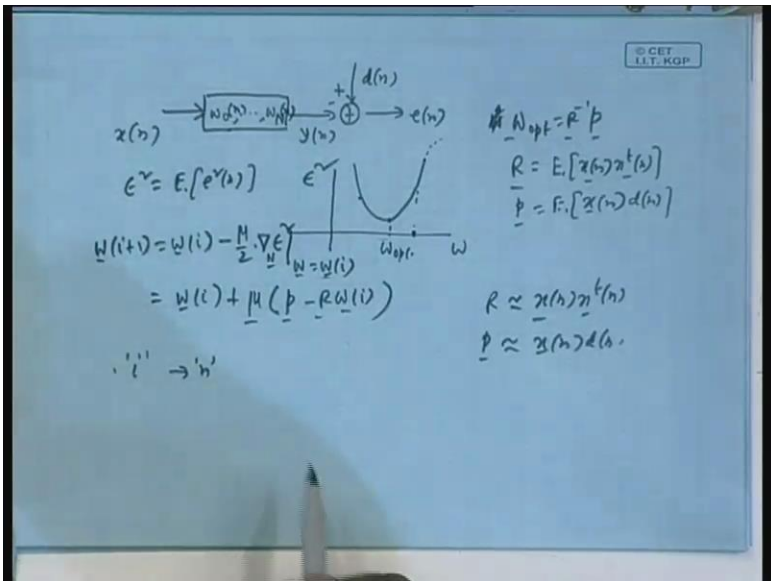
Adaptive Signal Processing
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 7
LMS Algorithm

(Refer Slide Time: 01:02)



Quick recap of what we did last time, our basic filtering that model was this. The input was a zero mean random process x n, w 0 to say w capital N to the filter taps where did you filter it y n filter such that if you take the difference between the filter output and desired response d n the error e n. It is also random process zero mean, because d n also zero mean its mean square value variance should be minimum that we saw to be quadratic function of the weights. Therefore, you can minimize them if I differentiating that mean square value of the error with respect to each weight and equate into zero and you get a solution, which in this case give minima you get the optimal filter.

Optimal filter was w w opt w opt was R inverse p R was e of x n x transpose n all real valued case that is why we transpose only no Hermitian transpose; p was d n cross correlation vector. Then we said that since it is a quadratic function. Suppose, I do not know how to compute the inverse of the matrix, but since it is a quadratic function has got only one minimum if you really plot this quantity epsilon square which is a variance. Remember variance is independent of n that you have already also shown when you took out the expression for e n and replace this by square n and all that; n disappear because of stationarity and joint stationarity between x n and d n.

Because, however the joint stationarity comes you see this p vector p vector is independent of n it is a correlation between x n vector d n, but its independent of n so because that is a joint stationarity. We said that time that else is epsilon if I plot epsilon square in an n plus one dimensional space where one axis epsilon square and other axis each of other axis is one weight or other. Then, it will be such function that is only one minima and if you were else it will going up up up it cannot have local maxima. So, it only can have can go up like in the sample case of single tap filter it can be something like this.

A single tap filter that is you know this could be the optimum it can only go up up it can never be like this it can never come down. If you tries to come down their will be a local maxima formed, which cannot take place because it is a quadratic function. If minima you differentiated with respect to weight you get only a linear equation with one one solution it will go up up. Now, if that be the case then we can follow an iterative procedure; we say that is suppose at i x type of iteration we are having some weight, so that time you find out the gradient here.

If the gradient is positive no point in going to this side; because only going further away going the opposite direction, if the gradient here is negative go in the right side; that means, you go in the opposite sense of the gradient. If gradient is increasing you go in the opposite direction; if gradient is decreasing go in the same direction. So, from that we derived that steepest decent algorithm from w i, we said we will go for w i plus 1 by taking going against the gradient. So, you did these we this is what we did.

This del is a derivative operator it is nothing, but just a vector form where all the partial derivatives of epsilon square with respect to w zero w one dot dot dot are puts one after another in the vector form that derivative does not gradient here just figure was only for one variable, when you have multiple variables multiple taps weight. Then, it becomes a vector of partial derivatives and that you are evaluating at the particular iterate w equal to w i and for the current iterate you are subtracting. So, the gradient is positive you are go in the opposite direction; if the gradient is negative you are going the same direction and mu by two is a proportionally constant mu is called a step size.

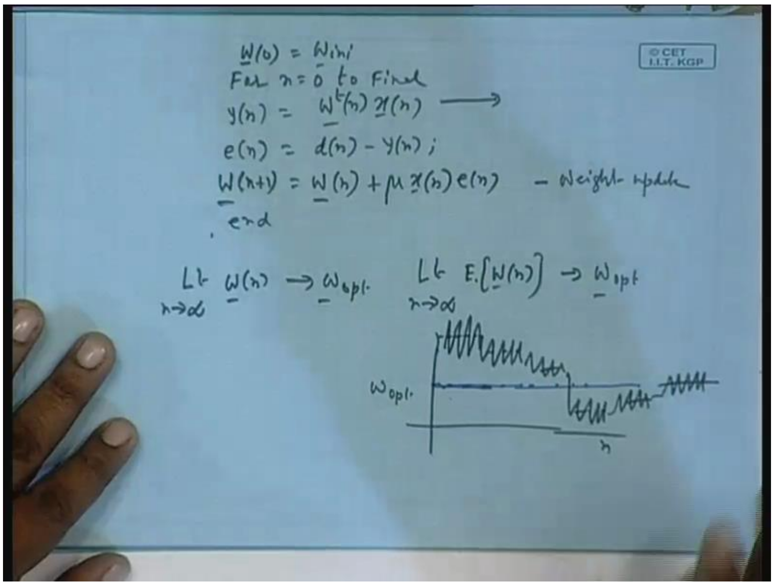
Then, we simplified it we knew the formula of gradient, which we worked out in the last time and we what we got was this; mu into p minus R w i this is what we got, but this is still an offline this was still an offline procedure, because given the p and R value you will just do it iteratively sitting at home. Then, I wanted to say that I want to do it in real time. So, first thing was to replace index i by n. So, at every clock cycle I have one i mean step of iteration. So, zero th clock cycle zero th unit of time means that time I am doing zero th initial state. Then first cycle means first iteration; second cycle second is like that.

