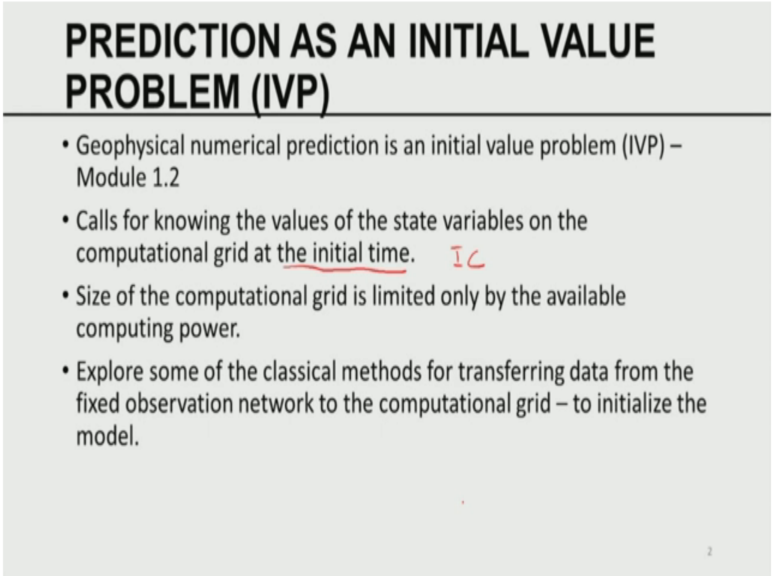
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture - 29
Initialization Classical Method

In this module 7, we are going to be talking about some of the classical methods for data assimilation.

(Refer Slide Time: 00:27)



These techniques look at prediction as an initial value problem. So, I am going to provide some background to motivate the kind of algorithms that has been used in the early 50's when data assimilation began to be used by national centers for weather prediction. So, it was realized long back that, prediction numerical weather prediction is an initial value problem, you may have we may have already alluded to this in module 1.2 when we talked about data mining, data assimilation and prediction, we argued their data mining, data assimilation and prediction are three different components of the predictive science.

So, the notion of prediction becomes important, if you are going to be using dynamic models to create prediction it has been known for a long time at least since early 1900's, that prediction as a mathematical problem is an initial value problem. This acceptance of the fact that prediction is an initial value problem for a given class of dynamic models calls for knowing the values of the state variable on the computational grid of the initial time for example; if I have a dynamical model, if the solutions are dynamical model are predictions or you are going to generate for cash product as a functions of the solutions of mathematical model to be able to pull the model solution I need to be able to get it started at a given time.

And, so if I assume T is equal to 0 as a given time, I need to be have the I need to have the values of the state variable on the entire computational grid at initial time and that in mathematics we generally call initial condition, so to get the initial value problem going I need an initial condition. In here, I would like to bring couple of other constraints the size of the computational grid there is often used in prediction is often limited by computing power, the larger the number of points in the computational grid larger is going to be the time required.

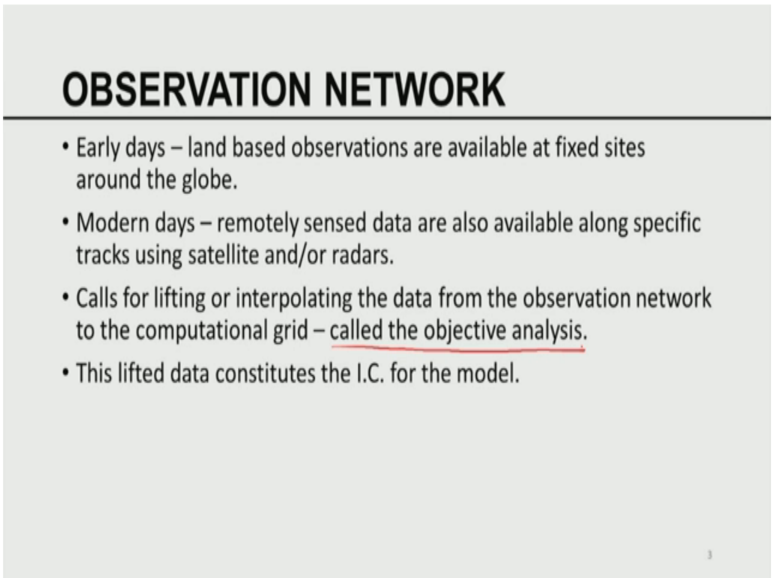
So, in before you deciding on the size the computational grid, in trying to solve the prediction problem as an initial value problem you need to make sure what kind of computing power you have available for you to be able to create the prediction. So, we have; so we will now assume some of the few things that are needed to get going I know the computing power contingent on the computing power I have already decided on the size and the maximum size the grid I could compute, I have the model dynamics, I have the model dynamics discretized.

On the chosen computational grid which is consistent with the computing power, so continuous time models have been discretized have been reduced to discrete time models, whether it is a continuous time model discrete time model so long as a dynamic model I need initial conditions. So, I am now going to explore some of the classical methods for transferring data from the observation network to the computational grid, so as to initialize the model.

In the early days, while the computing power was not very high, so they were limited to regional models or very course global models, the standard models are given by partial differential equations. So, you discretize the given model on the grid to be able to initialize this model I want to have initial conditions, initial condition is the unknown, we have been talking about estimating various quantities here the thing that I have to estimate is the initial condition. What is the input data from which I am going to have to estimate the initial condition that comes from the data observation? So, I am assuming they are given a set of observation stations.

In the early days, in the pre satellite pre radar day era they essentially had ground stations, they essentially had balloons in different parts of the world, so the sensor network was there is sparse the amount of the observation that were available for use in estimating the initial condition was much smaller, so these are all some of the background information one need to keep in mind to understand how? Before powerful computers came to be before powerful methods of dynamic data assimilations were used to be able to generate prediction. How they managed to create reasonably grid prediction? How they estimated the initial condition from the data? These are the class of problem that we are going to be dealing with.

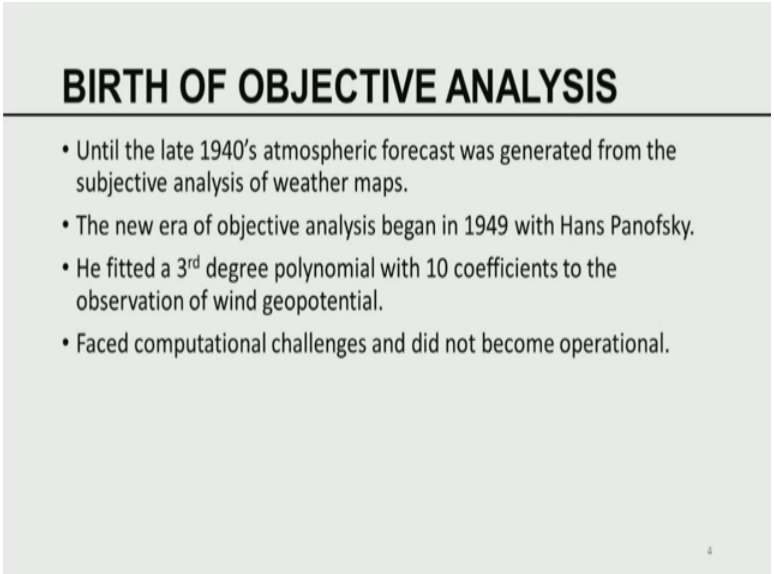
(Refer Slide Time: 06:24)



In the next in this and the next couple of modules, in the early days observation network very sparse, they are essentially land based observations available at fixed sights around the globe, moderns days remotely sensed data are available along specific tracks using satellites or radars, this calls for lifting our interpolating the data from the observation network to the computational grid. In the early day is this aspect of transferring the information from the sparse observation network to a denser computational grid was called objective analysis.

So, you can in some sense say objective analysis where the foreigners of the modern day data assimilation systems that are used by meteorological centers around the world, this lifted data will constitute the initial conditions for the model, so once the initial condition was fixed the model ran forward we made prediction.

(Refer Slide Time: 07:32)



At this time I would like to be able to recall the relation between this and what we did in 4 d VAR, in 4 d VAR what is that we assumed? Same kind of thing I have a model, I have observations at different times may perhaps in different locations I would like to be able to fit the model solution to the observation the least square sense, we used an objective function to be able to determine the initial condition once I have determined the optimal initial condition using 4 d VAR forward sensitive method we then ran the model forward in time, this run of the this forward run of the model starting from the optimal initial condition provided a reasonably good forecast.

