

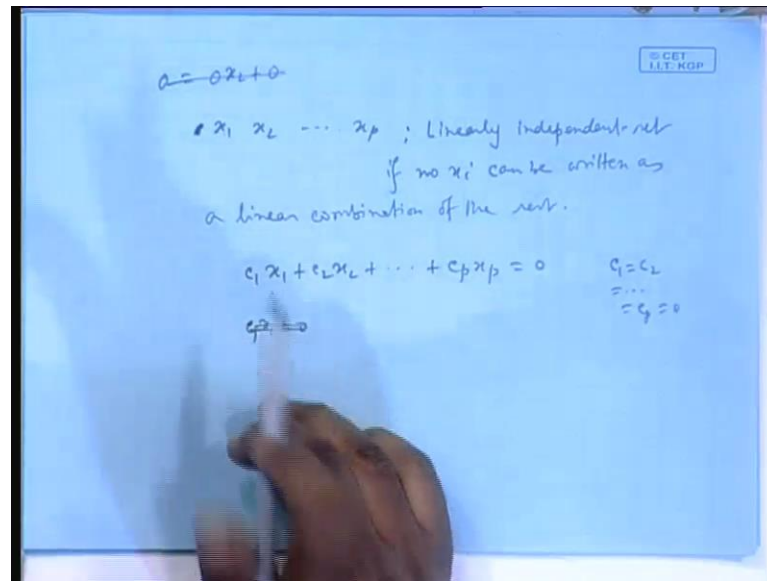
Adaptive Signal Processing
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 5
FIR Wiener Filter (Real)

So in the last class, we were discussing first we discussed Hermitian matrices and then positive definite matrices. Hermitian matrices, we have seen that they can be diagonalized; then Eigen vectors corresponding to define Eigen values are orthogonal. In fact, they can be normalized also, so that norm square of each Eigen vector is unity. Then, any Hermitian matrix we used showed that it can be replaced like you know it can be expressed as a product of three matrices $T D T^H$ Hermitian, where T is a unitary matrix or I do not which notation you used $E D$ is Hermitian possibly and D is diagonal matrix; consider the Eigen values Eigen values are always real.

Further, I consider positive definite matrices; Hermitian matrices we had a motivation because we have shown that all correlation and covariance matrices are Hermitian matrices. Then we considered another class we called positive definite matrices positive if a positive definite matrix has to be Hermitian and it should satisfy all property therefore, all non-zero x vector $x^H R x$. If r is the matrix $R x$ must be greater real and greater than 0 greater than equal to 0; if it is positive semi definite greater than 0, if it is positive definite that is where we stopped we proved that property also. That Eigen values are I mean for positive definite matrices Eigen values are not only real they are greater than 0 for positive semi definite matrices; Eigen values are not only real, but greater than equal to 0 that is where we stopped.

(Refer Slide Time: 02:34)



Now, suppose I give you a set of random variables x_1, x_2, \dots, x_p . This set of random variables; I will be calling them linearly independent if any of them cannot be written as a linear combination of the rest. Actually, this I will elaborate further when I teach linear algebra later, but why do not you say that they form a linearly independent set if no x_i can be written as a linear combination of the rest; that is it is not that x_1 is equal to some c_1 some you know some $c_1 x_2$ plus $c_2 x_3$ plus dot dot. You understand what I am saying that there is no linear relation there is no linear relation between this.

I mean this a set of variables x_1 to x_p there is no linear relation; that is you cannot have x_1 as a linear combination of x_2 up to x_p that is $c_2 x_2$ plus $c_3 x_3$ plus dot dot dot plus $c_p x_p$ or x_2 as a linear combination of $x_1, x_3, x_4, \dots, x_p$ so on and so forth. In that case, it is called a linearly independent set. One way to check whether they are linearly independent or not is you form a linear combination like this. This, I will elaborate I tell you when I am deal with vector spaces, but I am telling something beforehand.

Here, c_1, c_2, c_p they are constants not random variables, but x_1 up to x_p they are random variable. So, this summation is a random variable this random variable equated to 0 means 0 random variable; that is the random variable which always takes 0 value. If we form an equation like this and if I say that the only way this can be satisfied is by taking c_1 equal to c_2 equal to dot dot equal to c_p equal to 0. If that be the

only possibility the only solution for this equation because you see if you have c_1 equal to 0 c_2 equal to 0 up to c_p equal to 0 left hand side is 0, but if this is the only possibility then only they are linearly independent.

I will repeat this when I discuss vector space theory, but please see why because suppose you have a solution where c_1 c_2 up to c_p they are not 0 some of them at least are non-zero. Suppose, c_1 non-zero c_2 non-zero others are 0; then you can take $c_1 x_1$ plus $c_2 x_2$ equal to 0, which means x_1 is writable as it in terms of x_2 . Suppose, $c_1 x_1$ $c_2 x_2$ plus $c_3 x_3$ equal to 0, where c_1 c_2 c_3 not 0; then x_1 can be written in terms of x_2 x_3 so on and so forth. If suppose just now one of them is non-zero others are 0; say c_1 non-zero that mean $c_1 x_1$ is 0 vector 0 random variable, but c_1 non-zero; that means, x_1 must be the 0 random variable, but 0 random variable.

