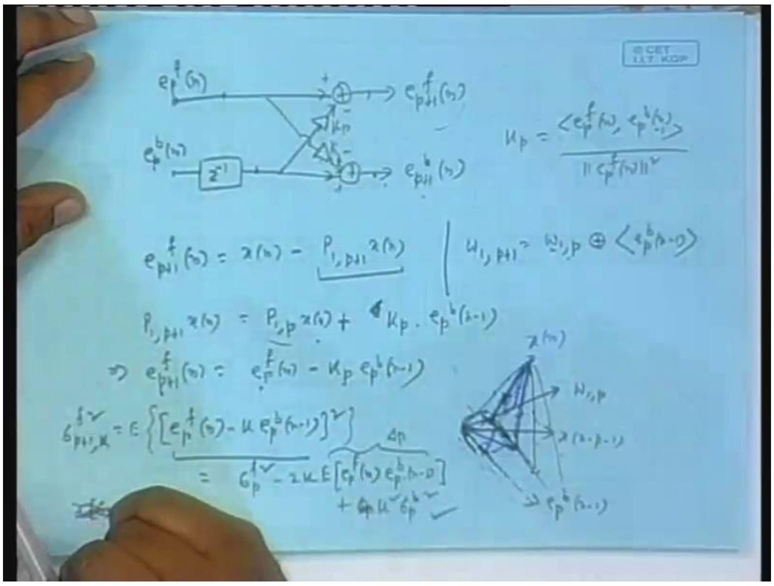
Adaptive Signal Processing
Prof. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 27
Gradient Adaptive Lattice (Contd.)

Last time I was discussing, I have very little time actually I just started discussing gradient adaptive lattice, but I just only mention, what is it I could not discuss it all. So, I will take that up today.

(Refer Slide Time: 01:22)



First let us consider the original lattice filter any stage p th stage. This is the thing, this you have familiar with, but to construct this lattice we must know only one thing that is K p. And K p I am assuming to be real for time being, let us assume that we are dealing with as far as the gradient adaptive lattice is concern, let us consider only the real valued random sequence. And therefore, all the correlations of variances a coefficients and all they are real, so that is why K p originally it should be K p and K p star, but I have made it K p and K p no conjugate because, we are considering all only the real case.

So, then K p was this we all know n minus 1 divided by any of the two both are same e p f m square or e p b n minus norm square this is K p. So, you need to construct this particular lattice stage, you need to have K p. And K p means you need to build the correlation between these two prediction error as given in the numerator and various of any of them as given in the denominator.

Now, if it is stationary and statistics is known you can even eventually get all this things in terms of the correlation of the original process x n that we have seen, we can recursively update these norm square. And also using the predictor coefficients, you can evaluate this numerator thing, we can called as delta p. So, essentially giving the statistics correlation values auto correlation values of the original process x n the input the very first input stage input x n, you can find out this K p.

But, suppose you do not know the correlation values or it is totally changing with time statistics. So, therefore, you need to have recursively by which this multipliers they learn their value, they adjust, they adopt themselves with types, so that they finally, I mean they learn from the data and added their values to the ideal K p. So, that if the input statistics changes afterwards, it will again track and readjust itself to the new set of K p values, there is a adaptive.

So, I want to make it now adaptive, when the input statistics is given input is stationery, say statistics is not changing and also correlation is known you can calculate this. But, when the correlation is not known or it is changing with time from time to time, then you have to have adaptive mechanism by which thus multiply learn, they adjust their values, you know regarding manner regarding mean time. So, that they finally, converged on the optimal values at least in mean as you did in the elements case.

And if there is a subsequent change in the input statistics, there is auto correlation values change it will again track because, it is a adaptive. So, I want to March towards that adaptive version of this lattice now, we know there original values. Now, before we do that let us see something, what is e p f, this quantity what is this after all x n minus the projection.

Let us for the time being concept on the this projection, here for this projection your this is the soft space, this you decomposes orthogonal span of what, what is the last guy in this w 1 to p plus 1 who is the last guy x n minus p minus 1 I take that out. So, it is what I am left with is w 1 comma p and the last guy is projected orthogonally on this and that error is taken. So, it is an orthogonal decomposition that derived as we have seen is e p b n minus 1.

This is a span of that component, this you have seen earlier the result I am just going back to the simple derivation, let us just try to recall how we derived. This total substance was decomposed as what w 1 comma p and the last element was projected orthogonal on this and the error taken that was e p b n minus 1. So, span of that and span of this they are mutually orthogonal and there is direction of decomposition of this.

