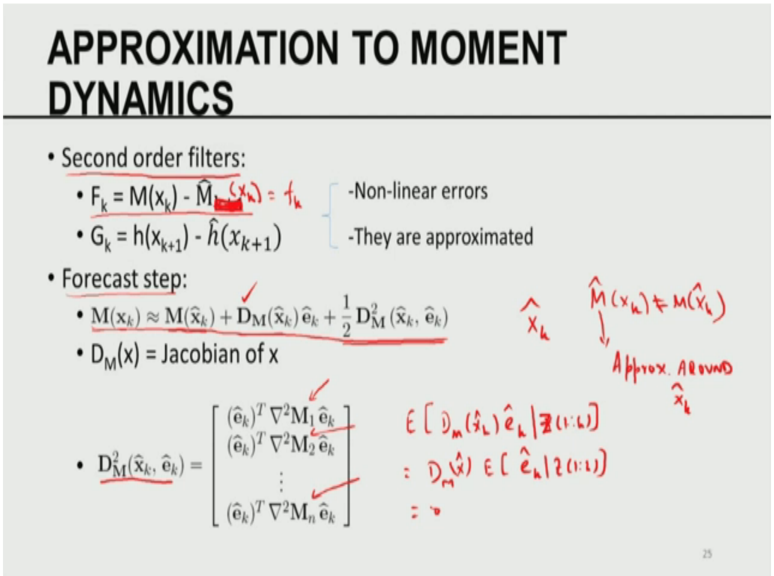
Pushing into the background the details of the or the difficulties of the conditional expectation computation, but giving them a name, they I we know that exists by giving them a name I can I do not have to worry about the computability at this time I simply can carry on the derivation. So, we have completed the derivation of the dynamics of the first 2 moments in any general non-linear filtering equation, which parallels the development in the linear carbon case, and how do we know if this parallels the development to the linear carbon case if you set M of X is equal to M times X, h of X is equal to h times X our derivation essentially reduces to the Kalman filter equation.

So, it is in that sense there is nesting, it is in that sense it is a parallel derivation of the filter equation especially the moment dynamics for the non-linear case, I hope this part is clear. So, in the earlier part we talked about the updating of the updating of the distributions now we have talked about updating of the first moment and second moment assuming such filter density assuming such prediction density exists they do we may not know it exactly, but I can handle it mathematically what we have done that is what we have done.

And this derivation again parallels the development of the linear minimum variance estimation it essentially rests on the fundamental principle, the conditional expectation is the best mean square estimate. So, that is the fundamental statistical fact the whole derivation rests on I think it is in sense it is a it is a unification of the derivation of moments both in the non-linear case as well as the linear case. I hope the reader will appreciate the parallels and the role of conditional expectations and so on.

With this as a background now I am going to consider specific approximations. So, that gives rise to approximation to moment dynamics. So, until now the moment dynamics I have considered are exactly in the sense are exactly in the sense even though I do not know how to compute them such a thing exists.

(Refer Slide Time: 104:59)



Let me proud through I got what I want, but once you realize there are certain quantities which cannot be actually computed we begin approximation, when we start doing approximations we get the notion of approximate moment dynamics.

So, approximate moment dynamics there are several degrees of approximation first order approximation depends only on first order Taylor series expansion of non-linear quantities, second order approximation rests on a second order Taylor series expansion of non-linear a conditional expectations. So, first I am going to derive approximations from the second order filter, what is the second order filter? The filter equations are approximate, but they are approximate up to the second order term, the second order term in the use of Taylor series up in the approximations.

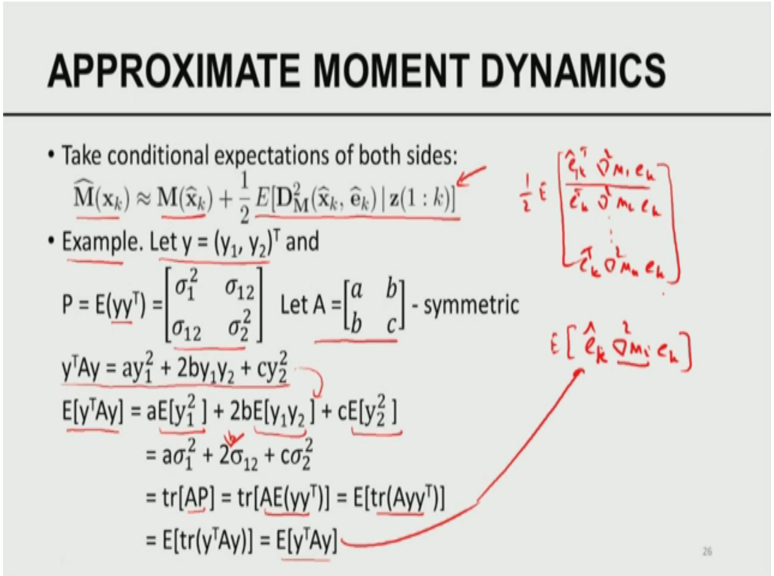
Now, you can see wherever there is approximation Taylor always comes to our rescue if you use Taylor what is the advantage I can cut the approximation at any order of accuracy in in all practices we generally are able to handle the first order accuracy the second order accuracy, because that is what mostly used in practice. So, to derive the non-linear errors we are going to approximate them. So, what are the non-linear error terms go back f of k is equal to I think earlier, we had we had denoted as lowercase f of k let me go back and talk to you about it in here in page 9 in slide 19 f of k is equal to M of Xk minus M hat of Xk that is the same expression that is given in here too.