If you have 0 random variable in this set becomes linearly independent, because any 0 random variable is a linear combination of any of set of variable say $0 x_1$ plus $0 x_2$ like that. If they have the coefficients equal to 0. So, any 0 random variable if you add to the any set that set becomes linearly independent; because 0 random variable can be written always as a linear combination of any set of random variables. So, when there is such linear relation existing between a set of a random variable; we say they are linearly related linearly dependent, but if there is none they are linearly independent.

Now, you know very rarely only we have situations in real life there is a linear relation involving random variables. In practice, most of the random signals we do not have such equation you know governing; I mean saying that one of the random variable is a linear combination of the other there is nothing like that. Occasionally that arises and they lead to beautiful algorithm and structures and all that that is a separate story; that is a separate story, but otherwise not. For example, if you give a sinusoidal signal you sample it say for three or four periods sinusoidal signal with amplitude A ; A is random.

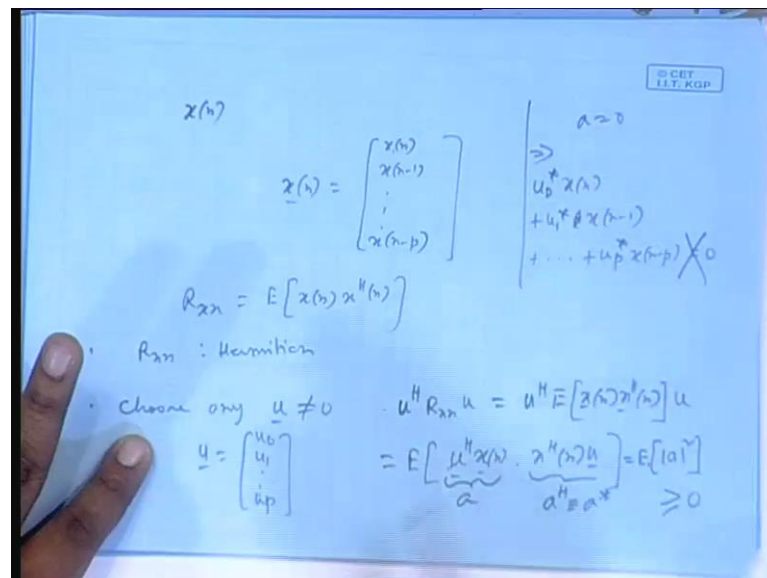
So, then every time you measure sinusoidal signal the amplitude fluctuates. So, it is a random process random signal because the amplitude A is changing. But the samples that we have in successive sample I mean suppose in three full periods they form a linearly independent set; because the sample that you have in first period that will occur again in the second period and occur again in the third period. So, if you want to

independent as a linear combination of the rest of the sample you we put 0 values to other coefficients take one value to the repetition case.

So, there you have got the linear relation linear independence, but in practice especially when you dealing with real life random signals unless there is a structure that is governing the generation of the random process. I will to leave there are such cases which occur rarely, but they occur and they are very well defined cases and have potential applications. Except for those cases especially for our case, we do not have situation, where one random variable is say summation of some other random variables because that will imply redundancy.

Suppose, x y two random variables and you say no x y is not enough you take x plus y equal to z also I say no x and y enough; in terms of them I can have z; so x y z in that case form a linearly independent set, but not x y not x y. So, in practice we will deal with we deal with sets set of random numbers or random variables, which are linearly independent there is not linear relation involving them. That is none of them expressible as a linear combination of the rest that is what happens most of I mean on most of the occasions.

(Refer Slide Time: 08:43)



Now, let us consider a random sequence x_n ; again I assume random sequence x_n and you form a vector call it x_n vector, in general complex valued you call the find out this correlation matrix R_{xx} . This autocorrelation matrix you all know we have defined; it

you could have taken covariance matrix also because, but whatever I am saying that is equally valid for both correlation matrix and covariance matrix. And also I will be most of with dealing with 0 mean variables, which means correlation covariance meaning the same no question of subtracting μ because μ will be 0. Most of n I will throughout this course rather, I will be dealing with 0 mean random processes.

That is a correlation or covariance they always would mean the same because mean will be 0 in our cases. Anywhere, take a form like this; we all know $R \times x$ is Hermitian about that there is no doubt, we have seen in the past that $R \times x$ is Hermitian. Now, so first thing is $R \times x$ Hermitian that we know we have seen earlier. Next, you take any vector choose any vector any u not equal to 0 any vector. Then, if you take and u equal to where u is equal to say $u_0 u_1 \dots u_p$; u_0 to u_p take any non-zero vector. Then, if I form this summation they can be any complex valued coefficients; if you form the summation $u^H R \times x u$.