So, you can see the 4 d VAR is the method that came to be used around 1980's the classical methods began in 1950's. So, you can see the similarity in the way in which they tackle the problem of trying to find the optimal initial condition those days in the pre 4 d VAR era, in the pre formal data assimilation era, so that is what I am trying to provide a background on. So, until the late 1940's atmospheric forecast was generated from essentially subject to analysis which are based on analysis of weather maps, so that is what is called subjective analysis, is the only tool they had is was available to them was the isobaric surfaces depicted as weather maps.

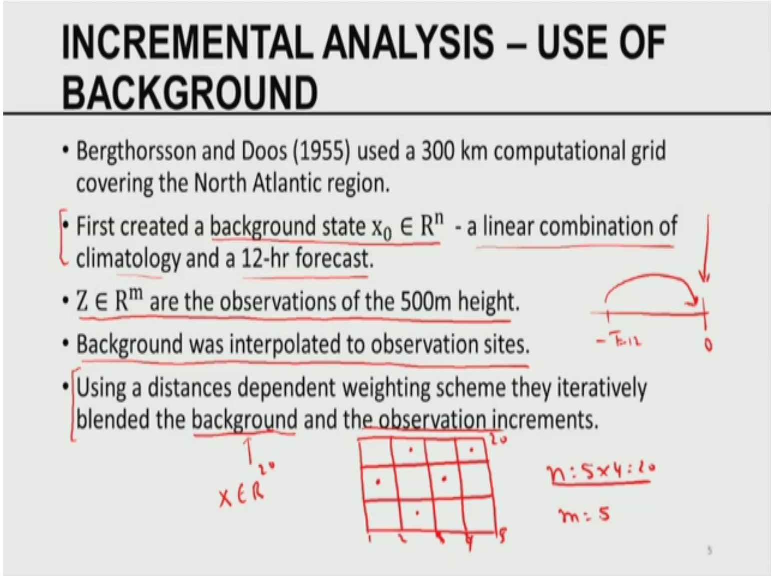
The new era of objective analysis began in 1949 with Hans Panofsky. He for the first time began this formal method of being able to transfer the information from a sparse observation network to a denser computational network to be able to lift our spread the information from observation network to the computational network to that end he fitted a third degree polynomial the 10 coefficients to the observations of wind and geo potential.

So, one could say within the parlance of atmospheric sciences 1949 the work of Panofsky may have been one of the early work which he called objective analysis which resembles the modern day data assimilation framework, but this approach in 1949 please remember that the stored program digital computers came into existence only 1951, 52(Refer Time: 10:42) this group used, the first available stored program digital machine in 1951, 52 to solve the geostrophic vorticity equation and they made a first 20 were forecast, this was pre computed era the computer era just about to break loose in 1949.

So, he faced lots of computational challenges and unfortunately while the idea is decent and workable it met with several operational difficulties, mathematical computational difficulties, nevertheless it is it is not out of out of consideration to consider to think of Panofsky as one who sowed the seed for an objective analysis of transferring information from observation to someone to estimating unknown state, so you can see the semblance of the ideas of estimation coming from the world of atmospheric sciences.

Shortly after that in the mid 50’s; Bergthorsson and Doos used 300 kilometer computational grid covering the north Atlantic region. So, they because of the computers were limited, 1955 stored program digital computers have come into existence I want to give you an idea of how sophisticated a computer technology at that time was? They only had a computer they did not have any programming language they did not have any compiler. So, when (Refer Time: 12:20) is group use the computers to make the 24 hour prediction they essentially program everything in machine language, anybody who has done programming in assembly level language you would know how involved it could be? To be able to code all the programs in the machine level language using simply codes using strings of 0’s and 1’s.

(Refer Slide Time: 11:48)



The Fortran was invented only around that time, the Fortran compiler came to be by mid 1950's, so this notion of being able to write programs in a general purpose language such as Fortran became a common place around the mid 1950's. So, it is everything was very raw people didn't quite understand how to utilize this monster call computers, so along so much have to be developed in terms of system software, programming languages, compilers, operating systems all these things were trying to develop it is at that time, Bergthorsson and Doos use a computational grid 300 kilometers to do a regional analysis covering north Atlantic region parts of north America as well as Europe.

They created; so they wanted to be able to bring in the observation on to a grid, to be able to have a very nice initial condition from which to run the model forward the model solution will become forecast, but they had a very good idea, so what did they do? They taught of two pieces of information. Why? They thought that, I already have climatology based information on the grid which is called the background state. So, they introduced the notional what is called the background state sorry the introduced the notional what is call a background state.

In modern language what is background state? Background state is a prior, it is the belief that you have about the state of the system before you took the very first observation. How did they create the first background? They created as a linear combination of climatology and the procedure that they use using which they had made a forecast the 12 hour forecast. So, this is the present time, this is minus T they ran the method from minus T to 0 forward, that is the 12 hour T is equal to 12 hours. So, minus 12 hours to 0 they made a forecast, so that forecast had information on the grid about the state of the system they are looking at.

Of course, they also have the information from climatology they simply took some arbitrary linear combination of the climatology and the forecast from 12 hours, to start with you may ask where do they get the forecast from, so to start with to it they did not have any forecast, so they essentially used climatology to do the things. So, this is one way of trying to initialize the model in this particular case no observation can be is used, but then nevertheless having a background information on the grid is one piece of information you can see the beginnings of Bayesian philosophy right there.

What is the element to the Bayesian philosophy? There is a prior is a belief I had before I started doing anything. Then I started making observations give me some new information Bayesian philosophy always looks at combining the prior belief with the new information to get the post area, you can see the elements of that Bayesian philosophy inherent in 1955 within the context of atmospheric science prediction in the work in the work of Berghtorsson and Doos. So, in other words they created a state from all using all the available information at that time you can think of it like that then they had access to observations of 500 mill bar height.

What did they then do? So, you can think of it like this now, this is the computational grid; so 1, 2, 3, 4, 5, 1, 2, 3, 4 so n is equal to 5 times 4 is equal to 20 computational grid, let us pretend the observation locations are in those points. So, given this grid and the location for the observation so you can think of the dots at the observation network, mathematically one can think of interpolating data from the observation network to the grid are from grid to the observation network, that is a simple mathematical interpolation scheme I can do. What is called? By linear interpolation scheme which are very simple.

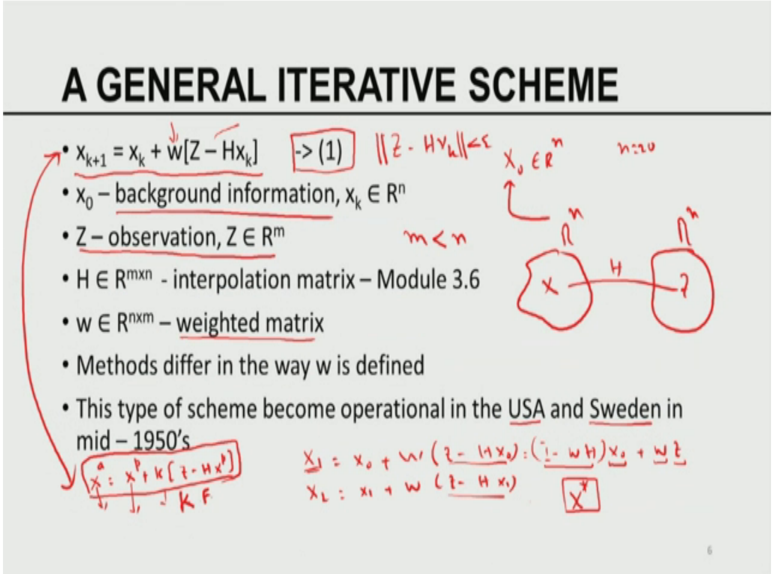
So, what did they do? They first built the background our information on the grid, then they interpolated this background information on the grid to the observation location, so at the observation location they have the background information, which is obtained by interpolating the background information on the grid to the observation network, then at the observation network you conduct observation, observations are available made available, so that is the second piece of information that you have at the observation of the observation locations. So, what are the observations here? The observations are essentially 500 mill bar height.