So, I am drawing that of this procedure iteration, but in real time. Even then it is not adaptive it is because you are still using the given value p and R, but suppose that is also not given. Then, I said that R and p normally what is R after all if you want to estimate R it will be what is R R is expected value of x n into x transpose n. So, you should take one x n vector multiplied by x transpose m. Take another vector say x n minus one. What is x n vector? Starts at x n then x n minus 1 x n minus 2 up to x n minus capital N; x n minus 1 vector it will started x n minus n minus 2 dot dot dot x n minus capital N minus 1, so so and so forth.

So, we will take x n vector multiplied x transpose x take x n minus one vector multiplied by x transpose n minus one dot dot dot may be you doing hundred times and then add an average divide by hundred that will be a good estimate of R. Similarly, for p but I said that suppose I will use a wiener estimate either do not take so many I take only one. So, R i replace by only this and let us see how it works. Suppose, p i replace by x if I can still prove that a convergence to optimal filter in some sense will work out we will take further amp through.

So, if I replace this from this I will not show the derivation it is very simple x p you replace by x n into d n; R x n into x transpose n x n and x n take out mu into x n. So, d n minus x transpose w n x transpose n w n is same as w transpose n x n; in this case it become actually function of n now which is a filter output and d n minus filter output is error. So, what we got was a LMS algorithm that is where I stopped last time. So, I start from that today.

(Refer Slide Time: 07:34)



So, you are for the n plus one th cycle iteration cycle or time cycle the weight filter weight vector is w n plus one that you get from w n by this. You understood how it came? You replace p by x n d n R by x n x transpose n and take x n common that is why this x n is coming up and then within bracket this d n first minus x n transpose n w, which is same as w transpose x w R n and w transpose x n is same as y n. So, d n minus one y n and which is e n. So, this is an algorithm but that means, what are the steps this is one step called weight update. Filter coefficients also called filter weights weight update, but before that I must have e n.

So, what is e n? E n is d n minus y n and therefore, I must have y n. What is y n? Y n in terms of given w w transpose n x n vector. So, it is like this for n equal to zero to say some final and then you start this iteration with some initial value w int. Normally, we take that initial values of the weight iterate to be zero vector. This is the LMS algorithm there are two main operations. One is a filtering operation; another is weight update operation, which consumes computation. We draw an architecture for this may be sometime late I will do.

In another class, I take the architecture and then pipeline and make faster version you know, but here I am not doing, but I might do it. This is the most celebrated algorithms; I spend some time on it this is the LMS algorithm basic LMS algorithm. It is for most popular algorithm in eighties another class came up totally different in approach if derivation called recursive least squares. We try to which gives more accurate estimate more accurate more I mean faster convergence like here w we will go on iteratively we will show that it will converge, but that algorithm RLS recursive least squares gives faster convergence, but that has some problem in structure and all that.

It is not easy to implement this is by far the most popular even I mean this has took that rest of time. This algorithm you see if I indeed that follow the exact steepest descent may be changing i two n also does not matter whether you are doing offline with iteration x i or online with iteration n as long as you are going this equation, this is an exact steepest descent exact p exact R. Then, obviously for proper choice of mu you understand you will go like this you will hang up on this. That is the filter weight we will directly converge on the w out directly the error between them will be exactly zero; it will exactly become equal to that.

But you have understood that I have not given you exact p exact R. I may have brought in some wiener estimate. So obviously, after that the algorithm that you get you cannot expect that to behave like you know pure steepest decent. So, there is some compromise, but still that will converge this convergence proof I will work out little later, but it will converge in what way that if you take say w n after all is generated in this algorithm from data. So, w n is also is random process; w n is also is random process because every time suppose you done the iteration start from zero to say five hundred.

You get a sequence of w w zero w one I mean vector zero for iteration zero w vector one iteration and like that you get a sequence. Next time, again you run the algorithm you get another sequence. So, you do not get the same thing because data is changing after all this weights are generated by data you put an initial vector zero vector this data comes up that generates the new vector new weights put that back again new data vector. So, basically filter weights are given by w generated by data. So, that is why a random process is. So, in that case if it is a random process in that case what I suggest that not suggest what happens is this.

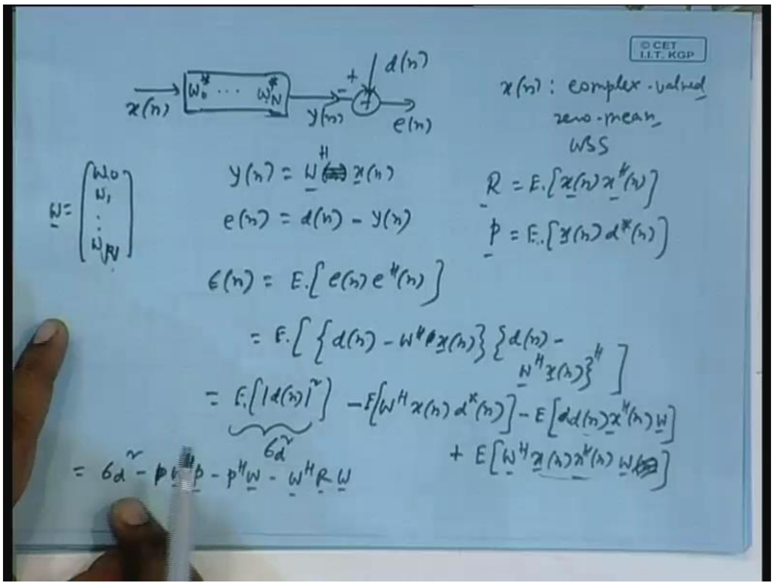
Ideally, I would be very happy if I could show that as n tends to infinity this vector convergences to w opt, but it will not happen because; obviously, you know we are not following the exact steepest decent. What happens if you take the expected value of w n. What is expected value of w n at each index n? You have got a set of filter weights take a particular one, so zero th filter weight just taken that only separately that particular one for the chosen n is a random variable, but a next time you run the algorithm you will get some different value for that n for that filter weight at that index.