Then I projected x n on this space means, projection on this and projection on this, orthogonal decomposition that is why I did. So, this p 1 p comma p plus 1 x n is nothing but, projection of x n on this space, projection on this and we add that two and then add the two and then you subtract from here, that is how we got this lattice stage, you remember or not you got it that way.

That means what we did there was P 1 p x n plus this component is that this was just K p I called it K p f , but latter I showed that K p f n ((Refer Time: 07:21)) mean conjugate of each other in this case they are same. So, K p times e p b n minus 1 and that you subtract from here, this is repetition mind you, but still I am doing it for some reason I want to go back to this derivation again, what we did you have to subtract. So, it becomes x n minus this which is e p f n minus this quantity, geometrically try to see this I mean I am not making use of any you know geometrically features.

But, just to have clarity as to what is happening, suppose we have a situation like this that, this line it corresponds to a soft space anything on this line, this origin, this line is a soft space. Suppose, this line denotes w 1 comma p, this is indicated by this and you have got this fellow the last component, what we will do, you will project this on this space and take the orthogonal this way, so that is this, this is the axis, this is this component.

So, anybody projected on this the plane will be what, the projection will be summation of the projection on this and the projection on this. Because, there are, but I am not making use of any angular on that in my derivation, just inner product and all, but just for projecting it I am just showing a figure. Now, suppose that external fellow x n is this guy, this is your x n, unfortunately you get 2D plane I have to I cannot show 3D plane you have to use your imagination and this is actually going lying outside.

And this have to project on this, this the plane span by this fellow and this fellow or if you only this fellow and this fellow. So, and that will be what we all know from basic projection that will be that if you project it this is that orthogonal projection, this is the projection, this is the corresponding projection error, this error is this e p plus 1 f n and this is that projection, this projection is given by this guy.

But, what is happening is, what we did here, we wrote that projection as a summation of two projections projecting this along this error, which is this component, projecting this along this that is this projection will be nothing but, if you project the perpendicular this much and if you take a projection on this, this might becomes little clumsy, this is a say here, suppose this is here this is 90 degree. Then these two projections, these two fellows this fellow and this fellow, they added together will give you this much that is a projection are you getting me.

Incidentally this competent is also orthogonal projection of this fellow, this final projection if you take it, project it on this guy you will get the same this much component because, after all this component is what total x n minus this error. So, this means x n minus error inner product with this will be what, this vertical component and this will go. So, again inner product with this divide by norm square of this into this you know, so you get the same thing.

So, whether you take this fellow and project it on this axis or take the total thing x n and project it on this axis we will get the same thing same on that side. So, then what is this error, this error is this component because, I asked them, so long talking of this projection, now the projection error means this fellow. So, clearly this fellow is what, this projection error this fellow and this much and this two, this actually this is the hypotenuse by the way, it might look to be shorter side.

But, this is the hypotenuse this is 90 degree, original you see this was the projection, this projection is a summation of what I am doing in this diagram is nothing but, it is a geometrical description of this equations. This projection is nothing but, this projection, this projection that is this, this, this, this, this, then I take the error, error means x n minus this projection that is e p f n that this fellow and minus this guy.

So, this guy minus this guy is this error, this guy minus this guy that is this error actually this is 90 degree. So, this the hypotenuse, so how much is this guy, this guy is this vector times K p vector, this is e p b n minus 1 that is spanning this line that times K p, suppose now instead of K p I have a gene any k. So, that can if I say that suppose instead of taking up to this I go up to this, then this error would become this much if I go after this error will become this much.

So, by varying K this error non will vary and as you understand as K defeats further and further from the optimal 1, non square of this error vector will increase, again if you go to the left hand side it will increase. In fact, we can now show that it will get quadratic function of that K non square of the error vector that will minimum only when we are taking at the projection and this, this is orthogonal to this. Because, after all that error normally minimum only when it is a orthogonal projection that we have seen in the general case are you following me.

So, if I take instead of putting a K p if I take a general expression e p f n minus some K times this, that is a general error vector I take the norm square of that, norm square has to be a quadratic function of K, which is seen algebraically, but I wanted to give you the geometric inside, quadratic function of K. So, that it has a unique minimum when K is correspondent to this point. And as you deviate further from this point, that non square of that error vector should increase I do not know whether you understand this is algebraically I can always.

