

Probability & Random Variables
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 39
Linear Mean Square Estimation –Wiener (FIR) Filter

Last class, we have been discussing this autoregressive modeling and then we gave you the equations, whose solution give rise to the AR parameters. Also we derived what is called that linear prediction lattice. Just few things little more things about this AR modeling will be considered today.

(Refer Slide Time: 01:11)

$$e(n) \rightarrow \boxed{A(z)} \rightarrow x(n)$$

$$x(n) + \sum_{i=1}^p a(i)x(n-i) = e(n)$$

$$A(z) = \frac{X(z)}{E(z)} = \frac{1}{1 + a(1)z^{-1} + \dots + a(p)z^{-p}}$$

minimum phase polynomial

First, that is suppose you have got a y -th process un I called it e_n es last time, maybe I will change it to e_n . So, e_n this is x_n . We are considering either AR modeling or linear prediction problem in any case we have got this equation. So, p -th order problem in to e_n . If it is a linear prediction problem e_n is orthogonal to the past p samples, that is x_n minus $1x_n$ minus 2 up to x_n minus p . If it is an AR modeling problem e_n is a white sequence and is orthogonal to all the past of x .

In the case of linear prediction n is not a white sequence n is orthogonal only to x_n minus 1 up to x_n minus p in generally it is not a orthogonal sequence. In any case, n will be a sequence. So, you can model this even in the case of linear prediction you can model this by a filter like this that. There is a linear prediction error sequence called e_n may not be

white and white in the case of AR process it is passing through a model AZ . What is AZ ? That is the transfer function of this AR system gives rise to x_n . What is AZ ? AZ is XZ by EZ if you take Z transfer to left hand side and right hand side; obviously, you get that all full transfer function and that is your AZ .

If you want you can put a constant here also, but I observe that in e_n . So, this is that model we shown methods by which you can estimate a l to a_p that is you solve the Yule-Walker equations. Ideal if you put the actual quotation values than you will get the actual solution, but of course, you may you not get the actual quotation value. So, you have to use the sample average best estimates of this correlation values and you get this.

Once you know them, you consider a e to the power j ω that is replaced z by e to the power j ω you get the transfer function this is the dtft. Mod square of that times some constant will be the output power spectral density. That constant is the input variance that also comes out of solution for solution of the Yule-Walker equations, that you have seen now, point is you are getting some solution from the Yule-walker equation.

So, I am I am now considering the ideal case where you have really put the ideal actual correlation values in that matrix. Question is the solution that you get from a l to a_p does it give rise to a stable and causal filter for az . That is, does it give rise to poles a system that has poles that is roots of this denominator polynomial lining within unit circle? Answer is yes, Yule-Walker equation guarantee is that and that is what we will prove now. And again you now there are various ways of proving it, I will follow a simple way which is easily appreciable.

I read it in a paper by P P Vaidhyathan and in fact, he obtained this profile just teaching in a class and I like that prove very much I will follow that. Though you saw that all roots of this denominator polynomial they lye within unit circle any polynomial whose roots lye within unit circle is called minimum phase polynomial.

(Refer Slide Time: 05:22)

$$A(z) = (1 - az^{-1}) C(z)$$

$$C(z) = c_0 + c_1 z^{-1} + \dots + c_{p-1} z^{-(p-1)}$$

Block diagram: $x(n) \rightarrow [C(z)] \rightarrow y(n) \rightarrow [1 - az^{-1}] \rightarrow x(n)$

Note: $= y(n) - q y(n-1)$

For that what I do giving Az I factorize it Az . What is the order of Az p ? So, they will be p first order factors the poles can be the roots can be complex I agree, but there are all first order factors. Now, out of that take out say 1 particular factor that has a root at say q . So, our style is to write like this qz inverse. So, at z equal to q this is zero; obviously, 1 minus and remaining is cz . Cz , what is cz ? Cz will be of degree what p minus 1 this is very important. So, it could be something like this c_0 plus $c_1 z$ inverse plus dot dot dot plus $c_{p-1} z^{p-1}$.

So; that means, the model can be written like this cz followed by 1 minus qz inverse here is your n you can call this intermediate signal as y_n output is x_n . Obviously, x_n is what if you see this differential equation this transfer function 1 minus qz inverse. Means in time domain 1 minus that is the, what does not mean y_n minus $q y_{n-1}$. So, x_n is equal to y_n minus $q z$ inverse amongst to a daily operator daily operation by 1 cycle; so q in to y_{n-1} . Now we now or may be. I just missed out something actually this is this is not az just was this this A prime z .

(Refer Slide Time: 07:23)

$$A(z) = \frac{1}{A'(z)} = \frac{X(z)}{F(z)}$$

$$A'(z) = (1 - qz^{-1})C(z)$$

Block diagram: $x(n) \rightarrow [C(z)] \rightarrow y(n) \rightarrow [1 - qz^{-1}] \rightarrow e(n)$

$e(n) = y(n) - qy(n-1)$

$$y(n) = c_0 x(n) + c_1 x(n-1) + \dots + c_p x(n-p+1)$$

$$y(n-1) = c_0 x(n-1) + c_1 x(n-2) + \dots + c_p x(n-p)$$

A prime z was the denominator polynomial this is your xz by Ez our purpose is to show that A prime z has all its roots within unit circle. So, now, look at this equation there is slight change from what I draw just a while back; that means, xz in to A prime z is equal to Ez . So; that means, if you have say an A prime z we write like this we take out 1 factor out in to cz cz we have will already given the expression. So, no point in writing, so A prime z in to xz that is Ez .

So, if you have system that is whose input is x_n and then followed by cz followed by 1 minus qz inverse and that is your e_n . Obviously, xz in to A prime z together is 1 A prime z . So, xz in to A prime z is Ez . So, I am just breaking A prime z as a factor cascade of 2 systems 1 is cz another is 1 minus qz inverse. So, output will be E_n intermediate variable is y_n . So, by the same logic E_n is y_n minus q y_n minus 1. Now, we now E_n is orthogonal to x_n minus 1 to x_n minus p now what is y_n ?

