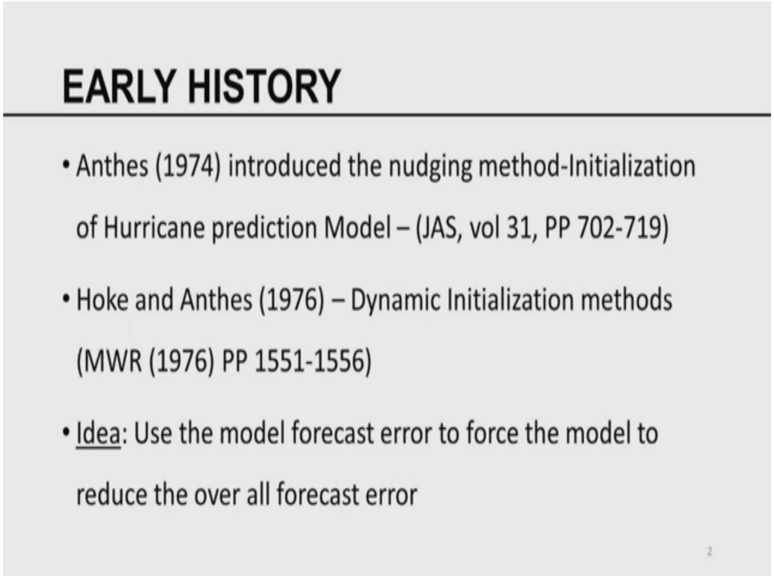
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture - 38
Basic nudging methods

In this module, we are going to be talking about another method that has been introduced way back in the early 70s called nudging. I am going to provide the basic principles of nudging and some of the associated questions relating to the design of nudging schemes. So, what is nudging? The model is always use to create a forecast, the goal of data assimilation is to make the model to fit the observations. So, this fitting was done by looking at the some of the square differences between the model predicted variable and the observation to decide on the values of the optimal values of the parameters. And the initial condition from which we started the model forward. And that is one of the themes that underlie 40-watt data assimilation are forward sensitivity-based data assimilation.

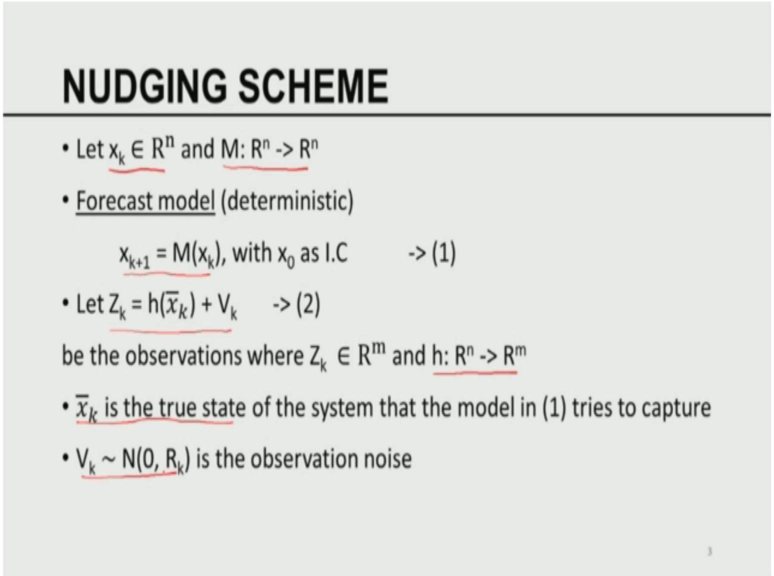
Nudging is an alternative method in nudging what you do you compute the forecast error, which is the difference between the model predicted observation and the actual observation. This error in the forecast is often used as the forcing, the forcing that makes the model move towards the observation, this ability to force the model by adding a force that depends on the forecast error is the fundamental idea behind the nudging scheme. So, to nudge to be able to force to be able to choirs the model towards the observation. These words force nudge choirs essentially captures the fundamental principle that underlie this notion of nudging algorithms or nudging methods.

(Refer Slide Time: 02:27)



A bit about a early history, Anthes in 1974 introduced the nudging method. He used this for initialization of hurricane prediction model, it was published in general atmosphere sciences in 1974. Hoke and Anthes in 1976 further explored the use of nudging schemes. Again, within the context of hurricane forecast, and that paper was giant paper was published in 1976. In monthly with the review the idea is to use the model for the model forecast error to force the model. So, as to reduce the forecast error. So, it is a kind of a feedback principle. So, the model makes a forecast observations are there is a forecast error, I am using the forecast error to be able to force the model to be able to reduce the forecast error. This is the fundamental principle that underlie any feedback control mechanism. So, it is a kind of a feedback control theory; that is, brought to focus by this nudging scheme within the context of data assimilation methodologies.

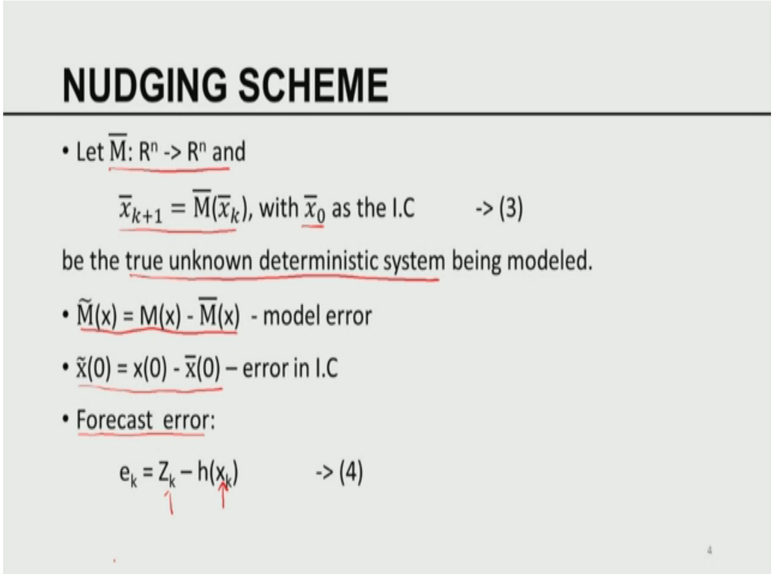
(Refer Slide Time: 03:54)



So, did it to get a feel for this nudging scheme. Let us consider a state, x k a time k which is an R n. M is a map from R n to R n the forecast model I am assuming to be deterministic with x naught does the initial condition. The observations are again non-linear function of the state x bar be the true state of the system that is not known I only have information about the true state through the observation. V ks are the observation noise, h is the forward operator. Again, I would like to emphasize the notion of a true state. And the observation noise is discussed here, it is a standard set up. So, what is the difference between this and the cognate filtering scheme? The model is deterministic.

So, this has commonality with 4DVAR the early in the early 70s within the meteorological literature. They consider the model to be a perfect. So, under the perfect model assumption noise is the observation. They would like to be able to use the forecast error to force the model which will in turn make the model move towards the observation.

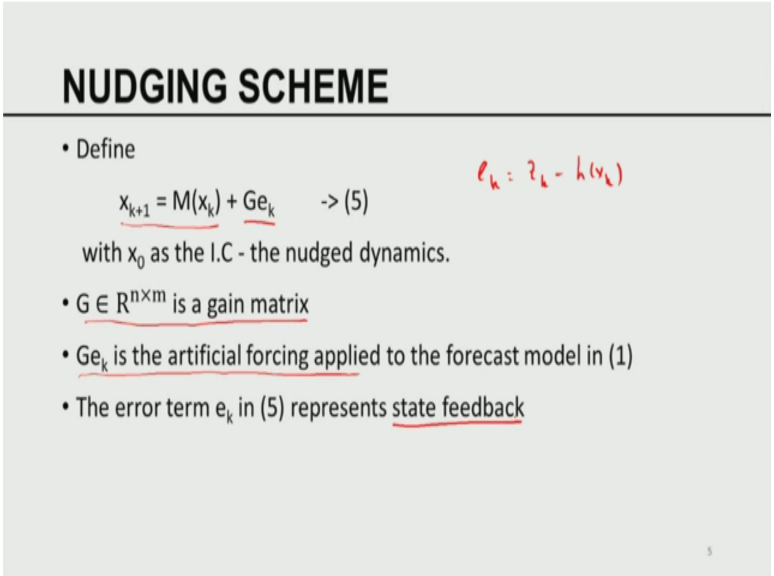
(Refer Slide Time: 05:18)



So, let M bar be the true model dynamics. So, we I am going to now develop a general theory. Let M bar be the true model dynamics. X k plus 1 bar is equal to M bar x k bar with x naught bar as the initial condition be the true unknown deterministic system being modeled.

If M bar is equal to M, the model is perfect. If M bar is different from M the model has an error. So, I am now going to consider a generalization of the nudging scheme, where I am going to think that the model may or may not be perfect. So, let M bar x be the model error. X I am sorry M tilde x is equal to M x minus M bar x that is the model error. X tilde 0 is equal to x of 0 minus x bar 0 is the error on initial condition. So, if I use M as the model to be able to generate the forecast x k. X k is the forecast generated out of the model M, M may have errors z k is the observation, coming from measurements in the real world. So, e k is the forecast error as given in 4.

