Dynamic Data Assimilation

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

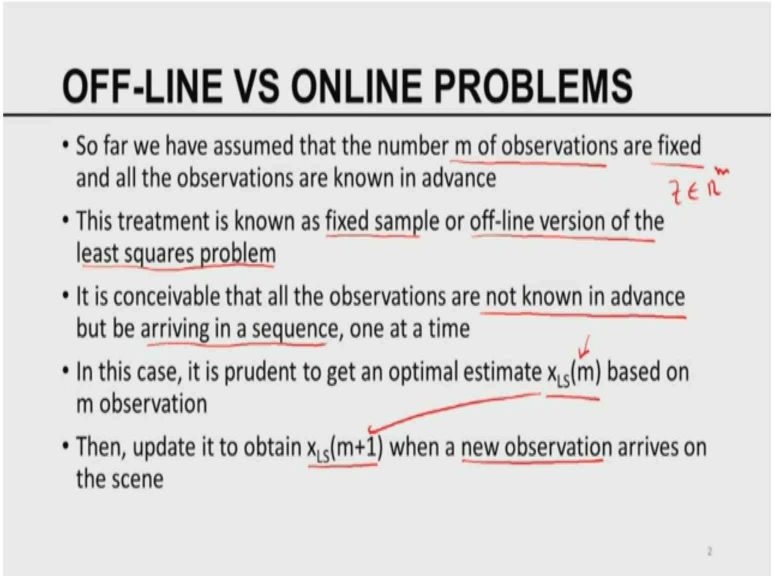
Lecture - 13
On-line Least Squares

So far, we solved deterministic least square problems both linear, non-linear well posed, ill posed in what is called an offline mode. What is an offline mode? An offline mode is a way of solving problems, where we assume all the observations are available at a given time. But in some cases, we may not be able to wait until all the observations are collected.

So, as soon when the observations come in, we would like to be able to make an estimate, contingent on a given set of observations. In such cases we would like to be able to improve the quality of the estimates, as and when new observations come into play. Such way of attacking the problem is called online as opposed to offline; is also called sequential methods because we are going to be updating the estimate as and then new observations arrived on the scene. They are also called recursive.

So, online sequential recursive these are common terms utilized to represent the concept of being able to update when a new information is available. We are still going to be solving the least square problems. And I am going to provide a broad overview of this class of online algorithms.

(Refer Slide Time: 01:46)



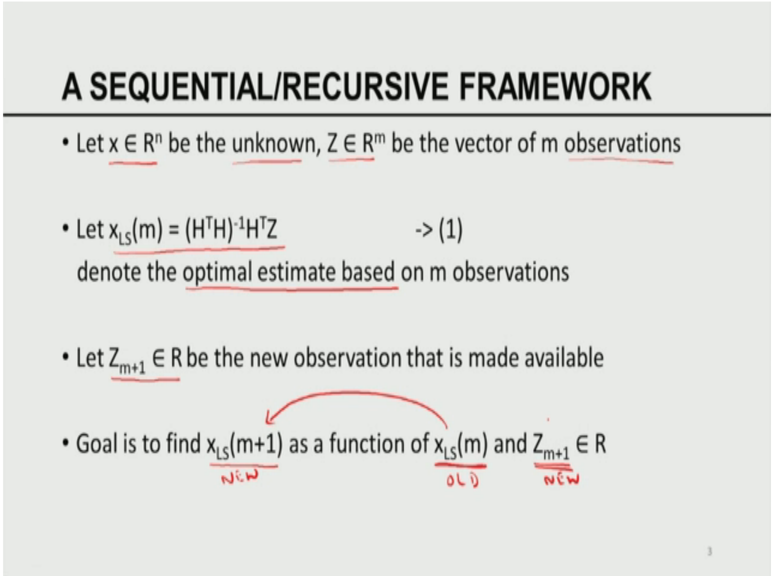
So, we have assumed another way of re emphasizing, this we have assumed that the number of observations m is fixed, please remember Z belongs to R m. So, what does it mean? Then there are m observation given to us, and that said with that we have to solve the problem. So, we assumed the vector Z to be a vector of a fixed length m. This treatment, was known as the fixed sample or offline version of the least square problems. Why is the fixed sample? Because m is fixed, m is the size or the number of observations, Z is the vector, observation vector has m components, you m is fixed everything is fixed.

So, what does it mean? I go to the lab make measurements, I make m individual observation, I collect them into a vector, I close the lab then, I come back to do the analysis, that is what called offline. It is conceivable on the other hand, the observations, all the observations may not be known in advance. They may be arriving in a sequence one at a time that could be a delay between the I am occurrence of each of these observations.

So, in that case it is prudent to ask it is prudent to ask a question what is the optimal estimate. What is XLS based on m observation? So, I am now going to associate the number of measurements I have used in estimating the least square value of x they are known. So, XLS is the least square estimate of x, but XLS is conditioned on having m observations. So, when a new observations of comes in, I would like to be able to update

XLS to XLS m plus 1. So, this gives you a flavor of what we mean by online, as opposed to offline.