So, I am seeking second order approximations look back even though I have call it capital F of k we have earlier defined this to be we had earlier defined this to be defined this to be f of k Gk is again the same kind of thing which is h, but at time k plus 1. So, I am now going to be concerned the forecast step M of X k. So, what is that I am trying to do I am assuming I know X hat k I also know M of M hat of Xk is not equal to M of Xk hat.

So, I am going to approximate M hat of I am going to approximate this in the neighborhood of in the neighborhood of X bar k. So, that is what I am trying to do now. So, M of Xk according to the second order Taylor series is M of Xk bar. So, I am trying to do everything around approximation around k hat. So, I have a Jacobean times the error I have this second order term in here where DM is the Jacobean and this is the DM, DMis the vector that depends on the hessian term. So, you can see this is the quadratic pop with respect to the hessian of M 1 quadratic form with respect to hessian of M 2 quadratic form with respect to hessian of mn.

We have already seen these things in a module on multivariate calculus how to have secondary Taylor’s expansion for maps. So, it is essentially comes from 1 of the early um slides on multivariate calculus now I would like to be able to take the conditional expectations on both sides I have to take the conditional expectations of both sides of this equation given.

(Refer Slide Time: 109:01)



Z k the condition expectations on both side given Z k is given by M bar of Xk is given by M bar of Xk and that is equal.

So, by taking the conditional expectation on both sides of this, I get the first 2 term I get the first term I get the expected value of the third term. Now if you go back to the second term consists of the Jacobean at Xk bar and ek. So, if I take the expectation of Dm Xk hat ek hat given a given observation Z, 1 to k that is equal to Dm Xk hat times e of Xk hat given Z 1 k and we have already shown that is 0 that this is 0.

So, in view of that even though there are 3 terms of the right hand side if I take the conditional expectation I get only 2, now I am going to give you a little example to be able to illustrate these calculations. So, let y be equal to y 1 y 2, let the like the covariance of y is given by expected value of yy transpose, which is given by this is assumed the given by this matrix let us also assume a is the matrix which is symmetric.

So, y is the random vector with this covariance matrix, a is a symmetric matrix I am now considering possible different quadratic form which is y transpose Ay, I am trying to compute the expected value of this possible different quadratic form and that by substituting this in here is the sum of 3 expectations because expectation of the sum of the sum of the expectations expectation of y 1 is sigma 1 square expectation of y 1,y 2 is sigma this must be 2 b sigma 1 2 plus c times sigma 2 square. It can be verified this is simply the trace of the matrix Ap trace of the matrix Ap is can be written as a times e of expected y transpose y.

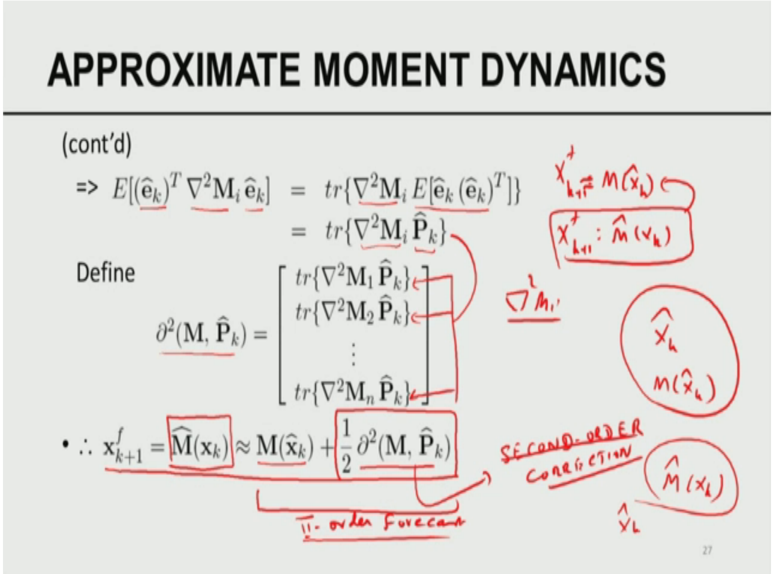
The trace of the expectation of the trays they commute therefore, this is equal to expectation of the trace of a yy transpose, the trace remains invariant under the cyclic permutation therefore, trace is equal to y transpose ay y transpose Ay is a scalar trace of a scalar is itself. So, that is equal to expected value of that therefore, we have come one circle around that tells you the details of the calculations with respect to compute in the expected value of this, we are we interested in this computation let us go back to this term.

What is this term from the previous slide this term is a vector. Each component of this is a quadratic form the delta square M 1 delta square M 2 delta square mn they are all hessian matrices they are symmetric. So, y plays the role of ek, a plays the role of the hessian therefore, expectation of this. So, the I am sorry therefore, the expectation of the vector of hessians are this vector of scalar product, let me let me talk about that once more sorry.

So, what is this expectation has got on it? This is 1 half of expected value of e 1 ek transpose del square M 1 ek, ek transpose del square M 2 ek all they are up to ek transpose del square mn ek. If I take the expectation of a vector that is equal to expectation of the individual component of it, expectation of the individual components which is of the form e of e k delta square mi ek and this is an hessian matrix this looks like this looks like this term that is why that particular term is this particular example is very meaningful I hope I hope the relations are very clear.

I am just trying to give this example to be able to manipulate the expected value of the second order term in the Taylor series expansion, e that is the vector each of the component is a quadratic form. So, if I know how to compute the expectations of quadratic form I can compute the expectations of individual elements of this vector, and hence this example provides you a handle on how to compute the conditional expectation and the second term. In the first equation on the top of the slide twenty 6 hope that is clear yes it is mouthful it takes 5 more than 5 minutes to be able to type this, but I am trying to spend less than half a minute, but you know the basic steps.