Y_n is $c_0 x_n$ plus c_1 dot dot dot total term is p I mean degree of cz is p minus 1 not p this is very important degree of cz is p minus 1 and not p . So, it will go up to p minus 1 x_n minus p plus 1. So, what is y_n minus 1 instead of n put n minus 1? So, you see y_n minus 1 consists of what the past p samples of x_n that is x_n minus 1 up to x_n minus p . Therefore, E_n is orthogonal to y_n minus 1, because n orthogonal to x_n minus 1 up to x_n minus p . So, E_n is orthogonal to y_n minus 1 that is correlation between E_n and y_n minus 1 is 0.

(Refer Slide Time: 10:08)

$$E[e(n) y^*(n-1)] = 0$$

$$y(n) - a y(n-1)$$

$$\Rightarrow a y(n) = a^2 y(n-1)$$

Block diagram: $x(n) \rightarrow [c(n)] \rightarrow y(n) \rightarrow [1-a] \rightarrow e(n)$

$$y(n) = c_0 x(n) + c_1 x(n-1) + \dots + c_p x(n-p)$$

$$y(n-1) = c_0 x(n-1) + c_1 x(n-2) + \dots + c_p x(n-1-p)$$

If that be the case that mean the in that case what I can write. What I can do is I know that $E[e(n) y^*(n-1)] = 0$. See it is important I told you that $y(n)$ consist of I mean c_z is a polynomial of degree p minus 1 that is why when you consider $y(n)$ minus 1. You start that index n minus 1 for x , but go only up to n minus p not any term beyond that, because that extra term. If you if you go 1 step further in the past say $x(n-p-1)$, that may not be orthogonal with n we have got strictly terms from $x(n-1)$ to $x(n-p)$ and that set is orthogonal to $e(n)$.

Therefore, $e(n)$ is orthogonal to n minus 1 and that is 0 if you now replace $e(n)$ by this expression $y(n) - a y(n-1)$. What do you get is that, is replace $e(n)$ by replace this by what I wrote earlier $y(n) - a y(n-1)$. So, if you substitute what you get $e(n)$ now I am since there are 2 few several processes involved, let me bring in the subscript. So, $e(n) - a e(n-1)$ that is equal to 0. So, $e(n) - a e(n-1) = 0$ mod $y(n)$ square or mod $e(n)$ mod $y(n)$ minus 1 square. So, this is of course we are considering wss process that is very much there.

(Refer Slide Time: 12:01)

$$E[e(n) y^*(n-1)] = 0$$

$y(n) = q y(n-1)$

$$\Rightarrow E_y(1) = q E_y(0)$$

$x(n) \rightarrow [c(n)] \rightarrow y(n) \rightarrow [1 - qz^{-1}] \rightarrow e(n)$
 $e(n) = y(n) - q y(n-1)$

$$E[|e(n)|^2] > 0$$

$$\Rightarrow E[e(n) [y(n) - q y(n-1)]^*] > 0$$

$$\Rightarrow E[c(n) y^*(n)] > 0$$

So, this is 1 relation find next is we also know that since we are not we are now considering a non deterministic process. You now x_n cannot be given by a linear combination of its past. And therefore, inversely e_n cannot be given by a linear combinations of the past values of x_n , so that means, this is strictly greater than 0. E_n cannot be 0 variable random variable, but e_n is 0 random variable; that means, x_n is exactly given as a linear combination of past p sample. That is not a deterministic that is not a non-deterministic process, then but we are considering purely non-deterministic processes.

So that means, n cannot be a 0 process n cannot be a 0 random variable which means variance of n variance is strictly greater than 0, but this means E of what. E_n in to $e^* n$ and what is $e^* n$. $E^* n$ is you replace again y_n minus $q^* n$ this is greater than 0 once again the second term second term is e_n in to y_n minus 1 star that is 0, because e_n is orthogonal to y_n minus 1 that we have seen earlier. So, forget that term. So, what you get is this leads to E of $e_n y^* n$ greater than 0.

(Refer Slide Time: 13:54)

$$E[e(n) y^*(n-1)] = 0$$

$$y(n) - \alpha y(n-1)$$

$$\Rightarrow \alpha y(n) = \alpha y(n-1)$$

Block diagram: $x(n) \rightarrow [c(n)] \rightarrow y(n) \rightarrow [1-\alpha] \rightarrow e(n)$

$$e(n) = y(n) - \alpha y(n-1)$$

$$\Rightarrow E[c(n) [y(n) - \alpha y(n-1)]^*] > 0$$

$$\Rightarrow E[c(n) y^*(n)] > 0$$

This means now you if you replace n again by yn minus qy n minus 1 and carry out the sum what you get. Ryy that is en again is replaced by this expression y of n minus y of q in to yn minus 1 q in to yn minus 1. If you do that what you get here ryy 0 minus q. Just a minute it will be better if we consider in a different way select the different way than the proof will be easier.

(Refer Slide Time: 15:35min)

$$E[e(n) y^*(n-1)] = 0$$

$$y(n) - \alpha y(n-1)$$

$$\Rightarrow \alpha y(n) = \alpha y(n-1)$$

$$\alpha y(n) - \alpha y(n-1) > 0$$

Block diagram: $x(n) \rightarrow [c(n)] \rightarrow y(n) \rightarrow [1-\alpha] \rightarrow e(n)$

$$e(n) = y(n) - \alpha y(n-1)$$

$$E[|e(n)|^2] > 0$$

$$\Rightarrow E[e^*(n) e(n)] > 0$$

$$\Rightarrow E[y^*(n) c(n)] > 0$$

That is we know the that E of mod en square greater than 0 this leads to the fact E of e star n en greater than 0, may be here you replaced here you replace by en by yn minus q this makes its simpler yn minus 1. So, y star n another term is the second term is yn

minus 1 star en that is related to the correlation between en and yn minus 1 that is the conjugate of that. So, that is 0. E of en y star n minus 1 equal to 0 means E of e star n yn minus 1 that is also 0 conjugate just conjugate.