So, they had an estimate of the 500 mill bar height from the background information, they have the observation of the 5 hundred mill bar height from reality, they have already interpolated the background to the observation side, so at each observation site they have two pieces of information. Then what did they do? They used a distance dependent weight function to iteratively blend the increment the background and the observations.

That is the very beautiful heuristics scheme it worked, it became to be used in operational centers you can see the fundamental philosophy of Bayesian inherent in here, so instead of working in the computational grid they worked in the observation space. So, in modern language you can say you can do assimilation in the observation space are in the model space, they had a two way communication between the observation network and the computational network which is the grid by developing this 2 dimensional bridge between these two worlds of observation the world of observation network and to the world of computational network they were able to go in between, so they converted everything on the computational grid to the observation network.

So, at the observation network they have the representative background information, they have their representative observation increment observation information they simply blended them in a heuristic way. How did they blend it? They essentially blended heuristically using a weighting scheme that depends on distance that depends on distance. So, that is the fundamental idea of some of the earliest work that happened.

(Refer Slide Time: 20:43)



I am sorry. Now, we are going to mathematically describe a general iterative scheme that embodies the vectors and those ideas that is what I am going to be talking about in a mathematical form. So, let us pretend, so you can do this iteration sorry you can do this iterative process as I mentioned in the previous slide either on the computational network which is the computational grid or the observation network. Now I am going to reformulate that problem as one of doing and the computational grid itself. So, let us start the process.

So, let us assume I have a computational grid, the computational grid such as the one that I had given in the previous one, so let us assume I have a 20 point computational grid, so in this case X is a 20 dimensional vector X 1 to X so I can number them as 3, 4, 5 all the way up to 20, it could be; so in this particular case it is the 500 mill bar height that is the state of the system that is being operated on. So, let X naught be a vector on this grid of size 20 that provides that gives that that represents the background information. What is this background information? This background information X naught belonging to R n, so n in this example is 20.

The background information is obtained as a linear combination of everything I know from previous forecast that are available climatology information so on and so forth. So, I develop a mosaic on my grid, so X naught background information is simply a mosaic a smooth mosaic of the state obtained from anything I can put my fingers on. Then Z is the observation that comes to us, Z is R m please understand R n and R m, m and n are different in general m is less than n, I had more number of grid points less number of observation that is how they started in 1950's until recently until remote sensing became possible such as radar such as satellite the amount of observation is always smaller than the size of the grid, with satellite and radar the amount of observation has become more and more computing power has become better and better.

Our desire to increase the number of computational size of the computational grid namely; to reduce the grid spacing becomes more and more possible, so there is a race between how finer I can make the computational grid making enlarger I can also there is also a raise but how much more observations I can have from satellite radar and everything else put together, so m is keeps increasing n keeps increasing as our ability to observe nature becomes better and better as our ability to create computers becomes better and better.

So, m and n are not fixed they are changing in time from era to era to era to era. So, H is the interpolation matrix, we have already talked about how to design H in our module 3.6 just to refresh your memory, H is a matrix so this is the model space, this is the observation space, this is X, this is Z, this is H. So, H is a mapping from from model space the observation space is called the forward operator, sometimes H comes out of the physics, sometimes H comes out of empirical rules, some time H comes out of interpolation formulas.

So, if I have a grid as in the previous case if I have a so in this particular example n is 20 in this case m is essentially 5, I have 5 observation locations and 20 computational grid m is less than n, so this is one of the situation that they were facing in there in those days the idea of this presentation is to make clear how clever they wear in the pre formal data assimilation era how they mimic some of the ideas and how they developed some of the ideas we know much more formally and they had a lot of intuition to be able to come up with these ideas.

So, H is an interpolation matrix, so I can go between the computational grid and the observational network. So, I have now everything I have a prior information I am sorry I have a prior information given by given by X naught, I have observation I have H. If either to have done Bayesian I would simply combine X naught with Z, either net well either using not either using it using Bayesian framework and had created a posterior and use the posterior as my new initial condition turn the model forward in time, but in those days what did they do? They picked awaited awaiting matrix this weight matrix is the one that has a distant dependent weighting scheme.

So, almost all the methods that were used in the 1950's and 60's in different parts of the world, essentially U S A and Sweden is in is spearheaded the development of these schemes, Berghtorsson and Doos they are doing in Sweden ah. In U S A the national whether center also developed similar iterative schemes, both the schemes have similar mathematical structure they deferred in the way in which the matrix w the weight matrix w was defined and used, this type of scheme became operational in the early to mid 50's.

So, what is the basic idea? So, now let us go to equation 1. X naught is the background Z is the information, H of X naught you can think of it as a model predicted observation, Z minus H of X naught is the new information that the observation has the background did not, so you can think of Z minus H X naught H of X naught has the innovation you multiply by a weight you add to X naught. So, what is that you get? You get X 1 is equal to X naught plus w times Z minus H of X naught.

So, X naught did not have any information about Z, Z did not have any information about X naught, but X 1 is the linear combination of X naught and Z, you could have also re written this to be i minus i i minus w H X naught plus w Z, so this is the weight matrix that is used for initial condition, this is the weight matrix that is used for observation. You can think of X 1 to be the linear combination of the background information and the observation, this is what I want to see the emphasis here is that you can see the Bayesian philosophy, so they are trying to use the Bayesian philosophy with are explicitly formalizing it in the Bayesian framework. So, they were very clever to anticipate some of the things to come.

Then what it is what did they do? They then did X 2 is equal to X 1 plus w Z minus H of X 1. So, what is the basic idea here? They were Z minus H of X 1 is the residual from X from using X 1, Z minus H of X naught is the residual from using X naught, but this second residual should be in principle better than the first residual, because the first residual did not has X naught did not depend on Z, Z did not depend on X naught, but in this case X 1 depends on Z and X naught. So, you can readily see this iterative scheme X k plus 1 is equal to X k plus w times Z minus H of X k became a very simple minded scheme.

You can see the beginnings of Kalman filter right here, we have already seen in the last lecture on Kalman filters, in Kalman filter what is the form? The posterior which we call analysis is equal to X prior plus Kalman gain times Z minus H of X prior a this is the form of Kalman filter. We have alluded to in the in the in the morning and we divided the basic equations. So, I have prior information Z minus H X p gives you the new information that the prior I did not have that comes from the new observation I multiply the new information by K k is called the Kalman gain, prior is also called background.

So, you can see previous information and the new information together makes X a which is called analysis, so posterior, prior, observation, innovation that we talked about in this language even though they did not talk in this kind of a language the equation 1 had all the underpinnings of the modern methodology except for all the mathematical artifice that goes with it, to me that is that is that is very interesting because without knowing many things they have good intuition. I also want to tell you one more Kalman filter was not invented until 1961, so in 50's when these folks were working in operational centers entrusted with the job of producing forecast they did not there is no Kalman filter to speak off, but there are clever people they know dynamics very well they also knew reasonably good statistics.

They were aware of good statistical principles for estimation and on the top of it they are very clever people they knew how if you have two pieces of information you need to be able to mix it that is a fundamental idea, that fundamental idea carries even today and is the centerpiece of any and every data assimilation scheme of all types. So, this became the equation 1 became the workhorse of the data assimilation industry and was used both in U S A and Sweden to be able to generate appropriate initial condition.