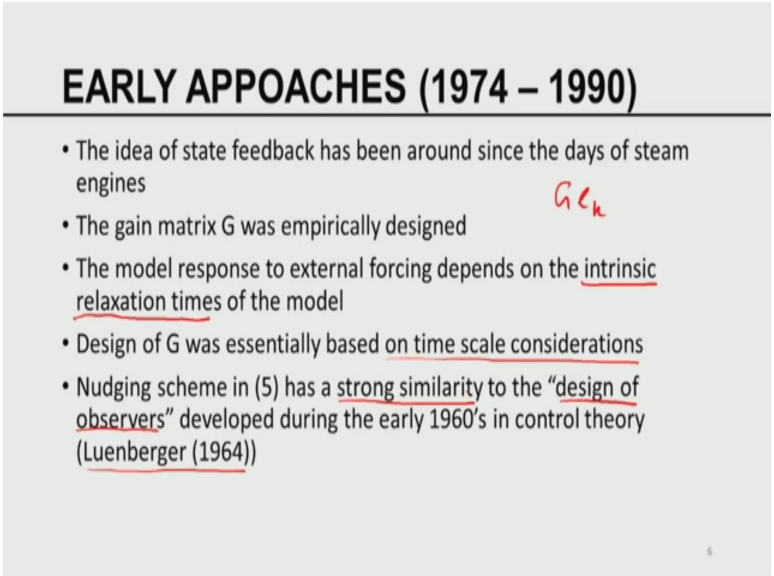
(Refer Slide Time: 06:54)



Now, what is the nudging scheme? Consider an otherwise deterministic model x k plus 1 is equal to M of x k, please understand x k is the forecast starting from x naught on the model M. M may have errors the initial conditions may have errors. So, the forecast x k generated or the model equation may have errors. I would like to be able to add a forcing term. Please remember, e k is a forecast error G is a matrix. So, G times e k is the forcing that is artificially added to the model. The forcing always makes the solution move towards a particular goal. Our aim is to be able to find G such that asymptotically the model state moves towards the observation. Which represents they true state of the model.

So, G is called a gain matrix G is called the gain matrix. G is again an n by m matrix. G of e k is a vector. There is an artificial forcing applied to the forecast model. The error term e k represents a state feedback. Why? Please remember e k is equal to z k minus h of x k. Therefore, I am using the state information to force the model. So, that is what is called the state feedback.

(Refer Slide Time: 08:42)

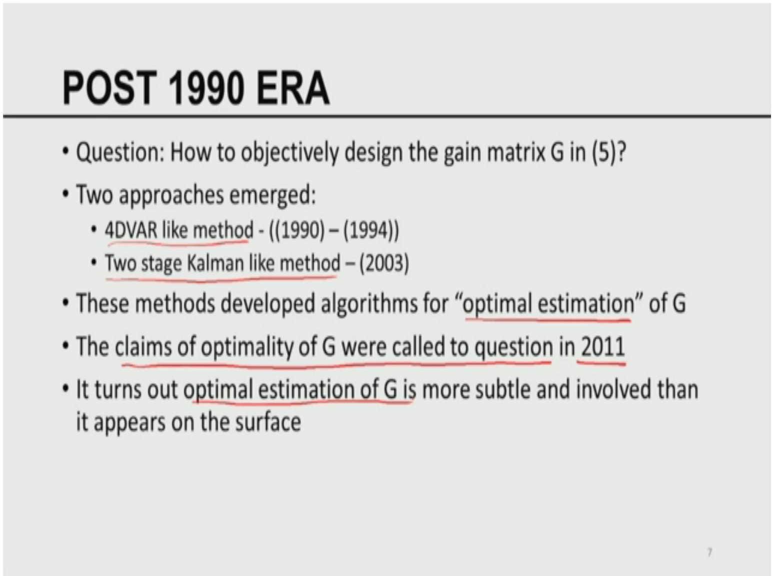


The idea of the state feedback has been around since the day since the early days of steam engines. The gain matrix G in the early days was empirically designed, I am now talking about some of the early approaches to design of nudging scheme the early approaches period from 1974 to 1990. During this period several people have applied nudging schemes to be able to make data assimilation schemes in other words you are trying to use the data to force. The model are the forecast error which involves data towards the model.

So, the model response to external forcing depends on the intrinsic relaxation times of the model. So, what is the basic idea? G e k is the forcing to the model. If you apply a force to a dynamic model, how long does it take for the model to respond to the initial force? That is that depends on what is called the relaxation times. The intrinsic relaxation times in engineering, we also call it time constants. So, the design of G in the early days was essentially based on the time scale considerations. What is the intrinsic time scale of the processes involved? How long does it take for the model to respond to the external force? So, the value of the matrix G was essentially heuristically decided, and considerations was essentially based on the time scales of the processes involved in in in the model solution.

So, the nudging scheme in 5 has a strong similarity to the design of observers. The theory of the design of observers in in in control theory. The theory observers was developed earlier in by Luenberger in 1964. It is not very clear whether Anthes and his group knew about this work on Luenbergers. Luenbergers work on observers, but the observing observer-based design as well as the nudging schemes had a very strong similarity structurally. So now, you can see Kalman filters came from control theory the notion of observers was already in existence in the early 60's in control theory, and maybe it was invented independently by anthes. But I would like to emphasize a very strong similarity between the nudging scheme as used in geophysical literature as well as the observer designs in in control theory.

(Refer Slide Time: 11:48)

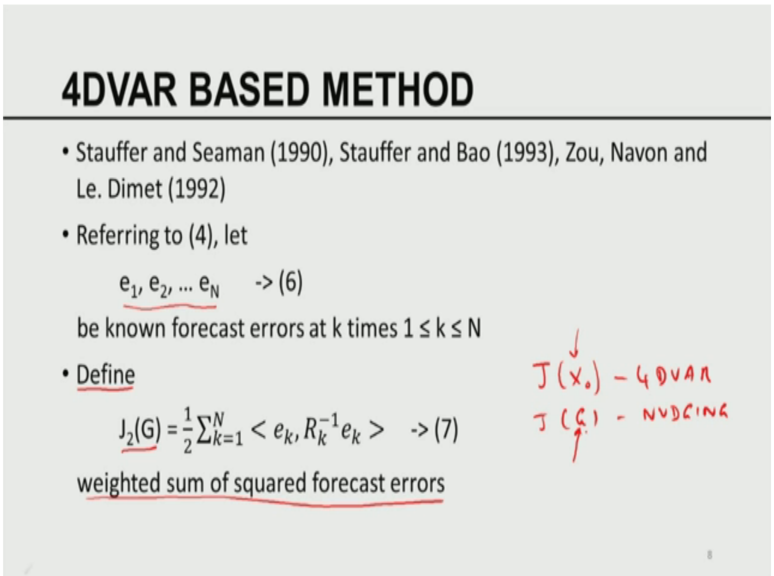


So, you can see there is a great influence of control theory in the design of data assimilation algorithms. So now, I am going to talk about the post 1990 era. What is the question? How to object so, before mainly people were heuristically designing values of the gain matrix G to be able to force the model towards the model state towards the observation so as to reduce the error. The only consideration they used was based on relaxation time. And that is all, but in many cases, it worked very well. But as the theory of 4DVAR was well developed, as the theory of optimal methodology was very well understood, the notion of being able to estimate the optimal state and the and the theory behind the strong constraint formulation we constrained formulation was well understood. Around the turn of 1990 the emphasis shifted towards trying to objectively design the gain matrix. That essentially dictates the amount of forcing that that was applied to the model equations.

So, 2 approaches emerged within this quest for optimal approach to the design of the gain matrix. One is the 4DVAR like methods. One is the 4DVAR like methods. Another is the 2 stage Kalman filter like methods. The 4DVAR methods sprang up in the early 1990 to 1994. The 2 stage Kalman like method arrows around was announced around 2003. These methods developed algorithms for optimal estimation for G. So, what did they do? Well, G is the known and they wanted to be able to optimally estimate the appropriate value of G. So, they brought the full force of least square based estimation theory within the framework of 4DVAR are within this within the framework of 2 stage Kalman like methods.

In 20 even though these theories were known in 2011, it was pointed out there are certain the claims of optimality that are obtained by these researchers were found to be a little defective. It turns out the optimal estimation of G is more subtle and involved then it appears in the surface, when you read the papers on 4D like 4DVAR like methods as well as 2 stage Kalman like methods. We are going to talk about both the methods as well as some of the problems associated with the methods, and ways to go around some of these challenges.

(Refer Slide Time: 14:55)



So, 4DVAR based methods. Stauffer and seaman Stauffer and Bao, Zou, Navon and Le Dimet were some of the earliest people who were working in trying to find the optimal value of G. So, referring to the equation 4 referring to the equation 4 let me go back.

Equation 4 at the bottom of slide 4 is the forecast error of the non-linear model. So, 4DVAR based methods were introduced by at least 3 sets of authors Stauffer and seaman Stauffer and Bao, Zou, Navon and Le Dimet in the early 90's. Referring to 4 which refers to the expression for the forecast error.