So, but you know I do not like that, you should see the geometrically meaning, this projection error is nothing but, this fellow which is e p f n minus this fellow, this fellow is K p e p b n minus 1. This is exactly this much when this coefficient is K p, but suppose instead of K p I make it just some arbitrary K times e p b n minus 1 that becomes sliding along this axis and instead of this some time if I take some K for which this much K into e p b n minus 1 is this much, then the error will become this, so on and so forth.

And as I deviate further and further this was a non square of the error vector; obviously, we will increase you can see, you can now if I geometrically also. So; obviously, it will be a quadratic function of K non square of the error vector and it will be minimum only when I am here. Because, there it will give rise to projection and that is what, you can easily see, if I now consider a case like this you know e p f n minus general K and take the non square means all are real, so just square it up and E.

What will you get because of stationarity, you will have sigma P f square K square term will come sigma P b square and another minus twice K this into this correlation that expected value. So, it will be a quadratic function of K that you can easily see from here, but I did not want to just rewrite it that way, that I have look I take a general K it becomes a quadratic function I wanted to give this inside. But, why it is quadratic because, the more actually the when you vary K you are sliding along this axis, keeping this part intact you are sliding along this axis.

And that is why instead of this optimal error, sometimes you can have this error, sometimes you can have this error depending on the K you choose. But obviously, from here also you can see that error norm will increase as K I mean deviate further from the point and that you can verify. So; that means, it is quadratic in K and it is in this quadratic in K, here itself you know that for there is only one unique K, which is optimal for this norm square is minimum that corresponds to orthogonal projection and with this mathematical ((Refer Time: 17:04)).

Now, suppose I carry out the steepest descent now, that I do not know what the optimal K is, like you know in the case optimal that adaptive filter, how do will you go to adaptive filter. First we started with the optimal filter, we not expression arrangements p and then I said suppose I do not know how to computer inverse or R is not although R will change from time to time. Suppose I do a any procedure I said, the steepest descent search, so what I did I did took an arbitrary with filter with coefficients I took arbitrarily.

And further I find out that mean square error of the output to the gradient of that with respect to the weight, this square will be more and more as you deviate, as the filter coefficient deviate further and further from optimal one. So, then I took the gradient and I went against the gradient, so that I reach the minimum point that optimal point, similar thing I can do here. So, suppose in this stage I have already know what is e p f n, e p that is this is crucial, please understand books do not discuss all this I am assuming that up to the previous stage of the lattice everything have been done nicely optimally without any error.

So, I am indeed getting the correct e p f n correct e p b n that's an assumption, which is actually not a valid do not a very practical assumption. But, suppose it is, so up to previous stage that is everything is fine, but here I do not know this K p value and I want to find them out the instead of carrying out the form instead of evaluating by the direct formula I want to do a steepest descent business to find out. That means what I have to do I formulate this quantity I call when it is K p I know what this quantity, this sigma p f square.

But, for general K it not sigma p f square it is sigma p f square sigma p plus 1 f square rather nothing sigma p plus 1 because, when it is K p, when it is this coefficient is not just K, but any arbitrary in the proper K p. Then I have this difference is this and it is non square there is variances nothing but, sigma p plus 1 f square, but for a general K it is not so, that is the minimum non, but general K it is not it could be anything it could be more than that.

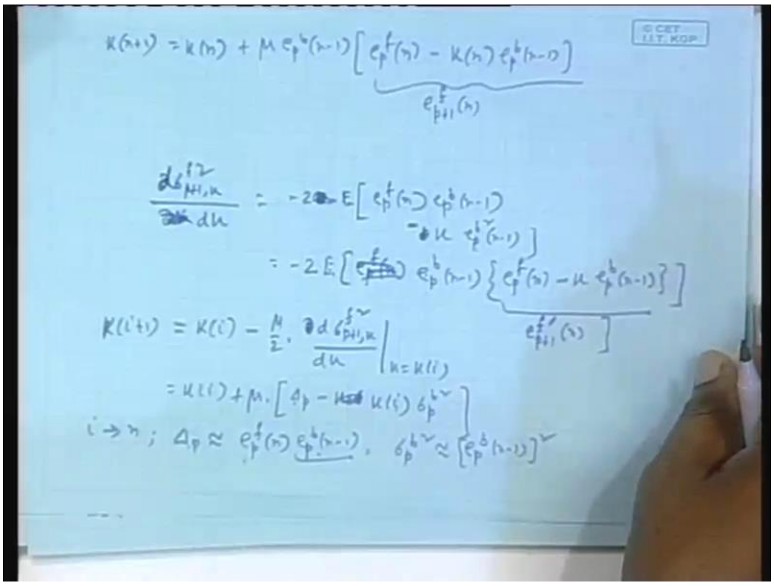
So, I call give it a name sigma p plus 1 is a function of K, say f square I bought in K, so this is a quadratic function of K, geometrically you have seen, this quantity this is nothing but, what is this quantity this errors. What is this quantity this difference, this quantity is either this error or this error or this errors these are function of K, depending on the K it depends on the you slide along this axis and you get this error this error or this error. So, this is whole thing is that error non square is a quadratic function of K of course, and only when K is the optimal K, you get the minimum norm error.