So, random every time you do this it will be the fluctuating that particular filter weight at particular index n. So, it has a mean then go to n plus one again at n plus one that filter weight we will have some value for this run of the experiments. Next time, you run again its value will fluctuate. So, it will have a mean. So, the mean will converge, this can be proved and this will become w out. That means this will prove; that means, suppose you consider only one tap filter to start with only one tap filter and this is your w opt only one tap filter this is time index and this is your w opt. So, this is your w opt. So, may be this w n it is a fluctuating thing may be it will be fluctuating like this here.

So, it is mean is here not on this there may be after a while it is it fluctuating may be here it is mean is here so and so may be here; may be after a while it will be here may be after a while it will be here and finally, it will like this. This much we can prove, but that is not enough; obviously, you can ask the question. Even if it convergence in mean if it fluctuate like this then what good it is. The next time after that we have to consider the spread the variance also then we will see how the variance will be kept bounded. That is a very lengthy analysis, but I am I mean after all I have to carry it out very lengthy kind of boring analysis lot of term lot of basic statistical analysis, but you would gross up roughly about how to carry out those analysis.

That is a very I mean that might take one or two basic event that I will do later. There is a mean square error analysis, but this proof is very nice. Now, before I do that I come back to more general case now. So, long I dealt with real valued optimal filter steepest decent LMS. Now, I want to go for more general case of complex case.

(Refer Slide Time: 15:02)



That is x n is complex valued x n complex valued. Other things are same complex valued zero mean WSS those are fine and given R, which is now a complex matrix, so Hermitian always. It is E of x n vector into x Hermitian; x n vector is a same as original one, I am not changing the definition of x n vector only thing is that x n now consist of your complex valued random samples more general case. Also given p x n d Hermitian n, which is same as d star n, because this is a scalar, no question of transposition; so d star n you can put conjugate n this star given to you.

Obviously, the filter weights also the complex valued y n complex valued, but zero mean because this is zero mean d n also complex valued and therefore, e n also complex valued fine. Now, we will introduce an introduce some notational a things you know I mean we will introduce some notation. You should have you put w zero dot dot dot say w n and as your filter weights and their complex valued in this case in general because I am dealing with complex data is a complex valued. But instead of putting w zero to w n in the filter let me take their star values. It does not matter complex; I will what I am trying to say that I will in n find out what is w 0; what is w 1 dot dot dot what is w p w n these vector I will find out.

If I find out w zero you know what is w zero star, you know what w 1 star is; you know what w n star is. So, I will find out w; I will find out the optimal filter version for w, but in the filtering I will use w star to w n star there is no problem no this is notation thing I want to make clear. My purpose is to construct the filter you could otherwise put w zero here w n no star and instead find out w 0 star w 1 star w s star, because that is how the derivation is. The derivation we will take the conjugate of this and find out, but after you find out you can construct the filter.

So, either you remove the star here bring the star here we put the star here and find out w. Once you find out w you know what the filter is this is just a notation so that means, if this is w what is y n here? Can you tell me? What time x n w Hermitian times x n that is the thing I want to do actually if I let us call this w. In fact, that will not bring n here because, so far it is not adaptive only when it becomes adaptive, then I will just get in the index n and all that. So, to start with it is not adaptive just I will clarify to find out what are the optimal filter optimal wiener filter.

So, some w i start with and w H. So, y n is w H x n no problem because if this is my w and said if the filter there will be star. This is just a notation nothing else there will be complex valued coefficient I either take them two with the way there you can take them to the conjugate of some other variables. Find out those variables w 0 w 1 and once you know w 0 you can construct w 0 star w 1 star dot dot. Essentially, whatever you have here you take the conjugate of that form of vector the derivation will give you that vector that is how the derivation is, but there is nothing wrong once I know the vector I have also known a filter weight just you have to conjugate that is all. So, y n is this.

Epsilon n this time is not E square this time it will be mod e n square the complex and mod e n means e H n; mod e n square scalar number. So, e n e star is same as e n and e j n after all and now we have got same quantity with Hermitian. Now, we expand if we expand d and d Hermitian that will give you E mod d n square, which is the variance of d n because d n is a zero mean process. So, mod d n square expected value that is the variance you can call it sigma d square it is of no importance to us here because it does not depend on the filter weights after derivation this will go. The two cross terms earlier was earlier were same; now this time the they will not be same one is the conjugate of the other.