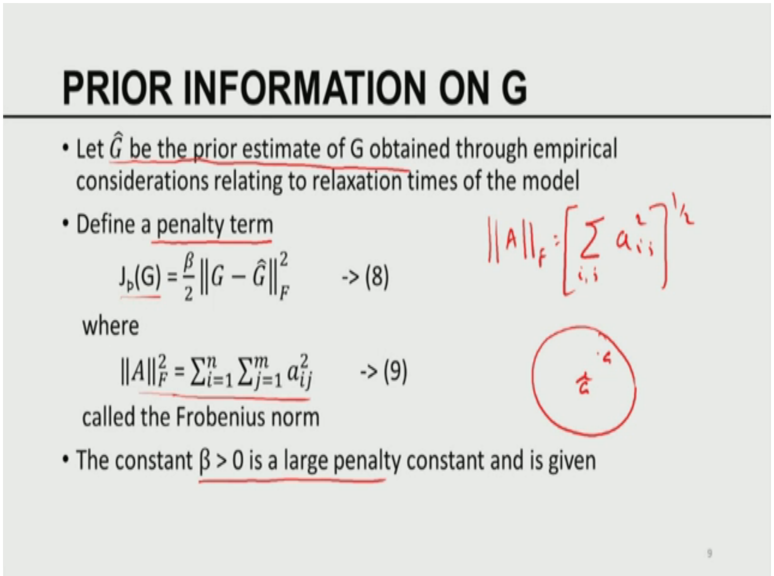
Let e 1 e 2 to e n be the set of forecast errors created by the model forecast are based on the model forecast. So, in other words, you can see like the 4DVAR like scheme here. So, while the original nudging scheme, involves essentially adding a forcing to the model and let the model evolve these people coming from the 4DVAR methodology. Their aim is to be able to design the optimal G. So, they are doing an offline experiment. What is the goal of this offline experiment? Let us pretend that there are n observations. If there are n observation then there are n forecast errors. If there are n forecast errors, then I can compute the least square cost function J 2 of G which is given by e k R k inverse the transpose, I am sorry the inner product of e k with R k inverse e k.

What is this? This is the weight of sum of squared errors. Please understand this is exactly the cost function that is used to minimize to find the optimal initial conditions and parameters in 4DVAR as well as forward sensitive based method. You can see the approach because 4DVAR is something they know very well. And they were they were part of the development of the 4DVAR techniques. So, they would like to look at this scheme, as though it were a 4DVAR scheme to be able to estimate G. Another difference is that in the classical 4DVAR they use this kinds of objective function to decide the optimal initial condition. In the nudging I am not going to worry about where they the model starts. Initially the model may have forecast errors, but as time goes on adding G e k forcing to the model will try to correct the model forecast errors, while the model is an evolution. So, that is the basic idea.

In the 4DVAR we wanted to be able to start from the initial optimal initial condition so that starting from the initial condition, the new forecast generator will match the observation as much as possible. Here that is not the goal. Initial condition could be anything. Our aim is to be able to simply move the model towards the observation as the model starts operating. So, they were able to understand that I may not want, I may not be able to correct some of the initial forecast errors, but in time the feature errors may go to 0. R may become very small, that is the idea. So, you can see in the 4DVAR the J 2 function was a function of if you if you recall we call it function of J x naught that is in 4DVAR.

In here we calling it J of G in nudging. So, the independent variable with respect to which 4DVAR was with respect to which the minimization was done is the initial condition. The independent variable with respect to which the minimization is going to be done in the optimal design of nudging matrices is G. The elements the metric them self G that is essentially. The difference, but rest of it are very similar. So, mathematics is not too different from what 4DVAR involves.

(Refer Slide Time: 19:33)



Again, they wanted to bring in the notion of background. Why this notion of background is useful? Because in the pre-90 era they have been very successful in trying to demonstrate the usefulness of nudging scheme based on some empirical values of G. So, that is the knowledge they knew what reasonably well in many circumstances. So, they do not want to throw that knowledge out of the window; say, they would like to be able to use some of the prior knowledge where nudging had worked. So, they would like to be able to give the benefit of doubt.

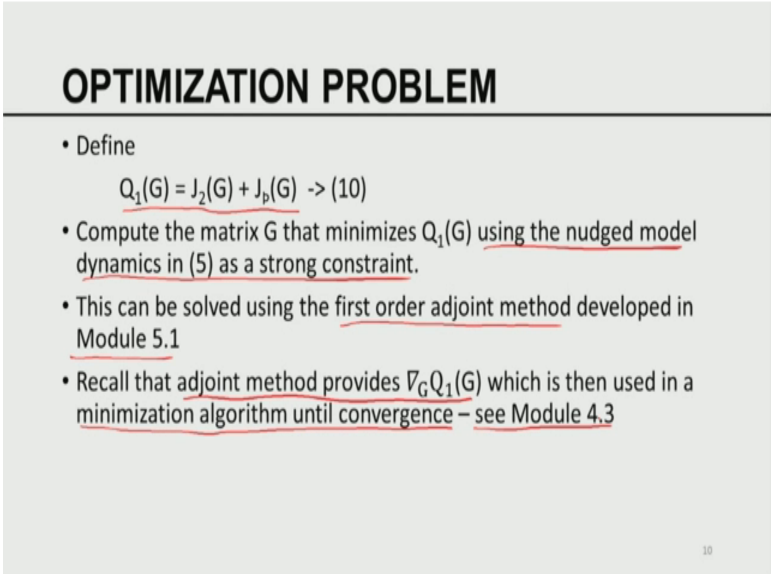
So, they said let G hat be the prior estimate of G obtained through empirical consideration relating to the relaxation times of the model. So, quite a lot of time scale analysis of the models have been done. So, they know which time scales respond to what type of forcing. And that knowledge they didn't want to go as a waste. So, they assume well we will also want to take advantage of some they earlier estimates. So, let G hat be a prior information. So, you can see now they are combining a prior and the new information coming from the forecast errors, they would like to be able to combine the 2 types of objective functions. So, the prior term gives raise to what is called a penalty term, a penalty term is as follows. J penalty of G is G minus G hat bit a times beta over 2 it is A square of the Frobenius norm. You may recall from our example from our module on matrices the Frobenius norm square I am sorry, I will simply say the Frobenius norm.

So, the Frobenius norm by definition is equal to sum of a i j square sum is over I and J. It is something like the Euclidean norm. I simply take the sum of all squares of all the elements the object if it is a vector. That is one sort of objects if it is matrices that is another sets of objects. So, it is simply sums of squares and the square root of that. So, that is called the Frobenius norm. So, the penalty term regards is relates to the difference between the 2. So, what does it mean? I would like so, what is the penalty term tells you? I am interested in designing an optimal G, I know they have been using G bar, but G bar I am sorry, G hat the G hat was obtained from heuristic consideration. I am trying to design G optimally. I am trying to design G optimally based on the forecast sums of forecast errors. At the same time, I do not want my G to be far removed from G bar, because G bar already worked.

So, it is a so, I want to find an optimal G. That is a compromise between minimizing the sums of forecast errors. At the same time going not too far away from G bar I am sorry G hat. So, we have defined the Frobenius norm in here. So, the constant b is called the penalty parameter. If the constant beta is large, since I am going to minimize the product has to be small. So, if beta is large G will be much closer to G hat. If beta is small, I am trying to relax the distance between G and G hat. So, by look by essentially picking the value of beta, one can have a variety of different range of G with respect to G hat. In other words, I have G hat here. Do I want my G to be in a sphere of radius dictated by the value of beta?

So, if beta is small. It will be you have more freedom for G. If beta is very large, because if beta is large and the norm is large, it will it will it will not come to minimum. So, if beta is large the only way to minimize it is to force G towards G hat. So, that G minus G hat will be much smaller. So, that is the idea of the penalty term. So, you have 2 terms now. One I do not want my new estimate to be too far away from the old estimate that they have used based on heuristic consideration. I do not I am, and I am doing myself some freedom by being able to choose the penalty term.

(Refer Slide Time: 24:31)

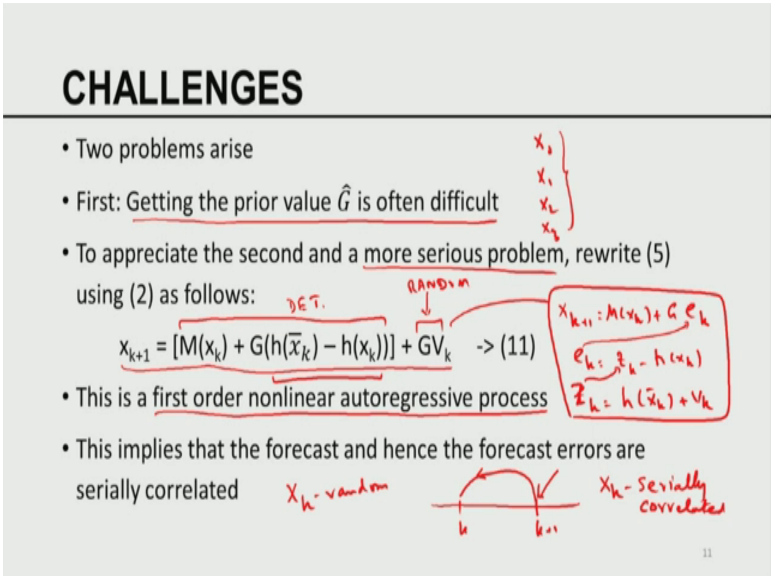


At the same time, I have the forecast error term sums of square errors. Therefore, I am not going to consider up a new criterion which is Q 1 G which is going to be a sum of J 2 plus J p.