And here also you consider will be quadratic function of K and what it is actually, if you break it up square it up and take expectation sigma p f square minus twice K let me write down instead of ((Refer Time: 20:24)) e p f n this what did I call delta p e p f n, this I called delta p, this quantity my lattice was delta p plus K square sigma p b square I am assuming stationary process and all that for the time being.

Because, when I remember when I move to steepest descent from optimum filter, I did not through stationary I assumed stationary p that is your gradient expression came they R matrix was there p was there. Afterwards when I switched to elements I approximated R by a very wild approximation x n vector into it is transpose no averaging and all that, p vector was this was approximated by x n vector into d n no averaging. But, there is a only where I move to elements from there, but up to steepest descent sationarity assumption was value it and my gradient was in terms of R and p.

Here also when I go to steepest descent stationarity assumption will be kept intact, that is why quite up I am taking expectation in whether it is n minus 1 or n it does not matter, that disappears here from sigma p f square sigma p b square, the only thing is it is a quadratic function. So, I know what is this gradient, gradient of this quantity sigma will be what I am writing here.

(Refer Slide Time: 22:09)



So, if now I guess there is no point in I thing partial derivative here because, it is a function of only one parameter K in the case of optimal filter I had, so many types w 0, w 1 up to w p and that output error variance was a function of quadratic function of each of the weights. You remembered e that b warm square that the variance of that error in the case of optimal filter there was what was a quadratic function of all the type ways w 0, w p.

Since there are, so many I have to use the notation of partial derivative, here I have got only one parameter K it is like a lattice stage one term another term, like there I had x n x n minus 1 it is just one component another component, they are linearly combined only one filter coefficient, which I am calling say K instead of K p, K p is the optimal one assuming I do not know. So, I am putting it arbitrarily K and by steepest descent I have to find out what the K p is that is the different matter.

So, you can drop this notation also it should be d according to me, you know this difference ((Refer Time: 23:24)). So, this should be according to me b, so what is this; obviously, minus twice K E K will not be there minus 2 E and this 2 K is here. So, that 2 and 2 I take common e p f n e p b n minus 1 plus K times I wrote it sigma p b square I am coming back to this form, I am putting them under the folder E operator instead of sigma p b square I am writing e p b n minus 1 square and putting it on the E operator that is fine.

This quantity square it up and there is E, so square it up b and then little trick here 2 E there is a minus inside is it.

Student: Minus K.

I took it right minus thank you, so and then I suppose I take this e p b n minus 1 common, then what I get is within bracket e p f n minus K, what is this quantity, this is nothing but, this quantity itself this we can call e p plus 1 if prime n prime because, I did not put K p here, the arbitrary K. So, it is n it will become e p plus 1 f n, when K is replaced by K p because, is a correlation is done that is this quantity is replaced by this.

So, it is just minus 2 if I mean correlation between these two terms, anyway I just wanted you to see this mechanic, that if you will take this thing I told you something in advance. If you take this common this comes out, these observation I will make use of later for the time being again I can switch over to this because, I do not need this at this moment, from here from here I can then derive one simple steepest descent equation as K i plus 1 as what this off line mind you this is not real time off line.

Before I start using the lattice I am just doing this off line business, like this all steepest descent minus, this minus some constant time, constant it is a mu by 2 I want the 2 and 2 2 cancel mu by 2 absolutely similar thing you know mu by 2 into this guy Del not del again d sigma p plus 1 K f square by d K evaluated that K equal to K i and you replace this value of the derivative here K i mu by 2, if you take this expression for derivative for the time being I will go back to that little later, actually I should never done that, that will come when I actually that is a more handling when I go to LMS from here.

So, for the time being let us stop here only and we derived this, so if I just take the derivative minus 2 delta p. If you want to write it compactly instead of writing, so much you can just write it minus 2 delta p and 2 K sigma p b square and then minus 2 and minus mu by 2 it will become mu. So, power plus mu delta p if I make any mistake correct me minus K, K means K equal to K i, so K i sigma p b square you are given the correlation values.