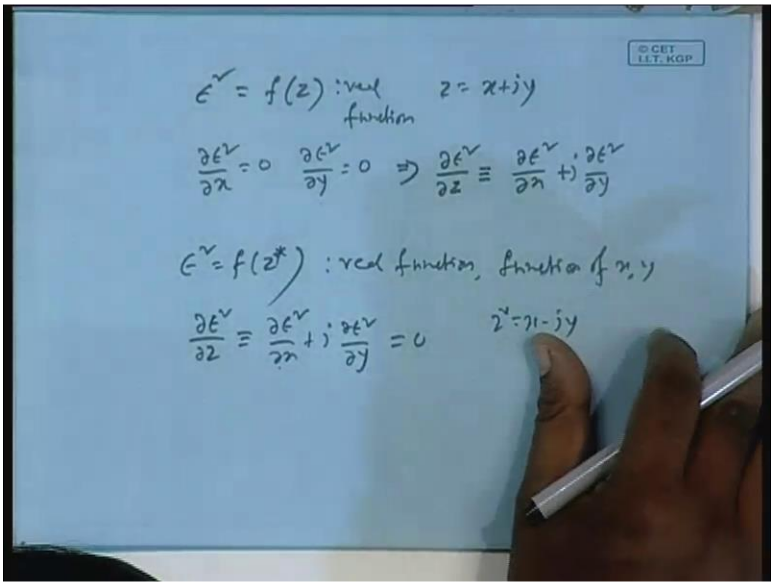
Earlier this conjugate fellow was not there I will repeat it. So, the two cross terms were same now this is because of this I have to write the things directly; w H x n d star n E of that minus E of d n x H n conjugate x H n w plus other term very simple. You see E will work only on this part or here only on this part w is not random. I am just trying to find out a particular set of w which is an optimal, but w is not random here unlike the LMS, w is just a constant and we have find out it is a unknown constant. Here also, we will work on this so; obviously, here we will get E of x n into d star n means the given p vector.

Here, we will get p Hermitian E what this is p Hermitian right and if it is the Hermitian of this Hermitian after expectation or Hermitian before expectation then then apply E you will get the same thing. You make it row and conjugate d n E or you take E and then make it row and conjugate this. So, expected value over this means p Hermitian times w. Here, E will work on this part only which is R, so w Hermitian R w there is no n I am again making a mistake. So, if you do that substitution very quickly sigma d square minus p w H p minus p H w minus w H R w. This epsilon square is a real quantity no doubt about it also it is you know independent of n because of stationarity as we show in see in R and p you know n is there.

It is a real quantity, but it is a real quantity is functions of set of complex numbers w zero to w n or w zero star to w n star, which one you want to see you want to see this way or that way. Each complex number has got both a real part and imaginary part. So, actually there are two n number of two n plus 1 2 n how many here zero to n means n plus one so twice that that many two into n plus one that many variables real imaginary real imaginary. It is a real function of two into n plus one number of variables. Each complex number has got two variables.

And then I have to then to find the minimum I have to differentiate it with respect to both the real part of each complex variable and the complex part. Again real part of another complex variable complex part of that variable so and so. So, it is more complicated than the previous case. Now, before I proceed further you keep this results let us do some again let us come back to some basic matrix facts and all that.

(Refer Slide Time: 24:08)



Suppose, I have a function epsilon square general function. It is suppose a function of say one random variable z and z is say x plus j y, but epsilon square is a real function; real function that is very important very very important. I want to find out del epsilon the minima of this also given it is a quadratic function or some function that is minima; I want to find out the minima minima maxima whatever I want to differentiate it and equate. That means, I have to take del epsilon square del x equal to zero del epsilon square del y equal to zero. So, if I instead I form a definition of you know del epsilon square del z; I do not know I do not know this is correct or not this notation.

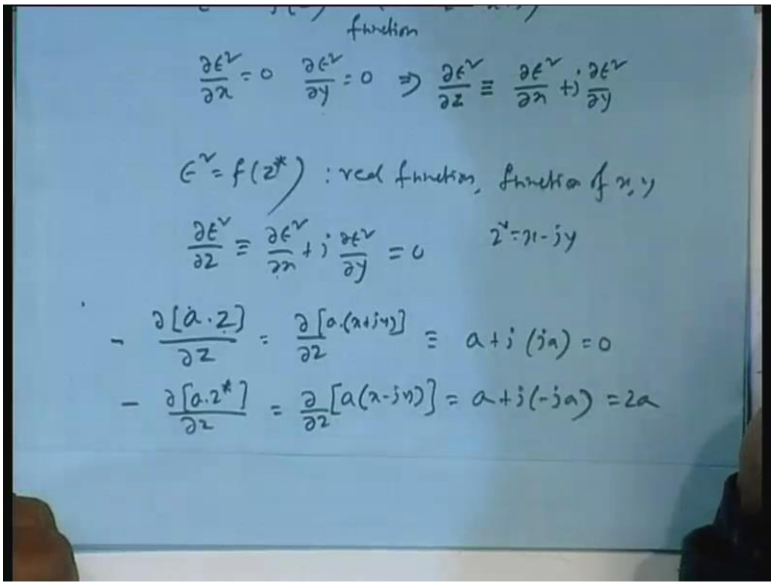
Strictly mathematically, it is correct or not I do not know, but suppose just define it this is equal to del epsilon square del x differentiate with respect to x also differentiate with respect to y and put a j here and then if i equated to zero I mean then this will be zero this will be zero. Zero means complex zero zero plus j into zero. So, instead of having two equations with this definition if I proceed with this if I proceed and equate zero still I will get the same minima. Then, suppose epsilon square it is not a function of z if it is a function of z star still it is a function of real function of x and y.

Suppose, z star comes here in this function, but finally, it is a real function. So, say still again if it is a still real function, but still function of x y only and again I want to find its minima. So, again I have to do this del epsilon square del x equal to zero del epsilon square del y equal to zero. So, even if it is a function of z star if I use the same definition and equated to zero I will still get the minima here because after all I am interested in this being zero this being zero. It does not matter whether a function of z star or z essential thing is it is a function of x and y. I have to minimize I have to form del epsilon square del x equal to zero this also equal to zero.