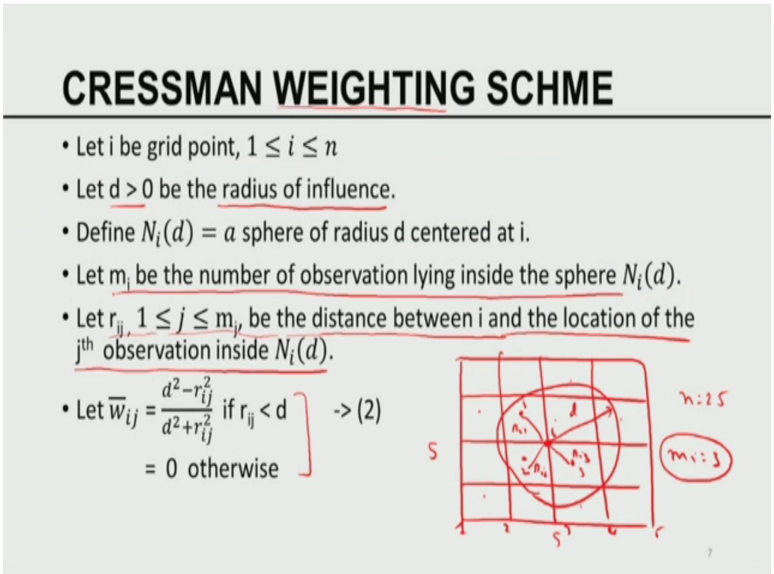
So, let us talk about that part now. So, if you run this iteration you need to truncate this iteration you cannot continue forever. When did you truncate? You truncate it the iteration when Z minus H of X k the norm of that vector became very small; that means, there is no more juice left in Z that has not been transferred to X k, if all the juice left in Z has been referred has been transferred to X k, Z might get closer to H of X k if Z H of

X k is closer to Z if H of X k is closer to Z the innovation becomes smaller and smaller if the innovation becomes smaller and smaller there is not much juice left I can truncate.

So, at the time when they truncated I am going to get a state, so let us assume they truncate it and at the result the state that results out of the truncation is called X star. What is this X star? This X star would now be used as the initial condition in the model from which they generate the forecast.

So, the X star is the initial condition that comes as an estimate by running one iteratively until convergence. The reason I wanted to bring this because even though this does not embody any of the mathematics that we have seen you can see the underpinnings of lateral date data assimilation methodology already inherent in these ideas that was proposed in the early and mid 1950's.

(Refer Slide Time: 34:21)



In the united states the person who has spearheading this scheme is cress man, in Sweden it was Bergthorsson and Doos the paper by cress man the paper by Bergthorsson and Doos even today if you start reading them it is very inspirational, how they thought of incorporating data with prior information to be able to have the good facility to be able to transfer data between observation network and the computational network.

Now I am going to talk about the weighting scheme, so what is the only thing that we have not specified how did they pick the weight matrix? How did the cress man scheme and the Bergthorsson scheme differ? They differ the way in which they describe are they picked the weight matrix, so essentially both they are running along similar tracks, but used different weights because of their belief in different philosophy of the influence of observation and the grid sorry yeah. So, now I am going to talk about the basic ideas the weighting scheme that is that was used by cress man, so weighting scheme.

So, consider the i-th grid point, the grid are numbered, so let us go back to the little diagram again this let this be the computational grid, this case; 1, 2, 3, 4, 5 I have 5 times 5, I have a set of 25 points so n is equal to 25, let me assume the observations are sparsely distributed like this maybe like that the grids are numbered; this is 1, 2, 3, 4, 5 and so on, so let this be the i-th the gird point, you pick a d greater than 0 and considered to be the radius of influence, the radius influences this was very good idea, so let us talk about this now.

Does the temperature in Bangalore India does it depend on the temperature in Delhi India? The distance between Delhi India's way too much, therefore you would expect the temperature between Bangalore and Delhi to be less correlated; however, if you consider a town which is only 10 miles away from Bangalore downtown there will be better correlation between that small town which is only 10 kilometers as opposed to over 2000 kilometers. So, what does what does it mean? If I consider a grid point i, if there are observations around the grid point which observation location will be affecting the quantity of state at the location i.

So, they said well like everything else in life I have to consider what is called a circle of influence are radius of influence this radius is called d. So, what does it mean? You take a circle with i as a center, d as a radius I am considering a 2 dimensional problem now, you can imagine a 3 dimensional version of this problem later. So, if I consider i as a center and d as the radius look at all the observation locations that are in within that circle we are going to postulate only those observations that are written a distance d from the point i are going to be influential in affecting the state at the point i.

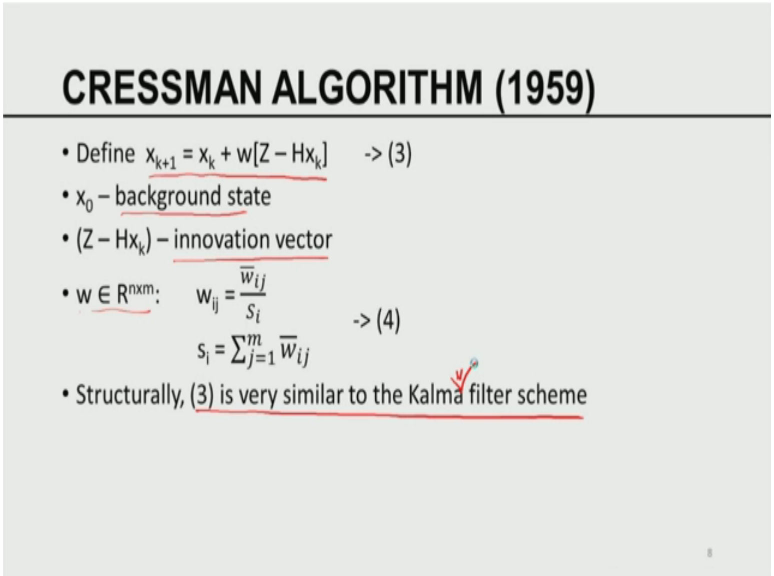
So, that is the very beautiful concept that arises out of the notion of influence of one grid point on the other grid point are one observation stations on the grid point. So, like N i d be a sphere are at a circle of radius d center that i, let m i be the number of observations that are inside the sphere centered at i, let r i j be the distance between the chosen grid point I and the location of the j-th observation inside the circle of radius d center at i. So, what does it mean? In this case this is an observation inside, this is an observation inside, this is an observation inside, so in this case for this m i is 3 in this particular example. And then I can also compute the distance, so let us assume this is the first observation, this is a second observation the third observation. So, this is r i one, this is r i 2, this is r i 3, so r i j is the distance from the i-th grid point 2 the j-th observation j running from 1 to m i, m i is the total number of observation that are bounded that there are contained within the bound defined by your sphere are a circle of radius d I hope that is clear now.

So, I am trying to divide the observation to two groups; one that influence is i, the other that does not influence I, if I pick d to be very large every observation would affect i, if a pick d to be very small only a smaller number of operations observations will affect i there is no formula for the choice of i, but it makes sense to think about the notion of observation influence and grid and the influence dependent on the distance. So, that is a very beautiful and a fundamental idea. So, what did cress man and his group do? They divided they divide a weighting scheme w i j please remember that w is a matrix w i j are the elements w i are the elements of the matrix.

So, w i j bar is the element i j-th element of the matrix that affects sorry that affects the i-th grid and so w bar i j is equal to d square minus r square i j divided by d square plus r square i j. So, this weight is less than 1, for all r i j less than d that was the non 0 weight for otherwise it is 0, so equation 2 essentially provided you a method by which they chose the weight. Once you choose the weight like this let me go back to my iterative scheme in equation 1. So, I have a means of choosing w once I have a means of choosing w everything else is given in here I can run this iteration I can truncate it at an appropriate condition I can get X star once I get x star I have initial condition use that as an initial state to solve my initial value problem, so that is how operational centers operated in those days.

Now you may want to discuss why this scheme why not any of this scheme? Again I want you to understand this is simply a heuristic. So, you probably cannot quibble with it too long too much, because it makes sense it makes sense and in what way it makes sense? When r i j becomes d the weight becomes zero; that means, for points on the circle the weight is 0, for points inside the circle weight is larger, for points out that circle weight is undefined. So, it is disability to give that cut off they probably decided the formula to be like this, there are very many differing methods for computing this weight; Bergthorsson and Doos picked wait according to one formula; cress man did by another formula, in those days you if you look in to the literature there at least half a dozen different ways of picking w, mathematically speaking all are equivalent that w essentially let us go back, w essentially decide the strength by which I am going to update X k to become X k plus 1 in other words how much weight it gives to the innovation.