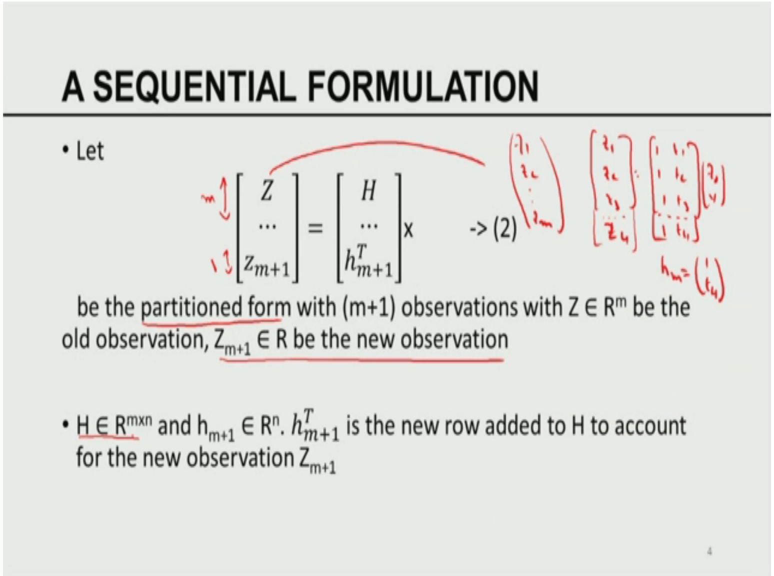
(Refer Slide Time: 04:01)



So, let us try to formulate this problem. Let x be let x be R n. Let x be a vector in R n. Let Z be a vector in R m. X is the unknown. Z is the known observations, in the previous lectures we have already seen XLS m is given by H transpose H inverse H transpose Z. I am assuming an over determined case, in here, and that denotes the optimal estimate based on m observations, that comes from the basic theory of linear least square deterministic inverse problems.

Now, let us pretend, Z m a real observation comes into play. I would like to be able to estimate XLS m plus 1 this is the new one that is the old one. I would lead to convert this into this, when the m plus 1 th observation come to the picture. So, what is the basic idea? As and when a new information is given, how am I going to update my old belief into a new belief, or a new estimate. So, this is the new observation. So, I have an old estimate I have a new observation, I would like it I would like to get a new estimate. So, that is the sequential problem.

(Refer Slide Time: 05:35)



So, the original vector Z is given in here, this is vector consisting of m, this is only one that that there is only one observation. So, I am. So, in the components of the vector Z are denoted by Z 1 Z 2 Z m.

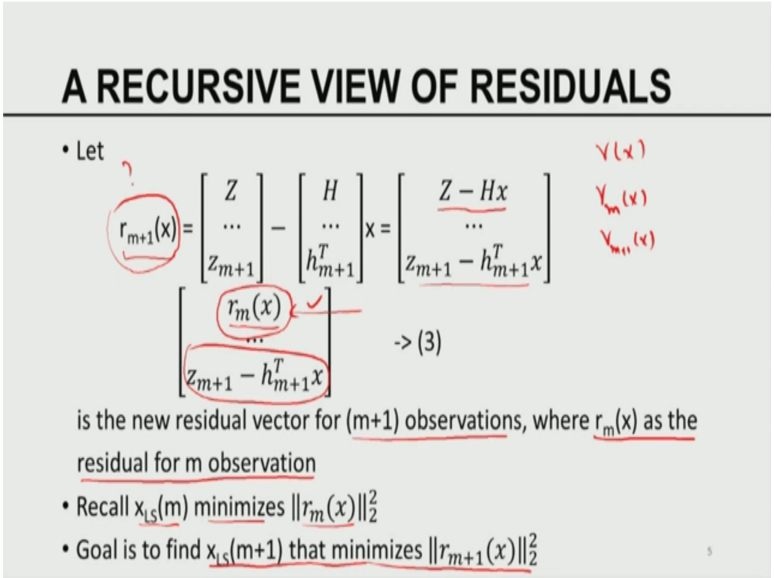
So, in that notation the m plus 1 th element is Z m plus 1. So, my model essentially say is the linear model. So, H Z is equal to H of x is the old model. Z m plus 1 is equal to H m plus 1 transpose x is the addition additional row to the matrix h, I need to be able to explain the new observations. All of you with me? Go back to my particle moving in a straight line. I had Z 1 Z 2 Z 3, I had Z 1 Z 2 Z 3, I had this, I had one t 1 1 t 2 1 t 3. I had Z naught v.

Now, suppose somebody gives you a 4th observations Z 4, the forth operation is given a time t 4. So, Z 4 is equal to Z naught primes v t 4. So, this is the new row, this is the new observation. Now, I hope the extensions are clear now. So, the last row of H is meant to relate the new observation to the unknown through the model equations. So, that is how Z m H m plus 1 transpose comes in to play. So, what does the main H m plus 1 is a column vector, transpose of H m plus 1 is a row. So, this is the row. So, H in the in this case H m in this case H m is equal to is equal to 1 t 4. So, H m transposes one t 4 I hope you understand the form of the H m transpose.

So, this is the partitioned form of matrix. Z m plus 1 is the new observation, I want to emphasize that. So, again please remember H is this matrix, H m plus 1 is this H m plus

1 is the new row added to H to account for the new observation using this example, I have already talked about that.