I will show that $R \times x$ is not only Hermitian; it is positive semi definite and if there is no linear relation between x_n up to x_{n-p} that is if they form a linearly independent set; it is not only positive semi definite is actually positive definite. Positive semi definite always, Hermitian done other property you have to prove that for any non-zero vector u if you take a product $u^H R \times x u$ this is a real quantity and greater than equal to 0. If I can prove that then it is positive semi definite or equivalent equal non-negative definite, but further I will show that if there is no linear relation involving this elements of the vector.

That is, if the elements of the vector form a linearly independent set; there is none of them no x_n is a writable as a linear combination of the others and that is what happens most of it. You do not observe a random process in lab, where one sample is some summation of or linear combination of the past few samples exactly. There would be an error, but is not an exact summation of some previous samples. So, then this matrix is not only positive semi definite. In fact, positive definite that is what we will show and this is independent of whether x_n is WSS stationery or not.

I am not remaining assuming that correlation depends only on lag and all that. So, whatever I say now that is independent of this is more general that applies both to the case, where x_n is stationary and x is not stationary. I am not using WSS property of x_n

here; whatever I say this is most general. Now, consider this $u^H R x x u$; $R x x u$ you know this, u is constant no random there is no random variable in u , u consist of scalar constants $R x x$ is E of this and followed by again u . So, both u^H and u can be pushed inside expectation operation because they are constants.

This $x n$ is random it consists of random variables, but not u consist of only constants. You know constants can be brought inside the expectation operation after all expectant value of two x is two into expected value of x two can go out or can come in the reverse way. So that means, this is same as $u^H x n$ and $x^H n u$. Suppose, $u^H x n$, what is u^H a row vector what is $x n$ column vector. So, it is a scalar and what is this row vector column vector scalar. If you call this a my claim is this a Hermitian which is nothing, but a star because a is scalar has no meaning of transposition.

You can see if you take this a if you take the Hermitian of this; x will come first with x^H Hermitian this will go as the second guy Hermitian on Hermitian the Hermitian goes you get back u . So, a Hermitian and a is a scalar; so a Hermitian means a star only no question of transposition. So, this quantity is real, because $\text{mod } a$ so is a real quantity $\text{mod } a^2$. So, non-negative quantity expected value of $\text{mod } a^2$ $\text{mod } a^2$ can never be negative; $\text{mod } a^2$ can never be negative. So, this is always greater than equal to 0 always. So, positive semi definite or non-negative definite is always true.

Why it is equal to 0? If it is equal to 0, then a should be the 0 random variable that is whenever you measure a you only get 0 then only expected value of $\text{mod } a^2$ will be 0. If a is 0 random variable; that means, $u^H x n$ is a 0 random variable that is a equal to 0 implies $u^H x n$; that is $u_1^* x_{n1} + u_2^* x_{n2} + \dots + u_p^* x_{np}$ that is equal to 0 random variable this 0 is random variable. What does it mean? You chose u as non-zero vector; that means, not all the coefficients of u are 0 that is say all suppose the coefficients are I mean not all the coefficients are 0.

This means there is a linear relation between the elements of $x n$; $x_{n1} x_{n2} \dots x_{np}$ that is they are linearly dependent. So, only if the linearly dependent I can find at least one u and therefore, many u for which this will be equal to 0, for other choice of u it will be greater than 0. So, when the dependent this is indeed greater than equal to 0.

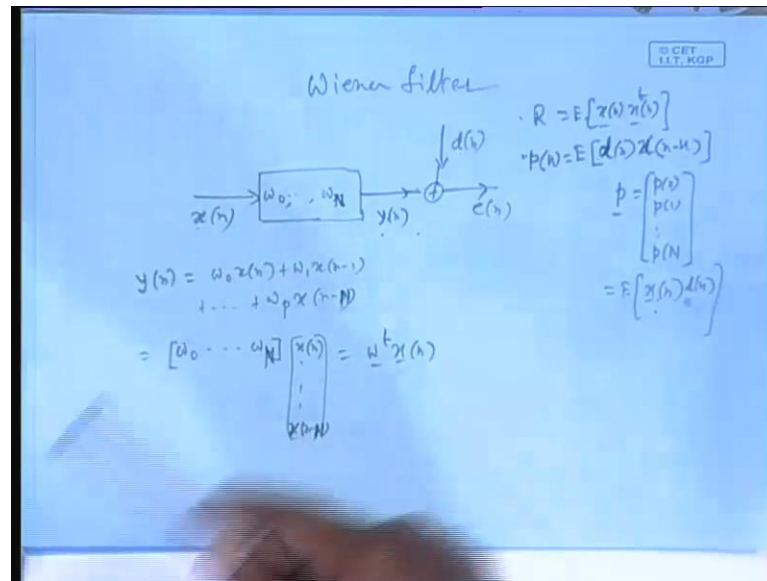
But when they are not dependent when this is this set extended to the up to x n minus p they are what I am saying you see they are linearly independent set. In that case, this can never be equal to 0 because you told me u consist of u is a vector, which is non-zero. So, all the coefficients are not 0 simultaneously.