So, the second term amongst to that and first term is y star n en and then you replace en by this expression yn minus q y minus 1. Then what you get? If you replace actually I need some space q yn minus 1 en that term becomes 0 and you are left with e of y star n en greater than 0. What is y star n, what is en? En again you replace by this. So, what you will get is ryy 0 minus q star 1 greater than 0 easy.

En is replaced by this yn minus qy n minus 1 first term is yn in to y star n this is ryy 0 it will give rise to second term q comes out and y star n yn minus 1. That is a ryy star n you can you easily see now and from the first equation you have you can put ryy 1 equal to q ryy zero. So, this means actually if you replace ryy 1 equal to q ryy 0. What is ryy star 1? That is q star ryy 0 ryy 0 is really is variance. So, it cannot be star. So, only it becomes q star in to ryy 0.

So, if you bring that here what you get ryy 0 in to within bracket 1 minus mod q square this is greater than 0. Now, ryy 0 as I told you this is this this quantity no, this is not. Ryy 0 cannot be 0, because what is yn cannot be a 0 random variable because; that means, what is yn after all yn is a linear combination of the past values of x. I mean xn and. So, its p minus 1 past values if that is equal 0. If that is equal to 0; that means, current sample can be again expressed as the linear combination of past p minus 1 value which is not the case for non-determinative process. So, this is greater than 0.

(Refer Slide Time: 19:45)

$$E[e(n)y^*(n-1)] = 0$$

$$\Rightarrow \lambda_{yy}(1) = a \lambda_{yy}(0)$$

$$\lambda_{yy}(0) - \epsilon \lambda_{yy}^*(1) > 0$$

$$\Rightarrow \lambda_{yy}(0) [1 - |a|] > 0$$

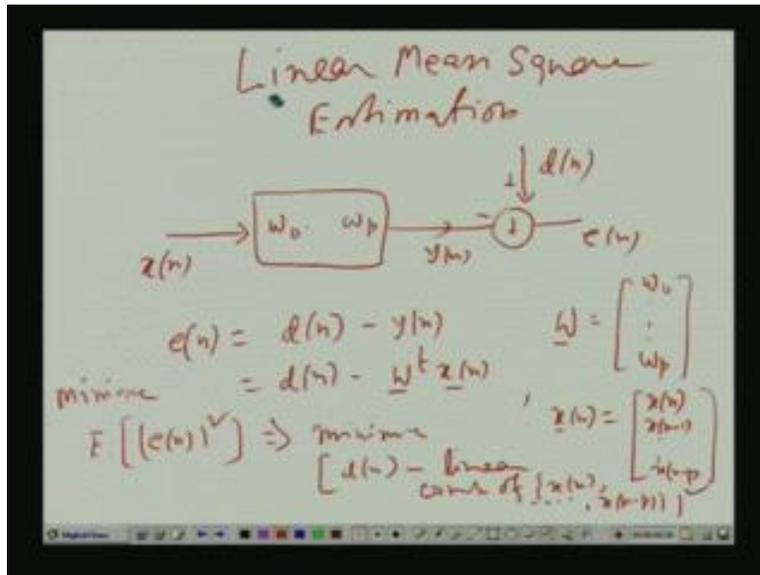
$$|a| < 1$$

$$e(n) = y(n) - ay(n-1)$$

That means what you have here what you have here is 1 minus mod q square that is mod q should be greater than 1, mod q should less than 1 which means the root is within unit circle. So, this I separate out this I did only for 1 root I took out 1 1 root. So, take out took out that factor 1 minus qz inverse separately and rest of them I put under another polynomial cz. Now, I can go on doing it to each and every root this only proves that every root of that polynomial is within unit circle.

So, provided you provide accurate estimates of the autocorrelation values in that Yule-Walker equation you get this. So, if guarantees the stability and causality of the AR model. So, that is all for here modeling. I now consider another application of this random process theory a very that is called Wiener filtering or linear mean square estimation.

(Refer Slide Time: 20:39)



Here, for simplicity I will consider only the real case real value signal and then I will give you the result for complex valued case, but I will not derive them. Here the again I will be considering only FIR filtering not IIR, because that is more complicated. Here you have got a random process x_n you want to designing a FIR filter so called optimal filter. May be you is you can denote the coefficients at w_0 to w_p . So, that the output y_n it becomes a valued estimate of some unknown signal d_n . So, target it is called linear response or target signal d_n .

That means, if you take the error what we do this error has you can see it is d_n minus y_n and y_n depends not only x_n depends on the filter coefficient we choose. So, if you take mod of e square n that is a variance of e_n and all process are assumed to stationary here than mod square of e_n will be a quadratic function of w_z to w_p . So, you minimize them, quadratic functions have got 1 unique minima and maxima in this case; obviously, it will minima, because you can only reach minima you cannot. I mean, there is no maxima here, because you can always deviate from your optimal coefficients and get more and more error variance here.

So, find out for which coefficients the mean square value of this output error e_n is minimized that will give rise to the optimal filter that is unique. And that is also called FIR Wiener filter in the name of the great mathematician and statistician Robert Weiner.

So, what we do we can first and I am assuming for the time being that all the processes and all the coefficients everything is real valued.

This can be generalized further to a case of complex valued case, but that I will do only later I might not derive re-derive it for the complex valued case I will just quote some reference, but I will give the results. So, here what I mean is this what is d_n minus y_n and what is y_n I can write as w vector transpose x_n vector where w vector is this coefficients x_n vector is called regressor vector. You can easily see w transpose x_n is what $w_0 x_n$ plus $w_1 x_n$ minus 1 dot dot dot w_p in to x_n minus p , which is the convolution actually between x_n on this w sequence and that gives you y_n .

So, what you have do, if you take mod n square take it is expected value expected value mod n square. Now, you now this will be this is clearly this will be a quadratic function of w 's, because after all d_n minus w transpose x_n it contains all this $w_0 x_n$ $w_1 x_n$ minus dot dot dot. If you conjugate the entire thing there is no question conjugation of course, because we are doing real valued case. If you just square it up you will have terms like w_0 square w_1 square w_2 square and $w_0 w_1$ $w_2 w_3$ like that.