(Refer Slide Time: 08:20)



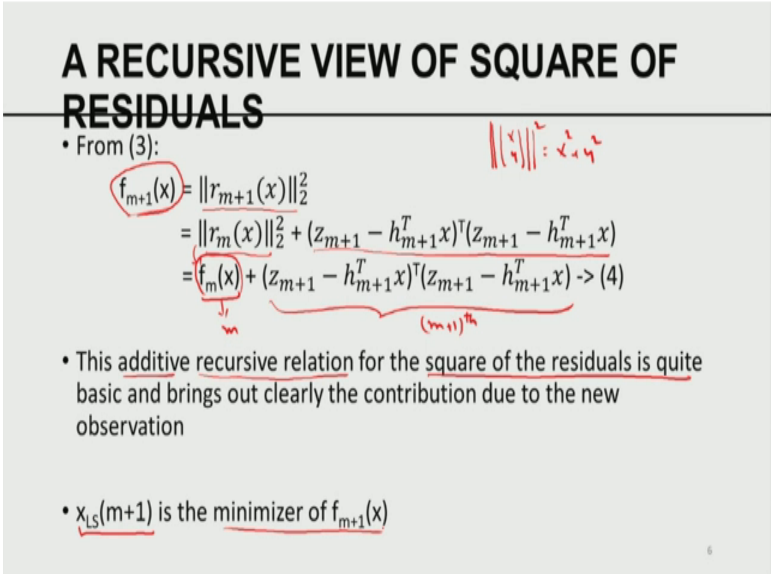
Therefore, now what do. I want to do I would like to be able to find the least square estimate as the linear problem. So, I am going to consider the residual the residual until now we simply said residual R of x because the number of observations are fixed.

Now, I am going to consider residual based on m observations, residual based on m plus 1 observations, that is the residual R m R m plus 1, R m plus 1 is equal to Z minus H x, Z m plus 1 minus H transpose this. So, observation minus the model, this essentially comes from the old, this essentially comes from the new. Z minus H m is R m of x, that for the new residual with the m plus 1 observation is the old residual the m observation and a new component in here. Therefore, I can you can see the recursive nature R m plus 1 depends on R m and the new information is the new information.

So, the new residual vector, has m plus 1 up it uses m plus 1 observation whereas, R m x is essentially the residual from m observations. So, I am able to read I am able to relate residual as a function of the number of observations. We already know XLS m minimizes R m of x. A x l m that is the optimal solution then I have m observations. So, the goal is to find x l m m plus the XLS m plus 1 that minimizes the norm of the new.

So, we already know how to minimize this, this is already know we want to know how to minimize, this say that is that is that is the key part of the problem formulation.

(Refer Slide Time: 10:21)



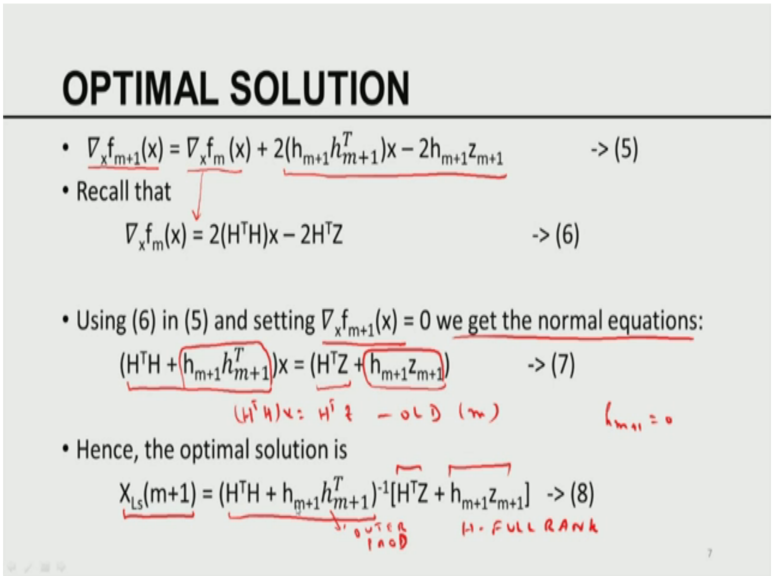
Therefore, I am going to follow the same route as, I have done in linear least square problems. The here f m plus 1 that the again the sum of the square there is residual when there are m plus 1 observation. The sum of the square of the residual from m plus 1 observations are given by this. If you have. So, so from simple um calculation of the norm, this norm is equal to the norm of the first m components plus the norm of the new component that is added I hope that is clear to you for example, but what did I have the norm of if I have a if I have if I am going to talk about the norm of vector x and y. We already know this is the norm of x plus the square of that is x square plus y square.

So, x plays the role of R m x m m plus 1 x y plays the role of the increment. Therefore, this sum of squared residual is equal to the sum of 2 terms, but this is exactly R m. This R m square essentially is equal to f m therefore, what is that we have got. We have got a recursive relation or a sequential relation the sum of squared error using m plus 1 observation is equal to sum of square error using m observations plus an increment. So, this is the sum of square observation using m plus 1 this is sum of squared observation using m and this is m plus 1 th observations.

So, you can see f m is recursively defined in m the number of observations. So, this is an additive structure, this is a recursive relation for the square of the residual it quite basic and it brings out clearly the contribution of the new observation. So, this is the contribution of the new observation this is the old observation. So, this is the total when I have all the observation together. So, what did I want I want like to be able to find XLS m plus 1. XLS m plus 1 must be the minimizer of f of m plus 1 of x. We already know the minimizer for this. So, knowing the minimizer for this I would like to be able to compute the minimizer for f of m plus 1. That is the mathematical problem. Again I would like to be able to find optimal solution let me go back.