So, even here if I make this I mean use this definition I equate that two zero I will still get the same minima even though it is not f z, but f of z star. So, even the function which is a function of z star if you want to find the minima; I have to take derivative of that as per this definition with respect to z not this complex z star, but z and equate that two zero nothing will change. I will still get this this equal to zero this equal to zero means the required minimum point because after all it is a function of x and y only a real function. Only thing is we have to find out whose j star you should say x minus j y that much j star at that z star this is minima; when you talk in terms of z or z star.

Then we say so it is a function of z star and at z star equal to this x minus j y this will have minima, but in terms of x and y there is no problem. Even if I differentiate with respect to z as per this definition equated to zero; I will still get the same solution for x y then if you say that at z star at which z star this is minima you see form this x minus j y at this z star this is fine; if that be the case using this definition you see some nice things.

(Refer Slide Time: 28:19)



Suppose, I give you two real numbers scalars, but complex valued a and z a is a complex value; z is a complex number a into z or z into a. I say you differentiate this with respect to z as per this definition. How do you add these? As per this definition you write z equal to x plus j y . So, a x a I do not need to expand because it is a constant a is complex, but I do not need to expand ax you replace z by I am showing those steps do not show those steps; a into x plus j y now apply that definition this quantity if you differentiate with respect to x this quantity with respect to x with respect to y.

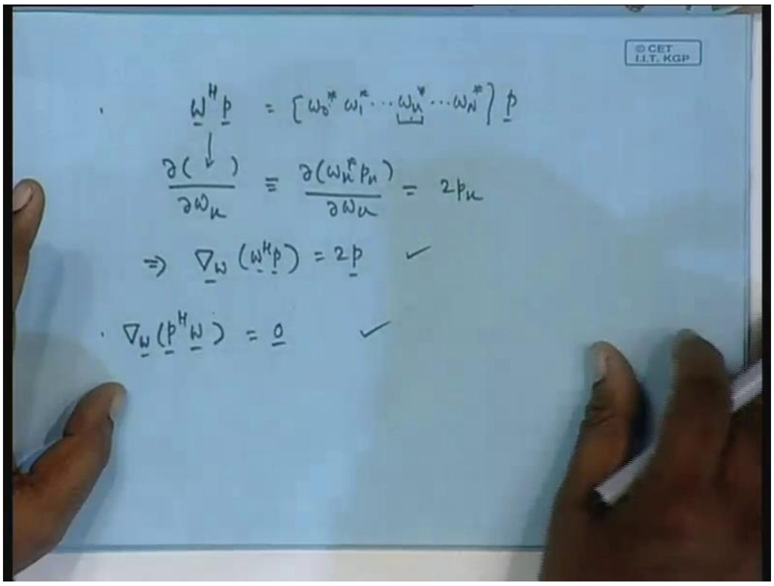
If you differentiate with respect x you get a if you differentiate with respect to y you get j into a and there is another j here j into a. So, you get zero. This you can extend here also what is the problem even if it is not real it is a function. I mean I start to this because that error was real, but we just consider a into z this is a good question a to z. If I extend the definition here it is extended I mean I do not say any problem. This is my definition on complex derivative whether it is a real quantity of complex quantity differentiate like that because of the left hand side I will follow that. So, I have to do the same thing on each of the term on the right hand side.

On the left hand side if I do that I will do the same thing on each term of the right hand side, but because of bringing this j and all that actually this is a smart mathematical trick nothing else the bringing in by j j here you can make this cancel by clubbing the two and bringing j some simplification. It is just a mathematical trick nothing else. On the other hand if it is not z, but say a into z star a into z star then; that means, and you get a and plus j times minus j a so you get 2 a. Again, mind you this is purely my construction I constructed a definition I am not deviating from my original goal of minima because ultimately it was partial derivative.

So, we want to write them separately equate to zero or under the same manner you form you club them like this and equate to zero you will heat up on the same minima because target is get that minima as far as that is constant we are not making any change mind you. I am only doing some jugglery some mathematical manipulations I brought in a definition so that, in some case j j this cancels in some cases. You may get zero and you get twice a that is all. But essentially I will be minima I will be taking the derivative of that with respect to real part imaginary part real part imaginary part of all the weights equal to zero.

I get some solution the same thing, I will get, if I take del epsilon square with respect to first weight as per this definition equate to zero with respect to second weight as per this definition equal to zero. Because del out you can separate these parts real part that is a solution that we will get. But by clubbing this is a mathematical trick I am applying to get some of this nice things, which real thing should be very well. Now, if that be the case now let me extend it little bit these are these are just scalars; now suppose I have got some vector.

(Refer Slide Time: 32:32)

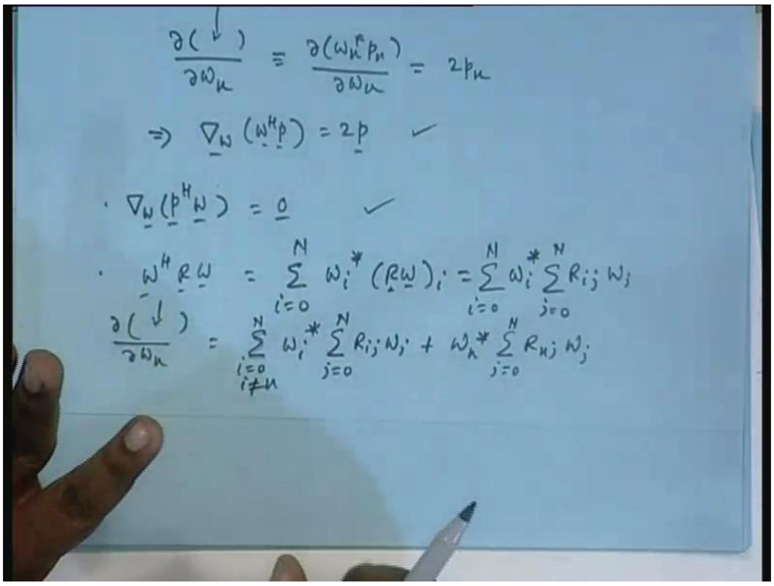


So, some vector w Hermitian some vector p that is w zero star w one star dot dot say w k star dot dot w n star dot p vector; p has p zero p one p two like that. So, this is a function of earlier I took the case of z star and not z here also this thing is a function of the conjugates of the complex number; earlier I had only one scalar complex number I took either that number itself or star z or z star. I have got number here; another complex number here; another complex number here it is a multi-variable case, but suppose I want to differentiate this above with respect to particular one w k.