So, it is quadratic expression in terms of w_0 w_1 w_2 w_3 . Any quadratic expression like this has got unique minima or maxima, in your case should be maxima minimum you can also verify by second derive derivative nevertheless. So, we have to find out the derivative is equated to 0 and that will best to the base filter. We will follow that approach a little later, but you can derive the same filter purely from orthogonal point of view.

That is if you orthogonally project for if you want minimize this; that means, minimize this means what minimize the error which error d_n minus the linear combination of of what x_n dot dot dot up x_n minus p right square of that. That is we have to find out we will consider the space subspace spanned by x_n x_{n-1} up to x_{n-p} . We have to find out an optimum linear combination. So, that that is we want to find out only a particular vector in that vector space. So, that the error between d_n that is an external value d_n and that vector the the error has minimum norm square.

Norm square is same as variance here, because inner product with itself which is this variance. That means, we have to basically do orthogonal projection of d_n on the space spanned by x_n to x_{n-p} . Thus we can show that this projection is unique it exists and

that project is such that d_n minus that projection, that vector that error vector will have the minimum norm that is minimum variance. So, this minimization of e of mod n square amongst to orthogonal projection of d_n on the space n by x_n to x_n minus p .

So, you can just find out that projection and get this expression for that Wiener filter that is the standard way it is a state forward way. We also do this direct minimization, because that will take us to what I will want to conclude this course with that is the case of adaptive filters. So, this linear combination is some coefficients say w_0 times x_n , then w_1 times x_{n-1} dot dot dot up to w_p times x_{n-p} there summation there is a coefficient.

(Refer Slide Time: 27:00)

$$E \left[\left[d(n) - \sum_{u=0}^p w_u x(n-u) \right] x(n-i) \right] = 0 \quad i=0,1,\dots,p$$

$$E \left[\underbrace{\left[d(n) - \sum_{u=0}^p w_u x(n-u) \right]}_{e(n)} \cdot \underbrace{\begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-p) \end{bmatrix}}_{x(n)} \right] = 0$$

$$E \left[x(n) \cdot \left[d(n) - \sum_{u=0}^p w_u x(n-u) \right] \right] = 0$$

$$E [x(n) \cdot d(n)] - E [x(n) x(n)] \cdot w = 0$$

This error must be orthogonal to that is d_n minus this must be orthogonal to whom x star may be n minus i it should be equal to 0 for i equal 0 1 dot dot dot up to p . You can do it in a systemic way either you can do it elaborate or in a very smart way. What you do we you expand this equation you will get E of d_n in to x star n minus 1. That is a cross correlation between d_n and x_n with lag i and you can take this product between the summation and x star minus y . There will be you will have correlation values of x in that equation.

So, you first you put i equal to 0 then i equal to 1 you get separate equations, but you can put the equations in a matrix vector form easily. What I do is if this be the case than I generalize this as this part remain as it is d_n minus w transpose x_n . X is a column vector

w transpose is a row vector this time if I put this way this is scalar d_n minus w transpose x_n is a just a scalar. If I put this way if you take it is product, if you take if you multiply it by the x_n star that also should be 0. If you multiple by x_n minus 1 star that also should be 0, because of the orthogonality.

If go finally, up to x_n minus p and star and there t product again that should be equal to zero, but what this vector this is x_n vector again. Again by every time putting star, but; that means, there is no scope of star here. So, you can as well erase the star, because you know I mean I have assumed every process to be real here. So, forget about this. Those things come in the complex case they are and you have to just generalization the derivations which I will not do, but I will give the results just to make life simple everywhere this is treated to be real.

In this case, I will just do little manipulation this a scalar number does not matter whether I put on the left hand side or right hand side. Scalar times a vector or vector times the scalar I know it is same, because scalar times a vector mean the scalar number multiplies each entry of the vector. So, that you put it on the left hand side or right hand side it does not matter I put on the right hand side and also w transpose x_n is same as this is same as x transfers nw . A transpose b and b transpose a when you had b vectors they are same.

So, I just substitute w transpose x by x transpose w x transpose nw and take this inter scalar factor to the right hand side. So; that means, what I get is x_n vector multiplied by this equal to a 0 vector all vector which all of length p plus 1 where all the elements are 0. Now, if you carry out the product what you get first term is x_n vector times d_n . So, d_n what does d_n do d_n is scalar desired response d_n in to x_n d_n in to x_n minus 1 d_n in to minus 2 dot dot dot d_n in to x_n minus p .

So, what I get we get a cross correlation between the process d_n and x_n cross correlation is lag 0 lag 1 lag 2 lag 3, and because of stationarity that becomes independent n . Because x_n and d_n are assumed to be say jointly wss. So, what you get is cross correlation vector p I will define p separately minus x_n x transpose n E of that. Or may be here only I define this is this is what comes from the orthogonality. What is x_n d_n ? Just for your sake I am elaborating here, otherwise I could have directly written just to avoid confusion in the first instance of the treatment.

(Refer Slide Time: 32:24)

$$E \left[\begin{pmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-p) \end{pmatrix} \begin{bmatrix} x(n) & x(n-1) & \dots & x(n-p) \end{bmatrix} \right] = \underline{R}$$

$$= \begin{bmatrix} r_0 & r_1 & \dots & r_p \\ r_1 & r_0 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ r_p & \dots & \dots & r_0 \end{bmatrix}$$