So, I would like to be able to compute the gradient of f m plus 1 which is four. The I also would like to be able to compute the hessian of f of m plus 1 in 4.

(Refer Slide Time: 13:16)



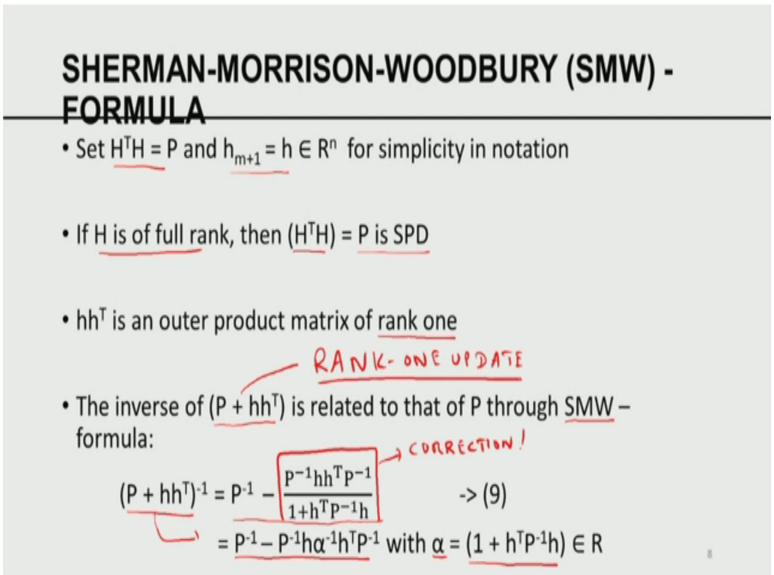
And that is what we are going to accomplish in this slide. So, if you consider the gradient. If you consider like the gradient, the gradient of this is equal to the old gradient plus an increment term. The increment term essentially comes from this the second term in 4, in equation 4. We already know the expression for the gradient of the first m terms from our previous analysis and that gives raise to this. So, if I substituted 6 and 5 I get the expression for the gradient using m plus on observation I set it to 0 if I set that 0 I get a new set of normal equations.

Now, look at this that normal equation is given by this matrix times x plus this plus this. Now if you set H m plus 1 is equal to 0 vector. This becomes H transpose H f of x is equal to H transpose Z, that is the old problem using m observations. If you bring in the m th observation I get this increment in here, I get this increment in here. Therefore, I can express the optimal solution using m plus 1 observation has the inverse of this matrix.

Now, look at this now this is an outer product matrix. This is an outer product matrix. So, using this out outer product matrix is added to the original matrix, I am assuming H is a full rank. So, this is the matrix of full rank and I am going to add a rank one outer product matrix, that inverse times this is the old this is the new. So, you can x you can see the beautiful structure you using the relation 8. So, I have an old term I have a new term the old term and the new term mix beautifully in a hue. So, if I can compute this I have an expression for the recursive way of updating the estimate when you go from m to m plus 1.

Now, in order to be able to express the sum the inverse of the sum, what is the real computational challenging in here, equation 8 it looks very familiar it is exactly the same equation we have seen except that there is one more term, that is all what it is. So, the question here is that am I going to compute the inverse that is needed in 8 from ground 0 all over again, or I can update the old inverse by adding a new correction term to compute the inverse of this new term, that is a question 1 to ask an answer.

(Refer Slide Time: 16:30)



And to that end I am going to invoke 2 a very well-known formula, that we have already alluded to when we did the module on matrices is called Sherman Morrison Woodbury formula. In order to be able to apply the Sherman Morrison Woodbury formula I am going to slightly change the notation to make it convenient to for discussion.

So, let us try to call the matrix. H transpose H as p. Let us simply call H m plus 1 as h. I am simply dropping the subscript here there I am replacing H transpose H s p in order to simplify the notation. We already know if H is of full rank, then H transpose H is p, that is s p d symmetric and positive definite again that is a result from matrix theory. H transpose is an outer product matrix. Outer product matrices are always of rank one. Therefore, the solution calls for computing the inverse of p plus H transpose.

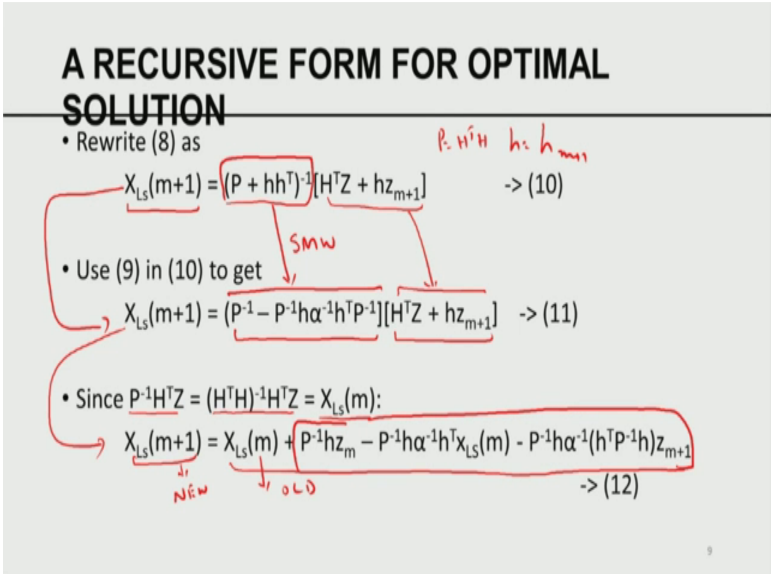
So, what is the question here. If I know the question is if I know the inverse of p can I compute the inverse of the sum and that is what is related through the Sherman Morrison Woodbury formula. The standard Sherman Morrison Woodbury formula that was given in the module on matrices essentially tells us p plus H H transpose inverse is equal to p inverse minus the ratio of these 2 terms.

