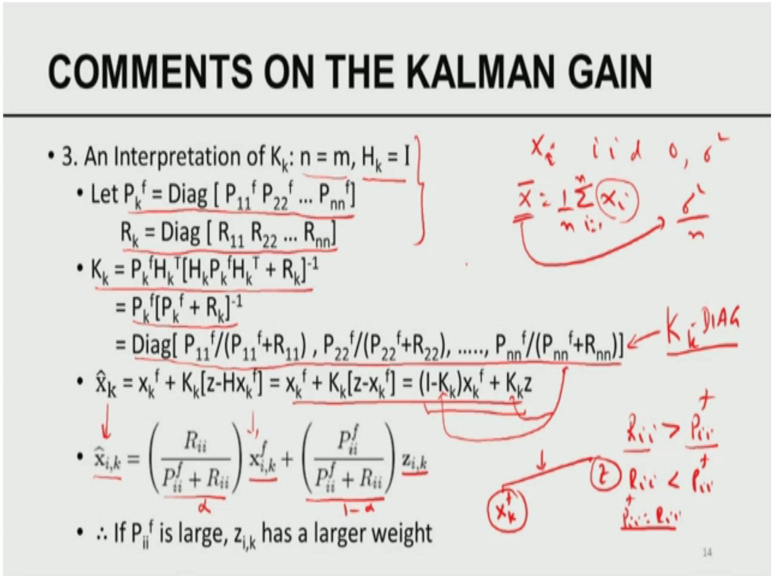
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture - 33
Linear Stochastic Dynamics - Kalman Filter Continued

Now that they have given an expression for the Kalman gain, we have expression for the forecast analysis, forecast covariance, analysis covariance. I am going to spend little time and understanding the structure of the Kalman filter in special cases, an interpretation if you wish of the Kalman gain.

(Refer Slide Time: 00:38)



So, let us try to consider a special case when n is equal to m in principle and practice m is not equal to n, but any anyway this is a mathematics sometimes if I can do some mathematical analysis under special cases that may throw some new insight into what I already have done. So, let us say what kind of insight I can gain by specializing some of the ah quantities in the expression for Kalman gain, let us assume m is equal to n, let us also assume H k is equal to I.

What does it mean? my Z k is equal to X k plus V k, because I am measuring the state itself that is very simple case. Let us also assume my forecast covariance is diagonal in general it may not be, but a I am trying to interpret things I am trying to specialize so I can special as many things. So, I am assuming R k is also diagonal look at this, there are lots of assumption going to that.

Why these assumptions? Why not, let us have fun if nothing let us have fun, to see what comes out of this and that is the curiosity. So, I am going to assume P k is diagonal R k is diagonal H k is I, m is equal n. In that case Kalman gain is a square matrix, this is Kalman gain I substitute all these things in here I get this expression, please remember P k is diagonal R k is diagonal therefore, P k f inverse a P k f times P k f plus R k inverse have an explicit expression in here therefore, if I substitute the KAL this form of Kalman gain into the analysis equation it becomes this. So, this is what I am after, now please understand k is given by this matrix so I know a special structure of k that is coming out of this that is coming out of this.

So, X k is a vector that vector is equal to this matrix times X k f plus K k z, but K k is a diagonal matrix right so this essentially tells you K k is diagonal, if K k is diagonal I minus K k is diagonal therefore, X k hat is equal to a diagonal matrix times the forecast plus diagonal matrix times the observation. If I consider the i-th element of X k that is X k I k; that means, i-th element of this vector the i-th element of this vector now has an explicit expression, this is not R this is P i i sorry there should be same as that ah oh I am sorry I am time that that is all I know I think I was a little R I correct, this is R i i is the is the diagonal element to the R matrix, P i i f is the diagonal element of the forecast matrix these are some of the two.