$$E \left[x(n) \cdot [d(n) - x^T(n)w] \right] = \underline{0}$$

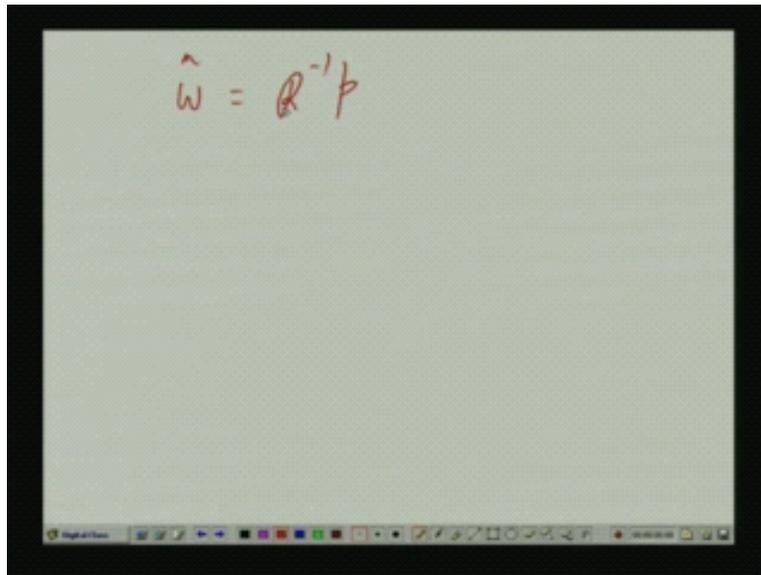
$$E \left[\underbrace{x(n)d(n)}_P \right] - E \left[\underbrace{x(n)x^T(n)}_R \right] \cdot w = \underline{0}$$

So, $d(n)$ is to $x(n)$ cross correlation with lag 0 $d(n)$ is to $x(n-1)$ cross correlation with lag 1 finally, up to lag p . This I denote as some vector and usually the notation in the book there is used actually notation uses p , but I have already used up p here as the order. So, what I do may be I put P this is a scalar vector mind you capital P cross correlation vector. What is $x(n)$ in to $x^T(n)$? That will give rise to actually correlation matrix $X^T X$ $x(n)$ is $x(n)$ it becomes row vector.

If you do that $x(n)$ in to $x(n)$ expected value of that r_0 than $x(n)$ in to $x(n-1)$ expected value r_1 dot dot dot up to r_p again $x(n-1)$ to $x(n)$ expected value of that lag is 1, so r_1 . So, what you get is something like this r_0 r_1 dot dot dot r_p stop this matrix and symmetric now. So, it will be this and this is called R input correlation matrix. So, this p vector this is R matrix suppose these are known to us this statistical things, and because of stationarity n disappears.

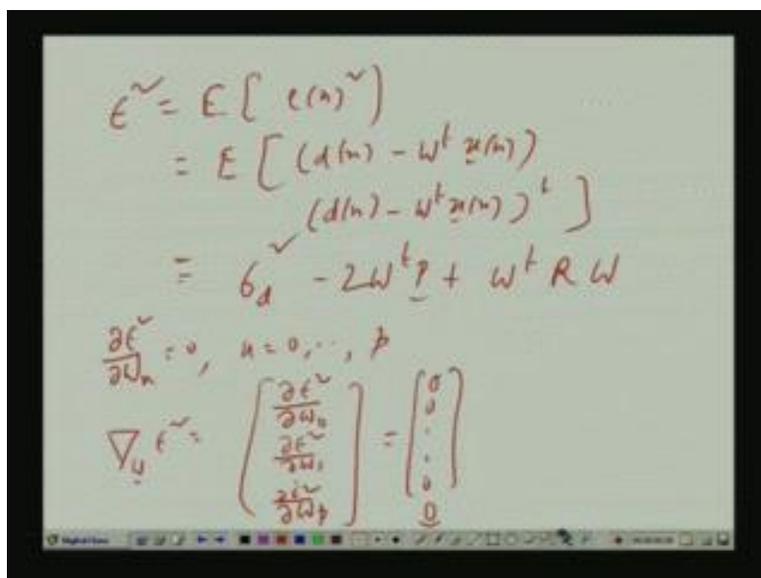
Suppose, p vector and R information about them is available to us than you have a simple equation $Rw = P$. And you get the filter as $w = R^{-1}P$ that is the Wiener filter expression.

(Refer Slide Time: 34:54)


$$\hat{w} = R^{-1}p$$

I mean put a cap here to indicate optimal filter Wiener filter that is r inverse p this comes from the orthogonality. So, if you r is know p is known that we put that, but there is another way of deriving, that is by computing by the differentiating that error variants directly with respect to each (w_i) and equate them to 0. We do that, because of some other reason, because that give rise to you now that you can be used further to develop what is call adaptive filter.

(Refer Slide Time: 35:28)


$$\begin{aligned} \epsilon^2 &= E[(e(n))^2] \\ &= E[(d(n) - w^T x(n))^2] \\ &= E[d(n)^2 - 2w^T p + w^T R w] \\ \frac{\partial \epsilon^2}{\partial w_n} &= 0, \quad n = 0, \dots, p \\ \nabla_w \epsilon^2 &= \begin{bmatrix} \frac{\partial \epsilon^2}{\partial w_0} \\ \frac{\partial \epsilon^2}{\partial w_1} \\ \frac{\partial \epsilon^2}{\partial w_p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$

That is you can write down epsilon square as that was. What was en en was $d(n)$ minus w transpose $x(n)$ and no point in putting mod in this case. So, I am still carrying out the

notions of complex case, but there not equate E square n and what is e_n . D_n minus w transpose x_n this square of that, but instead of square I can always write this in to this, that is a scalar number say a square a square can written as a in to a. Further in any scalar number a and it is transpose they are same a and its transposes are same.

So, I write this as this d_n minus w transpose x_n in to its transpose, because both are scalar. So, transpose of a scalar with itself. Now, we expand e of $d_n d_n$ that is d square n that will give rise to the variance of the process d_n this is stationary process. So, σ_d square independent of n than 1 \cos term is w transpose e of x_n and $d_n e$ of $x_n d_n$, we now this is equal to what p vector. Actually, I choose the order of the filter to be small p I should have used some other letter may be l or k or whatever, because this cross correlation vector p being a vector should have been denoted as lower case p .

But since, I have already used the p the letter p then letter p lower case p somewhere I am carrying on with capital P , but actually it is a vector. So, w transpose p another \cos term is w transpose $x_n d_n$ again w transpose p , because x_n in to d_n its a expected value is p . So, twice this and the last 1 is w transpose x_n and transpose of w transpose x_n . Now, transpose of w transpose x_n is what x transpose n first than w . So, w transpose $x_n x_n$ transpose w . W is not random.