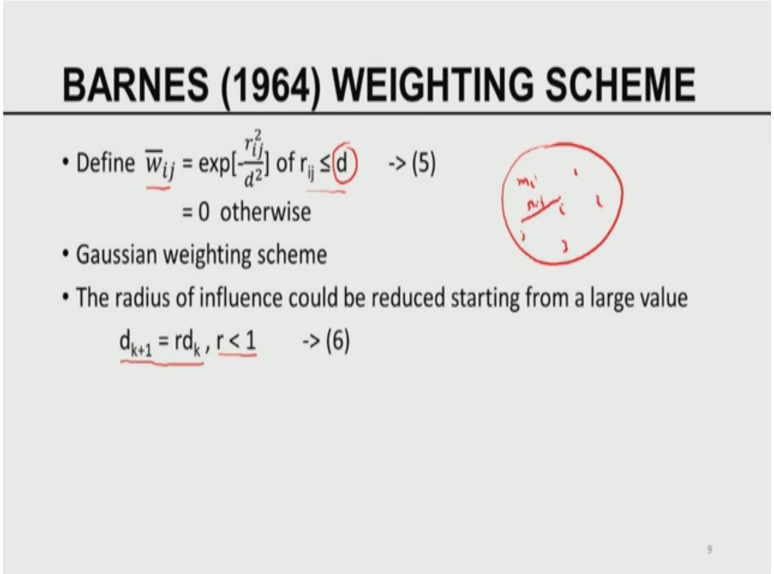
(Refer Slide Time: 43:48)



So, the cress man algorithm continued. So, I have X k plus 1 is equal to X k plus w Z minus H of X k, again X naught is the background state this is called the innovation. I would like to be able to create my w, we give w bar in the previous equation 2; I am now going to normalize my weights, so I am going to define w i j to be equal to w bar i j divided by s i, s i is simply the sum of all the numerator. So, this way I am going to make sure that the weights are normalized and the individual values are going to be between 0 and 1, so this is a very nice scheme 4.

So, if I used the scheme 4? I am sorry if I use the scheme 2 to define w bar if I use the scheme 4 to normalize it and if I use this normalized waiting 3 you get you get you get the cress man scheme. I want to reemphasize, cress man in the mid 1950's had anticipated Kalman filter, even before Kalman filter was invented to me that is the novel aspect of it this is Kalman Man sorry a Kalman filter was invented and that is a that is a very important information that one need to keep in mind.

(Refer Slide Time: 45:30)



Oh I am sorry we have done this sorry. Now I am going to tell you for the sake of completeness one more weighting scheme that game in 1964 that is called Barnes scheme; that is a little story I would like to tell, Barnes was working at the university of Oklahoma where I am currently working university of Oklahoma and national severe storms lab they essentially pioneered the use of radar me to use of Doppler radar in meteorology.

So, they would like to be able to develop schemes by which they can utilize the radar information and transfer the radar information onto the grid. So, they did something in the mid 60's similar to what cress man vectors has done the mid 50's, but instead of instead of 500 mill bar height they have the emphasis by Barnes was essentially using radar related data. So, he bonds essentially developed a weighting scheme, that still rust on the notion of a radius of influence all points at a distance r i j from i; that means, if this is the point i, if this is the region of influence if this is 1, 2, 3 m i, this is j the distance from here is r i j.

So, if r i j is less than d is the radius of this circle, the weight w i j bar is equal to is a Gaussian structure, you can readily see exponential minus of r i j square by d square. So, it is a Gaussian shaped function it was 0 otherwise, so it is a kind of a truncated Gaussian sitting over your sphere of radius d you can readily imagine that that weight function. So, this Gaussian weighting function was used and he was also a little bit more sophisticated.

What did he do? He did not keep the radius fixed he first started with the larger radius. What does it mean? He wanted to bring in as much of influence from as many observations as possible.

So, you if I pick d to be large larger number of observations are going to affect the grid point, so and then he continuously started shrinking the sphere of influence and what is the scheme he used to reduce the radius of the sphere of influence he decided r k plus 1 is r times r k where r is less than 1; that means, he started from a larger radius and kept on shrinking it, because he would like to get the benefit of as many as possible, but he wanted to make sure those who are closer had more influence than those who are far away.

So, he essentially he had an adaptive scheme by which he adapted the value of the radius in some fashion by which he shrunk. Did you shrink it to 0? No, when it became a particular value at that point he kept it because he wanted to make sure that he did not shrink the regional influence to 0. Why am I mentioning this? I just want you to think about there are all kinds of interesting heuristic ideas fixed d variable d Gaussian based weight function other heuristic base weight functions these are all the variations on the theme.

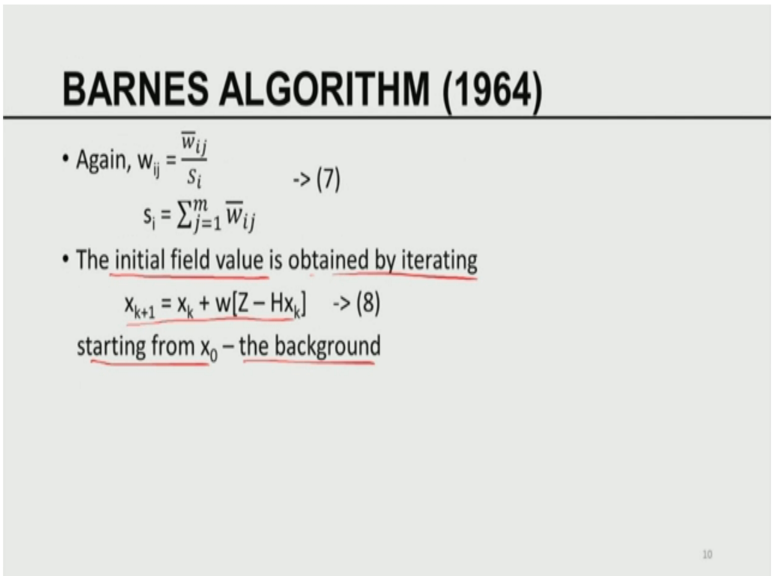
But the ultimate aim is to be able to consider a background control the observation try to transfer information from observation to the grid the grid is the one that I would like to be able to update. So, you do this iterative process this iterative process in some sense is the data assimilation scheme that were used in the 50's and 60's. It is here they combined the background information with the observation to be able to create the new mosaic which comes out to the iterative scheme the mosaic being a good initial condition from which they can run the model forward.

So, all these schemes were offline experiments they did to be able to create the initial condition X star. What is X star? X star is the value of X k at the time when you cut off the iteration and from X star then they generated the forecast. I think from a historical perspective and also to be able to appreciate how people did in the early days of computer era I want to emphasize that, the data the subjective analysis was the order of the day until 19 mid 40's, it is there was already anticipation that computers are just around the corner, Hence Panofsky described an interpolation scheme, but it was riddled with problems, computers scheme phenomena proved the use of computers and making 24 forecast.

If you have the ability to make a 24 forecast, the interest in getting good initial condition to run the model forward became center stage. So, phenomena's work provided the motivation to be able to create a very good initial condition and with that of the motivation research in Sweden and in other states sprang up Bergthorsson and Doos, cross men, Barnes these are various kinds of early examples of data assimilation schemes where they very cleverly combined background information and the data to be able to create this mosaic.

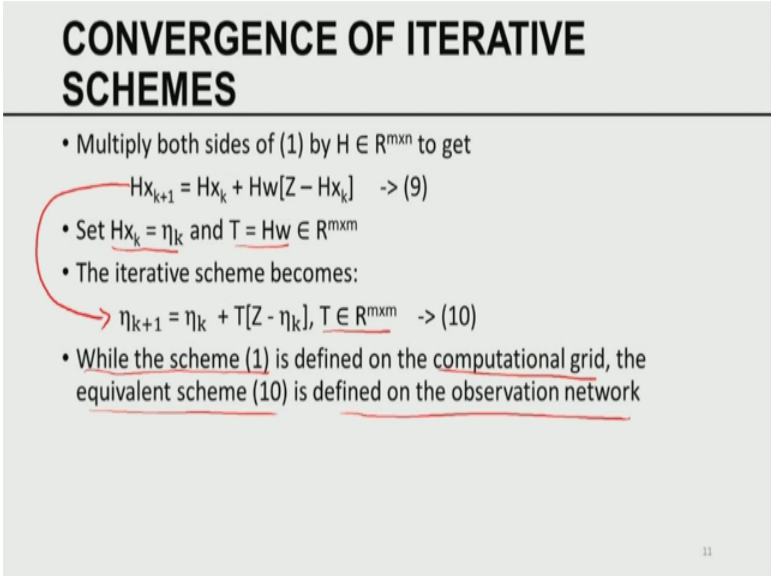
And what is this mosaic? In modern language in some sense we can think of that as an analysis, because of the combination of prior with the observation. Now if you think back what does data assimilation do in general? I have prior I have observation you get posterior in statistics they call posterior is called analysis in geosciences, so analysis and posterior exactly one of the same. So, I want to again reemphasize the governance of the idea the governance of the formula the anticipation of Kalman like scheme even before Kalman filter came to be I to be these are very important observation historically as well as these ideas are very inspiring, should be inspiring to anybody.