So; that means, delta p and this also known to you, like in the optimal just steepest descent case they are R matrix at p vector we are giving in that process and the gradient involved R and p remember p minus R w something that was the gradient in the steepest descent business also. I assumed input to be stationary and R and p were given to me and that gradient was available in terms of the R and p, in that p minus R w and w equal to w i similar thing here delta p and this are known to me.

And that gradient is evaluated K equal to K i that was evaluated at w equal to w i it is K at K equal to pi. Now, suppose I want to now move from here to LMS, so again I will see this procedure is very similar absolute similar. Then first thing was replace i by time index n as though I am carrying out the iteration by looking at my watch and getting time. You remember that is how you are going to do LMS from steepest descent or I was replaced by time index n and this correlations will be replaced by some wildly estimates same thing.

So, i you replace by n delta p, delta p is what I told you this quantity is delta p expected value of these two. That means, I should have several such products you know for I mean one sample times these, another sample of these times, another sample of these like that and then I should have average that will be a good estimate. But, suppose I am making a wild estimate, just this into this that is all, that will a very bad estimate, but suppose I replace that and sigma p b square is coming that is what expected value of the square of this term.

So, I should have several samples of this terms squared of add it an average, but again I replaced that by a wild average, just square of this itself right very much in LMS. So, delta p suppose I replace as, then from here I get the LMS version, what is that LMS version that is K actually I should have a K subscript p. Because, I am concerning p'th stage, but to avoid complication I am not bringing that subscript p also, but at the end when I give you the final expression, then there will be subscript with K.

Because, it is only for the p'th particular stage p'th stage of the lattice I am focus on, each stage will have it is own different constant. So, actually this is K p n plus 1 for the time being I have dropped that actually that dependents on p, but when I give the final expression and all that, that time I will bring that subscript. So, K n plus 1 will be K n plus mu times e p f n e b n minus 1 and e p b n minus 1 square and it is here I take e p b n minus 1 as common. Something which I was doing earlier I told you that I should do rather do it in the case of when I telling this.

If you put that delta p here, I will not do this steps anymore because, I have already done it once. You replace delta p by this and sigma p b square by this, take e p b n minus 1 as common because, it is present in this also and this also, so within the bracket what you are left with e p f n minus K i K n into e p b n minus 1, which is my e p plus 1 f n under this adaptive situation, not ideal e p plus 1 f n.

But, under this adoption situation means when the depletion coefficient is not the original 1, but whatever you got K n plus 1 or K n or in case you call you do not understand then I will write e p f n minus K n, this K n I replace by n K n square of this only one term remains. What is this quantity that is in the lattice e p f n what are the lattice, in this lattice suppose we have in a adoptive frame work, you get this for current values, but this coefficients are changing with time in the LMS manner. So, at any time it is K n, K n, so corresponding output is what we have in this expression that is...

So, please understand there is sight differences that when I write the same notation e p plus 1 f n it is not with the optimal K it is under the adoptive equation that I am just running an adoptive filter now this is the output. Like in the case of adoptive filter also during the turning phase just not converse to the optimal one, but I am not discarding the output that time, there is a still output y n at the filter output.

But, what is that output, the output with non optimal set of coefficient w n that corresponding is started as output is y n it is just that, it may converge to optimal one I mean when K n, K n be converge in this process to the optimal one. And then converge you will get this to be the actual optimal e p plus 1 f n as given by the optimal lattice we derived. But, as of now here it is just the one that you get at the lattice output with the available coefficient.

But, static assumption is that the two inputs are correct, that is why their variances were indeed those sigma p b square and all those things. Now, the problem is if I just stop there, problem is this will say converge ((Refer Time: 35:40)) converge though that converge analyze I am not doing because, that is more complicated because, that has to take care of the fact that I assume this to be optimal. But, this may not be the ideal one because, this also have been generated adoptive from the previous thing error as some error as come in.

So, is a overall thing is much more complicated, so I will not do, but again conversation analyzes has been done by people and they found out for that very low I mean mu as to be chosen in general. I mean you we have to take very less value of mu and all that some range also giving in some complicated expression, but it converges, but up till now the problem is, it will converge to some value. But, that some value is what converge will be mean it is not that we finally, exactly converge on the optimal one do we converge it in mean.