(Refer Slide Time: 114:31)



So, using that example I am interested in computing the quadratic form. So, this is the row vector, this the matrix, this is the column vector. So, from the previous example I can say this is equal to a trace. So, from the previous example what is the formula here expected value the quadratic form is equal to the trace of 8 times expected value of yy transpose, that is the equation I am trying to use here therefore, this is this is a, this is this is the matrix a, this is the expected value the expected value of this Pk.

So, this is the hessian this is the covariance matrix P hat. So, we have computed the expected value of the term on the right hand side of the first equation. So, with that I am going to derive a vector a vector of second order corrections. So, this is the vector of second order corrections, each of the terms are induced by term type the terms of this type the middle term was already 0 therefore, the forecast using the second order approximation is equal to is exact values M hat, it is approximate values M of Xk bar plus the second order correction that is the real kicker.

So, this is called the second order correction to the forecast let me write that down. You can see you learn a lot of probability manipulations, when you when you do these kinds of computation that further helps you to visualize the power of the statistical arguments and the interaction between statistics and matrix vector manipulations I would like to anticipate. Suppose I do not consider the second order approximation I only consider the first order approximation the delta square will be 0.

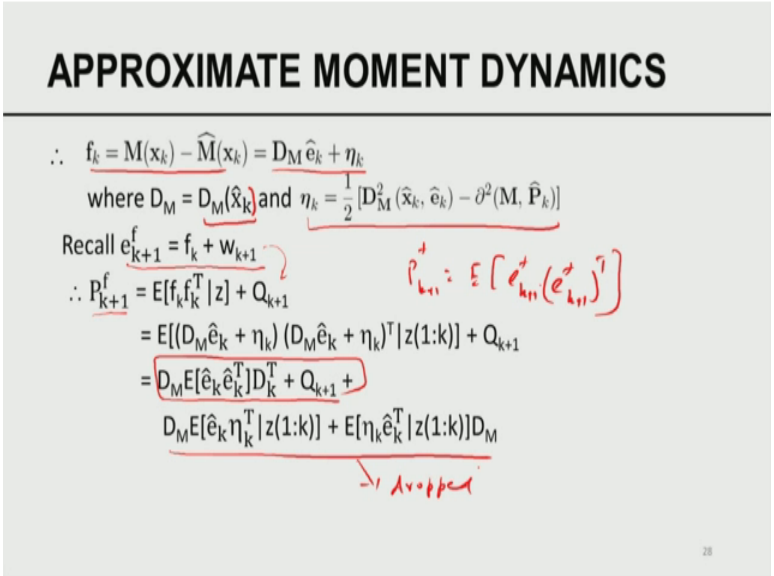
So, you can readily see if you make a first order forecast that is essentially the first order forecast is essentially equal to M of Xk hat. This is the approximation the actual value of the forecast is Xk plus 1 f is equal to M hat of xk. So, I am trying to replace this by this that is the start in the second order, what do I do I add the second order term therefore, second order forecast must be more accurate than the first order forecast.

Where do you when do you consider what is the right order for approximation, it depends on the degree of non-linearity. Now if you go back what does the second order correction term depends look at this. Now delta square mi is the hessian of the I th component of the model map if the model is mildly linear I am sorry mildly non-linear the second derivative may not be too high, in that case you can you can essentially get away with first order approximation.

If your model is such that the second derivative the hessian of each of the component is strong first order approximation would not cut it for making forecasts second order approximation is more meaningful. So, this essentially tells you by appropriately controlling the order of terms in the Taylor series expansion, I can improve the accuracy or the forecast. So, this is second order accurate forecast. So, this is the second order forecast. I also want to may remind you first order forecast is not easier while first order forecast is not easier I already know Xk hat from the previous step, I simply be able to evaluate the function map at the previous step.

So, that is the start, but to be able to compute the second order I have to do lot more arithmetic. So, a second order forecast is definitely more accurate, but it is computationally more expensive. So, what does it bring you it brings you the accuracy versus time trade off. So, anybody who is involved in approximating any quantity at the degree of approximation and the degree of approximation related to the quality of approximation, the cost of computing the approximation is will be the ultimate judge in trying to decide the order that we will feel comfortable with. So, using the derivation that parallels the linear Kalman filter by trying to approximate M hat of by trying to approximate M hat of Xk around X hat k we have tried to arrive at a second order correct forecast.

(Refer Slide Time: 119:54)

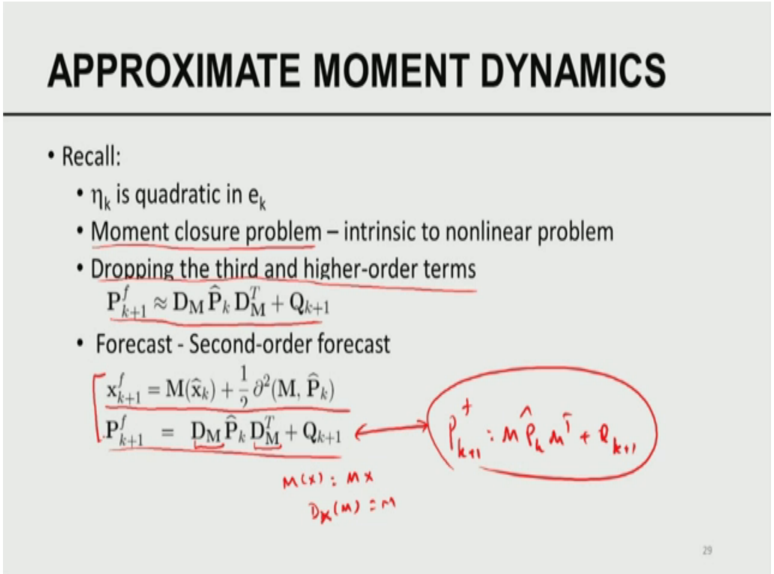


Now, if I am going to do a second order correct forecast, what is it is covariance. So, that is the level now I had talked about the covariance in order to be able to compute the covariance again I am going to go back to the error in this forecast. The error in the second order forecast can be written like this where DM is the model Jacobean which we already know eta k comes from the second order term eta k comes on the second order term we already know the error is equal to f of k plus Wk plus 1 therefore, the forecast covariance is equal to from here this equation follows very easily, the forecast covariance what is the forecast covariance Pk plus 1 f is equal to expected value of ek plus 1, f times ek plus 1 f to the power f transpose.