You also told me that this is linearly independent vector, that is all the elements are linearly independent and then this summation can never be equal to 0 because that will mean they are linearly dependent contradiction that is not possible. So that means, if there is no relation no linear relation involving the set of random variables at hand, then this quantity is actually greater than 0 positive definite. So, in any case a correlation matrix will have real Eigen values, which are always greater than equal to 0, but in most of n when there is no linear relation involving the random variables Eigen values will in fact be positive.

Therefore, the correlation matrix should be invariable because after all determinant of the correlation matrix will be same as determinant of product of the Eigen value that you have seen last time. Now, these kinds of processes are called full rank processes. If you take any arbitrary number p can be one p can be two; you can take two samples of the random process three or thirty or thirty thousand. If there is no linear relation amongst them if no random variable can be expressed as a linear combination of either one or two or three or any number of I mean any other samples in that sequence that is called the full rank process.

For our case, if it is a linearly dependent up to p it is where I am taking only correlation matrix of order p . So, you understand why I was seldom talking about Hermitian matrices, because correlation matrices or covariance matrices they are not only Hermitian they are positive semi definite always. So, you have seen it here and most of it positive definite; in fact, because the processes are full rank processes.

(Refer Slide Time: 18:00)



So, with that I know start the basis Wiener filter or optimal filter. So, long whatever we discussed this is our background, this is our background material; again we will develop background material afterwards, but this is the key. That is, suppose I have got a random sequence $x(n)$ 0 mean WSS 0 mean. So, mean will be stationary of course, because mean is 0 everywhere and correlation depends only on lag. So, start with assume that it is WSS 0 mean WSS process. I want to find a filter FIR filter may be with some coefficients w_0 dot dot w_p . So, what is the output $y(n)$? $y(n)$ is you all know $w_0 x(n)$ you know this.

Again, I forgot to mention one thing to make life simple to start with we will be dealing with only real valued case; we will develop Wiener filter then we will develop adaptive filter from this all for real value. Then, again I will come back this I will generalize the case to complex valued because the treatment has to be different there. Two different treatments exist for the complex cases things are defined in a different way. So, for the time being I am dealing with real valued cases. So, even when I use the Hermitian symbol, I simply would mean transposition nothing else, there is no complex conjugation required.

Because there is nothing complex for the time being this is same as w transpose vector $x(n)$ vector. Then, you understand what w vector is? I do not think I have to tell you w vector is a column vector w_0 to w_p ; $x(n)$ is a column vector $x(n)$ to $x(n-p)$. This is the filter output, now the filter output I want to be because input random means output

also random. So, I want this to be a good estimate of another signal target signal d_n some target signal. That means, what should I do you cannot say that let y_n be equal to d_n . Because for a particular index n you can equate the two you have only then you get you get y_n equal to d_n .

So, you can have one equation with so many unknowns you find some solution fine. But immediately the next index comes y_n will be not only be d_n ; there is a huge difference. So, y_n equal to d_n will not work and also these things are random; d_n minus y_n the error this is that also that error also is random. Only thing is x_n is stationary means y_n is stationary x_n and d_n they are jointly stationary. Jointly stationary means if you take the correlation between x_n and d_n again that will depend only on the lag they are jointly stationary.

I write down that is what is given R , I am dropping R_{xx} just R replace for the input E . I am writing transpose here no point input is Hermitian. Hermitian also will not do any go will not do any harm, but still for your conjugation I am doing this. R correlation matrix and covariance matrix are same here, because I am dealing with 0 mean cases this is given and we see this is independent of n ; obviously, because I assume WSS input that is why R does not depend on n . Another thing is its given is cross correlation; you excuse me instead of w_0 to w_p may be I just.

So, because p I will be using p using something else. So, instead of w_0 to w_p maybe I can make it just make this changes; E of say $x_n d_n$ minus say k . So, correlation but with not on with x with itself between x and d lag is k . If I say jointly stationary then this correlation also depends only on the lag; then the two processes are jointly stationary in the second order that is in correlation. So, that will be then a function of k only. You can call that p this p notation is used. This job there in all books for that that is we have to change from p here to something else; p_k and we define the vector p as $p_0 p_1 \dots p_n$ which is nothing, but these are my definitions.

You can make it d_n though it will be same in this case $d_n x_n$ minus k . For the real case you know that correlation between d_n and x_n minus k or x_n and d_n minus k they are same; correlation is Hermitian as such or if it is real correlation at a lag k at correlation lag minus k they are same, only when they are complex there is a problem there not only same they conjugate of each other. So, I could as well keep x_n here d_n minus k here,

but for your conjugation for your to avoid confusion I am changing it. So, that when you write E of x_n into d_n you indeed get this, what is x_n vector after all.