Now you can see the sum of these two, so you can see this is the weight. So what is that? Analysis is the weighted sum of the forecast in the observation that is very simple. What are the weights? If you call this alpha this is one minus alpha, so that is a convex combination. So, you can readily see in this special case I have forecast I have observation, so the analysis is simply a point in the line joining the two points analysis forecast and the observation; that means, analysis lies on the line segment joining the forecast point on the observation point.

M is equal to n observation space is equal to the model space, so these two points are lying in the same space. So, what does this tell you? The Kalman filter equation essentially tells you in this special case analysis is the linear is a convex combination our forecast and the observation. We have already seen this result earlier in a simple case therefore, Kalman filter equation is very con is very much consistent with everything we already know, it is the check of this internal consistency is the result of the analysis of the special case, not only that I would like to talk about that activity.

Suppose R i i is much larger than P i i f what does it mean? The observation is less accurate than the forecast, the variance worse means is less accurate or you may therefore, so if R i i is larger than forecast is more reliable than the observation therefore, R I is larger than P k I the denominators are same. So, this equation gives more weight to the forecast.

That means, analysis always favors the one that is more accurate, on the other hand if R i i is less than P i i; that means, observation is more accurate than the forecast, in which case Kalman filter gives more weight to the observation. So, what does it mean? I am it is like in a committee of 6 7 people every committee member have the same vote, we do not distinguish one vote is more valuable than the other if it did it that is not democracy in democracy all votes are equal, in a committee all votes are equal.

So, you can think of the analysis as a committee decision committee of what committee of two. Who are the members of the committee? Forecast and the observation, but do they have the same vote? No. Under what condition they will have the same vote? When R i i is equal to P i i p I, when P i i f is equal to R i i alpha is equal to 1 minus alpha, if alpha is equal to 1 minus alpha, alpha is half 1 minus alpha is half both of them have the same weight, the analysis is simply the average of the forecast and the observation, but seldom is the case where; P i i f and R i i are the same, R i i comes from the instruments p i i comes from the model, seldom the case p i i f and R i i can be the same.

So either R i i is more or R i i is less and that essentially tells you this Kalman scheme is very intelligent it is very adaptive it gives more weight to information with less variance, it gives more weight to the information which is more accurate that is the beauty that is the speciality. And we have already talked about this when we derived linear minimum variance estimate, we have already talked about this in the context of the Bayesian estimation, Bayesian structure linear minimum variance structure they all have this property.

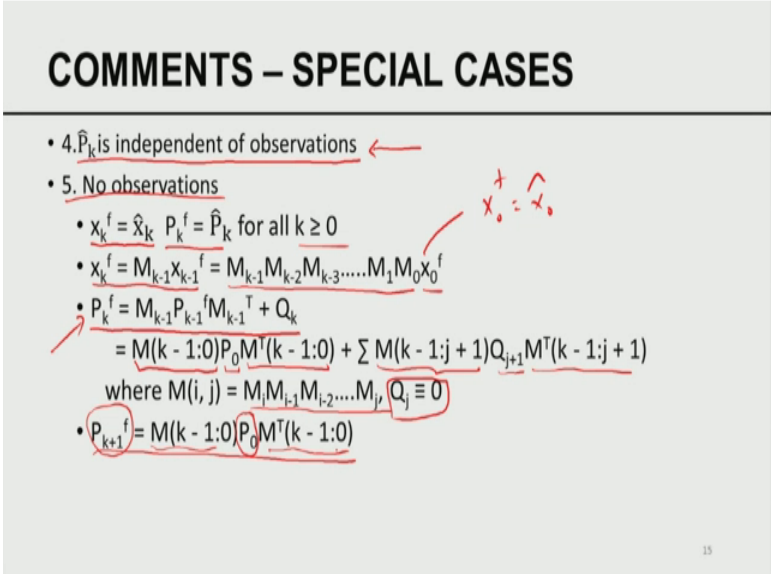
So, what is that other important thing? Forecast has some variance, observation has some variance, analysis has some variance when you combine these two random terms I am getting a new term which is random the variance of the combination analysis is less than the variance of the individual component; that means, I am having two bad decisions I am able to create a better decision from two bad decision, if you want to call bad in terms of variance in terms of the fact that variance is not 0. So, that is the beauty of data assimilation, data assimilation tries to improve tries to provide a linear combination whose variance is less than what goes on.