Firstly, others will go only you have to heat upon this others will go because w k. So, w k star and from here p k will come up k th term; w k star p k this will above to del w k star p k two scalars divided by del w k I mean this not divided by never assume that you know dx dy means dx divided by dy some and this we have seen is twice p k, when you have star there is no problem no zero comes twice p k. Obviously, this means del w that is one if on the other hand you have the other term p Hermitian w do I have to do it or you can see yourself.

If you do this take the general term because p Hermitian means p zero p one dot dot dot all star multiplied by w zero w one take the k th term because you are differentiating with respect to w k only the k th term. So, that mean p k star w k that if you differentiate with respect to w k as per the definition it will get to zero So, this will be a zero vector one result another result.

(Refer Slide Time: 35:10)



The other thing is now we take this quantity w Hermitian R w that again I follow the same procedure as I did earlier. What is this R w is a vector w Hermitian is another row vector row vector column vector. So, zero th into first element first element second element second element like that. Zero th element zero th element all you multiply and add; that means, i equal to say 0 to n w i, but because it is a Hermitian w star w i. So, row vector star, so w i star and what will be the R w thing i th element of that. Absolutely, same steps what will be the R w vector i th element of that and i th element of this guy star and you sum this is a meaning.

Then, what is R w? It is i th element means i th row of R times the vector this is a matrix this is a vector. So, i th row of what times the vector that will give the i th element of this resulting vector. R i j w j and now suppose you want to differentiate this you want to differentiate this this quantity with respect to w k for a particular k. That was before I will take this summation i equal to zero to N out of that i equal to k that case will be separate out, but i not equal to k; you can see this now and then the k i equal to k case; i equal to k means k j w j, I want to differentiate this with respect to w k.

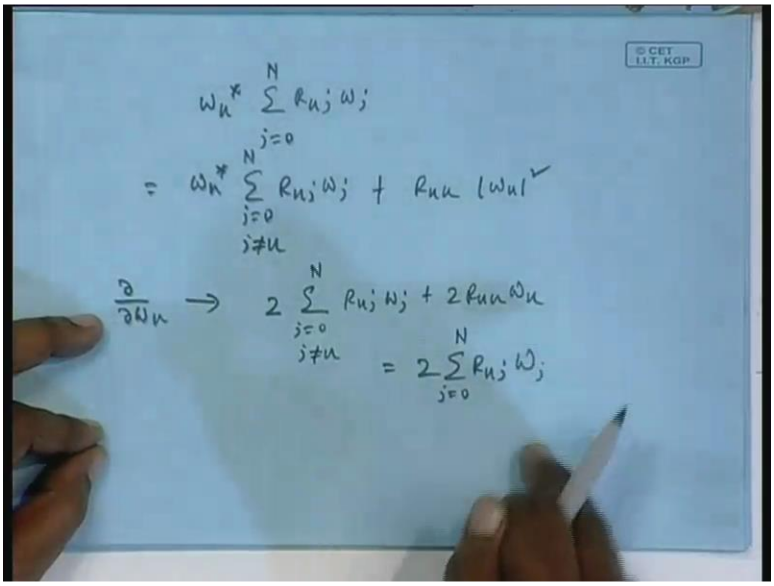
So, here w k does not occur, but for each w i star there is summation in which w k occurs when j equal to k. So, that times I have R i k w k R i k w k R i k could be complex or real I do not care R i k w k and w i star. So, w i star R i k is together is a constant independent k into w k. If you differentiate you get zero because no w k star are following these are the advantages I get by virtue of that trick i applied. It was this was not there in the real case because j factor was not there. So, I could not bringing j and multiply j by j and i have something canceled and make zero. This is the in mathematical you can do it bring in n and make it compact.

Otherwise, I could carry out the derivation you know every real derivative with respect to real part imaginary part separately and a more elaborate exercise. You will get you will get you will get you will get you will get you will get you will get the same the result agreed; you will get the same result will be same; same Wiener expression will come up provided I followed this model the w star here and w and this w I am finding out not w star I am finding out. That chain rule thing I know if you and then I have to prove that chain rule is valid under this definition this mod in I was thinking then I thought that first I have to prove it due to possibility of i could not work out the proof.

So, I thought of going down to the basic, but if you apply the chain rule what is say is you know that if you have two mod product f x into g x derive with x then f x into g prime x plus f prime x into g x. Same thing, you can do here w Hermitian R w; R w is a vector w r means an vector hold down R w and then multiply I mean differentiate w H with respect to w k and whatever and vice versa, but then that theory that derivative of f x dx is a prime x g x plus f x g x prime. So, I have to prove that that is exist that is valid also in this modified definition of since I did not do that I am deliberately keeping avoiding that I am aware of that one book that Furans book does that I remember now, but he does not prove that part.

He assume that that theory extends to this definition also it will, but since I am not prove I do not want to you know I mean do things where there is a mathematical gap that without proving some particular theory and all that. I do not want to use that result. So, that there is no point in I mean there is nothing wrong if I do this elaborate exercise. So, this part goes to zero all of you agree this part will go to 0.