First element is $x_n \times n$ into d_n p_0 second element is x_n minus one x_n minus one into d_n that is $p_1 x_n$ minus one d_n p_1 dot dot dot that is why I change it to confirm to this. But even if I had kept x_n here and d_n minus k things would not have been different because in the real case expected value between this two that should not change anyway this is the definition. Now, you see now come back to the physical problem what is our purpose I am trying to filter design a filter; so that this output becomes a good estimate of d_n .

As I told you, no point in equating y_n equal to d_n and solve the equation with so many unknowns and only one equation because even if you find some solution of that equation next for the next index data changes immediately equality goes. And see input is random this is random d_n is random d_n minus y_n , e_n is random so; that means, you have to minimize some statistical property of e_n , which gives the power of E_n . If the E_n power goes then; that means, e_n itself becomes a low power signal it never takes very high value. It has to be again I have to apply average because there is no point in taking says e_n square.

Because for a particular experiment, you observe one e_n you try to minimize e_n square and you get some set of coefficients. Suppose, that may not do good because next the time data changes you from the next experiment that time you get a new sequence e_n for that even this coefficients are not optimal. So, that is why it has to be some average power of e_n , which is called the mean square value. You understand one thing, if x_n is 0 mean y_n has 0 mean; d_n also I am telling 0 mean. So, d_n minus y_n that also has 0 mean. So, e_n has 0 mean.

So, around 0 mean how much is the average how much is the instantaneous power e_n square n . Capital E of e_n square n will be the expected power average power and because of stationarity; y_n is stationary and x_n and d_n they are jointly stationary. So, overall e_n of e_n square n we will evaluate that will be independent of n . That also then that is the measure that will give you the average power of p_n , I want that to go down not a particular e_n square not e_n square n for a particular observation observed by form of e_n .

Because that will not work for the next observation that is, why you have to bring in statistics here you have to apply the operator. So, you take the mean square error. Square means around the 0 mean deviation that is the power AC power you can say and mean of that. Now, you can say that will depend on what after all e_n depends on is d_n minus y_n and y_n is this this summation. So, if you take the square of e_n what will happen? d_n minus this entire thing that will be squared up then expected operation. Second thing will consist of terms of terms like w_0^2 w_1^2 dot dot dot w_p^2 $w_0 w_1$ $w_1 w_2$ $w_2 w_3$ is a second order term in w_0 to w_p .

So, it is a function of second order function of it is a quadratic function of the filter weight. Any quadratic function has got a either unique minima or a maxima in this case it will be a minima only $w_0 w_1$ dot dot dot consisting of what mean square error that will give you the optimal set of weight.

(Refer Slide Time: 30:27)

$$\begin{aligned}
 E\{e^2(n)\} &= E\left\{[d(n) - \sum_{k=0}^p w_k x_k(n)]^2\right\} \\
 &= E\{d^2(n)\} - 2E\left\{\sum_{k=0}^p w_k x_k(n) d(n)\right\} + E\left\{\sum_{k=0}^p \sum_{l=0}^p w_k w_l x_k(n) x_l(n)\right\} \\
 &= \sigma_d^2 - 2 \sum_{k=0}^p w_k p_k + \mathbf{w}^T \mathbf{R} \mathbf{w}
 \end{aligned}$$

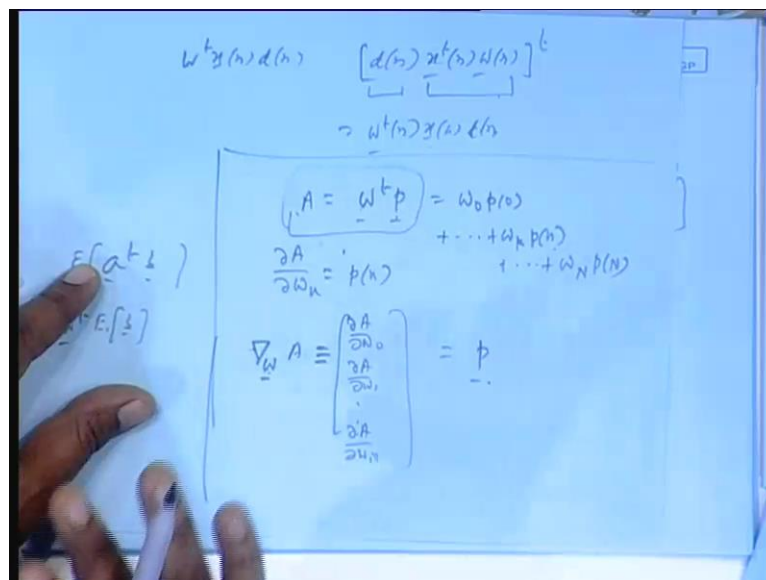
Anyway, in fact this e_n since I know that it will be independent of n , I have already calling epsilon square, but we will see that this indeed independent of n because of stationarity. What is epsilon square? Mean square error of the output error, which is a quadratic function of all the filter weights and I want to minimize it. So, what I do? and I suggested you will follow the way it teach you now, because some books might do in a very elaborate way and term by term definition all that do not do that. We will do things in a smart way. This is your e_n this is e_n , d_n minus $w^T x_n$ please see this; w

transpose $x \times n$ we should divide here forget this is the output $y \times n$, w transpose $x \times n$ vector w transpose $x \times n$ vector there is your output $y \times n$ minus that.