Now, look at the right-hand side every term in here is known. So, what is that assumption in here, if I know the inverse of p, can I compute the inverse of p plus H H transpose, the answer is yes, that is that is where the Sherman Morrison Woodbury formula comes into play. So, if I want to be able to compute the inverse of the sum I simply need to be able to have a correction term. That is a correction term. Please look at the numerator the numerator is a matrix. P inverse of the matrix H, H transpose an outer product matrix P inverses a matrix. So, it is the product of matrices the denominator one is a scalar. H transpose p inverse H that is a quadratic form that is a scalar. So, I am going to now to simplify notation call the denominator 1 plus H transpose p inverse H as alpha.

So, 1 over alpha is alpha inverse. I am going to interpose alpha inverse into this formula. Therefore, the inverse of the sum is given by inverse of p minus the correction term. The correction term has a very beautiful structure here; it is this structure it is, this formula that is going to enable us to be able to go from offline to online. So, you can see the importance of matrix theory comes into play, through this ability to compute the inverse of a rank one update.

So, this is this matrix is called rank one update. Again I want to emphasize if you are interested simply an application you can take this for granted, but what am I doing I am not simply giving you the algorithms that you could readily use I am also going behind the theory. Why if you understand the theory of these very well that will enable you to be able to strike out newer algorithms by reformulating the problem in several different ways and I am trying to provide all many of the mathematical tools that has proven useful in the past discovery of several algorithms related to data assimilation.

(Refer Slide Time: 20:30)



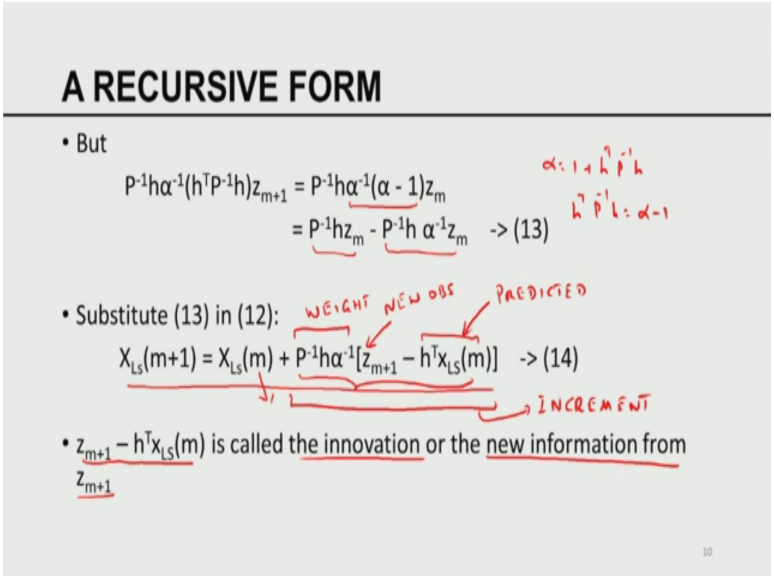
So, that is the scope of these lectures. Therefore, I would like to be able to reconsider my problem. I am going to be I am going to be able to express the solution using m plus 1 as this formula. This is the formula that was already derived in equation 8 I am rewriting equation 8 using the new notation please remember that p is equal to H transpose H my H is equal to H of m plus 1 it is a simpler version of the of the of the same formula

Now, I could apply the Sherman Morrison Woodbury formula for this and that gives rise to this relation. Therefore, the same solution in here is equal to the inverse computed explicitly please remember there is an inverse here there is no inverse and this is the same term that comes from here therefore, I would like to be able to, I would like to be able to relate the 2 terms, I would like to be able to relate the 2 terms this term comes in here that terms comes in here this is given by their Sherman Morrison Woodbury formula.

That is given by the Sherman Morrison Woodbury formula therefore, I have expressed this as a product of 2 terms each of which have 2 factors, here are 2 factors here are 2 factors. You can multiply these and I also I would like you to be able to remember that p inverse H transpose Z is equal to this that is equal to x l m. So, if you multiply this and use this fact, this equation x l m; now becomes, this expression becomes this expression.

So, what does this say this is beautiful expression. This is the old estimate this is the new estimate. So, the estimate using m plus 1 observation is based on estimate using m observation plus the one that comes as a correction term, this correction term looks pretty complex here, but it can be very easily simplified.

(Refer Slide Time: 22:56)



To be able to simplify this I am going to use the proper I am going to insert alpha in here. So, alpha please remember alpha, alpha is equal to 1 plus H transpose p inverse H therefore, H transpose p inverse H is equal to alpha minus 1. So, I have substituted those terms in here. If I substitute those terms and simplify I get this term. So, if you substitute and simplify 13 I substitute 13 into 12 and simplify you get the new estimate.

