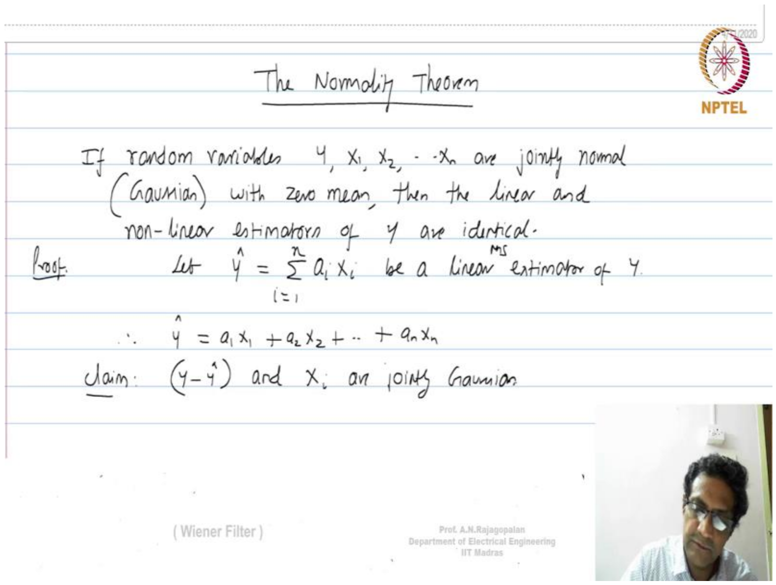
Image Signal Processing
Professor A N Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology Madras
Lecture 80
Wiener Filter

(Refer Slide Time: 0:17)



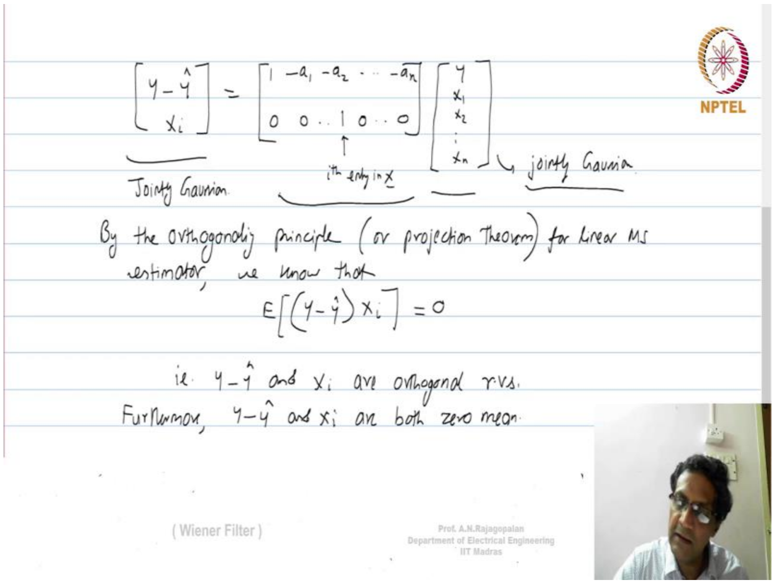
Let us examine or let us study the normality theorem within the realm of linear and the nonlinear estimators what is this normality theorem have to say. What is says is, if random variables, if random variables Y, X1, X2 all the way to Xn are jointly normal that is jointly Gaussian, normal means, normal in the sense that they are jointly Gaussian, normal means Gaussian with 0 mean with 0 mean, then the linear and nonlinear estimators of Y, estimators of Y are in fact identical, are identical.

This is the statement of this normality theorem and let us get of look at the proof, guess what we wish to show is that the, that if you were to express Y hat is a linear function X1 to Xn then it turns out that, that is also will be conditional mean under the condition that Y and (())(1:51) observe random variables were all jointly Gaussian. Now let Y hat be a linear estimator of Y that is it comes summation ai xi where we are assuming that xi is carrying information about Y, be a linear estimator.