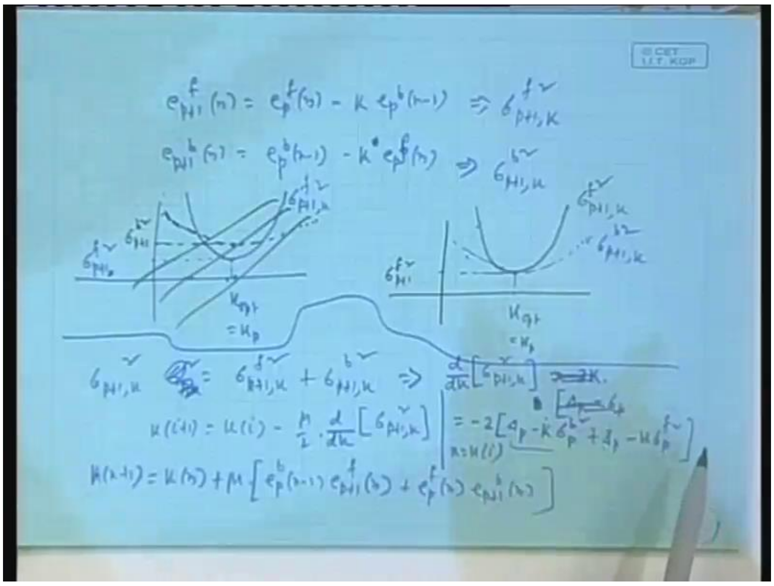
And on the mean it will be oscillating and then you have to see how much the variance, the variances to be suppressed by clever choice of mu and all that as we did in the LMS case. Now, suppose I can do the similar analyses, but with this guy that time I took the norms this fellow e p plus 1 f n this projection, this error this will give a geometrically interpretation some all that, it become a quadratic function of K and carried out that steepest descent followed by LMS.

So, that this guy converges at least in mean to the optimal one, but I can do a similar thing here also, it is a same projection kind of frame work, only the names will differ names will change here. This will become x n minus p minus 1, this will become x n, this will become w 1 to p it will remain, x n w 1 comma p x n minus p minus 1 similar thing. So, there also I can carry out a same thing you know I can put a general K, this norms square will be that is this fellow with a general K here it is norms square of variance will be quadratic function of K again.

So, if you go the steepest descent and then LMS you will get another similar set of equation. So, that will also give rise to that convergence and that K there will converge to the optimal 1, but only mean, but there is no guarantee that the two case, in the two branches will always remain same and after convergence this also will remain same. But, the lattice once in the lattice filter I have proved that this two must be same, they are same under real case they are same that is never assured.

So, I have to generalize this business somewhat, so that it is guaranteed that in the after converges also, if any 1 converges only in mean and not exactly on the optimal K p, this two figures are locked each other, to do that it is very simple to generalize this. You see one thing, if I replace this by K p thus K p by K I take a general error like this error, this error take the norm square variance of that that is of course, a quadratic function of K geometrically you have seen mathematically we have seen.

(Refer Slide Time: 38:47)



Similarly, if it is e p plus 1 b n this I will write down, we have seen this much that suppose if you bring a general K here it is a norm square of that that is of course, quadratic function of K the physical meaning, geometric meaning have giving all that, for this half. But, this is a problem is I from that only if I go for steepest descent and LMS, I have that problem that I will not be I mean I will not be able to assured that the two case will be locked to each other.

Similarly, if I do for the lower half that is this guy minus a general K time this, say some K prime or say the two case are different let me I mean just even though I use a same notation e p f n is here also. If I take the norm square of that, that will be the quadratic function of K by the same geometric understanding an algebraic expression and you can go for a steepest descent here and finally, LMS in same manner, only when the two case they will even after convergence, this will be replaced by K n this also replaced by K n or K prime n if you want to use different notation.

They will not the remain lock to each other, they will not be same after the convergence also not only before convergence, after convergence also because, convergence only is in mean not exactly on optimal K p. So, two guys have put land on K p and therefore, remain equal that situation will not come, they are only oscillating around K p the optimal coefficients.

So, there is no guarantee, but now see one thing, if I just take this fellow with general K this was given I take the norm square of this the variance of this that I gave the name sigma p plus 1 I bought in K f square. And if I take the norm square of this or I can give the name sigma p plus 1 K b square, now this guy is a quadratic function of K and there is an optimal K, K optimal, what is that K optimal the 1 K opt which is same as K p.

So, this fellow would be a function like this sigma p plus 1 and comma K f square, it will be a function like this, this much will be sigma p plus 1 f square that is when K equal to K p. This will be having the minimum value and minimum value is the variance of this, similarly if you pop plot this it will be a quadratic function again of K that minimum will occur at the same K from our theory. Because, we know from the theory that the two case two K p's are same under real case.