And that should not be new to any one of us who have done anything in probability theory for example; in probability theory we have this follow that mean I am sorry let me cut this result, we have already seen that earlier if X i is a random variable, if X i is i i d random variable, if X i have 0 mean and variance sigma square. If I compute X bar as summation X i i is equal to 1 by n, i is equal to 1 to n, X bar is also random but the variance of rank this is sigma square over n. So, individually they all have a larger variance, but I combined them linearly as a linear combination with equal weights average is a linear combination are you may so the variance of the average is sigma square over n when n goes to infinity the variance goes to 0.

So, what is that we are trying to do? We are trying to develop a quantity for random which is unreliable, a quantity which is more reliable than what goes in. And that is the general context of central mid theorem. What the central mid theorem says? That if you have a sequence of i i d random variables if you compute the average of that the average has a variant that goes to 0 as time goes to infinity and I can also normalize this average the normalized average. So what is the normalized average? If I sup if you divide by the variance it tends to have a normal distribution, a that is what is called central mid theorem. So, the idea what we are seeing is that is very similar to what central mid theorem says except that central mid theorem is an asymptotic theory we see an embodiment of the principle.

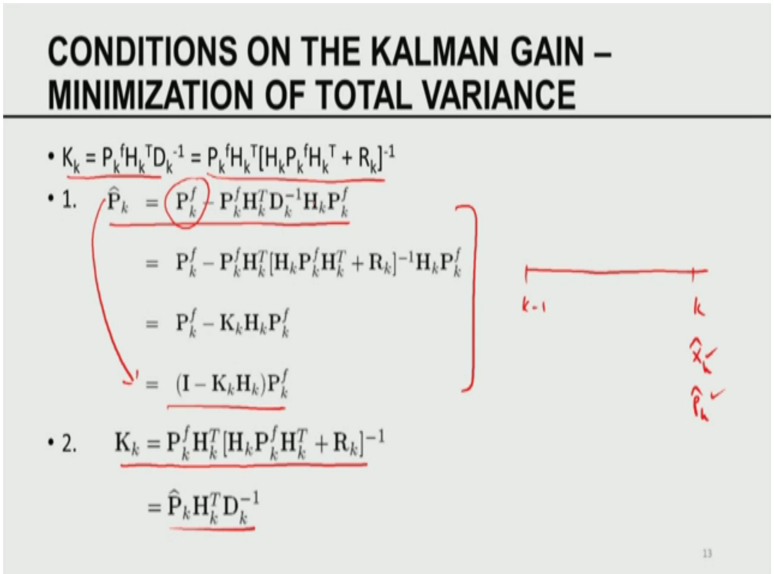
What is the principle? I can travel if I can if I if you give me two random quantities i can combine them in a clever way the combined quantity is random, but it is variance is less than the variance of the two quantities that went in, that is the basic principle of data assimilation and that is borne by this analysis.

(Refer Slide Time: 11:13)



I also want to make another ah comment P k; P k hat, to be able to do P k hat I do not need the observation I can pre compute them, their independent observation. I hope you recognize let us go back to the expression I think that can be talked about only when I look at the expression

(Refer Slide Time: 11:50)



Look at the P k hat in 1 in here sorry look at the P k hat in 1 in here, the expressions were P k hat involves forecast i already know the forecast needs only the model look at the other term, they all need only the forward operator they do not really need the observation itself. So, what is it what does it mean? The expression for the analysis covariance while it depends on the observational covariance forward operator it does not depend on the actual values of the observation. So, if you set up a problem if you know the forward operator even before you take the first observation you may be able to analyze the structure of some of these covariance's offline out of time that is idea. Now I am going to talk about the case when there is no observation.