So, that is the that is the formula. So, when I apply this formula using this because f of k and Wk plus 1 are not correlated it reduces to 2 terms the cost and vanish the f of k is given by the expression that we have seen earlier. So, if I substitute all these things and bulldoze all the details I get these terms. So, this is the expression for the forecast error covariance when the forecast is second order accurate.

Now, let us take some time to be able to compute all the expectations all the conditional expectations in here.

(Refer Slide Time: 121:39)



Eta k is quadratic in e k. So, now, let us look at this now. This is the most important part of the whole step sorry. So, this is where we are eta k is quadratic in ek, this gives a this gives what is called the moment closure problem the this gives us to what is the moment closure problem.

So, what is the moment closure problem, if I want to compute the second moment if that depends on higher order moment if I want to be able to compute the first moment, that depends on the higher order moment. So, let us look back in here Xk plus 1 f that is the conditional expectation for the forecast. So, the conditional forecast depends on the second moment. So, first moment depends on the second moment, second moment depends on third moment fourth moment.

So, what does that tell you, you cannot compute these moments in a closed form because each 1 lower moment depends on the higher moment this provides the computational difficulty that difficulty has been around for a long time in all non-linear problem especially in turbulence they always deal with this problem, which is called moment closure problem. So, what does it mean? If a second order moment depends on the previous second order term in the higher order term they will simply approximate by dropping the higher order terms that is what is called a simpler moment closure approximation problem.

The same moment closure approximation problem comes in here. So, dropping please understand I am trying to compute Pk plus 1 f Pk plus 1 a second moment where are the third moment terms comes in. So, let us look at this now this term ek plus 1 times eta k eta k it depends on the second order term ek depends of the first order term the product of ek hat and eta k, third order term. Therefore, you can readily see the moment closure problem coming and derailing our ambition to be able to improve the accuracy. So, what is our aim our aim must be able to use second order term to improve the accuracy for the first moment which you have already accomplished.

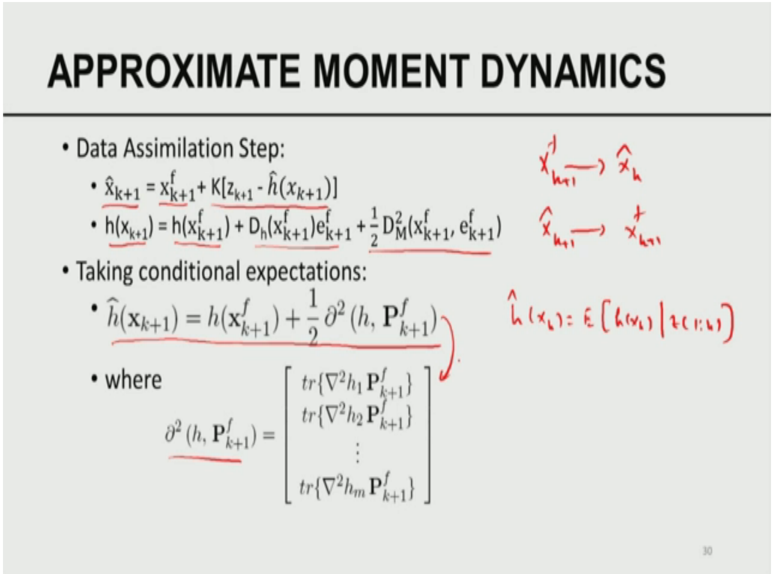
Now, using that approximation, for the first moment I am trying to develop an approximation for the second moment, but the sec expression for the second moment depends on second moments of other quantities and third moments of related quantities. So, I am going when I am trying to compute if I need third moment if I do not know the third moment I can complete the second moment. So, that is the issue in here. So, what are we going to do approximation ideas are very clear, moment cursor problem this is intrinsic to non-linear problem and how do we tackle this we simply close our eyes and drop all those terms that we do not know.

So, in trying to get a second order approximation, drop the third order and higher order approximation that actually leaves, that essentially leaves only these 2 terms I am sorry that essentially leaves with only these 2 terms the following terms are dropped . So, once I drop that, my approximation to the second order I am sorry second order approximation to the second moment is given by this. So, there are 2 things to do consider 1, what is the moment that is being approximated second what is the order of approximation.

So, first moment depends on certain second order term second moment, dependent on higher order terms. So, the moment closure problem shows ugly face. So, by considering simple solution to this moment closure problem by dropping the third degree and higher degree term we get the expression for the forecast which is second order accurate what is that this is the forecast dynamics this is the moment dynamics I would like to remind you, this is very similar to the Kalman filter equation in the Kalman filter equation what is that we have forecast is equal to M times P k, M transpose plus Qk plus 1.

So, this is very similar to that DM is the Jacobean of the DM transpose is the Jacobean transpose. So, when M is M of X is equal to M times X, the M DM of X is equal to M therefore, this relation even though we say this is second order accurate it looks like the Kalman filter Kalman filter a linear case. So, you can see the analogy between this equation and that equation. So, I have derived the expression for the approximate evolution of the forecast and the forecast error covariance.