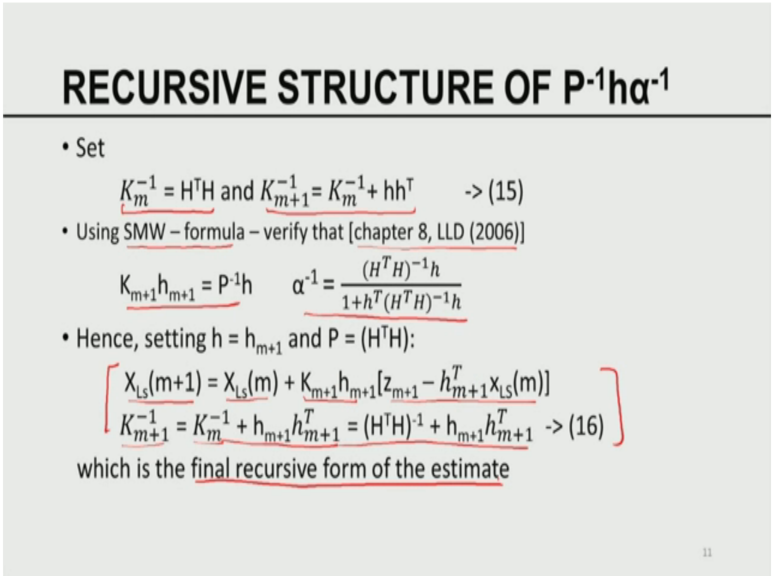
The new estimate structure is absolutely beautiful structure. This is the old estimate this is the correction term the, correction term has a weight. So, this is the weight. This is the weight term and what is the Z m plus 1, Z m plus 1 is the new observation. The new obs and what is this. This is the model predicted is the predicted model predicted observation using m plus 1 using the previous estimate.

So, this term Z m plus 1 minus H of m is called the innovation or the new information from observing the new observation, and this is called the increment to. So, this is the increment to the estimate. This is the increment I this is going to be the increment. So now, you can readily see the recursive nature coming into play.

So, if I have estimate based on m observation, if you give me the new observation, I can extract the new information from that observation weighted by a matrix. I get the new estimate. This structure is a beautiful structure. This structure will occur again in Kalman filtering, but Kalman filters are essentially discussed within that stochastic framework, but here I am doing everything deterministic linear least square problems with sequential update. So, you can readily see a precursor to Kalman filter even within the even within the deterministic framework. We will show later when I am I am in in one of the later lecture is that; Sherman Morrison Woodbury formula is also used in the derivation of Kalman filter equations.

So, you can see Sherman Morrison Woodbury formula is fundamental to the derivation of sequential estimates of the unknown and that is important thing I would like to be able to emphasize in here.

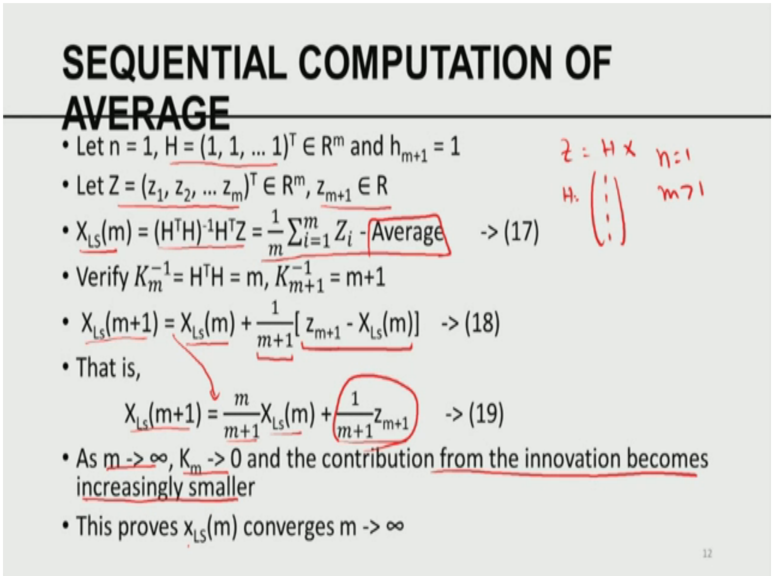
(Refer Slide Time: 25:59)



I am I can rewrite this in several different ways, I can call the matrix H m k m as H transpose H, and I can rewrite k m inverse to be that using Sherman Morrison Woodbury formula for this, you can verify these things these are given in chapter 8 of the book on dynamic data assimilation by Lewis Lakshmivarahan and Dhall, is the basic textbook from which all these lectures are derived. You can verify the relations that is given in here I am going to leave these as simple homework problems.

So, with that I am now going to be able to say the final way of recursive estimation is given by, the estimate using m plus 1 observation is equal to estimate using m observation a matric times the innovation the matrix itself is going to be updated and this update is exactly similar to the Kalman filtering update. So, what does it mean the new observation comes in first update the Kalman filter the gain this is called the gain matrix. Once you update the gain matrix this is the new information times the gain new gain matrix plus the old estimate that gives us to a new estimate. This is the final recursive form for the estimate as a function of the number of samples involved.

(Refer Slide Time: 27:33)



I am now going to see how we can utilize this required recursive set of to solve a simple problem. Suppose you want to find your weight. You made m measurements of your weight, the measurements are given in the vector Z 1, Z 2, Z m, it could be that you made your weight in m different scales or you use the same scale, but made measurements of your weight in the morning in the afternoon and evening for several days together. So, we have a set of m weights. This is the m plus 1 th weight, if I am going to consider the estimate of your weight based on the first m observation, the least square solution in this case I have this simple form Z is equal to H of x, H is simply a vector of all ones.