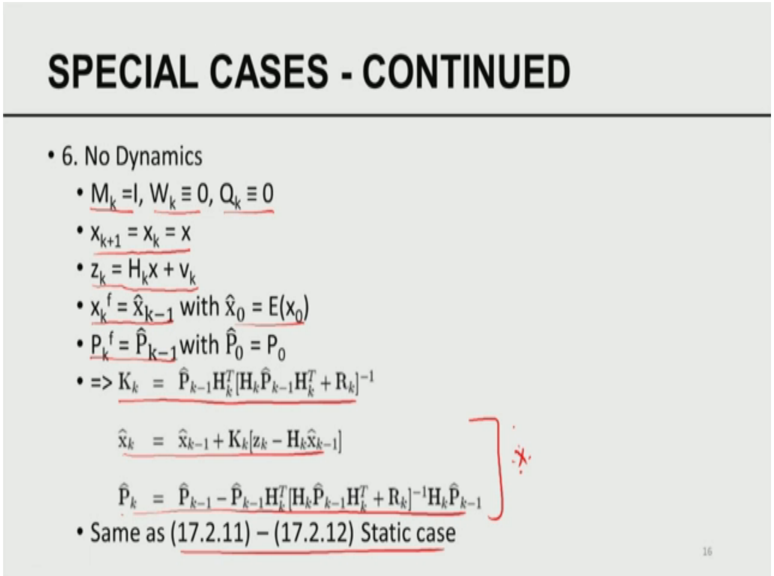
What do you mean there is no observation? I am simply have the model I am going to run the model for work what happens let us have fun. So, in this case X k f is equal to X k analysis is equal to forecast because analysis differs from the forecast only when there is observation, when there is no observation analysis forecast is analysis in that case analysis covariance is forecast covariance for all k I hope that becomes very clear. The forecast itself is generated from the previous forecast, so the forecast at time k is M k minus 1 times X k minus 1 f because there is nothing else, that essentially tells you X k f is equal to the product of all the matrices times the initial forecast, but the initial forecast initial analysis therefore; X 0 f is equal to X 0 hat. And what is X 0 hat? It is the mean of the initial distribution.

Now let us consider the forecast covariance, forecast covariance look at this there is a recurrence k depends on k minus 1, so if you open this up there is a product there is a product of model matrices times P naught, the transpose of that product plus the product of model and Q j this where M i j is given by this, all of with my place when. So, when there is no observation this is what is the forecast covariance, now when there is no modern noise I can set Q j is equal to 0 no observation no model noise.

So this is the variance of this is the variance of the forecast when there is no when there is no observation then there is no modernized please understand the forecast covariance at time k only depends on the initial covariance and the product of the model along the line along the trajectory. So, I want you to be able to look at all the special cases then there is no observation, when there is no model noise, what happens to these expressions? This is simply a sidekick arrangement of analysis which is which is helpful to recognize the role of each of this.

When m is not equal to n, when H k is not equal to I then there is no observation, when there is no model noise, these are all these are all these all provide different ah interpretation of the derivation.

(Refer Slide Time: 15:19)



Now let us consider when there is no dynamics static case when there is no dynamics mean that study case; that means, M k is I, W k is 0, Q k is 0 there is no model noise there is no model noise covariance there is no M k M k is I in which case X k plus 1 is correct X k is equal to I then which K z K a is equal to H k X plus P k there is a static case, because there is no dynamics that is only 1 time k. So, in which case the forecast is equal to analysis, this is the initial covariance the forecast and the present forecast is equal to the previous analysis.