This is your $e \times n$, but it is $e^2 \times n$. So, I have to repeat it I have to repeat it, but what I will do I will repeat it, but if it is a scalar and these are the smart tricks I will apply this is scalar and scalar quantity or its transpose they are same. So, I will write it as its transpose. Now, you expect there are this multiply there will be four terms, first term will be $d \times n$ into $d \times n$ transpose that is $d \times n$ square itself, which does not depend on your filter weight there is external signal $d \times n$, it will give the power of $d \times n$ first term. This is independent of filter weight this is the power of the $d \times n$ and if $d \times n$ is a stationary process; I am assuming $d \times n$ also to be WSS it will be the power average power of $d \times n$ it can remain 0.

This is independent of n because of stationarity of $d \times n$, you can call it σ_d^2 square minus there is a cross term linking this two and there is a cross term linking these two. Now, can you see that w transpose $x \times n$ into $d \times n$ if you take and if you take the transpose of that you get back this other one. Say such things you get accustomed this kind.

(Refer Slide Time: 33:10)



I mean, w transpose this cross term we are in the initial phase of the course. So, I am doing step by step, but very often this matrix tricks and all I will apply at random; w transpose $x \times n$ into $d \times n$ and the other cross term is $d \times n \times$ transpose n transpose $w \times n$, my claim is both are same. See, this is a scalar w transpose $x \times n$ w transpose row vector x and

column vector so scalar scalar into scalar scalar. Here also, x transpose row vector w is column vector. So, scalar scalar into scalar scalar and my claim is they are same; all are real my claim is they are same.

That is not difficult, you take this guy the scalar n transpose if you take the scalar or take its transpose they should be same, but if you take the transpose transpose of this and transpose of this. So, first transpose of this means w transpose $n \times n$ and then d transpose, but $d \ n$ is a scalar. So, transpose of $d \ n$ is $d \ n$. You can do it by elaborately you know expanding $x \ n \ w \ n$ and term by term, but do not do that. You got this idea. So, the two of cross diagonal elements the two cross elements; actually they are same are you now convinced.

So that means, twice I can write E plus the last term $E \ w$ transpose $x \ n$ and transpose of this; w does not come here I have to minimize it as a function of w all the weights, but w does not come here w comes here and here this is where you will get the quadratic terms because two w occur here only one w occur.

Anyway w is a constant, I can take w transpose out and E of $x \ n$ into $d \ n$; w transpose are you getting me w transpose into $x \ n$ into $d \ n$ then E . Why not take E on $x \ n \ d \ n$ and then pre-multiply w term you will get the same thing. You understood my logic, suppose a transpose b and b consist of random vector a consist of constants E of that. One way is to multiply all the elements term by term and add and E applies, but constants remain constant. So, you will see we will get the same thing. If you take out a transpose and do this expectation first and then to a ; we will get the same thing.

So, w transpose can come out here similarly w transpose can come out here this w can go out here. First, you take out w transpose here remaining thing on that $E \ x \ n \ x$ transpose n consist of random data that times w ; on that E operation. Why not take out the w to the right hand side just apply E on this is that difficult or you cannot see. Whether we multiply first and take w take expectation operation or E on this part and then multiply, you will get the same thing because w is independent w is not random w transpose is not random.

So that means, this will give rise to twice w transpose, you remember I gave a definition two things are given to us R and p ; R consist of the R is the input autocorrelation matrix; p is the cross correlation vector between this. After all you are trying to estimate $d \ n$, you

should have some kind of knowledge between d_n and x_n ; there is a cross correlation, what kind of cross correlation they have; d_n and x_n are you following me. This vector p suppose given to you; R is given p is given what is $p^T p$ is E of x_n into d_n . It consists of the cross correlation terms.

So, you get w^T transpose out E of x_n into d_n that is $p^T E$ of this on this that is your p and w^T transpose comes out w go out E of only this part which is R , R is we are assuming positive definite it is always positive semi-definite, but we are also assuming positive definite so invertible and all that. There is a full length process no the linear relation amongst the samples of the process x_n . Now, here you see $w^T p$ is a column vector real data all real w^T transpose row vector consisting of w_0 w_1 up to w_n . So, if you multiply get only first order terms w_0 only w_1 only no second order w_0^2 .