So, e will work only on x_n in to x_n transpose which is r matrix correlation matrix. So, what you get is w transpose Rw this is the expression. You can see this w transpose p this is linear it has only first order terms of the weights. If you put w as a vector row it becomes a row vector here. You will get only first order terms of each, but w transpose or w will consists of the square terms second order terms this I have to derive. That is I want to break it this way for k equal to 0 up to p I define a notation δ δ with respect to w of epsilon square. It is nothing I just put the partial derivatives in a vector form this must be equal to zeros is called 0 vector.

(Refer Slide Time: 39:01)

The whiteboard shows the following handwritten equations:

$$\begin{aligned}
 \tilde{e} &= E[(n)^2] \\
 &= E[(d(n) - w^T z(n))^2] \\
 &= E[(d(n) - w^T z(n))^2] \\
 &= \sigma_d^2 - 2w^T p + w^T R w
 \end{aligned}$$

Below this, the gradient is set to zero:

$$\nabla_w \tilde{e} = 0 \Rightarrow \frac{\partial}{\partial w_k} [w_0 p_0 + w_1 p_1 + \dots + w_p p_p] = 0$$

The vector p is defined as:

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_p \end{pmatrix} = p$$

The result of the differentiation is shown as:

$$\Rightarrow p_k$$

So, erase this I have given you the meaning of del w should be 0 vector. Now, w transpose p now. Firstly, del w sigma d square it is independent of the filter weight. So, that will give rise to 0 this will give rise to 0 0 vector. If we differentiate this quantity sigma d square with respect each tap weight each derivative is 0 than w transpose p what is w transpose p. This is w 0 say p 1 p 0 zeroth entry w 1 p 1 dot dot dot wp p b if you differentiate this with respect to wk a particular term what will you get is pk.

If you differentiate this if you differentiate this say with respect to wk only the k th term will be picked up that has only wk others do not. So, wk pk, so that will give rise to pk. So, we put back the derivatives partial derivation of vector form this del del w 1 del del w 2 del w 3. What will you get is, p 1 p 2 dot dot dot say p 0 dot dot dot up to P p that is equal to a p vector.

(Refer Slide Time: 40:42)

$$-2P + \nabla_w [w^T R w] = \underline{0}$$

$$= \frac{\partial}{\partial w_n} [-2w^T p + \frac{\partial}{\partial w_n} [w^T R w]]$$

$$\Downarrow$$

$$-2P$$

So, that means, this will give rise to minus twice p vector complicated is here. What will this give rise to? So, here you just remember the terms. Let me write down the entire thing on top we have got minus 2 p plus this del w of this quantity w transpose. R w this must be equal to 0 vectors. Sigma d square I am forgetting, because that is already giving rise to zeros.

(Refer Slide Time: 41:30)

$$-2P + \nabla_w [w^T R w] = \underline{0}$$

$$w^T (R w) = ? \quad \frac{\partial}{\partial w_n} [w^T (R w)]$$

$$= \sum_{i=0}^p w_i \cdot R_{i,j}$$

$$= \sum_{i=0}^p w_i \sum_{j=0}^p R_{i,j} w_j$$

$$= w_n \sum_{j=0}^p R_{i,j} w_j + \sum_{\substack{i=0 \\ i \neq n}}^p w_i \sum_{j=0}^p R_{i,j} w_j$$

So, w transpose R w and R is a symmetric matrix. What is R w transpose R w, What is R w? Firstly, this is R consider the i-th row what is the I-th row. i-th of this time is this that will get the multiplied by w_i w_i times i equal to 0 to p that is this w transpose vector

if you take it i th entry wy that will multiply. What after you take this Rw vector, you can call it w prime i-th entry of w prime isn't it. So, i-th entry of w just a minute, if you call that matrix Rw R is a matrix w is a vector Rw is a vector.

If you call that vector give it any name say x or say z than w transpose z means this zi and what is zi, i-th row of R times w vector. What is i-th row of R? Rij wj j will vary from 0 to p. So, replace that here all right. Now, you see suppose you want to derive it with respect to wk you want to derive this quantity, this quantity with respect to wk for a particular k. What I will be doing in this outer sum I will be first separate out the k case wk. This i equal to 0 to p in that range when i equal to q i becomes equal to k that case, I am separating out and than wi i equal to 0 to p, but i not equal to k this.

(Refer Slide Time: 44:22)

$$\sum_{j=0}^p R_{ij} w_j + R_{iu} w_u$$

$$+ \sum_{\substack{i=0 \\ i \neq u}}^p w_i R_{iu}$$

$$= w_u \sum_{j=0}^p R_{kj} w_j + \sum_{\substack{i=0 \\ i \neq u}}^p w_i \sum_{j=0}^p R_{ij} w_j$$

I have to differentiate it with which is respect to wk. First consider second term in the second term in the outer summation wk never comes for each i, because i equal to k has been separated out. But for each i in the outer summation in the inner summation wk comes 1 when j becomes k. So, I get Rik wk and there if I differentiate with respect wk I get Rik. So, this give rise to wi Rik simple i equal to 0 to p and i not equal to k.

This term here outer also I have got wk on the outside and inside also j become k and then you have got wk. So, it is a product. So, first you differentiate with respect to wk hold this inner summation as it is a you know differences of a product. So, that will give you and then hold outer sum 1 wk as it is differentiate inner quantity the summation with

respect to w_k . So, get only 1 term $R_{ik} w_k$ that is differentiated you get R_{ik} . So, R_{ik} and outside you have got w_k , so $R_{ik} w_k$.

Now, consider this second sum here i equal to 0 to p , but i equal to k was not included i equal to k case means $w_k R_{kk}$ that will be here. Actually, I meant once just a minute well I made 1 mistake actually when I took out separated out i equal to k case i equal to 0 to p was here. And i equal to k was separated out. So, this time I cannot put as any independent i is already k isn't it, i varies from 0 to p out of which i equal to k was separated out. So, in this left hand side wherever I had i is to be replaced by k . So, actually this is not R_{ik} , it is R_{kk} .