Clearly ai is are all constants this we have already seen, the linear estimator of Y. Linear means square estimator of course, therefore Y hat is equal to we know is a1 x1 plus a2 x2 plus, plus,

plus up to an xn. Now, the claim let us claim the first claim is that Y minus Y hat this and this random variable and Xi, any Xi are jointly Gaussian it is the first claim.

(Refer Slide Time: 2:56)



How does one show this, this simply comes about because of the fact that if you write Y minus Y hat and if I have Xi and suppose I express this two together as one random vector then I can express this as a linear transformation on applied on the vector that is of the form Y X1 X2 all the way up to Xn. Now, this we know is jointly Gaussian because it is already mentioned in the statement, theorem statement has this that Y Xi's are all jointly Gaussian or jointly normal. Now, if you look at Y minus Y hat I can actually write this as 1 and then minus Y hat.

So, Y hat is a1 x1 plus a2 x2 and so on therefore we can simply write this coefficient, these entries is minus a1 minus a2 all the way up to minus an, as far as Xi is concern, we can always write this is 0 0 all the way up to maybe 1 where at the ith. So, this is the ith entry, entry in X, so somewhere you will have an Xi corresponding to that you will have a 1 and then again followed by all 0's, ith entry well let me say ith entry here. ith entry in X correct.

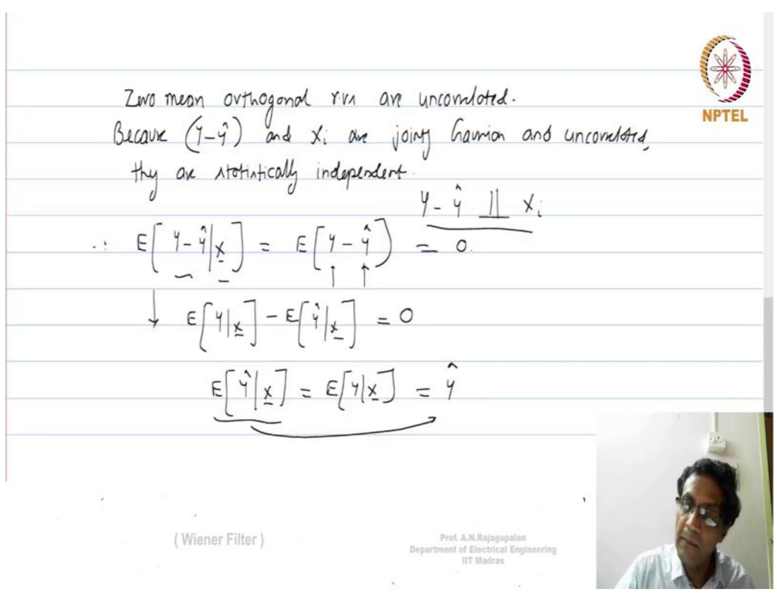
So, out of here X1 to Xn somewhere you will have an Xi with respect to that you will have a 1 here. Now, what this actually means is that, now because of the fact that this is a linear transformation because this just consist of constant because of the fact that this is a linear transformation applied on a random vector which is, which is originally jointly Gaussian therefore this is also jointly Gaussian, linear transformation on a Gaussian random vector yields again a Gaussian random vector.