(Refer Slide Time: 40:51)



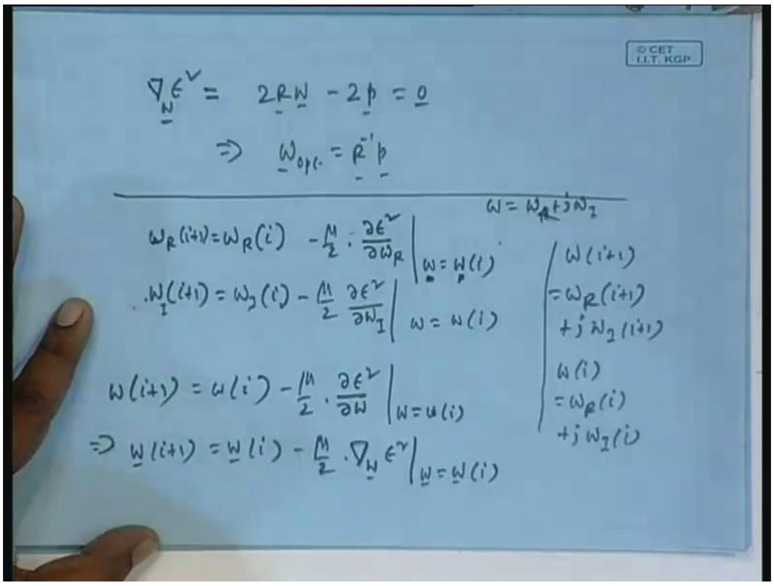
So, this part I work out separately w k star. Once again if you take out the k case separately out of this Hermitian; what we have is j not equal to k w j plus j equal to k k R k k; w j k here w k star here which means mod w k square first differentiate this first differentiate this quantity here. This quantity is a w k; obviously, twice this will come up that is del will then real it was first twice this will come up. See, what I do we will not find any gap mathematically you know it might be looking elaborate R k j w j. Here, mod w k square; mod wk square you see if you differentiate after all, what is this you forget about this once consider z is a function of x and y; z is x plus j y mod z square is x square plus y square.

So, if you differentiate with respect to x you get two x with respect to y two y, so putting that formula two x plus j to y that is two into z. So, same thing two into w it will come up two R k k w k and we combine the two k was missing here k has come back here k case. So, this is twice R k j j from zero to n w j and what is this k th row of R you are scanning; scanning the row k th row j is equal to zero means first columns j equal to one means second column and multiplying also simultaneously by this. So, k th row of R times is vector w that will give you derivative with respect to w k.

So, with respect to other w also other rows time the vector so; that means, if you want to do together in the del form of that factor w transpose w Hermitian R w. What will get is twice R w w this will give rise to w vector this was a k th row. So, k including first row second row third row depending on the twice of k putting all together it becomes twice R w. So, now we put that in that formula derive the full filter this is what we are this where we are. You have to differentiate this with respect to that is del I have apply now del del epsilon I mean I should not have written epsilon n because it is actually you can see now that time I wrote epsilon n.

Because I did not prove stationarity, but you can see it is independent of n. So, now we back to epsilon square. So, we have apply del epsilon square with respect to w. This will go independent of w this is going no this is not going this is giving rise to two p. This is giving rise to 0, this is giving rise to this is plus. This is giving rise to twice R w that we get to 0.

(Refer Slide Time: 44:41)



That means will give rise to will give rise to what twice R w minus twice from here; you get this part twice p and that equal to zero vector; obviously, you get that optimal filter same as before R inverse p. Only thing is do not use this filter, but use the conjugate components of then things will work. So, in the complex case mind you there is a difference actually filtering you are not using this optimal one, but conjugate values, but really everything this you get the same expression. Have taken d n complex, which d this d n; d n is complex in general which d d n is complex everything is complex.

I have taken mod d n square you see I do not know I mean everything is complex here; zero mean stationarity is valid joint stationarity w H all those things are fine. No no no no, you can construct. You see in communication systems, most of the modern modulations you know they use two components i component and q component and together it goes. So, that symbol is a complex symbol when you transmit basically you transmit one waveform, but when you recover the signal out of it there will be two component and they are treated in the complex variable way i and j. So, there if you want to develop in equalizer and you have complex equalizer, which will be like this.

Complex cases are our construction for our convenience that both are observing one here now one is here. So far, I derived I give you the complex more general case complex version of wiener filter. Now, again I want to do steepest decent here I do not want to I do not know how to compute a inverse I have to do steepest decent. Once again, epsilon square is a real function of what of how many variables two into n plus one variables two into n plus one variables; real part imaginary part real part imaginary like that. So, again there it is a quadratic function all those real and imaginary parts. So, it has got unique minima and everywhere will be going up.

So, I can apply the steepest decent. So, I have to take the gradient with respect to real part also imaginary part also for each variable and go in the opposite direction. So, suppose this is epsilon square suppose I have got only one weight to tell you I have got only one weight w, which is say w R plus j w I capital R w R plus j w I epsilon square is a real function of this, but it is a quadratic function. So, at iterate what you have to do at any point of iteration I have to differentiate with respect to because it is a real function of w R also w I also. So, I have to differentiate with respect to w R with respect to w I and going the opposite direction going opposite direction; that means, w R I have to go in the opposite direction. So, mu minus mu by two times this this term will come up.