And if you go back to the theory and no LMS now, proper lattice and all, we know that this also will I mean b quadratic function it is norm variance will be quadratic function of K and that will be minimize at K equal to the same K p under real case. Because, under complex K p and K p star, but now K p and K p, if I again take this expression and I mean plot it you can even work it out is a square it up and differential and all you get the same K p.

Then it will again be a quadratic function of K and minimum will be under the there is important thing the same K. So, it could be something like this, this much is your sigma p plus 1 b square, but actually this two are also the same, so it should hit the same point, this figure is not correct because, even these two are same. So, I should not, so them to be different K opt equal to K p this is sigma p plus 1 comma K square with f of course, and this is the other one could be like this.

What is your problem, the other one could be like this sigma p plus 1 K b square both have the minimum value, both are quadratic both have the minimum value of the same K from the theory it appears the because, under real both are same K p. And that K of is K p and again what is the minimum value in this case sigma that is the norm square of this and norm square of this, which are same we have proved that is why it is like this.

But, earlier in the steepest descent I forgot about one I consider this on this I did steepest descent and then LMS approximation I got one recursive equation for K n as I did that time this one. But, ((Refer Time:44:17)) you can do at this also, you can get another recursive equation on K n and there is no guarantee that the two K sequences will at least after convergence will be same. But, in the lattice I want them to be same because, I know from theory that both the arm should have the same coefficient that is not guaranteed.

In that separate business of you know I mean LMS updating the either the multiplier of this branch or multiplier of this branch, how to get rid of that problem. Now, question is in the each is a quadratic equation of K, this is also the quadratic function, so summation also is a quadratic function of K that is how only one minima. So, then what will be instead of taking of this or this, we add that two and that we minimize with respect to K, that is what is done.

So, neither this fellow nor this fellow, but they did the addition, so; that means, now I am coming to the actual one, you form sigma p square as sigma p plus 1 K square as f square just addition of that 2, just add that 2, this function is also quadratic in K with the same minima this you minimize. So; that means, your steepest descent will become K i plus 1 as K i minus mu by 2 into evaluated at K equal to K i and what will this give guys to this norm square is if you square it up this and this.

And differentiate this guy, this guy norm square, norm square this square it up, square it up e over this e over this, what we will get actually after differentiation. This is give us to what d d K of will be equal to what, you can mentally I mean you can visualize it know I mean there is no point in doing that, first you take the square of this expected value and differentiate and here also. Here, we will get in both cases minus 2 K will come out, then E, E or maybe you can write it in terms delta p and all that, so delta p minus please correct me if I make any mistake sigma p.

Student: K 1 will be not common ((Refer Time: 47:44)).

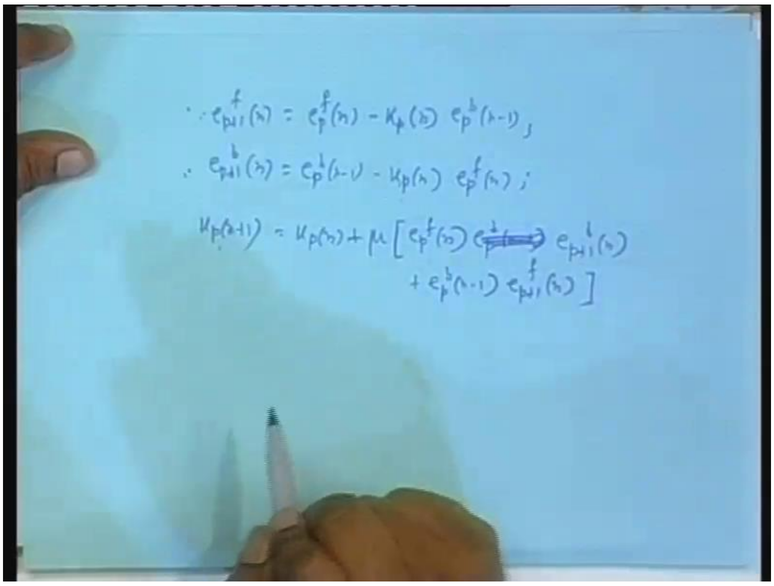
K will go know square it up yes K after differentiation K will go yes minus 2 delta p minus K into sigma p from here sigma p b square and then from the lower half again delta p, please correct me delta p is a right minus K sigma p f square ((Refer Time: 48:25)) am I correct these are thing. So, this you have to bring back here minus 2 and mu by 2 cancels, you get plus mu into this entire quantity and evaluate it to that K equal to K i.