But here, $R w$ first $R w$ this is a column vector each element will be function of w_0 up to w_n then multiplied from here. So, you will have square terms on that side that is why it is a quadratic function. So, what I have to do I have to now differentiate it with respect to w_0 also; w_1 also dot dot dot; w_n also equating each of them to be 0 get a set of equation solve that will give me w_0 to w_n the optimal one. I will get a unique solution because it is a quadratic function; it has only unique minimal or maxima all that theory we know. But instead of doing like that again, let us do things in a smart way. So, we will derive some matrix results.

Suppose, I have a thing like this; $w^T p$ you call this quantity A ; I want to differentiate $\text{del} A$ with respect to particular weight say w_k . Now, what is $w^T p$ $w_0 p_0$ dot dot dot $w_k p_k$ plus dot dot dot $w_n p_n$. So, very simply what is $\text{del} A$ $\text{del} w_k$ that this is p_k . So, I will define a vector now $\text{del} w$ del this is my notation please follow the notation, $\text{del} w$ is a operation $\text{del} w$ working on A by by definition this is the definition is nothing, but derivative of A with respect to w_0 put in the first place derivative of A with respect to w_1 put in the second just arrange them one after another in a vector form.

I am doing nothing new I am only deriving collecting all the derivatives, but putting it in a vector form that is why I am giving it and then giving a compact name here $\text{del} w$ of A . That is A is differentiated by the w vector basically, this will be $\text{del} A$ $\text{del} w_0$ $\text{del} A$ $\text{del} w_1$ dot dot dot $\text{del} A$ $\text{del} w_n$. So obviously, what is this vector here? In general case you

have seen $\text{del } A \text{ del } w_k$ is p_k $\text{del } A \text{ del } w_k$ is p_k . So, what is this vector first element is p_0 p_1 dot dot dot p_n which is nothing, but p vector. So, whenever you have got a form like this A equal to w transpose p first order; if you take the del of A del of that term with respect to w you get back this this fellow always remember this.

I will be doing this kind of operations every now and then in this course for a while. So, w transposes p this kind of first order term; if you differentiate with respect to w you should get back p if you differentiate with respect to p you should get back w , w transpose p and p transpose w are same. This was simple so remember this formula, but there is another result I have to prove.

(Refer Slide Time: 42:10)

Handwritten mathematical derivation on a blue background:

$$B = W^T R W$$

$$R^T = R \Rightarrow R_{ij} = R_{ji}$$

$$\frac{\partial B}{\partial w_k} = \sum_{i=0}^{N-1} R_{ki} w_i + \sum_{j=0}^{N-1} R_{kj} w_j + R_{kk} w_k$$

$$= 2 \sum_{i=0}^{N-1} R_{ki} w_i$$

$$= \sum_{i=0}^{N-1} w_i \cdot (R W)_i$$

$$= \sum_{i=0}^{N-1} w_i \sum_{j=0}^{N-1} R_{ij} w_j = \sum_{i=0}^{N-1} w_i \sum_{j=0}^{N-1} R_{ij} w_j + w_k \sum_{j=0}^{N-1} R_{kj} w_j$$

If you have a thing like a w transpose, $R w$ is a scalar row vector matrix column vector matrix column means one column row into column scalar you call it B . Given that R is Hermitian in fact, it is real it is a real matrix. So, R is given. Since you are dealing with real case R Hermitian means basically, R is just symmetric R transpose equal to R R meaning R_{ij} and R_{ji} they are same. R_{13} R_{31} R_{14} R_{41} they are all same; this is given to us and we are evaluating this.

This is a scalar again a function of all the weights w_0 , w_1 , up to w_n . So, I can differentiate them differentiate with respect to each put them in a vector form. So, that will be my $\text{del } w B$. Any scalar is a function of the weights w_0 to w_n and if you differentiate with respect to each of them differentiate with each of them put them in a

vector form I say compactly that this is $\frac{\partial B}{\partial w}$, which is nothing. But again the same definition $\frac{\partial B}{\partial w_0} \dots \frac{\partial B}{\partial w_k} \dots \frac{\partial B}{\partial w_n}$ by $\frac{\partial B}{\partial w}$ fine this is what you have to find out we have to find out this.

Now suppose, I want to find out this this general guy generally you have to find out $\frac{\partial B}{\partial w_k}$ k can be 0 k can be one up to k can be capital N this I have to find out. So, let us see this B fellow let us evaluate what is this B fellow; B first you form this vector then 0th element of this guy 0th element of this vector first element of this guy first element of this vector so on and so forth. That means, B is w_i , i equal to 0 to n and then this vector $R w_i$. So, simple $R w$ is a vector that times I mean the i th element of that any w ; $R w$ it is i th element what will be the i th element?

Now, consider $R w$. First row of R times w will give the first element; second row of R times w will give the second element dot dot dot. So, i th row of R times w will give the i th element. That means this summation remains as it is here i th row. So, R_{ij} will vary from 0 to n and w_j this much i th row. So, $R_{i0} w_0 + R_{i1} w_1 + \dots$ you that scanning that row. $R_{i0} R_{i1} R_{i2} \dots$ say $w_0 w_1 w_2 \dots$ that is what I am doing. These quantities then you see this is how B depends on all the weights including w_k and you ask me to find out this derivative.