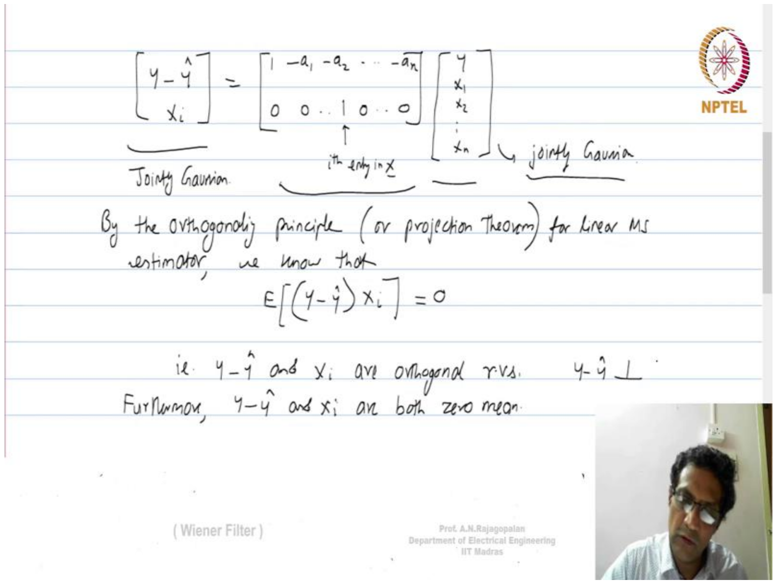
So, one of the left is also jointly Gaussian, by the way when you say that something is jointly Gaussian what it means is if you take any linear combination of this random variables every linear, any linear combination will again give you back a Gaussian random variable. At any linear combination you take they can return a Gaussian random variable. That is the most general statement regarding jointly, regarding random variables that are jointly normal.

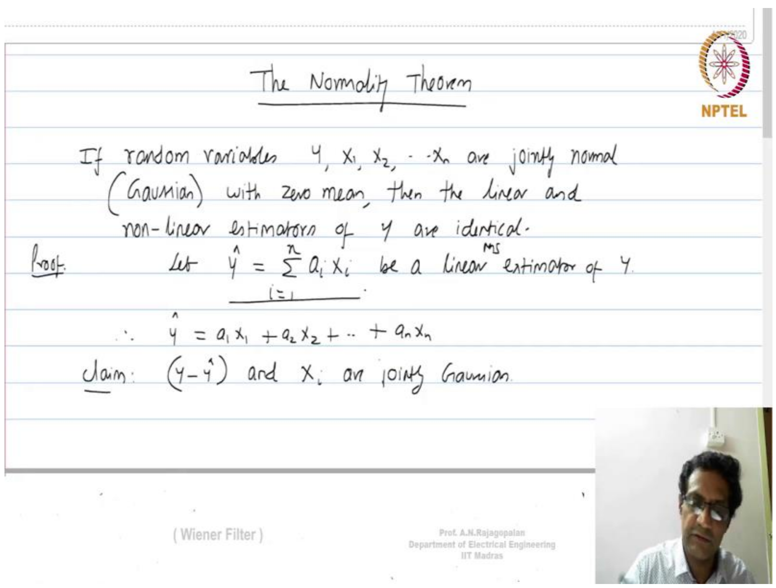
Here which a linear transformation on this random vector which is directly Gaussian and therefore Y minus Y hat makes I have jointly Gaussian. Now, moving forward by the orthogonality principle or the projection theorem for random, for linear means square orthogonality principle or what is equivalently called the projection theorem which we have seen, projection theorem for linear means square estimator or projection theorem for linear means square estimator because our Y hat is still a linear estimator.

We know that the error which is Y minus Y hat, expectation of the error Y minus Y hat with Xi is equal to 0. And this comes from our earlier this one, earlier result which means that is Y minus Y hat and Xi are orthogonal random variables. Furthermore, both Y minus Y hat and Xi are both 0 mean because Y is 0 mean, all the Xi's are 0 mean therefore Y minus Y hat is 0 mean and Xi is of course 0 mean, the Xi are both 0 mean.

(Refer Slide Time: 6:49)







And we know that from standard random variable theory, we know that 0 mean orthogonal random variables are uncorrelated. And because Y minus Y hat and Xi are jointly Gaussian and uncorrelated they are statistically independent. This is, this follows because of the fact that you have a Gaussian, Gaussian situation but if does not Gaussian we could not make this statement.