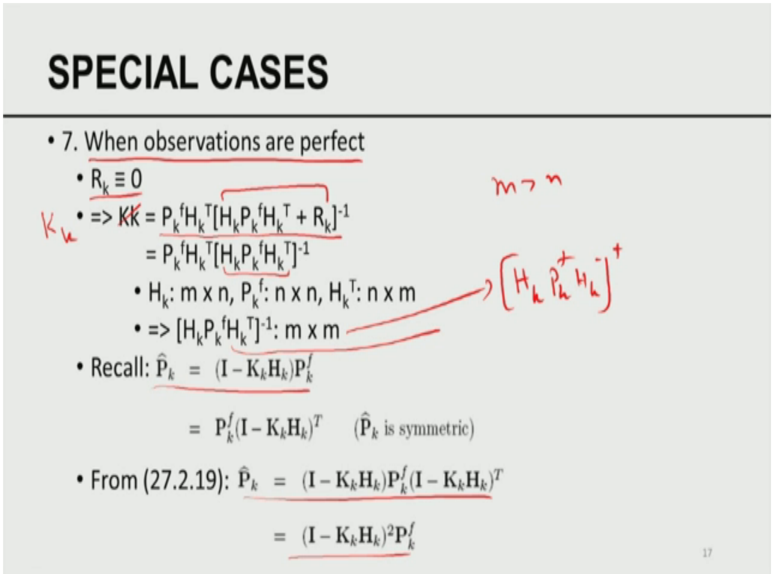
These are the Kalman gain in this particular case if you do that they essentially take this following form these are exactly the same as in the static case. I am going to leave the verification of these exercises, why is that? I am going by doing this exercise we are now going we are going to establish I have already done static deterministic data assimilation, I have already done static stochastic data assimilation we have direct formulas for the optimal estimate their covariance and everything else. Now I want to understand are they and the Kalman filters are related? Yeah if you if you take the dynamics off if you take the model noise off it becomes a static case the Kalman filter equation reduces to your form there is already discovered within the context of static analysis and all these exercises are meant to show the beauty the nesting.

What is the nesting? when you when do you say some two results are nested? when you special when you specialize one you get the other, linear and non-linear result are nested in the same when you assume non-linear is especially kept linear you get old results back. So, in mathematics getting a set of nested result is beautiful in itself.

Why? yes a set of nested result is beautiful in itself that gives you a room for consistency check if nothing we can check the consistency, if nothing we can understand the static theory and dynamic theory are one is called extension the other another is called the specialization of the other the as animals they are not too different from each other it is that realization of nesting that enables you to see that this theory is not a ah an amorphous collection of ideas it is a monolithic structure.

It is essentially understanding appreciating this monolithic structure, monolithic nature of the ideas that the underlying concept of least squares brings to the forefront, it is the least square that ties everything; static, dynamic, special cases, generalization and that is the beauty of the whole discipline of dynamic data simulation and static case as a special case.

(Refer Slide Time: 18:28)



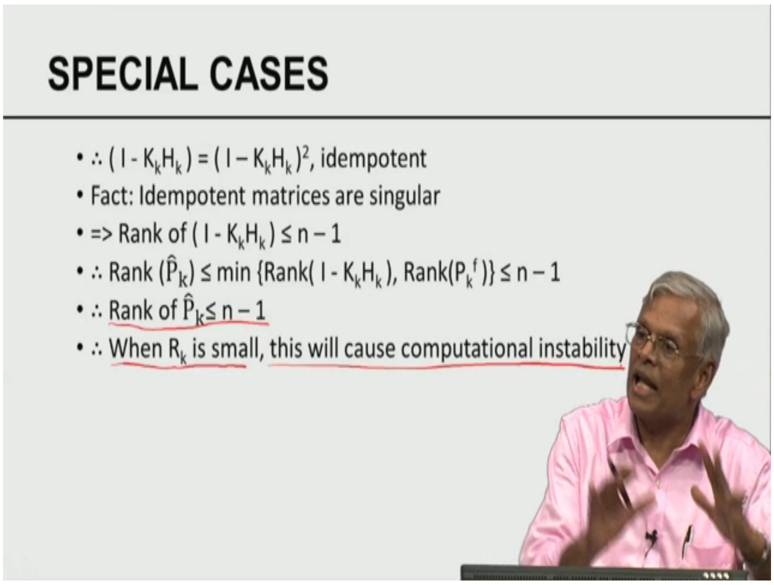
Now another last one you can see there are lots of important special cases to consider; when the observations are perfect R k is 0, in this case the analysis covariance take this far take this farm ah the P k takes this particular form.