So, the way I will go about is this first consider this outer summation in this outer summation I will become equal to k ones that case I will separate out and i not equal to k that case I will write separately. So, this I will write as and i equal to k means w_k , i equal to k case separately I am writing. So, there is w_k summation R_{ik} here R_{kj} wherever I had I am replacing that by k . Now, you differentiate B with respect to w_k . First here in this outer summation w_k is never occurring, but for each w_i here there is an inner summation in which w_k will occur only once that time I will get R_{ik} so $R_{ik} w_i$.

Here there is something more it is a product. So, first you differentiate with respect to w_k . So, you get one this remains as it is. So, that you write. And next time you hold it as it is differentiate this with respect to w_k ; you get R_{kk} only where w_j becomes w_k you get R_{kk} otherwise 0. So, $R_{kk} R_{kk}$ so $R_{kk} w_k$; this $R_{kk} w_k$ if I club here you see this first summation i equal to k case was missing; if i equal to k this becomes $R_{kk} w_k$, but that has come here. So, this two can be clubbed and this fellow will go this thing will

go. Here, if it is I really will become k if I included this i equal to k case there we have got a term $R_{kk} w_k$, but that that has come already so; that means, I can club the two and remove this.

Secondly, here in this summation R_{kj} , because of the Hermitian property is same as this you can write same as R_{jk} and then why carryout with carry on with j call it i or R_{ik} with i you get the same summation and have a local sum local index; R_{kj} I am calling that R_{jk} because R is Hermitian symmetric matrix j k th element k j th element are same. So, R_{jk} and j k they are same. So, this is $R_{kj} w_j$ and then why call j you again call it i $R_{ik} w_i$, which is same as what you get here.

So that means, this will give rise to twice this summation $R_{ik} w_i$. In fact, let me call it instead of R_{ik} rather you call it R_{ki} ; I should have instead of changing here instead of changing here, I should have changed here R_{ki} that makes it better. R_{ik} instead of that R_{ki} ; $R_{ki} w_i$ again $R_{kj} w_j$ call this j as i , so twice $R_{ki} w_i$. So, look at what is an i is changing from 0 to N . So, look at this now $R_{k0} w_0$ $R_{k1} w_1$; that means, k th row of R is scanned $R_{k0} R_{k1} R_{k2}$ dot dot and similarly $w_0 w_1$ so; that means, k th row of R times w vector.

So, this is nothing, but what if it is R matrix this is the k th row k th row times this vector k th row of R into w that is what this is this is that will give you $\text{del } B \text{ del } w_k$. So, if it is w_0 ; that means, first row times this vector that will give you derivative with respect to w_0 , if it is $\text{del } B \text{ del } w_1$ second row times this vector. So, if you put them in a vector form this will be nothing, but R time w . So, this is equal to $R w$ are you getting me. So, this is twice R twice $R w$ rather there is a two here twice $R w$. Things are not so easy in the case of complex case in the complex case. So, I have to generalize in the real. So, I could say just simply think R_{ij} and R_{ji} no conjugation there because real symmetric. So, then this is what you get here?

(Refer Slide Time: 51:39)

$$E^2 = E[e^2(n)] = E[(d(n) - \hat{W}^T x(n))(d(n) - \hat{W}^T x(n))^T]$$

$$= E[d^2(n)] - 2E[W^T x(n)d(n)] + E[W^T x(n)x^T(n)W]$$

$$= -2W^T p + W^T R W$$

$$\nabla_W E = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} = -2p + 2RW \Rightarrow W = W_{opt} = R^{-1}p$$

So, now I want to do this equal to 0 0 dot dot 0 that is derivative with respect to w 0 0 with respect to w one 0 with respect to w n 0; because I am deriving I am trying to minimize. So, all the partial derivatives must be equal to 0. So, del w epsilon square equal to 0 you can call it 0 vector. But we have already defined derived this derive this and this is independent of w it will go. So, this will give rise to minus two p w transpose p if differentiated with respect with respect to w will give you p if differentiated with respect to p with give you w. So, this two p and this fellow gave twice R w.

So, you take two two cancels and take p on the other side this will give rise to that optimal filter. W equal to w opt or sometimes in some books denoted with a cap w opt which is nothing, but R inverse p. I assumed R to be positive definite invertible. So, I can write inverse R inverse p this is the Wiener filter Wiener FIR filter. So, if you want to design the filter you must know R and p, but often you do not know R and p because R and p. R will depend on the R is given the input autocorrelation matrix that will depend on the circumstances, which is generating the input like in the case of your this thing you know equalizer.

The received signal depends on the channel property also. So, its correlation involves the channel information. Moreover, R and p their value may change their statistics may change from time to time. So, you cannot design it once for all that is why you have to

do it adaptively. So, from here I will march towards the basic elements algorithm in the next class.

Thank you very much.