Why I call it J 2? Because it is it isn't 2 norm. J p, p for penalty term. So, compute the matrix G that minimizes Q 1 G, where the nudge dynamics is going to be used as a strong constraint. Please understand, that that the nudge dynamics is the one that is going to be ultimately used in the forecast. So, then I cannot decide G are minimize this independently I have to find an optimal G within the context of the dynamic models. Therefore, it is formulated as a strong constraint problem. So, we have we have now essentially formulated the problem. So, what is that we can do we can essentially develop the first order adjoint method. That was developed in 5.1. So, we can use the first joint a first order adjoint method of module 5.1 to be able to decide on the optimal G and the optimal G. So, what is that we do, we start from we start with the G, run the model forward in time. Compute the forecast errors. Compute the objective function. And evaluate the gradient of the objective function. And then once you evaluate the gradient of the objective function numerically, I can use it in a gradient method.

So, as to minimize the elements of so, as to minimize the objective function Q 1 G. Recall that adjoint method gives raise to the gradient. Which is then used in some minimization algorithm until convergence. And we have already talked about minimization algorithms in module 4.3. So, by because we have done lot of things relating to adjoint methods relating to optimization methods. Our discussion becomes simpler, because we do not want to repeat the entire derivation. An interest reader can simply apply. The methods and derived the up, and derive the expression for the gradient. So, this was the theme of people who wanted to be able to estimate G optimally using 4DVAR like method.

(Refer Slide Time: 27:12)



But there are a couple of philosophical challenges. What is the first one? Getting the prior value of G hat may not be as simple as one deems with why is that. If you change the model if we change the process for hurricane prediction the dynamics is of one type where the time scales are of one time. If you change the model equation and go from hurricane to other physical processes, again the time scale analysis will be different. Therefore, the choice of G hat depends very much on the process that is to be used are the process that is captured by the model. So, in general there are no specific guidelines for choosing G hat. So, the only thing that we can fall back on is that for those processes for which nudging method have applied in the early years.

You have a very, very well-defined G hat, but in general there is no clear-cut algorithm for generating G hat. So, the difficulty with respect to getting the prior value is one problem. To appreciate the second difficulty, and this is a more serious problem, we want to be able to write 5 using 2 as follows. So, let us go back to 5 and 2 what they are. So, 5 is the nudged dynamics, 5 is the nudged dynamics. And 2 is the observation equation. So, I am you know going to combine the 2 to be able to come up with the exact equation for the nudged dynamics. So, when I use when I use 2 and 5, they explicit equation for the nudged model becomes this. Why where is this coming from? Please understand, x k plus 1 is equal to M of x k plus G times e k. And e k is equal to z k minus I am sorry, z k minus h f x k and z k is equal to h of x bar k plus V k.

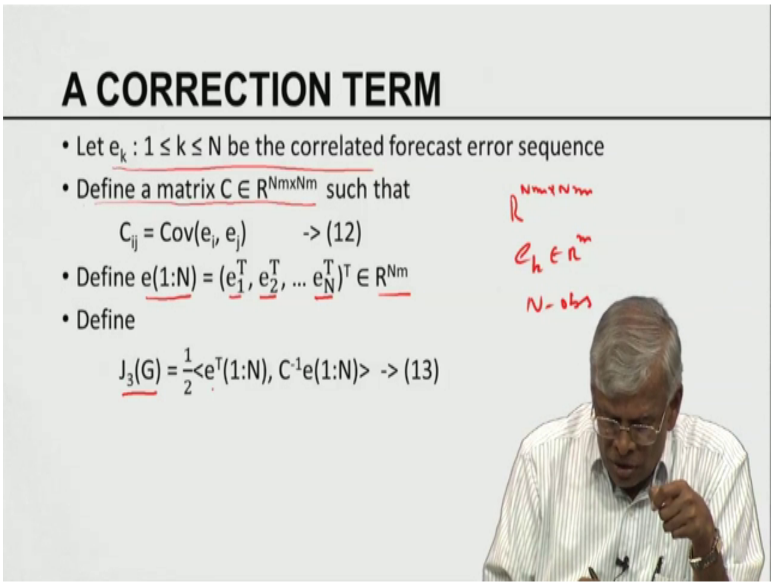
So, these are all the various quantities that are involved in here. So, z k is equal to h of x bar k plus V k x bar k is the unknown true state. So, I substitute z k in here I get e k, I substitute e k in here. If you do these substitutions the resulting equation takes this form. So now, let us look at the structure of this it consists of a deterministic component. Why this is deterministic component? M of x k is that M of x k is a deterministic model forecast, h of x k bar the 2 state that is a deterministic. H of x k that is the forecast an observation counterpart of the observation. So, this term h of x bar k minus h of x k is there is the actual expression for the forecast error true minus the actual G is a multiplier constant. So, this term is essentially a deterministic term.

The V k are term now occurs as G V k. So, that is the random term. So, nudging method in fact, induces a stochastic dynamics, because observations are stochastic. Even though you are model your model is deterministic, the process is stochastic because of the observation always have observation noise. If you look at this carefully, you can readily see this is the first order non-linear autoregressive process. So, what does it mean? X k is not deterministic x k is stochastic. So, in the previous approach to 4DVAR they didn't realize that that is the stochastic term that is affecting the evolution they simply assume that there is no such thing as a stochastic part in the forcing, they applied the 4D 4DVAR like scheme. And they found optimal within the framework what they had built. But a clear examination of that of those ideas essentially it tells you, a correct formulation has to take into account this autoregressive process. Why this is autoregressive? X depends on the x k plus 1 depends on the previous x k. So, what is an autoregressive process, this is k plus 1, this is k.

So, k plus 1 depends on the previous values of the state. So, I am dependent on myself at the previous time plus a random noise, plus the random noise. So, once you recognize this is an autoregressive process x k becomes a random. If x k is random what does it mean? X k the trajectory is a random process. So, this essentially tells you x ks are serially correlated, x k serially correlated. Why they are serially correlated? Let us go back now. I have x naught I have x 1. X 1 has the effect of effect of then x 2 x 3. X 3 x 3 depends on x 2. X 2 depends on x 1. X 1 is random x 2 is random. Therefore, x 1 and x 2 are not totally independent they are x 1 is a random process R, R is a realization of random process x to the realization of random process x 2 depends on x 1. So, there is a serial correlation that is induced.

So, this serial correlation was neglected in almost all of the treatment of nudging schemes. And this observation was first made by us. So, we would like to be able to make amendments to the optimal estimation of G by taking this serial correlation into account. So, that is one of the second problem. Not only second problem, but also, we propose a solution to go around and solve the second problem.

(Refer Slide Time: 34:32)

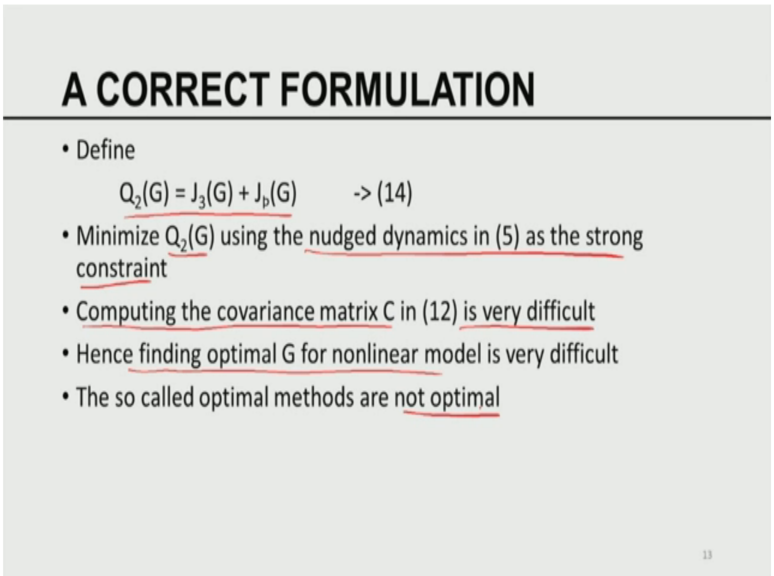


So, let us let us talk about this now. So now, that we have established that the errors are serially correlated. So, what did I want? I would like to be able to define a matrix C. What is this matrix C? The matrix C is R N m times R N m. Why N m? M is the size of the forecast error. Please understand, please recall this decides the forecast error. I have

N observations. So, if I consider all the observations to all the observations together as well as I consider a gigantic vector. The set of all forecast errors from time 1 to n. I am sticking them all together to get a gigantic vector which is R n m.

If I have a vector R N m, it is covariance must be R N m times R n m. So, C is the represents in M a gigantic matrix that represent the serial correlation between all the all the states. So now, we are going to define in now you may ask a question, how do you know C. We have to estimate C, because once we know that they are serially correlated we have to take that correlation to account in trying to define your weight function. And that is what J 3 G is all about. So, J 3 G is an alternative to J 2 G that we saw earlier, J 2 G neglects the serial correlation. Now this is a new objective function, that involves the weighted sum of squared errors, the waiting is related to the serial correlation matrix.