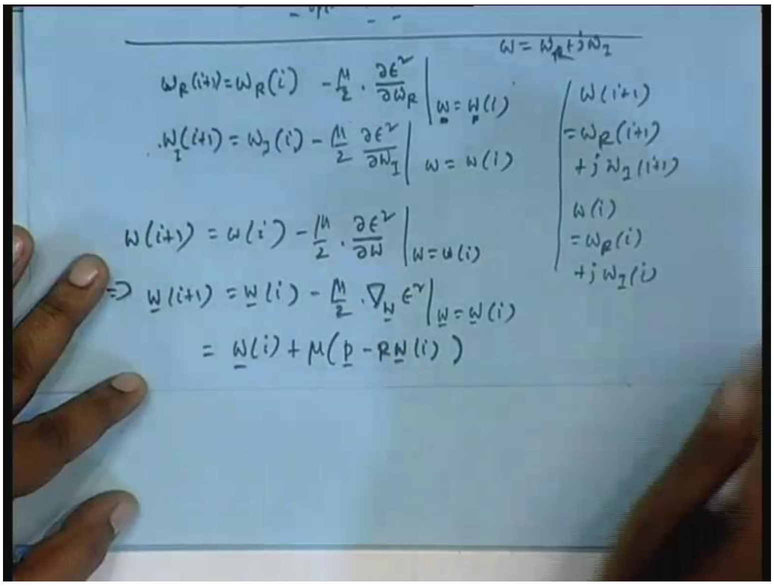
So, w real part if you want to do real part you want to update the real part you should have things like this from i th iterate you go to i plus one th by this and this will be evaluated at w R equal to w R I or you can put w equal to w I does not matter; because when w I obviously, it will be w R I. Similarly, there is no vector here I am talking only single case no point w I , but once again instead of having two equations I have to club them. My i plus 1 th net weight is this plus j times this; net weight net filter weight at the i plus 1 th iterate the net filter weight w i i plus 1. What is that? W R I plus 1 plus j w I i plus 1.

What was w i? i th iterate w i was w R I plus j w I i. I dealt with them separately because after all epsilon square is a real function of both w R w I. So, you going the steepest decent manner consider real part separately i th to i plus one th imaginary parts separately i th to i plus one th and going the opposite direction of respective gradients, but suppose I do not want to spend. So, much of space I have to write the two equations simultaneously. So, I multiply this side by j and i mean right left hand and right hand side both by j here and add from top add with top so what I what I will get on the left hand side total weight.

Here, it is total weight and mu by two is real mu is real; that is important mu is real and constant same important thing mu same for the real and imaginary part. So, that is common and then use by definition del epsilon square del w R plus j times del is epsilon square del w I. So, that is same as del w at w equal to w i. So, this the underlying thing I could we will just written down this expression I follow the same steepest decent procedure and going the opposite direction of gradient, but that is for really mathematically it is summed I have to see what it is.

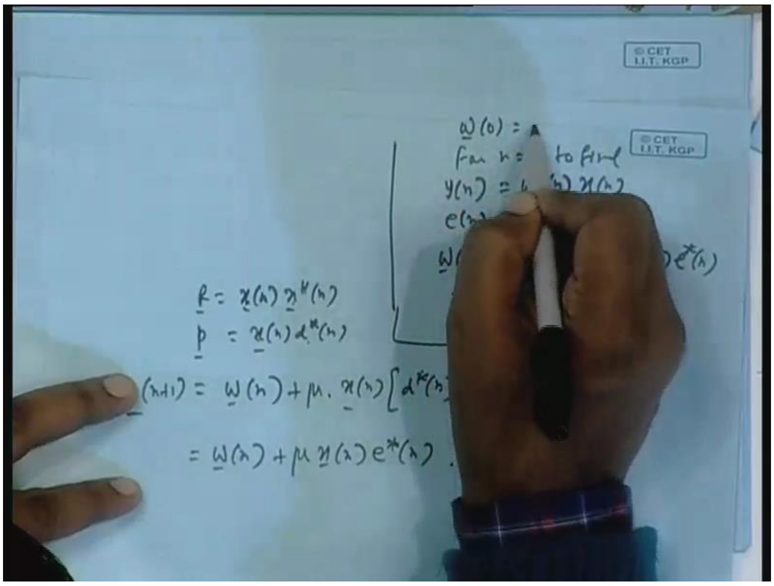
This is not a function of this is actually this derivative is my creation actually it is epsilon square is a real function of, so many variables real part also imaginary part also. I must do the steepest decent on each, then I am saying instead of doing like this. I want to combine them and I get the steepest decent thing. So, you can extend it w i plus 1 is a multi-variable case multi weight case w i minus mu by two del comes now. I get this is generalizes I am generalizing and this gradient also you have found out in the previous minimization procedure. This is the gradient. You put that back two two cancels.

(Refer Slide Time: 52:32)



So, what you get is simply mu into p minus R w I and then to go to LMS from here. What we do? You can first make it online by switching from i to n, but still you are using p and R no real data only thing you are getting the direction and looking at your watch, but I am saying I want to use real data.

(Refer Slide Time: 53:06)



Again, replace R by x n x Hermitian n please see R is not x n x transpose x n x Hermitian n please see some difference now. This is x n d star n right if you put them back here you can quickly w n plus I am I will just finish in one or two minutes plus mu x n d star n; x n d star n here and x x Hermitian. So, x is common bring it out in both x n is in the you know common thing that is a first. So, you take x n the x d star n remains here minus x Hermitian n and w, but w n now I has been replace by n. What is this quantity? We know by our this definition w H x is y. So, x H w is y star Hermitian of y and y is scalar so y star.

So, this quantity is y star n; that means d n minus y n star of that. So, e star n. So, this is a complex LMS algorithm e star n. Other steps are same filtering step that is write out those steps quickly just it will take one minute only. You first find out y n; y n is w H n mind you have to write w H n filtering requires w H into x n e n d n minus y n. Next step is w we have used current weight. Now, go for the next weight from this that is mu times x bar n a x vector e star n for n equal to one to final end. You can take some initial value w zero is w in it. So, that is all. So, that is all for today is a complex LMS Algorithm we will start from here in this class.

Thank you very much.