(Refer Slide Time: 126:52)

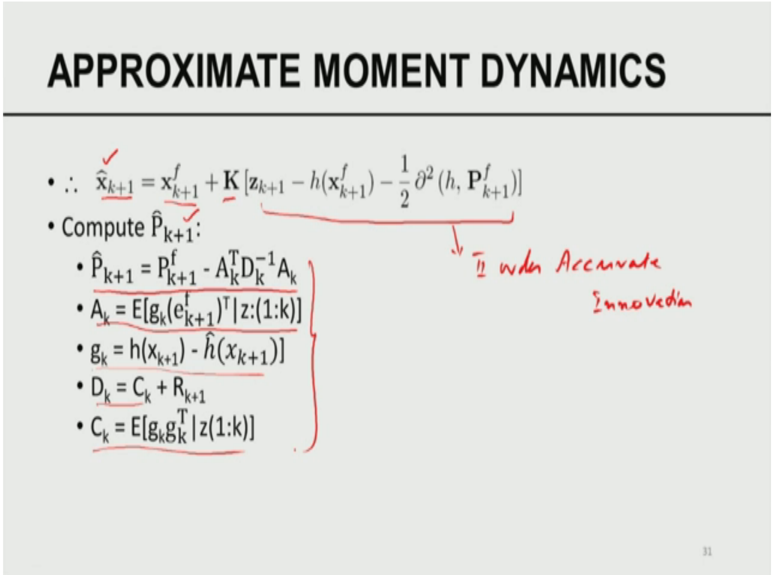


Now, we can do the same thing for the data assimilation step, I have the expression for the analysis that is equal to forecast plus the Kalman gain times the innovation. Now h k plus 1 I can express h k plus 1 in the form of forecast now look at this now to be able to make a forecast I am anchoring on the previous analysis. To be able to make the analysis I am going to make I am banking on the previous forecast. The forecast depends on the Xk plus 1 forecast depends on the previous analysis and the analysis depends on the forecast we can we can even put it Xk plus 1 is equal to Xk plus 1 that that relation still holds good

Therefore I can approximate h of Xk plus 1 is equal to h of Xk plus 1 f plus this again this is the first order term this is the second order term taking again the conditional expectation conditional expectation, we already have h bar what is h bar please remember h bar of Xk is equal to e of h of Xk given Z 1 to k all these are conditional expectations the conditional expectation of ek plus 1 f is 0 therefore, this is the second order accurate expression for h bar.

Where delta square h again follows the same second order Taylor series approximation by in view of the example, I have already incorporated all the information. So, this is the second order correction term this is the second order correction term, I hope that is clear I am doing exactly similar to what I did in the case of model map except here the map is h.

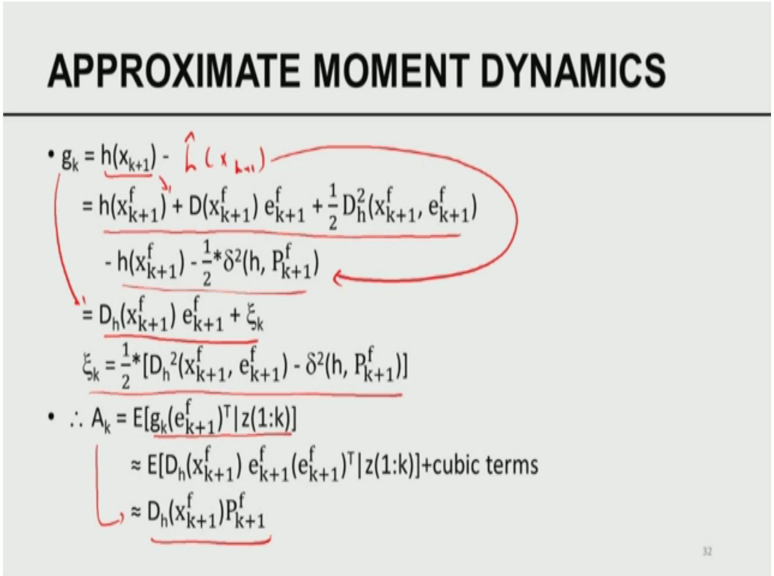
(Refer Slide Time: 128:43)



Therefore I want to substitute back. So, the analysis, what is the second order accurate analysis? The second order accurate analysis is equal to second order accurate forecast plus Kalman gain times the second order accurate innovation. So, this is the second order accurate innovation second order accurate innovation, this is the second order accurate innovations.

Now, we want to be able to compute. So, I have already approximated the analysis, I would like to be able to approximate the analysis covariance analysis covariance is given by this expression from our definition, we already know Ak, we already know Gk, we already know Ck, we already know Dk, please remember these are the derivations I had already given.

(Refer Slide Time: 129:39)