Now, you see in the right hand side summation I have got $w_i R_{ik}$ i equal to 0 to p , but i equal to k is excluded. If i is equal to k is included I would have got 1 more term what is that $w_k R_{kk}$ and that is what I am getting here. So that means, this term can be included there, I can simply remove this constraint.

(Refer Slide Time: 47:17)

$$\sum_{j=0}^p R_{kj} w_j + \sum_{i=0}^p w_i R_{ik}$$

$R_{ui} = R_{iu}, R: \text{symmetric}$

$$\Rightarrow 2 \sum_{i=0}^p w_i R_{ki}$$

Now, you see again this is R_k , because i equal to k case was separated out this correction should be d 1 here also it is k . As I told you i equal to 0 to p was here, but i equal to k was separated out. So, whatever is separated out there I must be replaced by k . So, I need not do that. So, here again I replace i by k this is the these are corrections to be done. Now, consider the 2 summations R_{kj} and R_{ik} . Firstly, permit me to replace j by i here

does not matter. So, it becomes $R_{ki} w_i$ and summation over i and R_{ki} , because of symmetry is R_{ik} , because R is symmetric.

That means, what I have here I can write as again $w_i R_{ik}$ and that is what I have here. So, twice $w_i R_{ik}$ I from 0 to p . In fact, instead of R_{ik} I would prefer the other 1 R_{ik} equal to R_{ki} why not put R_{ki} . So, since I am minimizing this derivative, now what is this R_{ki} what does it mean.

(Refer Slide Time: 49:31)

The image shows a whiteboard with handwritten mathematical expressions. At the top, there is a matrix multiplication: $2 \underbrace{\begin{bmatrix} R_{k0} & R_{k1} & \dots & R_{kp} \end{bmatrix}}_R \begin{pmatrix} w_0 \\ w_1 \\ \dots \\ w_p \end{pmatrix}$. Below this, the partial derivative with respect to w_k is shown as $\frac{\partial (w^T R w)}{\partial w_k} = 2 \sum_{i=0}^p w_i R_{ki}$. Finally, the gradient is given as $\Rightarrow \nabla_U (w^T R w) = 2 R U$.

I am I differentiated that quantity where from it came actually. It came from deriving this with respect to w_k , but I am deriving this will respect to $w_0 w_1$ up to w_p . So, you can take value from 0 1 dot dot dot. What is this summation? You can easily see if you take the R matrix k th row $R_{k0} R_{k1} R_{k0} R_{k1}$ dot dot dot R_{kp} and $w_0 w_1$ dot dot dot w_p . This row times this column multiplied by 2 of course, this is what you get to obtain the derivative the k -th derivative. Now, k can you find k is 0 you will get the first row than second row like that. So, this very simply this is the R matrix isn't it this is R matrix easily. So, what you get is twice Rw ; that means, $\text{del twice } Rw$.

(Refer Slide Time: 50:46)

$$\nabla_W [E^2(m)] = -2P + 2RW = 0$$
$$\Rightarrow \tilde{W} = R^{-1}P$$

↓
Optimal / Wiener filter

So; that means, if you sum up what is the net variant. We earlier had minus twice p in to rest class I kept it upper case. In the next class I will again I will convert in to small p I will change that small p from the order. I used the letter small p to denote the order of the filter that I while change to some other index, because you know let us skip to the conventional followed universally. Never the less 2 P and here I got twice R_w remember this is the gradient expression I will use it in the adaptive filter.

This gradient should be 0 when I have got the minima that; obviously, will give rise to the same Wiener filter you can see w cap equal to R inverse p this is the optimal filter. In fact, is if you know R_{np} fine you can calculate this R inverse p, but problems is r_{np} are not always r_{np} are not always known in that case, you have to supply information these information externally. What we do these actually give rise to the interesting case of adaptive filter. When r_{np} are not known or when the input statistics or the joint statistics between d_n and x_n the change from time to time you cannot have 1 filter design once for all.

So, you what you should have you should try to have some adaptive mechanism by which your filter learns from the data that is coming in and tries to adjust itself by some iterative method. So, that finally, it convert this on what should be the corresponding Wiener filter for a given r and given p. Then again after a while if R_{np} changes than again that adaptive mechanism will be used it will further read from the data and readjust itself. So,

that you know you get a new R and new p and you get another corresponding Wiener filter.

So, there will be continuous adaptation mechanism if input R and p changes it will track and it will have an iterative self adjustment procedure and it will finally, converge. Every time that iterative procedure should converge the corresponding wiener filter. There are plenty of adaptive filter algorithms, but most celebrated is that least mean square algorithms which we will consider in the next class. And I will just give 1 or 2 applications. You know lecture on equalization in communication and all that, that will end this course.

Thank you very much.

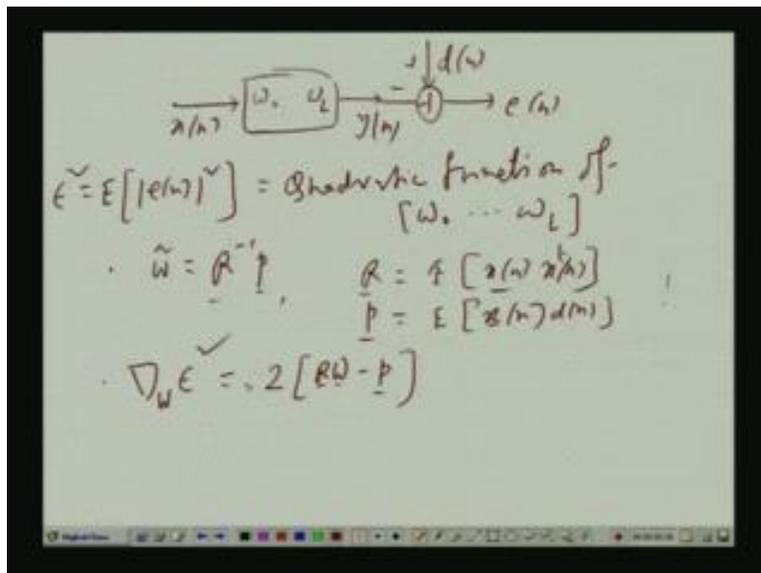
Preview of Next Lecture

Lecture - 40

Adaptive Filtering - LMS Algorithm

What you have seen in the previous class in the nutshell was this we are considering a filtering optimal filtering problem like this.