But, I will not stop there I will go to LMS from here I will go to LMS simply same thing actually what I did earlier I am just generalized. So, delta p will be approximated by and while the estimate there is e p f n into e p b n minus 1 itself, no sample average and all that, sigma p b square sigma p f square there will be replaced by what, this will be simply square of the e p b n minus 1 and this will be square of e p f n while we estimates.

If you put back here, what you get and i replace by n, so K n plus 1 plus mu into you take e p f n that thing common, you will see e p b between this two you take e p b n minus common. So, e p f n minus K n times, so this is and other one you take e p f n common is it this between these two, this is e p f square this e p f b, so take e p f common, this e p b n minus 1 minus K n...

So, this guy into this guy and this guy into this guy, this is the basic form of this called gradient adaptive lattice algorithm of course, along with this we have to have the equations for e p f n and e p b n that I have wrote up in equations.

(Refer Slide Time: 50:40)



So, that the p th stage actually will be like this e p plus 1 f n plus simple lattice equation, these are I am assuming to be correct one minus K and that instead of K I am calling it K p I told you I bring the subscript p to indicates its p th stage K p f n into this you are familiar with. Similarly, the other half also we are familiar with again I am assuming the that this is correct e p b n minus 1 correct one again K p n.

And then K p n giving we are giving K p n, we calculated these two we are giving K p n, this we are giving K p n, we first calculate these two then we march towards K p n plus 1. This adaptation, during adaptation also I continued to get these values, like in the case of error elimination algorithm for optimal filter for that transversal filter, during adaptation also with those incorrect weights, which are under training process is to get the filter output. And using that you get the error that error feedback will be the LMS algorithm get better coefficient.

Then, similarly here also is it incorrect K p’s you get them using this you get new coefficient. Then you again use that back to filter in the next time index n plus 1 you understand, so this will be e p f n I am making all kind of mistakes no, no e p f n I told you see diagonally know e p f n e p plus 1 b n, this will converge. But, usually you have to take mu to a very small, there is some expression for mu and all that I mean upper bound that I am not giving here.

So, this is gradient adaptive lattice simple, in the intermediate stage I assume these to be these two to be correct and you are getting these. But, these are not correct because, not optimal K p has come up, but still this data will be used to in the next stage to update this K p’s, but those training will be useless. But, suppose for the p'th stage that the training is over and then the K p’s are identical to K p are optimal one are very close to one for that moment onwards, whatever is generated you can assume that to be correct.

And from that moment you wants the training for next stage will become meaningful they actually for all the stage it will take more time are you understanding. So, that is all for these this kind of business, you know it is called stochastic gradient treatment we will basically we will take the gradient of that error thing and then you know stochastic gradient. Because, it is depending on the data finally, in the LMS case it is not just dependent on the constant condition like R matrix, p vector.

But, when you move to LMS it inverse data and therefore, it becomes a stochastic gradient because, it is just taking stochastic process and evaluating a gradient. So, these is one side one side of the adaptive filter, where mean square error is minimized mean square is minimized and gradient is computed, we will minimize by first competent a gradient, then approximate the gradient by data giving quantities not are my correlation values, but data giving quantities and then analyzing the converges and all.

But, there is another side another approach that is called recursive least square, where is more direct, they are we do not do any you know we do not evaluate the mean square value of some error and all that and minimization and all that. Whether we take, suppose you want to find out to you are looking at the error variance, you just take several samples of the errors square them up and average, that average will be minimized with respect to the combiner coefficients directly.

And that will be done with time recursively, that is suppose you have taken 100 samples of an error, error E 0, E 1 to E square 0, E square 1, E square 2, like that you are added up to 100 and divided by 100, these errors are function of those linear combiner coefficient filter. So, you minimize it you get a set of filter coefficient fine, but suppose now 101 100 first data coming 100 and first data come in, so you have to again compute the same square of error to minimize that we just to filter coefficient, if you minimize you will get a new set of coefficients.

Question is how to generate it recursively, that suppose I got that set of coefficients which are obtained by minimizing sum of square of errors up to 100 terms. If other term coming new set of if a better set of coefficient will come up, how to obtain it recursively from the previous stage. Exactly, there is no question of when you expectation values anywhere not that is, just numerical linear algebra. And it is more tricky and more difficult, but again far more interesting than this, so tomorrow next class we will be giving this recursive least squares.

Thank you very much.