(Refer Slide Time: 52:08)



So, again if you have w bar i j given by 5; I can normalize it as in 7, so the initial so the initial field please understand. What is the idea here? I am interested in creating the initial field; initial field is obtained by iterating this starting from X naught the background and the observation, so that is 8.

(Refer Slide Time: 52:35)



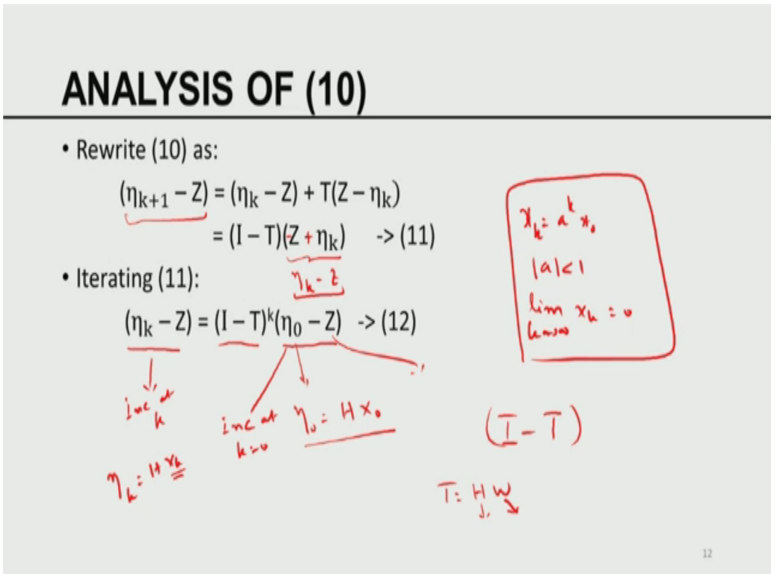
Now I would like to indulge your interest in a theoretical question. Suppose you can you iterate it like this (Refer Time: 52:48) converge under what conditions this iterative scheme will converge; that means, I am a guaranteed to get an analysis if I ran the iterative scheme of cress man as an example unto asymptotic time. I think the convergence results and the convergence question becomes very fundamental and important that has already been worked out.

So I am going to provide you a glimpse into the concept of convergence of iterative schemes used in the early days in data assimilation. To do that; you multiply the equation one you may remember the equation 1; X k plus 1 is equal to X k plus w times Z minus H of X k, now multiply both sides of the equation by H now I change the variable, so I am now trying to bring in, so until now everything was little heuristic.

Now I am going to talk about some mathematical structure relating to convergence of the iterative scheme to understand weight ever converge, if I know if it is going to converge after a longer period of iteration I can truncate, it if I truncated it that truncated should not be too far away from the convergence because I will lead to convergence. So, to be able to understand the quality of the state that it will get then a truncated before at some after some finite iterations one need to be able to inquire about will it converge had I gone had I not stopped but continue. To examine this, i am now going to concoct a new variable eta k is equal to H of X k, I am going to change the name of the product metric H and w to be T, with this change of notation this equation becomes eta k plus 1 is equal to eta k plus plus T times Z minus eta k, here T is a m by m matrix.

Look at this now, why the scheme originally the scheme one originally defined on the computational grid? The equivalent scheme 10 is defined as the observation network. Why is that? How did I bring the iteration from the computational world to the observation network? By the matching of H, please remember H is the bridge between observation world and the computational grid world or a model space. So, you can think of convergence in both the domains because I can translate the results from one domain to another domain. So, I would like to understand the asymptotic properties of the iterative scheme on 10-th given by, you can now rewrite 10 as you can subtract Z from both sides let us go back, you can subtract Z from both sides.

(Refer Slide Time: 55:43)



And after a little bit of an algebra you can see you get the equation 11, eta k plus 1 minus Z is I minus T times Z minus eta k. So, iterating this I think I yeah it is theta k I think just make the theta. Did I screw it up? I think it must be this must be I am sorry one second let me check on that, Z eta k minus T Z now this must be minus this must be plus I am sorry this must be eta k minus Z the same structure is this sorry for that for that for that error.

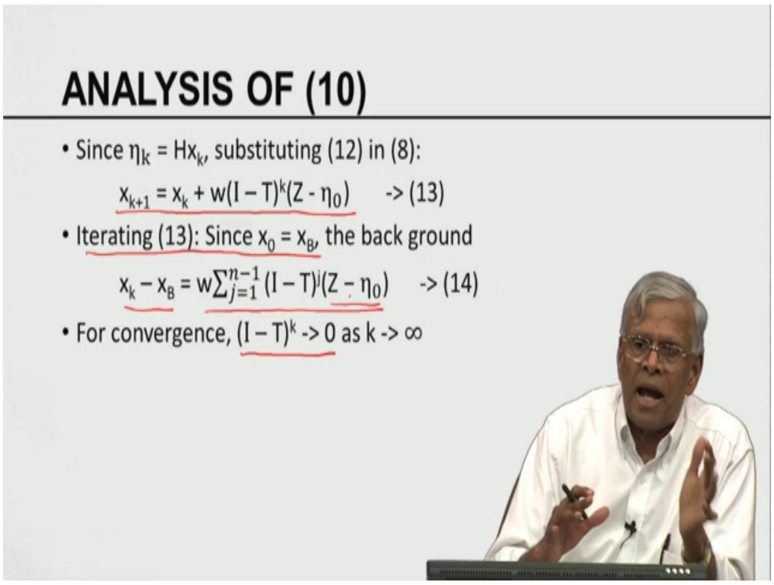
So, you got you got 11; now I can iterate 11, if I iterate at 11 eta k minus Z is equal to I minus T to the power k eta naught minus Z. So, what is eta naught? Eta naught is equal to H times X naught. What is Z? Z is the observation. So eta naught is my predicted observation X naught is my background, so is a predicted observation from the background using H, Z is the actual observation the difference between the two is the innovation.

That innovation is to going to be multiplied by I minus T to the power k to get eta k minus Z. So, you can think of eta k minus Z as an increment. So, this is the initial so this is an initial increment at k equal to 0, this is the increment at a general time k, the increment at time k is the k-th power of the matrix times the increment at time 0. So, if I had an equation X k is equal to a to the power of k X naught, if absolute value of a is less than one you readily know limit k tending to infinity of X k is 0.

Therefore by in analogy with this, now you can see; if the k-th power of I minus T goes to 0 as big as k goes to infinity, eta k minus Z will go to 0 so eta k will match the observation. So, what is eta k? please remember eta k is equal to H of X k; that means, the state of the grid will match the observation if there is going to be convergence. So, the whole convergence now rests on the properties of the matrix I minus T to the power k or in principle it rests on the properties of the matrix I minus T, please remember what is T is T please remind let us remind ourselves of this, T is equal to is equal to H times W.

What is H? H is the interpolation matrix. What is W? W is the is the weight was created heuristically, W is again created heuristically by way of interpolation there are infinitely many ways of choosing H there are infinitely many ways of choosing W, now we are going to ask what particular choice of W and H will give you a T that will induce the property namely I minus T to the power k will go to 0, that is the mathematical question here. So, what is it depends on? It depends that is correct.