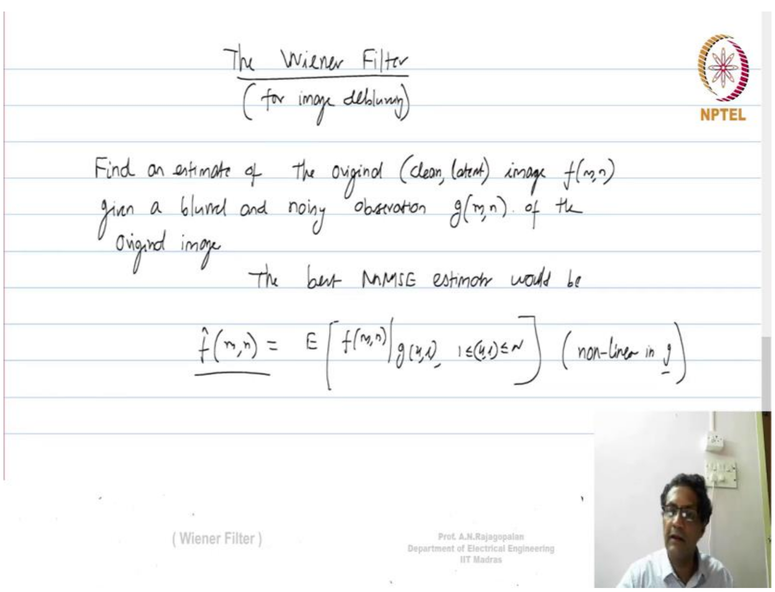
And for a Gaussian random vector and correlatedness implies, implies statistical independence, they are statistically independent, they are statistically independent. So, to indicate it normally so what we do is indicate this is Y minus Y hat is orthogonal to Xi and here we will indicate that Y minus Y hat statistically independent of Xi. This is simply a notation, therefore if I compute expectation Y minus Y hat given Xi X let us say that given the random vector X, where X consist of X1 to Xn this we could expectation of Y minus Y hat.

Because of the fact that this is independent of X, Y minus Y hat we just now showed that this guy is independent of X and then because of the fact that Y is 0 mean and Y hat is 0 mean, this will be equal to 0. Now, this further splits as expectation Y given X minus expectation Y hat given X and therefore in the right hand side we have 0 or in other word we have expectation Y hat given X is equal to expectation Y given X.

However, Y hat is simply a function of X therefore here it is all, it is all random vector X all the X1 to Xn and the Y hat, given X is simply equal to Y hat. Because this one is simply equal to Y hat once you give X because Y hat totally depends only on X and therefore the linear estimator that we started off with because if you realize we started with Y hat being a linear estimator and this orthogonality principle also that we are used for was also for a linear means for estimator.

And now what we have eventually concluded is that the Y hat that we started as a linear estimator is also the conditional mean therefore that kind can of proof is equal that means when with when what the unknown and they observe and the observe random variables are all jointly Gaussian. Then the conditional mean is linear in the in the this observe variables that means it will be a linear function of X which is what, which is what it is in this case because you know that it is a linear function of X. That is the normality theorem

(Refer Slide Time: 10:12)



Now, given all this background we are now equipped to study the Wiener filter given all that we have done till now we are now equipped to study what is called the Wiener filter. And the Wiener filter again now had comes in various context including communication and all that then we are going to look at Wiener filter for the image deblurring problem. The Wiener filter as we are as I said in the beginning our focus is going to be on the problem of image deblurring for image deblurring the Wiener filter for image deblurring.

Now, the Wiener filter until now whatever we did, whenever we kind of showed a relation between no whatever the rate when we said that we have a linear estimator or a nonlinear estimator we did not make use of the, make use of the exclusive relation between Y and X. We only said that X carries information about Y then this what can be done. Now, so as it in deblurring problem is concern we have more information, because of the fact we know exactly what we observe and we know how it is related to the unknown.