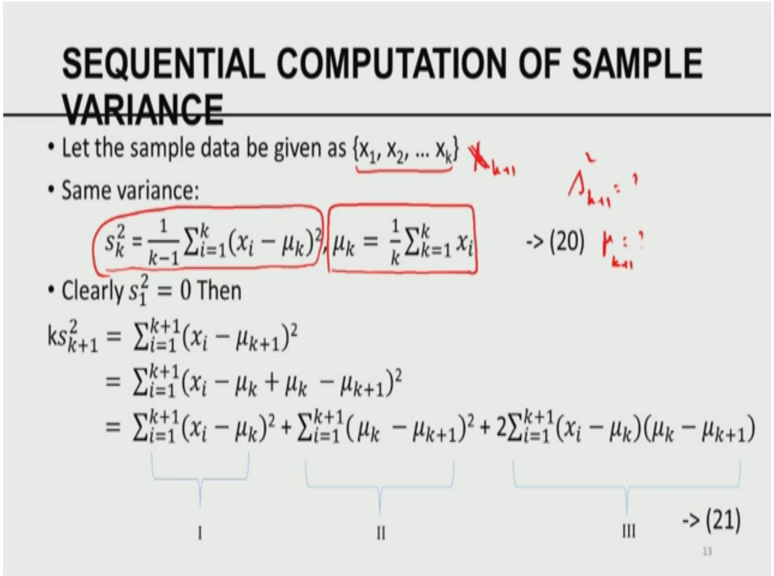
So, Z 1 is equal to I do not have to go there. So, H is all ones we have already represented that. So, H is all one, I have an unknown x the unknown x is your weight m is a set of m measurements of your own weight. So, H is all ones. So, what is the best estimate of your weight the average of all the weights. So, average. So, what does it bring it brings is a very fundamental result. Average is the best linear estimate of your unknown weight. So, x is your weight x is unknown you have m observation unknown is one n is one m is much greater than one is an over determined problem. So, in this over determined problem your weight varies from different time to different times. So, what is the intrinsic wait the best estimate of the your intrinsic w8 is simply the average of the measurements.

So, that is what it says the average is the best least square estimate. If you are going to get a new weight they after tomorrow morning and if that weight is Z m plus 1 this is the old estimate of the w8 this is the new estimated 8, I can update your w8 based on this information and in this case the gain is simply m plus 1. I can rewrite this equation 18 as x these times m by m plus 1 times the old estimate plus the new estimate these 2 are the same equations written in a different way.

Now, you can really see when m goes to infinity m goes to infinity, k m tends to 0 and the contribution of the innovation terms becomes increasingly becomes increasingly smaller, if the contribution of the in of the innovation term this is the contribution of the innovation term if that goes to 0 your w8 has stabilized in other words XLS converges as m goes to infinity. So, this is a very simple illustration of the recursive linear least square set up the set up being one of being able to find our estimate your unknown w8 based on m observation. This also further brings out the inherent optimality property of the averages have again this intrinsic behaviour of being the optimal estimate based on unweighted linear least square problems.

So, this is perhaps one of the reasons why whenever there are multiple opinions or multiple measurements we take the average to be the one that would utilize or we used to interpret the unknown.

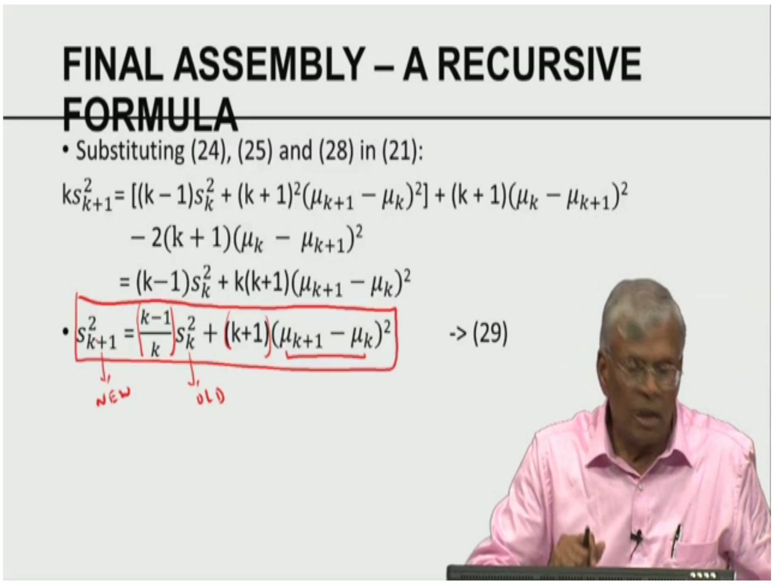
(Refer Slide Time: 31:39)



One can also compute sequentially the notion of sample variance. So, what is that I am going to state the problem, but I am given all the derivations I am going to leave this reading assignment, but let me tell you quickly. Suppose I am given a I am given a set of data x one to x n. I want to be able to compute the mean and the variance the mean of x and the variance of x.