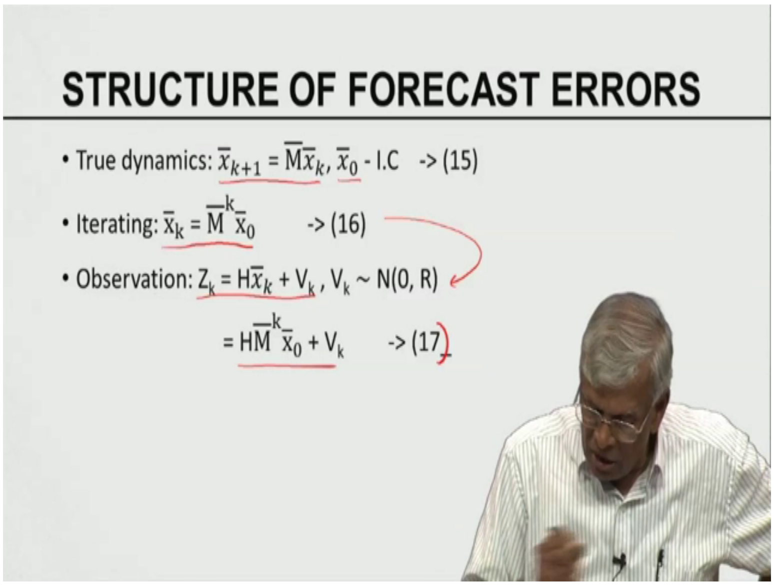
(Refer Slide Time: 36:39)



So now we can minimize, now we can minimize with respect to a new function instead of Q 1 is Q 2 Q 2 G is equal to J 3 G plus J p. G if you minimize J 2 G using the nudge dynamics as a strong constraint model. You try to take into account the serial correlation, but computing the covariance C in 12 is not easy.

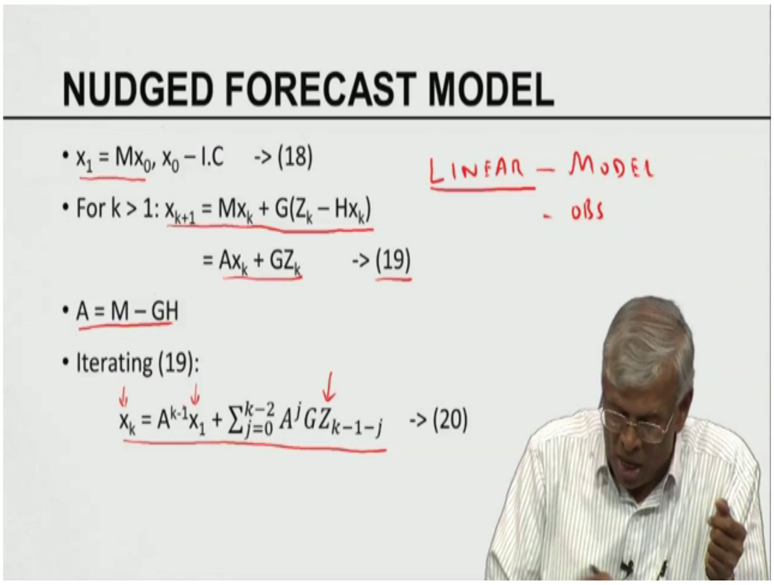
Hence in principle finding an optimal G for non-linear model is a difficult problem. So, in our view the so called optimal methods that are proposed between 1990 and about 2005. Covering a period of about 15 years. They claim that their optimal is indeed are not optimal. So, that is that critique about the methodology. And we are also going to suggest a way out of this critique, but in general. So, what is that mean? Nudging you can implement heuristically. More often that did not works, but if you want to move away from heuristic methodology to an optimal methodology, where I am trying to design a G which is best, then you have to take into account all the processes that are involved. If you look at it carefully, there is a serial correlation trying to minimize the errors without taking that serial correlation to an account always leads to leads to results that are not optimal. That is the, that is it that that is their simple way of looking at what is happening in here.

(Refer Slide Time: 38:12)



So, in order to be able to define the matrix e I have to look at the structure of the forecast errors, because if you want to be able to understand the serial correlation I need to understand the structure of the forecast errors. So, let us spend few minutes on trying to understand the structure of the forecast errors, within the context of being able to consider the nudging scheme as the first order non-linear autoregressive process. That is the next step. So, the let this be the true model. Let x not bar be the true, but unknown initial state. So, what does it mean if you use the true model, and starting from the true initial state the model forecast will be perfect? It will match the observation modular noise. Please understand, when we say match, match only the deterministic sense, we cannot match the random process anytime. So, whenever we say something matches something, it is always modulo noise. Mod noise is something we may have to live with. So, if I iterate this equation x k bar is equal to M bar k to the power k x naught bar. So, observation is given by this equation. So, I can substitute 16 in here to get 17. So, the expression for z k the observation at time k is given by equation 17.

(Refer Slide Time: 39:51)

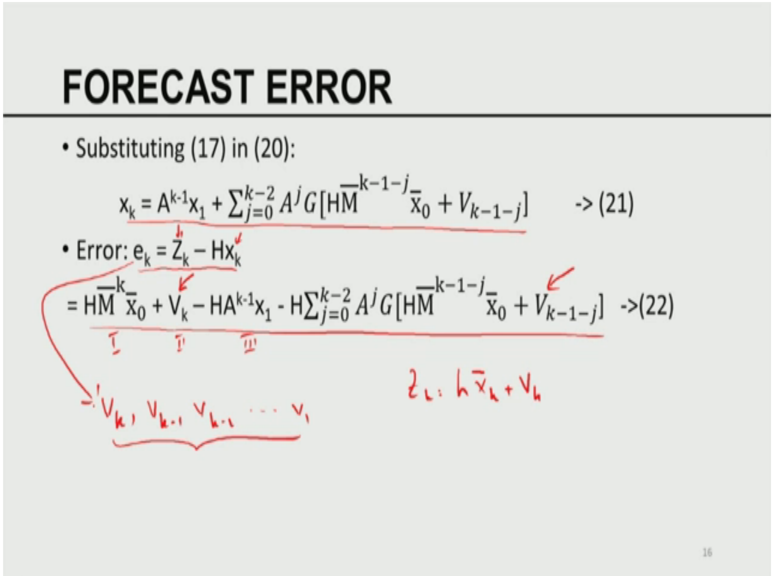


So, x 1 is equal to M of x naught, sorry. So, I am now going to.

Talk about the nudged forecast model. In the previous case we talked about the unknown true forecast model. M may not be equal to M bar; that means, that there is a model error. So, what is that I mean? I am going to talk about I would like to be able to simultaneously arrange G such, that it not only corrects for the model error also it corrects for the initial condition error. So, I am trying to kill 2 birds in one stroke. So, let x 1 be equal to M of x naught. X naught be the initial condition for any k greater than that M of x k plus G times z k minus h k. That is the nudged model. The nudged model can be written like this. Because I am considering a linear model to be able to do things a little more precisely. For nominee I would like to be able to expose the difficulty for the linear model, if the linear model is difficult the non-linear model at least is one nudge more difficult than the linear model.

That is the aim in here in this discussion. So, my nudged model becomes 19, where the matrix a is equal to M minus G h. G is the gain to be determined, h is the forward operator, M is the 1 1 step transition matrix for the linear model. So, I am assuming the model is linear, the observations are also linear function of the state. So, iterating this equation 19, I get this. And z k contains the unknown truth plus noise. So, the noise is the embedded within the z k term. The noise is embedded within the z k term. I hope that is clear. So, 20 is obtained by iterating 19, a simple iteration gives you the expression. So, what is this expression the nudged state at time k depends on the state at time x 1 plus anything beyond substituting 17 in 20. So, let us go back to this is 20. What is 17? 17 is an expression for the observation based on the models solution.

(Refer Slide Time: 42:25)

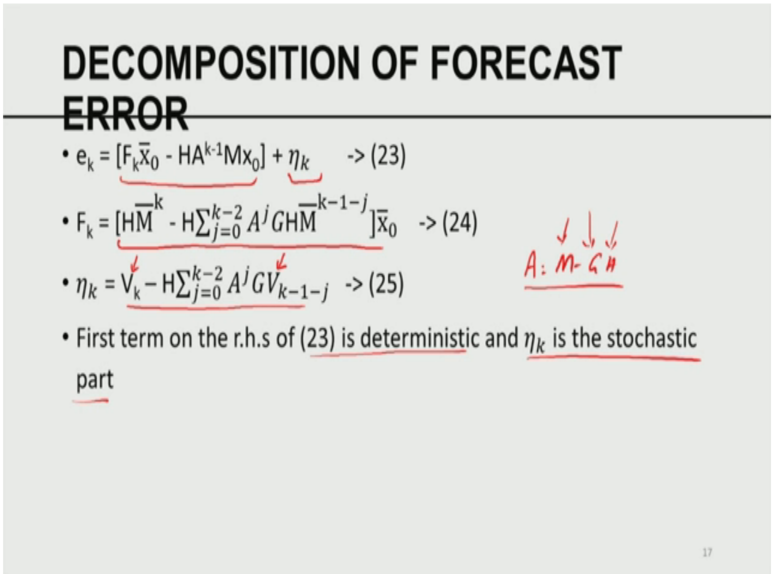


You get an expression for x k which is the model state the forecast stage given by this expression, is a funny looking expression. A little complex expression, but I do not think there should be any difficulty in trying to verify that. Why am I trying to find that? Because I would like to be able to pin down the forecast error. So, z k minus h of x k is the forecast error. Now, therefore, I know z k z k please remember z k is equal to h of x k bar. Plus, V k h of x k bar we have already computed in equation on 16. Based on h of x k bar I also have in relation for z k from 17. So, I can substitute for z k from those equations in here. I have already computed x k in the previous slide let us look at it once more that is the equation 20. So, substituting all these things, what do I get? I get an explicit expression for the forecast error in the nudging model. Why is that? If I want to be able to compute the forecast error covariance, the serial correlation I need to be able to get an explicit expression for the forecast error itself that is the first step. And that is what we have accomplished. So, look at the structure of 22.