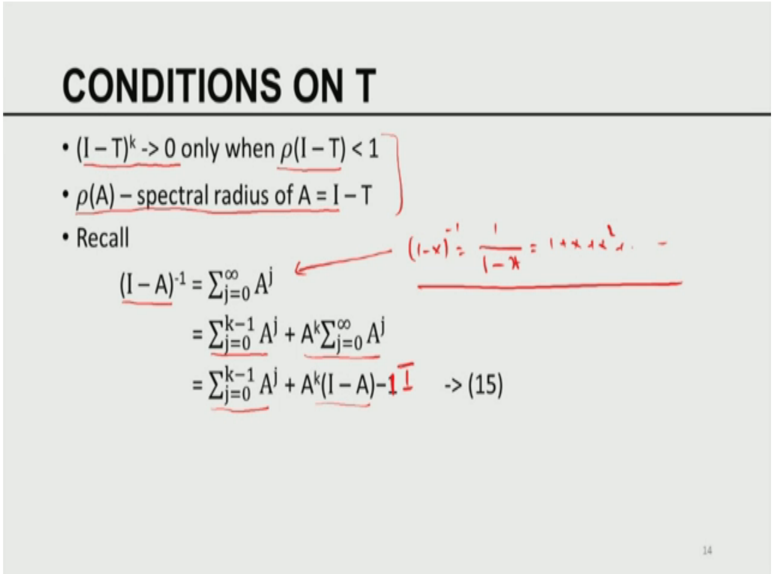
(Refer Slide Time: 59:53)



Now, I am since eta k is equal to H of X K, now I am going to substitute 12 in 8. So, if you did this substitution now I have come back from the observation network to the computational network, I have transferred from the observation network the computational network; iterating 13 since X naught is equal to X B. What is X B? X B is the background, X K minus X B what is X K minus X B? That is what is called analysis increment X K is a current analysis at the result of k-th iteration X B is the background. So, that is the analysis increment you can think of X K minus X B as an analysis increment, Z minus Z naught is the initial increment in the prior information, so you can see the analysis increment is expressed as the sum.

So, for convergence as we have already pointed out I minus T to the power K must tend to infinity, if this goes to infinity convergence will happen, when convergence happens we are guaranteed we will get descent combination of the prior and observation, descent combination of the prior on the observation.

(Refer Slide Time: 61:15)



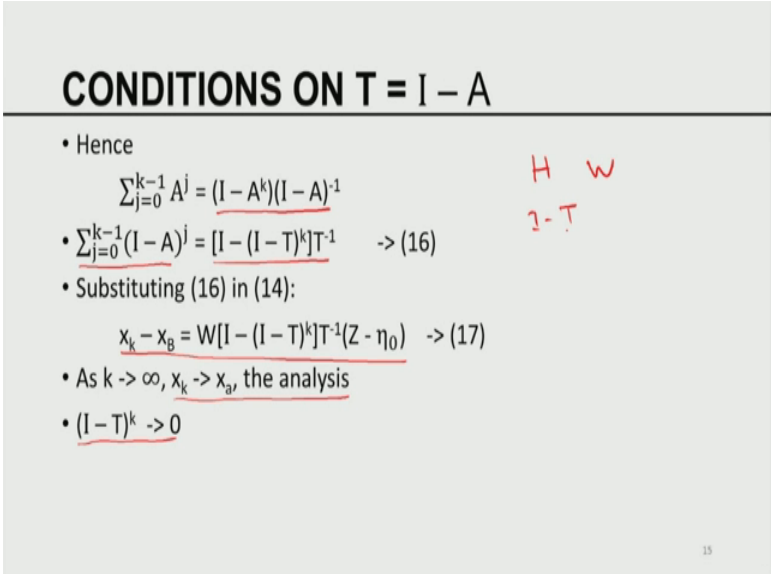
So, under what condition the convergence will happen? So, consider the matrix I minus again yeah now we are going back to the matrix the need to be able to formalize some of the convergence results, under what condition I minus T to the power k 0? It will happen only when the spectral radius of the matrix I minus T is less than 1.

Please remember spectral radius is related to is related to maximum eigenvalue, maximum singular values, eigenvalue, singular values there they are related concepts, eigenvalue becomes the spectral radius for symmetric matrices singular values becomes this spectral radius for general non symmetric matrices, we already know rho A is called the spectral radius of A is equal to I minus T. So, if the spectral radius is less than 1 it is it can be shown that the power of the matrix is going to be tending to 0.

Now I am going to provide few other steps in here. In general; if A is the matrix I minus A inverse can be express in the power series and this series is the matrix analogue of 1 over 1 by x is equal to 1 plus X plus X square, so that is an infinite series we all know, so this is equal to 1 minus X to the power minus 1 this is a standard series you all should know about it from basic calculus.

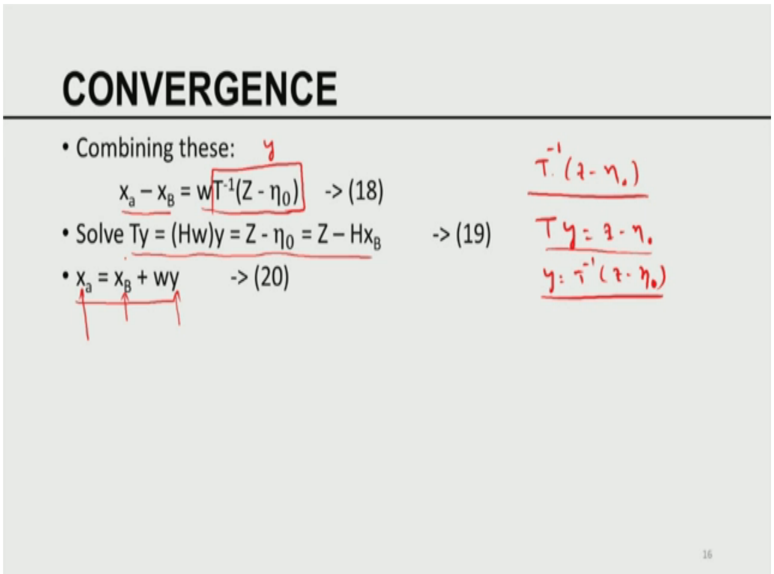
So, if I put X is equal to A I get this series this series can be divided into two parts the first k terms the rest of the k terms and the rest of all terms, I can take A to the power of k as a common factor from the second; again this sum is infinite therefore, this is equal to finite sum plus A to the power of k I minus A minus this cannot be 1 this could be I i am sorry this would be I.

(Refer Slide Time: 63:18)



Therefore summation A to the power of j is given by this equation, again a little bit of algebra is given by this equation therefore, this sum is equal to this sum, this sum is equal to this sum and substituting 16 in 14 I get X k minus B is given by this; therefore, if I minus T to the power k tends to 0, X k is tends to X a that is what is called analysis the limiting value of X k is called analysis.

(Refer Slide Time: 64:07)



Therefore we have now found conditions under which this crisp unlike scheme will converge, the condition for cress man like scheme to converge is simply that the spectral radius of I minus T must be less than or equal to 1, so that gives you a design criteria. And I am going to spend a couple of minutes on that now go back, one can do a lot of simple experiments you can fix a particular method for H, you can pick a particular method for W you can multiply this you can get a T and then you can do an Eigen analysis for I minus T and recover and see when.

If I so you can keep H fixed you can change W or you can keep W fix change H, for which combination of H and W one gets a matrix I minus T such that the spectral radius less than less than less than 0, if you can strike a combination you got the gold, so that is the basic idea, that is the design question. So, here we are not interested in talking about how to design we are in place interested guidelines through the design.

So, I minus T the rho they the spectral radius being less than 1 provides the broad guidelines to the design of H and W. So, then when this term goes to 0 again come back, again this term goes to 0; 17 reduces to 18 and so X the analysis minus. So what is that? x a minus x B is the prior x a is the analysis, so you can think of x a minus x B is called the analysis increment, the analysis increment is simply calculated by w T inverse Z minus eta naught, you know what eta naught is and how do you solve this in order to be able to do this you do not have to compute the inverse of T.

So, if you want to compute T inverse Z minus eta naught you simply need to do the following solve the equation T y is equal to Z minus eta naught, then y will be equal to T inverse Z minus eta naught. I would why I mentioning this? I want to tell you a couple of things now, your interest is not in trying to compute T inverse, you can do this by computing T inverse and multiplying, but I do not want T inverse I only want T inverse times Z minus Z naught.