So, let mu k be the mean of x. Let x k be the variance of these quantities. So, given k quantities I know how to compute the mean is that is given by this algorithm the variance is given by this algorithm, but what is that I want suppose I give you a new observation cape x k plus 1 the question is how I am going to be able to update. X square k plus 1 mu k plus 1. So, that is essentially the least square problem that we are interest interested in and that is the sequential update as I give you a new information I would like to be able to update the old information by adding the new information, I have exactly I have derived the formula for the recursive estimate I am not going to go over that you can readily read this it is a reading assignment for you. I am now going to show you the calculations in the next pages.

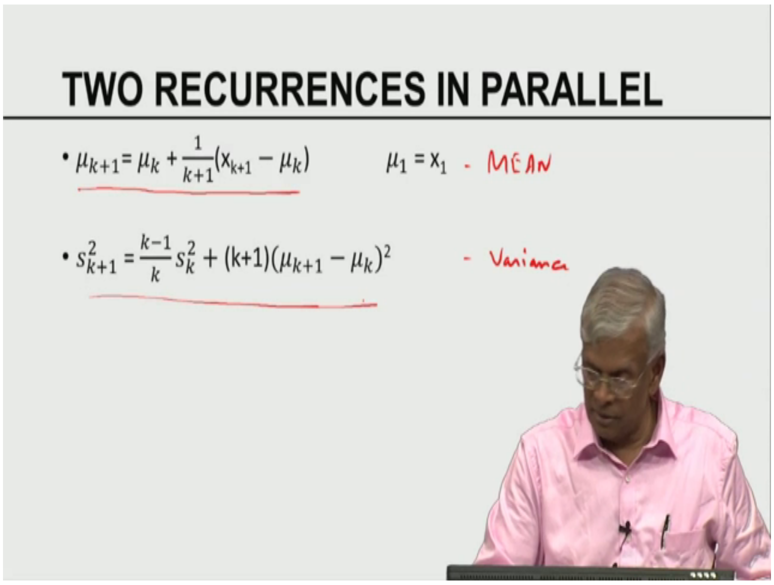
(Refer Slide Time: 33:23)



That ultimately gives raise to a formula that I am looking for; the formula is this one it is a beautiful formula what does it say the sample variance using k plus 1 sample that is s k plus 1 square x k square is the sample variance with k items.

So, this is old this is new I am going. So, this is the w8 function for the old this is the w8 function for the new this is the innovation term the innovation term is mu k plus 1 is the average of the k plus 1 items average of the k items the different square. So, this gives rise to a very beautiful recursive formula for computing the sample variance as well as a sample mean. Sample mean we already updating the sample mean we already saw in the previous example. So, this is an example where we can recursively compute the sample variance as well. So, these 2 together very beautifully illustrate the notion of sequential or recursive or online algorithms with.

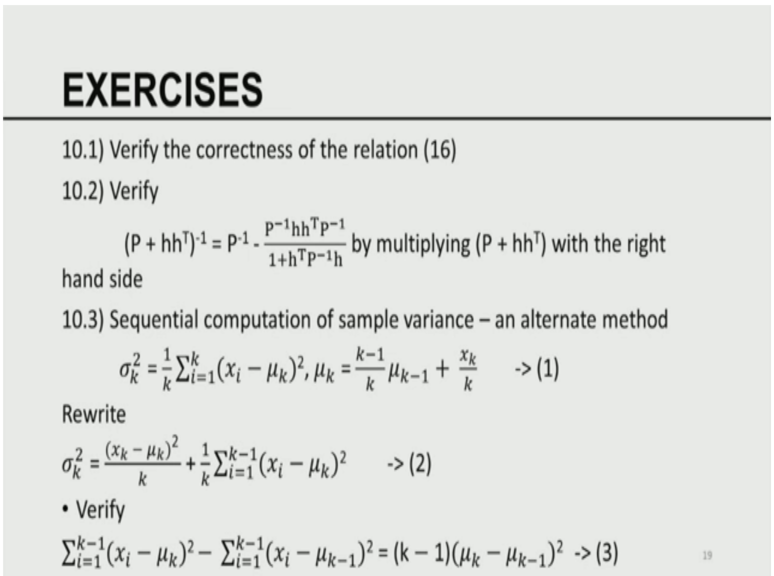
(Refer Slide Time: 34:45)



So, the 2 recursive algorithms are going in parallel this is the one for the mean I am sorry this is the this is one for the mean and this is the one for the variance.

So, what is that we are illustrating we are illustrating the application of recursive estimation by solving a very simple problem in statistics namely, if you get a newer sample if you get a newer data item how do you update your mean how do you update your sample variance sample variance. So, these 2 together illustrate very beautifully the notion of sequential update or online update.

(Refer Slide Time: 35:34)



With that we come to the end of the discussion of sequential estimation. So, the notion of sequential estimation is absolutely fundamental it can it arises naturally in solving statistical problems in computing many of these statistical standard moments, the first moment mean, the second moment variance centered second moment variance, and so on the notion of a recursive computation also occurs within the context of geophysical inverse problem.

So, we considered a set of linear inverse problem and illustrated how one can utilize the recursive solution to be able to solve the linear deterministic inverse problems with respect to availability of observations and being able to update my old belief to incorporate the new measurements there by deriving a class of online sequential recursive algorithm and we also mentioned this idea is inherent to Kalman filters and there is a considerable similarity between both Kalman filter and this algorithm, but one works in the stochastic domain another version is sequential domain and both the applications rest on the fundamental matrix formula namely Sherman Morrison Woodbury formula with that we conclude our discussion of sequential estimation.

Thank you.