22 has several terms. This is the first term. This is the second term, this is the third term, this is the 4th term is the summation. The 4th term itself is a sum of 2 terms, but among all these V k is a noise term V k minus 1 minus J is a noise term. So, there are 2 nice terms, the rest of it are deterministic terms. So, the error is noisy. You can also see the error of time k depends on V k. As well as error I am the; I am sorry, the noise at time V k as well as the noise at time 0 to yeah, 0 to timetables k minus 2. So that means, there is a serial dependency among all these noise expressions. So, when J is you can you can really see. So, I am I am I am depending on V k.

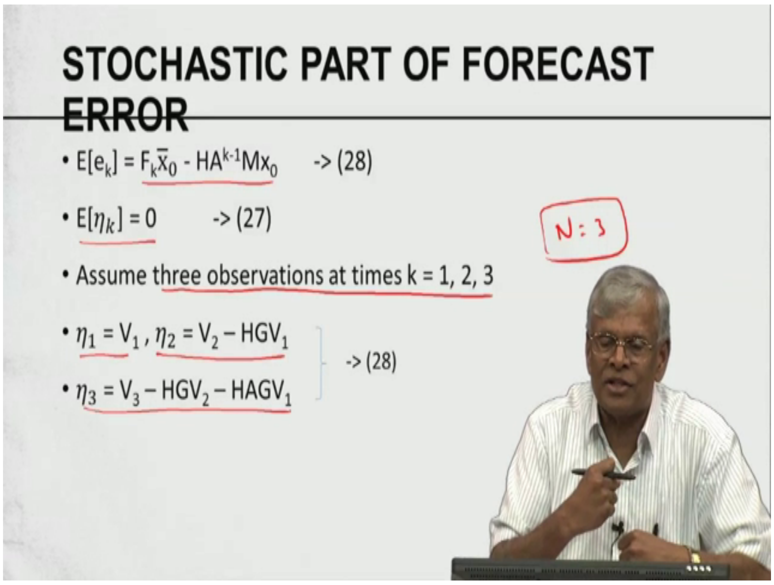
Then when J is 0, that depends on V k minus 1. So, let me write that down. Then J is equal to 2 that depends on the V k minus 2. When J is equal to k minus 2, that depends on V 1. So, therefore, e k the error d k does not only depend on V k, but also the entire sequence. In the past it is this dependency of e k, on the entire not noise sequence up to including time k, that it uses the serial correlation. I hope that part is clear to you. It is this serial correlation I am now going to have to extract.

(Refer Slide Time: 45:43)



I am now going to rewrite this expression e k for the sake of convenience, your deterministic part plus the random part. The deterministic part has this expression the random part has this expression. So, 24 and 25 correspond to the domestic part of the forecast error, and the random part of the forecast error. So, you can see 23 is deterministic eta k is stochastic. What is the stochastic part? Stochastic part is again the noise. From the past the noise from the present, the past noise are weighted by the powers of A and you may remember A is equal to M minus G H. So, so what does it mean? The random part of the forecast error depends on the model dynamics depends on the forward operator it depends on the model dynamics A it depends the forward operator H, it depends on to be chosen the matrix G which is the gain matrix to be used in nudging. And of course, all of the errors in the observation starting from time 0 to this time to the time k.

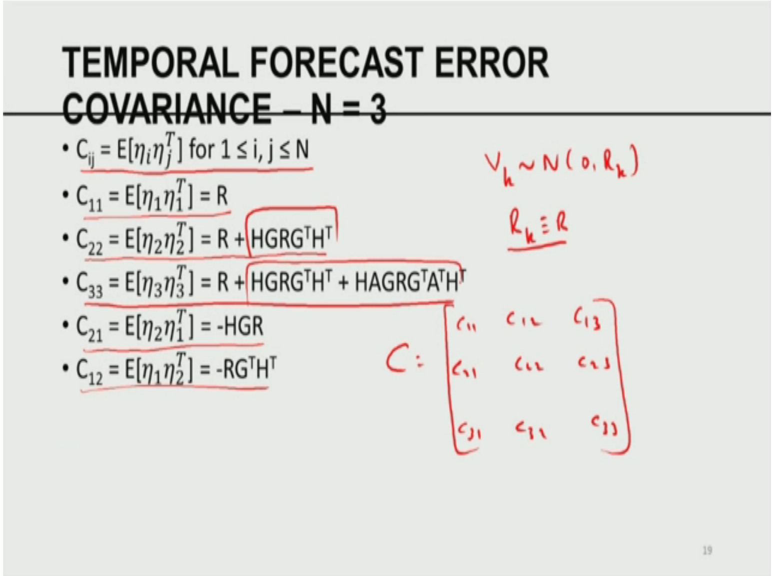
(Refer Slide Time: 47:06)



Therefore, the expected value of e k is equal to the deterministic part. The expected value of eta k the random part is 0. Now I am going to further. If I assume the a general value of n, the expressions are more gets more complex. So, instead of n, I am thus assume n is equal to 3 to just to get a feel for the expressions in this in this in this quick discussion. So, let us assume I have 3 observations at time k is equal to 1, 2 and 3. And so, by specializing this. Now look at this now the previous expressions, they go for k 1 to n. N is the last observation time. To simplify to get of to get that aha feeling I am simply going to assume n is equal to 3 with our loss of generality.

So, if I substitute n is equal to 3 and simplify, the expression there the expression for the random part of the forecast error which is eta. Eta that is look at this, now eta one depends on V 1 eta 2 depends on V 1 and V 2 eta 3. Depends on V 1 V 2 V 3 eta 4 will depend on V 1 V 2 V 3 V 4. So, that this means etas are correlated. It is this correlation makes the 4DVAR scheme, little defective in the sense they have not taken, the entire weight function that accounts for the serial correlation.

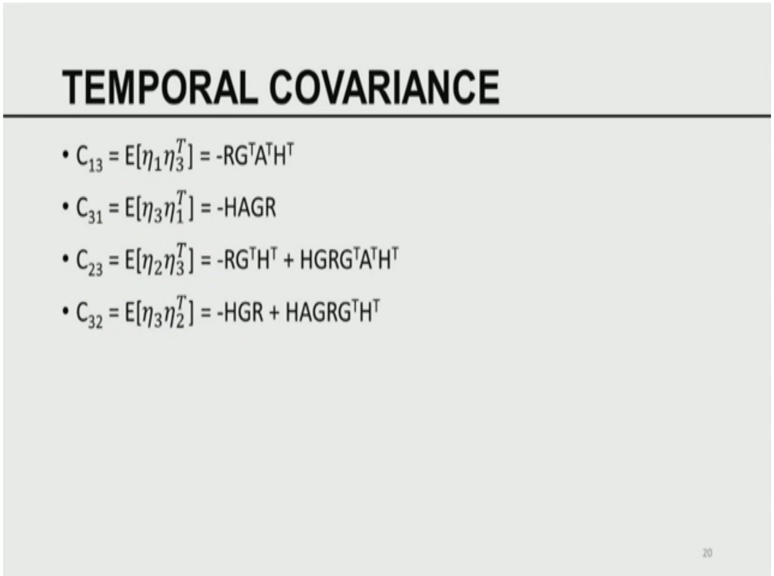
(Refer Slide Time: 48:46)



A simple exercise in statistics, computation of correlation tells you C i j which is the correlation between errors at time i and errors at time J is given by expected value of eta i eta J transpose. In the case when n is equal to 3, eta 1 1 is given by R. What is R? R so, I am please go back now. V k is equal to M of R k, I am it is not is generally assumed R k is identically equal to R, why vks are coming from instruments. When we buy instruments, we buy instruments in bulk. So, I am going to assume all the instruments, that make measurements are the same type; that means, the covariance of the error the error characters of the instruments are the same. So, R k does not depend on k R k simply is R. So, that is a very useful assumption. Even though the theory I can continue with R k, I do not want to unnecessarily complicate the expressions by trying to be general. There is no loss of generality in assuming R k is R.

So, that is R C 2 2. So, what are this? So, C is a matrix, which is C 1 1, C 1 2, C 1 3, C 2 2, C 2 3, C 3 3 everything because it is symmetric, I do not have to worry about the bottom part. But I can continue the C 2 1 C 3 1 C 3 2. So, I am going to compute all of these elements. You can see I have computed all of these elements. Like this, look at this now R, R plus this R plus 2 terms. C 2 1 C 1 2, now look at this C 2 1 2 1 2, they are that transposes of each other we will talk about the symmetry of the resulting matrix in a minute. But I am trying to give you the exact expressions for the C s. So, let me go back and remind you once more. I would like to x, I would like to be able to understand the presence of serial correlation, n observations to make life simple I assumed n is equal to 3. So, I substituted an n is equal to 3 in the expression for the forecast covariance. Especially the random part of the forecast covariance using the random part of the covariance, I am simply computing expressions for these co-variances.