So, when the observations are perfect you can readily see you are going to have to require to compute the inverse of H k, P k f plus H k, please understand this is the quantity that decides the Kalman gain, so this must be K k sorry this must be K k. So, go back this is the Kalman gain expression R k is 0, if R k is 0 I have to compute the inverse of this H k is the m by n, P k is n by n, H k transpose is n by m therefore, this matrix is a m by m matrix the whole question comes in how do you know this matrix is invertible? So, that could be trouble. So when you have a stochastic dynamic model when there is a perfect observation Kalman filter computationally could be could face difficulty because this matrix may not be non singular.

Why? Let us look at this now let us suppose m is greater than n. Let us suppose H is a full rank matrix. So, the rank of H is n the rank of P k f is n, but this matrix is m by m, m is larger than n. So, your matrix is made out of matrices of smaller rank, a larger matrix built out of matrices of smaller rank and that is a kick. So, even though; the expressions are simple H k, P k f, H k transpose is a m by m matrix, m is larger than n, H k even reproduce the full rank this matrix need not be a full rank if this matrix is near need not be a full rank I cannot get the inverse. So, in this case what is one way computationally to deal with I simply take the generalized inverse which you have seen, so generalized inverse.

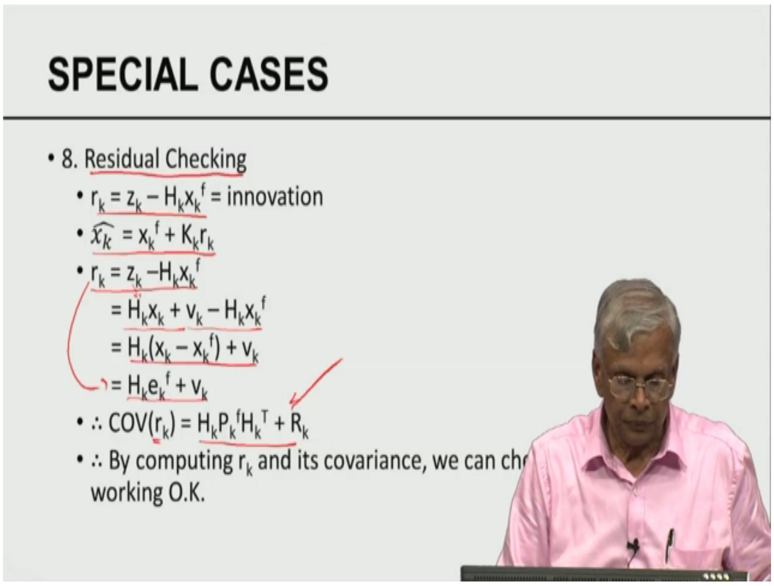
So, that could be that could be numerical difficulty, computing the generalized inverse is not easy therefore, what is the story? We would tend to think that in the observations are perfect mean there is no observation there is no error means that will help you. So, within the context of Kalman filter equation perfect observations are a nuisance perfect observations of nuisance in this particular case ah.

(Refer Slide Time: 21:31)



Therefore, the rank of P k so that is exactly what I what I have talked about sorry, so from here you can see the rank of P k is less than or equal to n minus 1 that that matrix may not may not be may not be s p d and if that matrix is not a s p d there could be a computational difficulty therefore, what is the story? When R k is small this could cause computational instability. That is again comes from the analysis of the special case.

(Refer Slide Time: 22:04)



The last one is called residual checking how do you check the correctness of? Suppose you have written a filter, you think your program is right how do you check the correctness of that? That is what is called residual checking. What is the residual? Residual is z k minus H of X k f innovation, x k hat is given by this; the Kalman filter equation is the forecast plus Kalman gain times the innovation therefore, r k is equal to z k minus H k X f I am going to substitute for z k from here I have the forecast I can be re written like this it can be re written like this therefore, residual has this equivalent expression. The covariance of the residual is given by this you can readily see that.