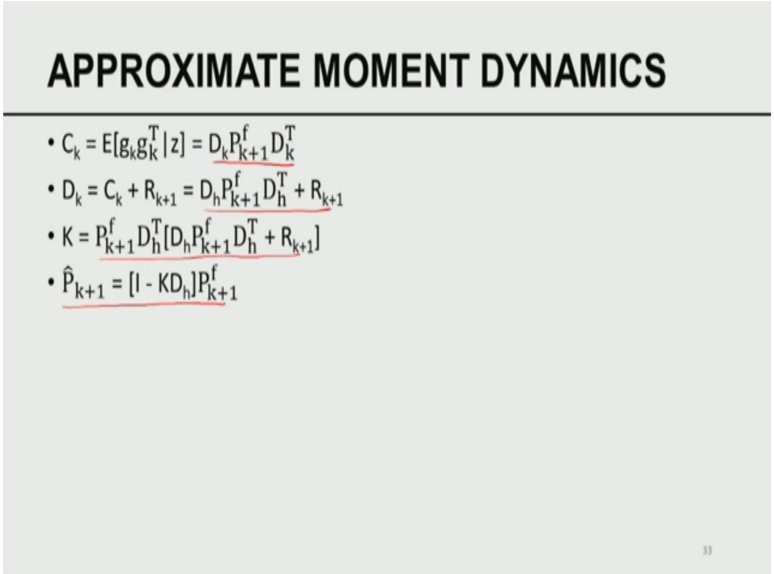
I am now going to substitute manipulate the whole thing therefore, g of this is given by accept k plus 1 there is no and here and I have I am going to express this by a second order Taylor series minus this . So, Gk I am sorry that is a the error term in here I will I will I will conclude. So, this must be please go back I would like to go back and tell g there is a term in here that is missing, I would like to go back to the definition of Gk. So, look at this now at the top of the page 23 Gk was defined h of Xk plus 1 minus h hat of Xk plus 1 and that is what I am now going to have to pull in here.

So, this is going to be h hat of Xk plus 1, let us go back to 23 once more that is right. So, this is please remember that this is the expected value, which is being subtracted from h of Xk I want to remember that. So, Gk is a kind of an anomaly in the non-linear conditional expectation and that is that is what being done in here that is what being done in here.

So, if I expanded this on the Taylor series expansion, I know the Taylor series expansion is given by this, I know the expansion for this is given by this. So, I have to subtract these 2 quantities. If I subtracted these 2 quantities my Gk now takes the form which is given by this now my psi k much like the eta k previously is given by this gobbly goo expression this Gobbly goo expression essentially relates to the error in the second order approximation, which is again you can think of it is second order anomaly if you wish to call it.

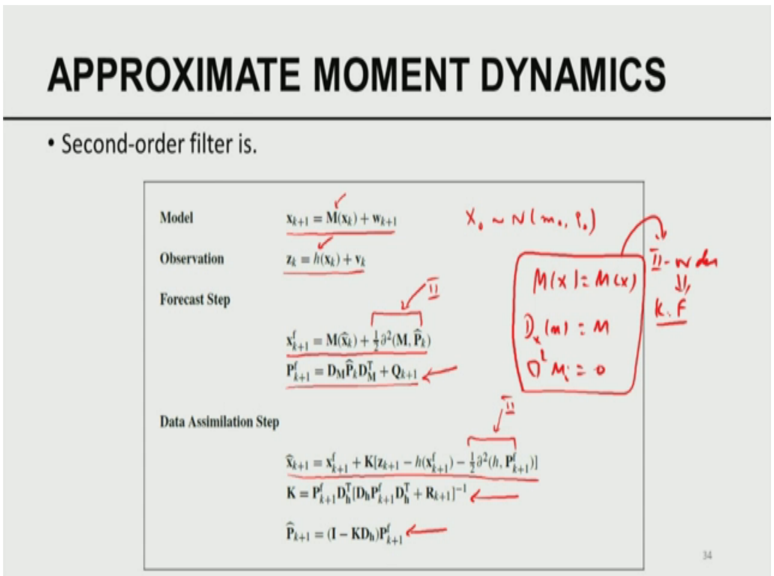
So, Ak is given by definition is this if I substitute the value of Gk from the previous consideration Ak can be seen to be equal to this quantity.

(Refer Slide Time: 132:05)



Now I would like to be able to compute a show that Ck is given by I know I am going a little too fast for many of you, but I want you to understand these calculations are very simple I want to keep repeating that. So, Dk is given by this expression kk is given by this expression Pk plus 1 in the end is given by this expression that is the expression for the analysis covariance for the expression for the analysis covariance.

(Refer Slide Time: 132:35)



So, with this I am now going to summarize this is the final summary of all the things we have done so far. So, I am going to give you the expression for the second order filter, the second order filter. So, it is given in a form that you can right away right away the program. So, this is a model equation this is the observation look at this now model is non-linear observations are non-linear we have the standard assumptions about w kvk I also have standard assumption assumptions over X naught, X naught is given by M of M naught P naught.

Therefore what is the second order accurate forecast; this is the second order accurate forecast what is the second order accurate forecast covariant that is given by that, what is the second order accurate analysis that is that now please understand what is this term that makes the second order accurate this is the second order term that affects the forecast likewise this is a second order term that affects the analysis.

So, this is the second order term that affects the analysis there is a second order term affects forecast that is why both the forecast and analysis are second order accurate. These the even though the expression looks like the first order expression because of the closure we ended up having the second order accurate forecast covariance like this second order accurate analysis, covariance by this and the second order accurate Kalman gain is this.

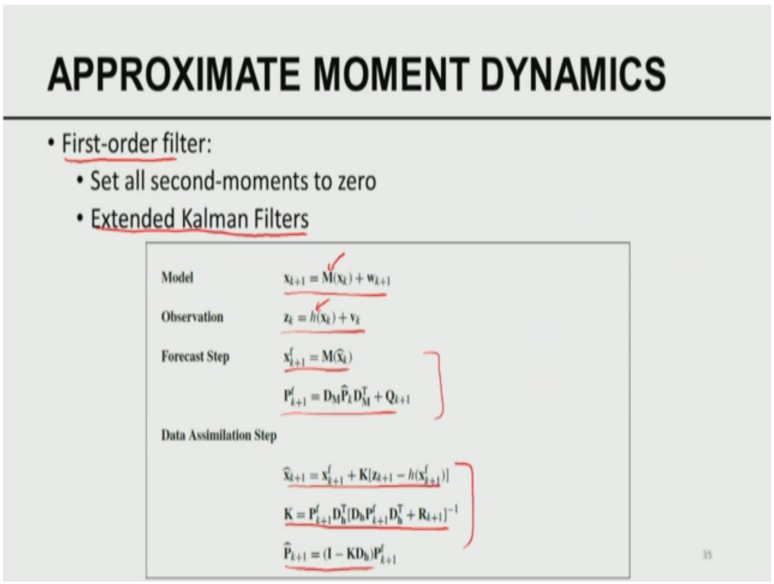
Now, look at this now everything is approximation forecasts are approximation, forecasts covariance is an approximation Kalman gain is an approximation analysis is an approximation. So, if you think back hey this is the best you could do in the case of non-linear system, second order approximation is the best you could do. So, this filter in the literature has come to be called second order filter that helps you to approximate the evolution of the state as a function of time. So, you can readily see forecast step and the and the analysis step they go hand in hand much like much like the Kalman filter equations are. So, this is the sequential second order accurate moment dynamics for the non-linear filter.