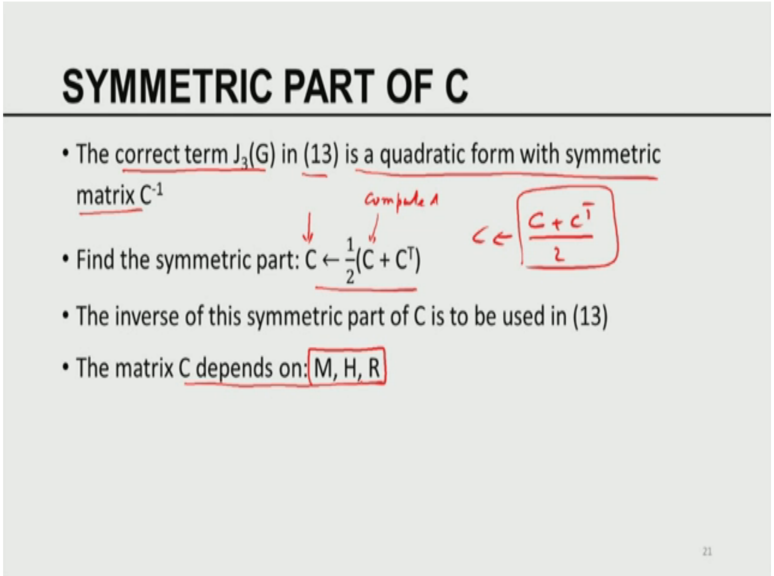
(Refer Slide Time: 51:21)



So, C 1 3 C 3 1 3 C 3 2. You can you can really say can compute this. From this you can readily see they are all related. So, what if I, if they so, in the early methods based on 4DVAR what did they not use? They did not use this term. Sorry, they did not sorry, they did not use this term, they did not use this term. So, these are the new term that comes into the picture.

And it is these new terms we are interested in in incorporating. I hope these expressions are clear. These expressions can be very explicitly evaluated by the closed form expression for the random part.

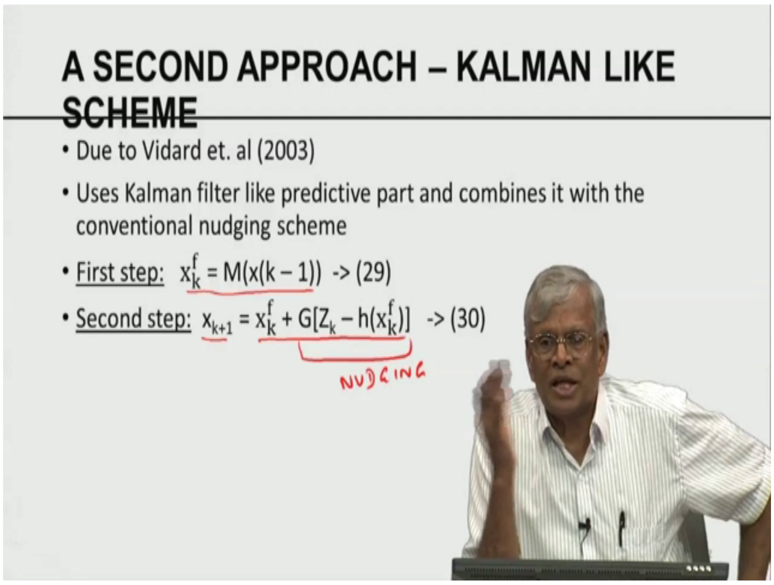
(Refer Slide Time: 52:17)



So, I have I have compute a matrix C. So, let us go back now. I have computed the matrix C. Please understand, quadratics in trying to consider a quadratic forms only the symmetric part of the C matters. So, the symmetric part of C. So, C is equal to C plus C transpose by 2. So, we compute C as I have done. And compute the kind of concern the symmetric part of C. So, this the new C is called the symmetric part of C. Now the correction term J 3 now look at this now the correction term J 3 in 13 let us go back. So, 13 a; this is the new J function. Z inverse is to be used here please understand, C inverse is used to be here. It is generally the k is that I need to know the weight matrix if I am going to consider the weighted least squares, I need to know the weight matrix weight matrix is C inverse I have computed C. So, in principle I can compute C inverse. Therefore, the correct term the correct term J 3. So, J 3 the correct term in 13 is the correct form with the symmetric matrix C inverse, where C is the symmetric part of the computer. So, what is this C? This is the computed C.

This is the C that is used in the left-hand side C is the one that is used in 13. So, you compute the deterministic part of the computed C, and that becomes the new C, who C inverse is the one that is going to be used in the quadratic function. The inverse of the symmetric matrix of C is the one that is used in in in in in 13th. And I would like to reemphasize the fact which I have already mentioned this matrix C depends on M H and R. What is M model? What is H observation operator? What is R noise property? So, you can readily see it the serial covariance is a function of the model the forward operator, and the noise all the 3 players in the game.

(Refer Slide Time: 54:47)

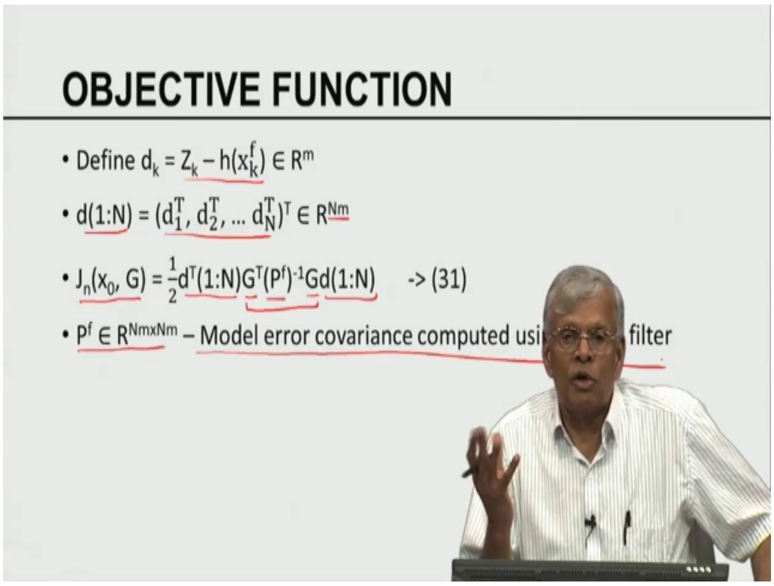


So, I have now completed one aspect of the estimation problem for the optimal G. This is using 4DVAR. So, we talked about what they did, and we talked about what is wrong with it, and we also talked about what is the meaningful way to correct it. Now I am quickly going to provide a review of the second approach that was used in the post 90 era based on 2 stage Kalman. Like scheme it was introduced by a group of French atmospheric scientist vidard et al 2003. This uses a Kalman filter like predictive part that combines it with a conventional nudging scheme. So, you can readily see so, you can now see the following idea. If you knew some of the basic approaches to assimilation that we have covered in this class you can hybridize these methods to be able to generate newer methods.

So, in this course, we are not going to be talking about all possible methods of hybridization. We are going to be we have described all the methods in their purest form. Because before you can hybridize you need to understand what is the power of each of these techniques. Therefore, this being the first level course at the graduate level we have emphasized. All the basic tools which if well understood, not only be applied directly also can be used to devise newer schemes for data simulation and other problems; that is the idea. So, this is an example of such hybridization. So, what is the thing in here. The first 2 step is the following. Let x k minus 1 be the state I know I am using the model to create a forecast. The second step is I am going to do an analysis, which is forecast plus G times z k minus h of f. So, what is this? This is the nudging part. So, you make a forecast. And then you create an analysis. The forecast comes from the model.

The analysis comes from the nudging the nudging uses a G. G plays a role of a Kalman filter. And they would like to be able to determine G. Using method similar to the arguments of Kalman filter. You can see the how the hybridization comes into play. So, that is what I am going to quickly describe.

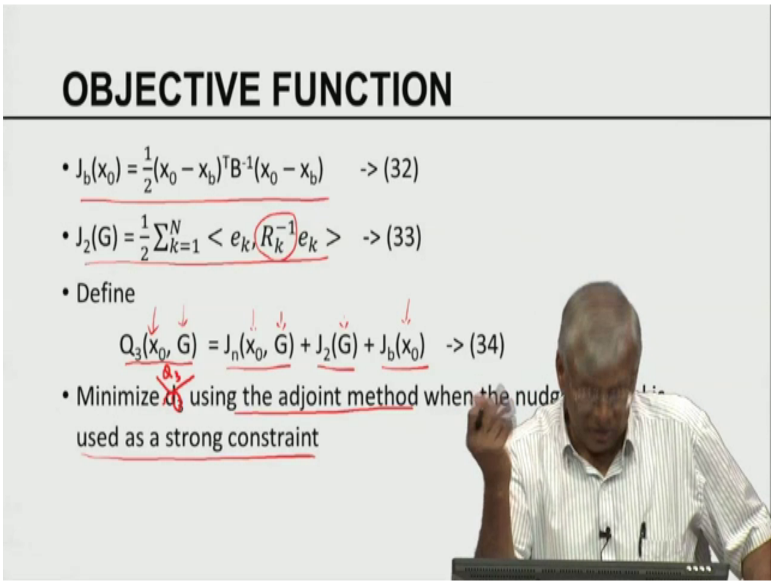
(Refer Slide Time: 57:18)



So, define d k which is equal to z k minus h of x k f, which is the innovation. The new information that G k contains other than what the observation gives you. So, d is the vector of such innovations. D is again a vector of size N m now. I am going to concoct N m m J, n x naught of G, J n is called the nudging induced a cost function. So, that is equal to the transpose of this.