So what is that one could do? So, if you are trying to implement if you are trying to implement the Kalman filter at every stage you can extract from your program the residual. So, you will have a time series this residual, if you have this time series that there is a residual from the time series you can compute the covariance of the time series. So, the computed value of the time series from the residual must match this. So, you know H k you know the forecast covariance, so you can theoretically compute this you can practically compute this if these two agree; that means, implementation is pretty good that is the way to check for the residuals.

(Refer Slide Time: 23:37)

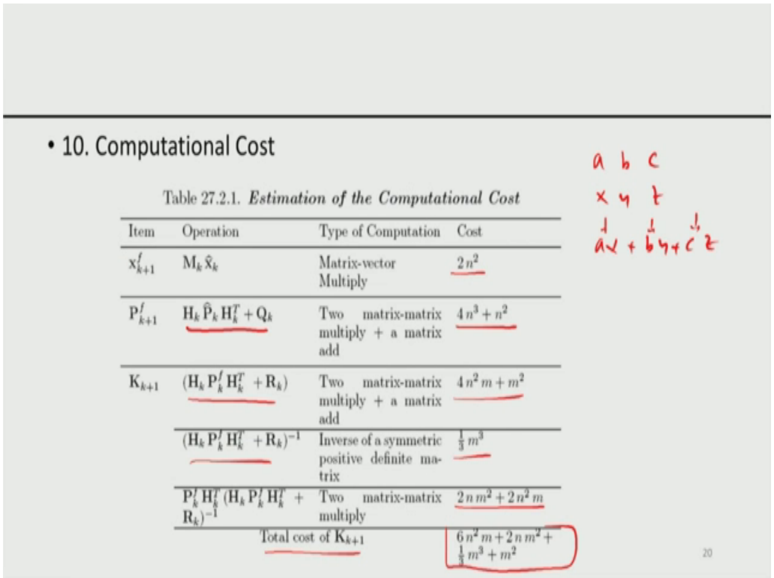


Table 27.2.1. *Estimation of the Computational Cost*

Item	Operation	Type of Computation	Cost
x^f_{k+1}	$M_k \hat{x}_k$	Matrix-vector Multiply	$2n^2$
P^f_{k+1}	$H_k \hat{P}_k H_k^T + Q_k$	Two matrix-matrix multiply + a matrix add	$4n^3 + n^2$
K_{k+1}	$(H_k P^f_k H_k^T + R_k)$	Two matrix-matrix multiply + a matrix add	$4n^2 m + m^2$
	$(H_k P^f_k H_k^T + R_k)^{-1}$	Inverse of a symmetric positive definite matrix	$\frac{1}{3} m^3$
	$P^f_k H_k^T (H_k P^f_k H_k^T + R_k)^{-1}$	Two matrix-matrix multiply	$2nm^2 + 2n^2 m$
		Total cost of K_{k+1}	$6n^2 m + 2nm^2 + \frac{1}{3} m^3 + m^2$

So, now I have talked about computational cost; The forecast step let us look at very quickly, the forecast step matrix vector multiply because to multiply matrices you have to multiply and add. Let us talk about this now, if I am going to compute the inner product of two vectors; so a b c, x y z if I want to compute the inner product of these two a x plus b y plus c z I need to multiply also I need to add, so I have 2 3 vectors I have to multiply three multiplication two additions. So, if you have to do the inner product of 2 n vectors I have to do n minus 1 multiply n and in my I am sorry n multiplied and n minus an addition or it may therefore, there are basic operations a basic number of operations. So, matrix vector multiply 2 n square to compute this quantity is going to take 4 n cube plus n square, to compute this quantity is going to take this much time, to compute the inverse is going to take me that much time, to compute the Kalman gain is going to take this much time, so the total cost of computing the Kalman gain is given by this expression.