So, that is the whole description of this, it is the second order accurate evolution of first moment and second moment of the forecast and the analysis within the context of within the context of non-linear model and non-linear observation it turns out that this equation reduces to the Kalman filter equation, when the model is linear I am going to approx I am I am going to establish this now. If the model map M of X is equal to M of X, D x m is equal to M the del square M is 0 with me. Please I should say del square mi is 0 because each term if del square mi is 0, the second order terms are 0 in and M and the whole thing reduces to the Kalman filter equations.

Therefore in this case second order filter implies reduces to the classical Kalman filter. So, in the sense this is an extension, when do you say a is an extension of b a is said to be an extension of b, when you specialize a becomes b that is called nesting. If an extension does not have this natural nesting property, then then the extension does not have much much much it cannot hold much water the extension argument cannot hold much water. So, to say something is an extension of something else, I should be able to get that something else from the extended value if I set certain parameters if I specialize set certain parameters to extreme values.

So, in that sense you can see the consistency. Please understand in my derivation of the second moment in my derivation of the approximate moment dynamics I showed that our derivation parallels Kalman filter derivation, and it reduces the Kalman filter when you make appropriate choices again I am trying to demonstrate the same thing. So, this this allows us to be able to maintain that beauty of nesting when you go from special to general or general to special.

(Refer Slide Time: 137:20)



If you set all the moments all the second order moments to 0, you get what is called a first order filter, first order filter in the literature is called extended Kalman filter. So, extended Kalman filters which many of you may have heard of what is extension Kalman filter once Kalman filter was announced in 1960-61 very soon they were interested in extending to non-linear cases very they met with lots of difficulties. So, they started approximating the first approximation that was developed within the context of space travel was essentially extended Kalman filter. Extended Kalman filter in our notation is essentially a first order filter first order filter is obtained from second order filter simply by setting the second moments to 0.

So, this is the model dynamics this is the observation both are non-linear, this is the forecast step this is the forecast analysis the forecast covariance, I would like to remind you 2 things now the forecast covariance is the expression for it exactly the same in the first moment in a second moment.

But the trajectory of the second moment dynamics, second moment approx are going to be different therefore, even though the expressions look the same the actual values will be different because the forecast trajectories. In this case and in the previous case are slightly different because the second order term is going to affect the forecast trajectory

So, one thing we had remembered while the expressions may look the same the actual trajectories will not because the second order term; alter the trajectory compared to the first order term. So, the forecast step again is similar to what we have except the second order correction term, there data assimilation step is again that that looks very much like the linear Kalman case this is the Kalman gained this is the analysis forecast. These are gain very similar to the second order filter what is the only difference the only term that I marked by second order approximation terms are absent in the table in 35 compared to the 1 in 34.

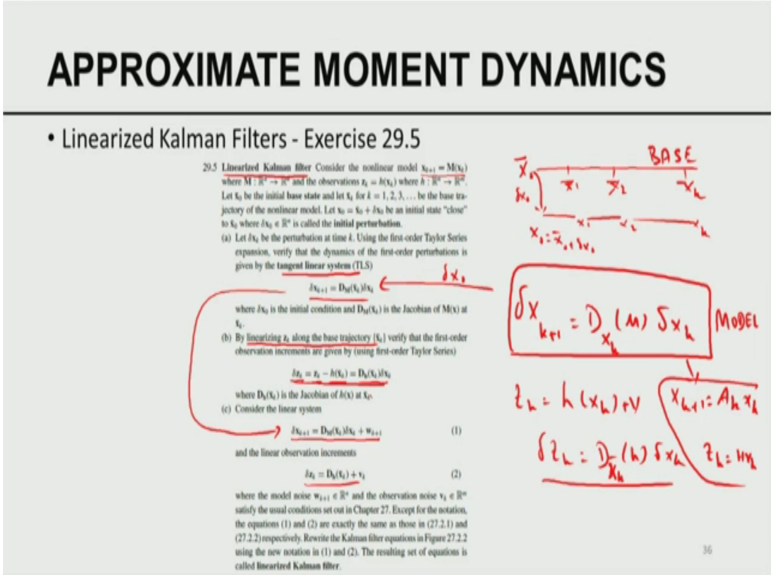
So, that essentially completes our derivation of moment approximation to the non-linear filter. So, general order general expression from the moment dynamics second order approximation first order approximation. So, what is that? So, if you are given a non-linear system if the system is not too big you can apply the first order moment equation the first order dynamics second order filter. So, you can experiment with by taking a small simple problem you can solve the problem by second order filter, you can solve the problem by first order filter, you can plot the trajectories of forecast from first order versus second order you can also plot the trajectories from first order filter analysis of it the analysis from both the cases.