G transpose p f inverse G d f n. So, you can think of these 3 as part of the weight matrices. P f is the forecast error covariance. So, that is very similar to the one that comes in 4DVAR like scheme. Here, because of the way that G appears in in in equation 30. 31 is a very natural way to be able to consider a J function. So, the model error covariance. So, here look at this now in here p f is the model error covariance. Using Kalman filter, you may realize even if the model is linear, computing the forecast covariance involves 2 matrix multiplication. So, computationally it is it is much more expensive than the 4DVAR based idea 4DVARbased idea.

(Refer Slide Time: 58:57)



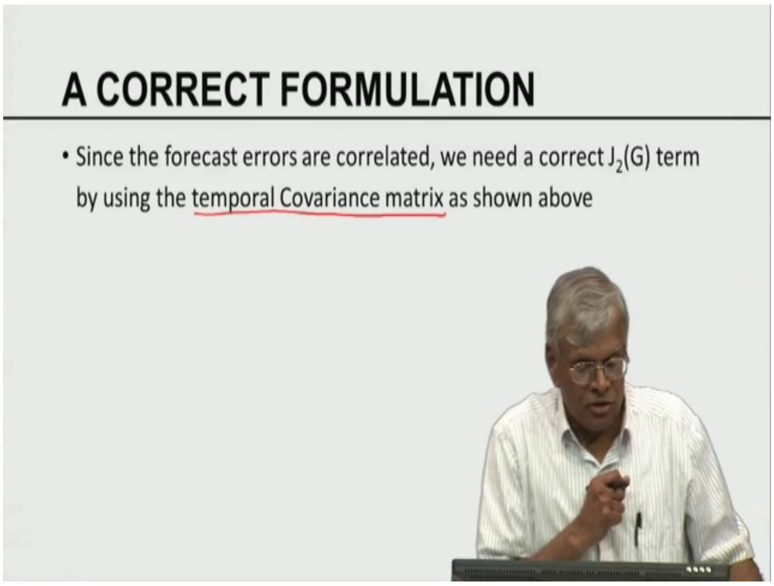
Therefore, they concocted several components for the overall minimization. One is the background term. What is background term? X naught until now we did not worry too much about x naught. So, initially I may allow the error, but as the system picks up in time the error will reduce that was the basic idea. Now they would like to be able to start with some background information for the model initial state itself. So, that goes to show you the flexibility of how many such terms if you knew you can add to the objective function to be able to create um a solution that takes care of several pieces of information that you may want to bring to barrel the problem. So, J 2 is essentially the sum of square of a criterion. Please remember, they used R k, yes. We have now argued use of R k is not correct because the forecast errors are correlated.

So, in that sense the use of R k in 331 essentially closer that I to the presence of serial correlation. So, what is the best way to do it? You still need to be able to get an expression for the forecast error, and need to be able to compute the serial correlation. Essentially, you are trying to talk stochastic dynamic models with stochastic observation within the context of nudging. So, when you are trying to do everything stochastic you need to call spade a spade and using R k does not fit that pattern. i that is one of the observations that we have made. So, you can cut a new cost function. Please understand, until now we only consider the cost function is a function of G. Now the cost function is a function of x naught and G.

So, it is a slightly extended formulation. So, J and G we have already seen J 2 G we have already seen J b x naught. So, there is a penalty coming out of x naught, there is a penalty coming out of G there is a penalty coming out of x naught and G. So, our objective is to minimize not d 3. I am sorry, it is Q 3, sorry minimize Q 3, using the adjoint method. When the nudging model is used as a strong constraint now look at this. Now again they concoct a common like scheme, but they want to be able to find the initial condition, and optimal G by method similar to the adjoint method when using the adjoint the nudged model as a strong constraint. You can see the power of the 4DVAR like principles where you can apply it repeatedly whether it is initially whether it is estimating the initial state or parameters or anything else.

So, in this case G is a parameter, you can think of it for the nudge model.

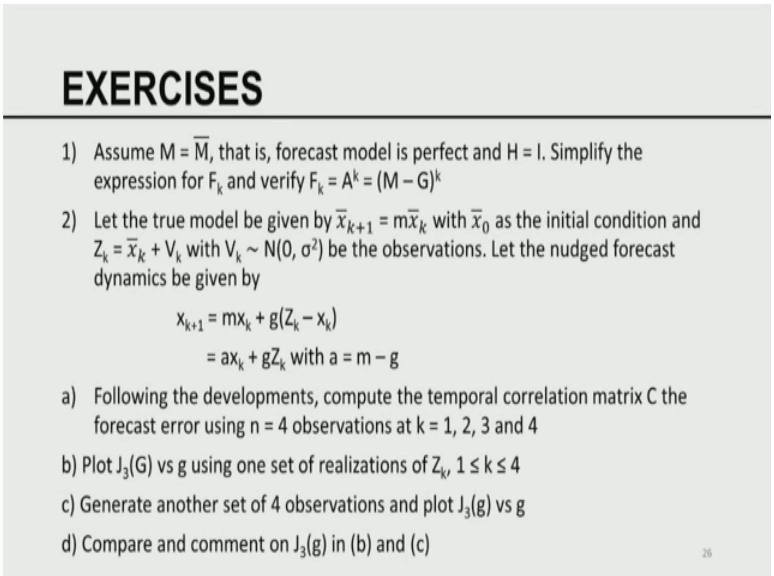
(Refer Slide Time: 62:05)



Since the forecast errors are correlated, we need to correct J 2 G. Please understand, we need to be able to correct J 2 G J 2 G is given by 33. So, this is what we talked about. R k the use of R k inverse. So, all the other terms are closer. The only term that does not fit the bill is; because, 33 relates to the total sum of squared errors weighted sum of squared errors the waiting is not appropriate, the waiting is incorrect. So, we can again correct the weight function by appropriately computing the seal the serial correlation. So, the temporal covariance estimation is an important part of this. We only cited the need for computing this temporal correlation.

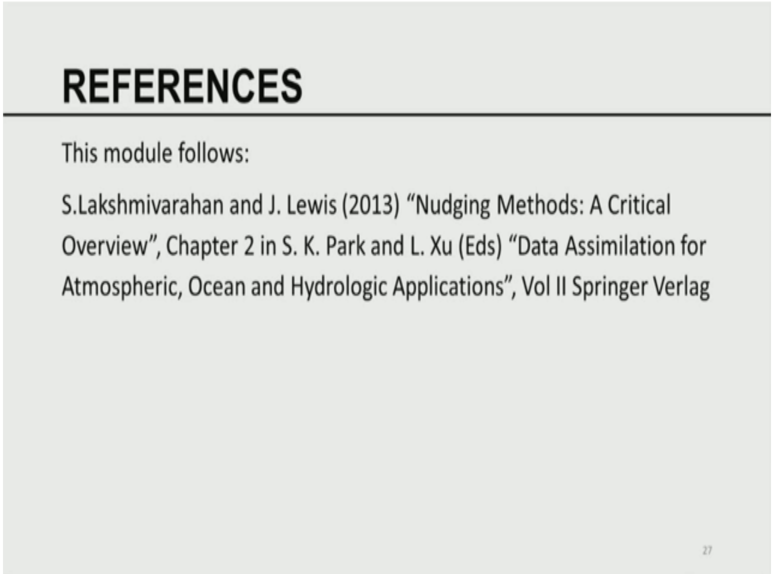
We have not done this explicitly. I think it will be an interesting exercise for somebody to be able to take up this 2-stage nudging scheme, that involves 4DVAR and the Kalman like scheme. And compute the serial correlated errors. And if you if you use that you should be able to find out what is a good scheme, what is the good method. So, that could in my view a good starting point for probably a master's thesis.

(Refer Slide Time: 63:29)



With this we have provided you a major all the major ideas that relates to the development of nudging schemes. We have given several exercises there are extensions of the discussions that we have had.

(Refer Slide Time: 63:46)



This module follows a paper that we wrote in 2013. Lakshmivaraham and Lewis, nudging method a critical overview. This appeared as a chapter 2 in a book entitled data assimilation for atmospheric ocean and hydrological application. It is the second volume published by springer Verlag in a series, edited by Sangee Park and l x. And that paper contains lot more information about basic nudging. We also leaded to the relation between observer theory as was developed by Leuenberger in 19, the early 1960 62 63 64 in the time frame work. And in that in our paper in our critical review, we have talked about the intrinsic relation between observer theory and the nudging theory to be able to see how observer theory can help to be able to design better nudging schemes. Nudging schemes are general schemes, which are which are very useful class of methods for forcing a model towards the observation by using the notion of this state feedback. So, with that we conclude our discussion, an introductory discussion of nudging methods.

Thank you.