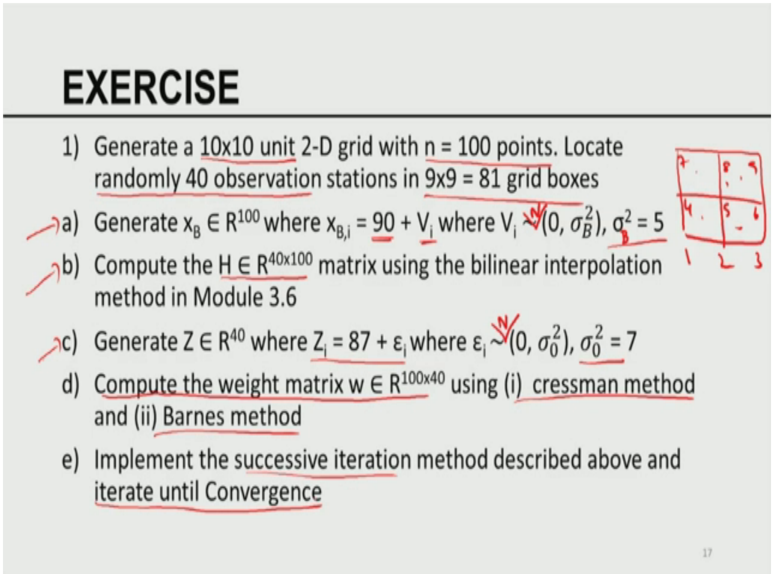
So, what is the best way to compute T inverse Z naught? please do not indulge in inverting the matrix T because it will take a long time you simply write the linear equation T y is equal to Z minus eta naught solve it. Why? is indeed equal to T inverse Z minus Z naught Z minus eta naught, so you solve the system you get so you got this is what we call as y, so x a is equal to x B plus w y so I have now given you a pathway to compute the analysis starting from the background and an increment y, the y depends on solving the equation in 19.

So, that is the pathway to proving convergence and this is very educative because it tells the role of w, it tells the role of H it also indirectly provides conditions on not on H alone w alone, but the product H w that creates T that creates T. So, what was thought to be a good heuristically method can be put in a in a firm mathematical background once we do the convergence analysis.

So, with this convergence analysis we have equation 20 is the essence of the scheme that were used in 1950's and 60's. So, what is the basic scheme? You give me an x B background on the grid based on the observation you compute w, you compute I am sorry based on the grid and a rating scheme create a matrix w you create an interpolation matrix H you solve for y from 19, x a the initial condition mosaic that I am going to have to use to generate the model forward in time is x a it is simply the sum of x B plus w y.

So, that is a very nice and important methodology that was used and very successfully 50's and 60's using these schemes. They repeated this experiment every day to be able to create an analysis of ready to be able to run the model forward in time and they were very successful in the in the early computer age with respect to creating good forecast.

(Refer Slide Time: 68:53)



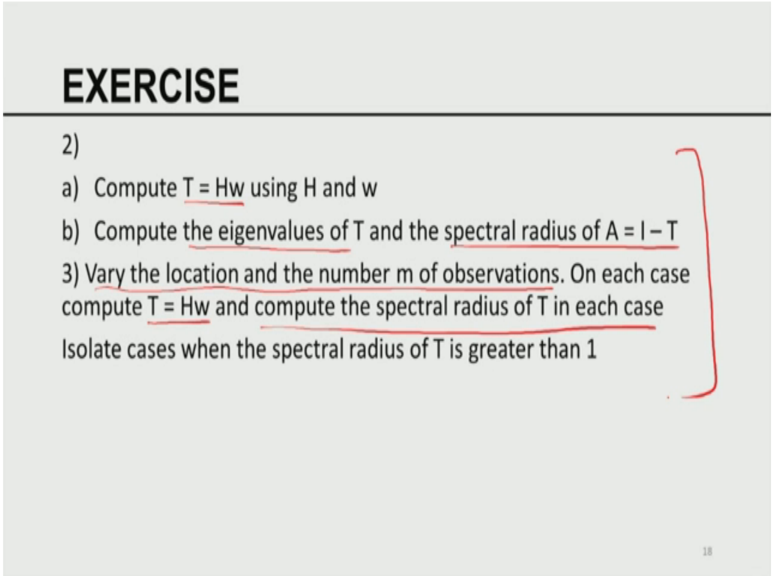
Now I am going to leave you with a couple of very good problem these are very important problems. So, I am going to spend a couple of minutes on this; generate, generated 10 by 10 unit 2 dimensional grid with 100 points locate randomly 40 observations. And the; If I have 10 grid points there are only nine times 9 grid boxes. So what is the grid box? That let us come in here. So, there are 3 grid points here there are 3 grid points here, there are two times to 4 grid boxes. So, 1, 2, 3, 4, 5, 6, 7, 8, 9 there are 9 grid points there are 4 grid boxes your observations are going to go within the grid boxes.

So, you are going to have to distribute 40 observations in 81 grid boxes generate x B which is r 100, I would like to be able to ask you to generate the background mean is 90 w i is a noise you this must be normal this is this is normal 0. So, V i is normally distributed 0 mean and sigma square V as a variance, I am going to ask you to pick sigma B square to be 5 that the variance. So, you generate the noise add to 90 you get you generate 100 random vectors you have the background information. Then compute H which is the 40 by 100 matrix using the bilinear interpolation that we developed in module 3.6 ah. So, that is that so we have so here we have generated the background, here we generate H.

Now, I am going to generate observations are generated by this equation again, this is again normal observational covariance is different from the background covariance. So, I have all the components now I have H, I have x B, I have Z you now need to compute the weight matrix using the cress man method using Barnes method there are two methods for doing this implement the successive iteration discreet about and iterate until convergence.

So, it is a very beautiful scheme if you can do this on a computer then you can say you thoroughly understand some of the classical methods, these methods are still could be used they are very powerful they are very meaningful and these are heuristics developed, but it is for the power it is for the simplicity it is for the elegance I believe these kinds of exercises should be done by anybody who is involved interested in trying to learn the tools and techniques for data assimilation.

(Refer Slide Time: 72:10)

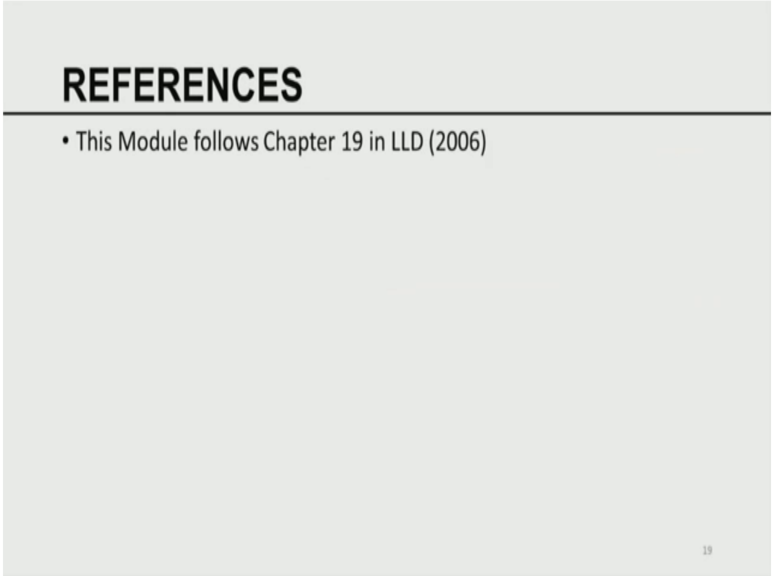


Again the apart 2 years I would like you to continue, I would like you to be able to compute the matrix T which is equal to H times w using the H and w that we picked in problem the previous problem compute the eigenvalues of T, compute the spectral radius of A is equal to I minus T and check whether the spectral radius is less than or equal to 1, if it is 1 you are done if it is not the scheme still makes sense, but it may not converge.

So, what is it I would like you to do now very the location and the number of observations in each case compute T is equal to H of w and compute the spectral radius and find out for what kind of combinations the spectral radius of T is greater than 1 for what kind of combination the spectral radius or less than 1, I believe it is a very interesting worthwhile research it.

One can even develop the master's thesis are perhaps part of a P H D thesis out of it. So, we only provided the general theory for convergence trying to utilize the general theory and transferring to the actual design questions of w and H, I believe would be a worthwhile interesting research topic at the M S P H at least part of the P H D it can be one full master's thesis project.

(Refer Slide Time: 73:32)



This in here this module follows chapter 19 in our book Louis (Refer Time: 01:13:40) that we conclude our discussion of the classical methods for data assimilation.

Thank you.