So, how do the analysis differ with the order of approximation to me that is a very good in an interesting exercise. It could be a part of a classroom computer related project on a model chosen, and it turns out this if you change the model the quality and the quantitative differences between this approximation may not hold across various models therefore, when you want to be able to apply non-linear filters to small dimensional problem, it is better to do it in slightly different ways first order filter second order filter and then compare the performance and then compare the performance. So, that will be a very very nice interesting class project, I often in my teaching, I give these projects in in in my class and they are extremely very educative. The systems I give I do not give large dimensional system 2 dimensional 3 dimensional.

So, what is the typical system I will like suppose there is a object falling freely from the sky it has been falling for, long that the acceleration is countered by with the friction. So, stokes law comes into effect and the particle is descending with a constant speed. So, the vertically it has attained the terminal. So, called terminal velocity if a friction filled medium if a bar if a particle is dropping down, a stokes lie essentially tells you it will reach a terminal velocity where the acceleration is going to be countered by the friction I am putting a radar at the bottom I am trying to observe the position and the velocity of the particle through the radar, and I am asking them to be able to assimilate your non-linear model for the free body and radar observations. So, it is a very simple educative model using which one can bring out various discussions relating to non-linear filters non-linear approximation quality of non-linear approximations.

So, we have talked about second order filter, first order filter I have now I am now going to take couple of minutes in talking about another Kalman filter equation these exercises are taken from our book this is exercise 29.5, because this exercise is imparting is important I am going to talk about this this is called a linearized Kalman filter.

(Refer Slide Time: 143:10)



What is the linearized Kalman filter I have a model equation Xk plus 1 is equal to M of Xk.

So, what do I do? I pick a state X naught I compute the non-linear trajectory X 1, X 2, Xk then I induce a perturbation to X naught. If I want to call it the unperturbed state it could be a bar state it could be a base state the perturbed state X naught is equal to X bar naught plus delta X naught it comes to X 1, it comes to X 2, it comes to Xk. So, what is that we now know delta Xk is equal to D at Xk of M delta Xk plus 1.

So, let me write that equation clearly sorry let me write that equation a little bit carefully. Delta Xk plus 1 is equal to d at Xk of M delta Xk this is the propagation of the perturbation this is called in mathematics variational equation meteorologists calls a tangent linear system we have already come across this equation in the context of 4 D war and that is this equation

So, for this linear system delta X naught is the initial condition. So, what is that I do now? I am having a non-linear system I control the base trajectory this is the base trajectory I am I superimpose perturbation; I superimpose an initial perturbation and talk about the dynamics of evolution of perturbation that is a linear equation excuse me. This linear equation essentially tells you how the superimposed initial perturbation propagates at the top of the base state.

So, now forget about the original non-linear model consider this perturbed model this is a linear model. Now let us assume I have been given observations originally. So, let us assume originally I have been given observation which is zk is equal to h of Xk plus v. So, now, what am I going to do I am going to consider increments the observation. So, I am going to linearize ek along the base based trajectory.

Therefore delta Z k is equal to Z k minus h of X bar k, what is X bar k is the base trajectory to a first order approximation, I can say del delta Z k is given by this first order quantities. So, look at this now I have this as the model equation, my observation equation is going to be delta Z k is equal to h times I am sorry this is not right my this is equal to D of Xk bar of h times delta Xk.

So, look at this this equation looks like Xk plus 1 is equal to A k Xk, this equation looks like Zk is equal to h of X k. So, you have a linearized observation, you have linearized model. Now add some noise to this linearized model to get this. So, this becomes a linearized stochastic model this becomes the linearized observation if you have a linear model and the linear observation I can do a classical Kalman filter, that filter going to give you an approximate estimate of the forecast and the approximate estimate of the analysis for the perturbed system by adding the perturbation the perturbed forecast on the analysis that the perturbed analysis to the base I will get the actual.

So, if I know the increment if I know the base by adding the increment to the base I know the actual. So, what is the what is the order of approximation? Here this is called zeroth order filter. So, we talked about second order approximation we talked about a first order approximation now this is called a zeroth order approximation to the non-linear filter. So, what do we do we simply a create a linearized variational equation for the evolution of the model of the perturbation across the base state then we considered a linearized version, of the observation we throw the original models out you are you consider the linear model, a linear observation as your given model you do a linear Kalman filter or a classical filter, you compute the analysis you compute the forecast you add them to the base state you get the actual forecast and the approximation to the actual forecast and approximation to the actual analysis.

So, this is called the zeroth order filter or the linear Kalman filter is a very interesting exercise. So, you can take a simple non-linear problem and do it in 3 ways, and look at the quality of approximation in and what do I get what do I lose what is the computational cost this could be a very interesting exercise and this module has been a summary of our chapter 29, and that completes our discussion of non-linear filter yes this module is very dense because non-linear problems are not easy. So, we have on 1 part the stochastic dynamics the Markov property, on the other hand we have non-linearity on 1 hand we have conditional Gaussian distribution, on the other hand the conditional Gaussian distribution do not transfer itself as Gaussian for the predictive density.

So, we have tried to deal with that exactly as far as we can and derived the update in the infinite dimensional space, then we want to come to the real world of finite dimensional computations. When we came from infinite dimensional space to a finite dimensional space we had moment approximation, but moment approximation even though is an approximation even this approximation is riddled with what is called a closure problem.

So, I have to tackle the approximation at several levels. So, if you superimpose one level of approximation to another level of approximation, other level of approximation ultimately, when you when you when the when the fog clears you can have essentially a second order accurate filter, first order accurate filter, zeroth order accurate filter. These 3 are considered to be meaningful approximations to the non-linear filter problem; with this we come to the end of the discussion of non-linear filters.

Thank you.