So, this is the price they are to pay, that tells you the amount of addition, multiplication, subtraction, division you have to perform your computer must be able to perform all these in a short time; that means, I may need more powerful computers, so this talks about the workload. And who should be interested in this? If anyone who is going to be able to write a program to develop a system they have to worry about these things.

(Refer Slide Time: 25:20)

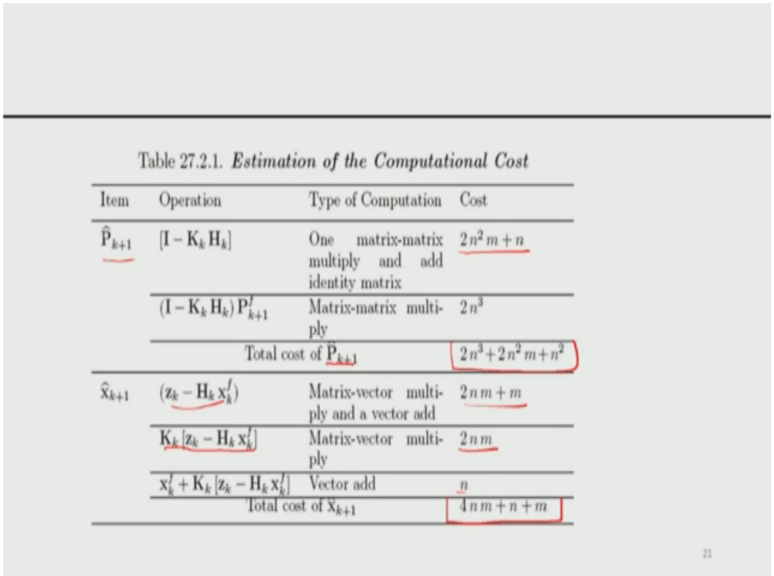


Table 27.2.1. ***Estimation of the Computational Cost***

Item	Operation	Type of Computation	Cost
$\hat{P}_{k+1}$	$[I - K_k H_k]$	One matrix-matrix multiply and add identity matrix	$2n^2 m + n$
	$(I - K_k H_k) P^f_{k+1}$	Matrix-matrix multiply	$2n^3$
	Total cost of $\hat{P}_{k+1}$		$2n^3 + 2n^2 m + n^2$
$\hat{x}_{k+1}$	$(z_k - H_k x^f_k)$	Matrix-vector multiply and a vector add	$2nm + m$
	$K_k [z_k - H_k x^f_k]$	Matrix-vector multiply	$2nm$
	$x^f_k + K_k [z_k - H_k x^f_k]$	Vector add	n
	Total cost of $\hat{x}_{k+1}$		$4nm + n + m$

Again that thing continues this is the analysis covariance I am sorry, this is the analysis covariance analysis covariance computation takes this much, the total cost of computing the analysis covariance these are the various steps of this that is n cube, this is the residual computation that takes this much time, this is the multiplying the residual by the Kalman gain takes this much time, then I have to add the forecast with that is n time so the total cost for the analysis computation is this much. So, in the previous we talked about the total cost of computing the Kalman gain.

Now, we are computed talking about the total cost of computing the analysis. So, I believe we all have a good handle on what things are happening; yes it is a very dense set of lectures, these lectures use many of the results from the previous analysis this way a representation is a modular way that way I can present many results ahead of time we become familiar with individual components of these results. Now we are trying to ah assemble all the results within the context of Kalman filters, so with this we have in essence completed the analysis of linear Kalman filter equation this is a classic problem is the sequential data assimilation scheme even though Kalman did not call it data assimilation that is exactly what he was doing.

And I want folks in geophysical sciences to realize that Kalman coming from an engineering discipline actually solve the data assimilation problem for the first one to solve it, but the first one to solve it that is the importance of this. Now where do we go from here? I am going to illustrate these things by some examples and I am also planning to give a table with a complete algorithm and that is what we will do as the first order of business in the next lecture.

Thank you.