(Refer Slide Time: 54:04)



There was a random process x_n it was passed through an FIR filter w_0 dot dot dot w_p . Instead of p as I told you I will be changing it to some other change out, because I mean using p as a vector for some other or take w_0 to is w_l . So, l plus 1 coefficients. Output is

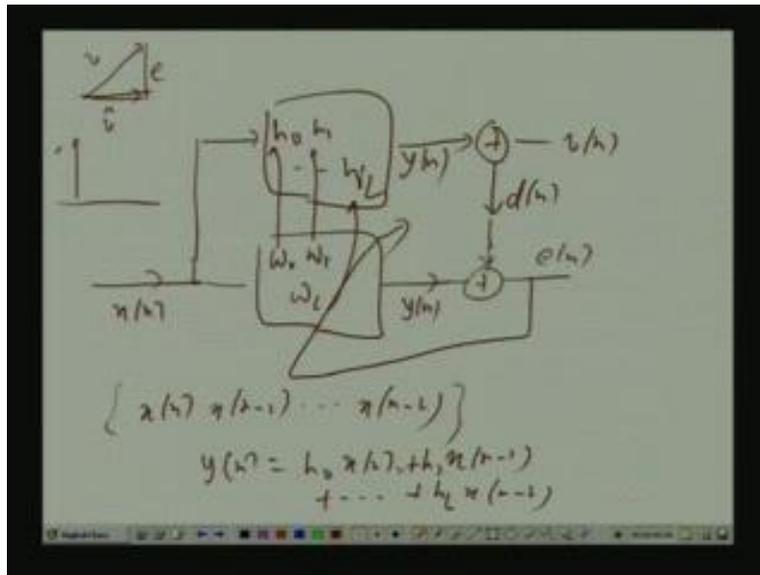
yn there is a dr response dn the error is. So, you try to minimize the mod square of this error with respect to the filter taps, because this e of. In fact, you can call it epsilon square is independent of m n, because of stationarity.

This quantity you have seen is a quadratic function quadratic function of w_0 dot dot dot up to w_l . We have also found out 2 things. What is if you calculate the optimal filter directly w_{cap} is $R^{-1}P$ where R is $E_{n \times n}$ transpose P is $E_{n \times d_n}$. Other thing you have seen the gradient this in the morning you have seen for a particular w it was twice RW minus P vector this you have seen. Now, 1 thing is that you can find out the optimal filter directly by $R^{-1}P$, but suppose you do not want to follow this path suppose you do not how to compute the inverse of a matrix.

We can propose an alternative method which is iterative method from different directions. So, I want to receive only 1 and I do not want to hear others. So, I will be you known I mean adjusting my antenna weights. So, that you now a μ is formed that is antenna gain is maximized is only 1 direction. The direction which I want to hear and noise, are created in other direction from which other signals are coming. So, thereby you induce some kind of special filtering.

Now, here the persons who are receiving the transmitter signals there in motion, so there direction changes. So, this business of giving you noise in certain directions and applying beam in certain direction, this needs to be again adaptively adjusted that give raise to adaptive beam former.

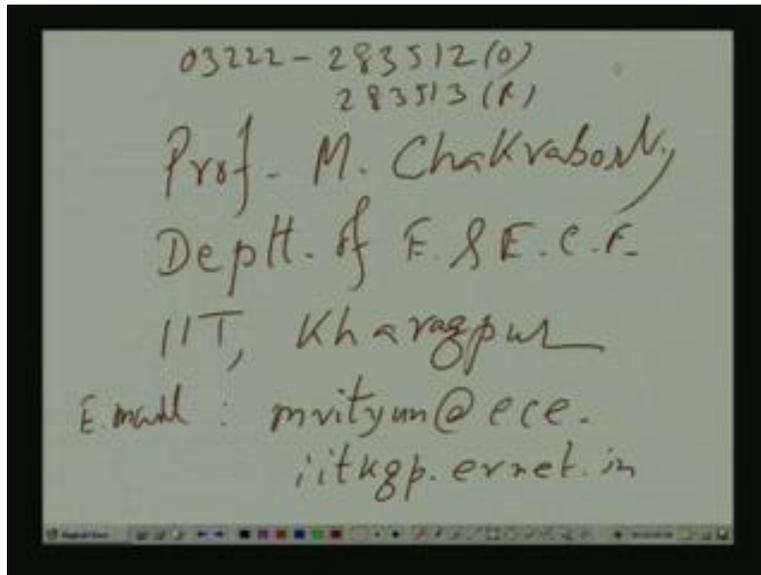
(Refer Slide Time: 56:56)



So, there are many applications which you will find in this book. So, I will not discuss them any further I hope you enjoyed this course. Now, you know probability and random process there are too many things in this so; obviously, in a unique course like this all cannot be covered all. But, I suggest I feel that what has been covered will serve as a good background and you can now going to you now I mean mark more advance topics or else you can go in to topics like Mart Angels.

You can go in to topics like queuing theory and this and things like that this all given in I mean in a at a introductory level in Populaces book and the background is required for you know reading those for studying those that is already been covered. So, I hope you are enjoyed this course and for any feedback you can contact me, you can write to me or you can contact me it will be nice to have a communication with your end.

(Refer Slide Time: 57:46)



So, my you can contact me here, my email number email address is mrityun, mrityun at ece.iitkgp.ernet.in. My phone numbers 03222 283512, that is office 283513 that is residence. Than hope you enjoyed this course and in case, you learn something from this that will be good result for me. So, thank you very much for all you attention good luck and good bye.

Thank you.