Because for example the observed image is blurred and noisy and we know that it is related to the input image through a PSF then basically if you have some noise and in this case of noise statistics which you might be aware of, therefore if you like to incorporate all of that, I mean that is something you should not done (())(11:31) till now, we only showed that the conditional mean exists and then conditional mean is that MNSC and then ((())(11:36) can also be linear under certain special condition and so on.

But until now we have not made use of the fact that, the fact about how Y and X are related and for the deblurring problem we are going to further utilize that relationship also in order to arrive at the Wiener filter. And the Wiener filter can be the spatial domain, we will first derive in this spatial domain and then kind of look at it, this look at its Fourier interpretation.

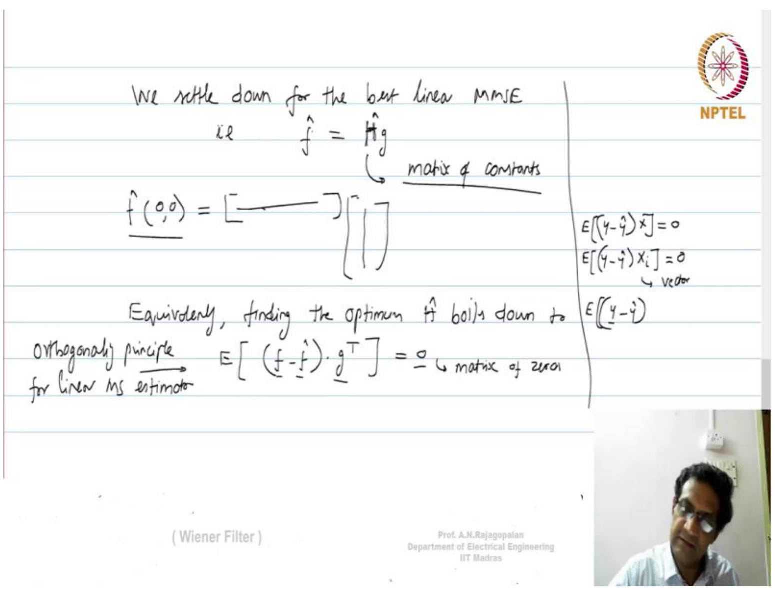
Now, like I say that we want defined now coming back to our image restoration problem for image deblurring problem what we have is find an estimate of the clean image or original image, Original whatever clean latent, it comes it is called by different names latent image f m, n given a blurred and noisy observation and noisy observation g m, n see all the apparatus that we have seen till now really did not have per say anything to do with image processing.

The condition (())(12:51) of course indicate, if you have deblurring problem how you would probably hold your problem, write down the cost function and so on but really now we are kind of, when we reach CLS at the time I indicated to you that how you would be able to use the observation model, now again it is time for a Wiener filter now wherein we will try to make this these are the observation model.

Given a blurred and noisy, given blurred and noisy, noisy observation g m, n noisy observation g m, n of the original image of f. So, these are noisy version of f. Now we know that the best MMSE, the best minimum means square estimator, MMSE estimator would be the conditional mean which will be f hat m comma n is equal to conditional mean f of m comma n given g let us say k comma l where k comma l will both are from let us say 1 to n.

Given the entire blurred and noisy observation g k n the conditional mean of f given g k l would be the best than my mean square, mean square estimator but then this will be typically nonlinear, this will be nonlinear in g, this will be nonlinear in g, which is what we are observing and the since the nonlinear estimators are kind of difficult to deal with, so we kind of restrict ourselves to a linear estimator.

(Refer Slide Time: 14:27)



So, what do we do? So, since the conditional mean is in general nonlinear, we settle down for a linear estimator for the best let us say linear MMSE for the best linear minimum mean square error estimator MMSE estimator which is that is we settle for f hat is equal to some H hat, H hat into g, this is simply a matrix of constants, matrix of constants.

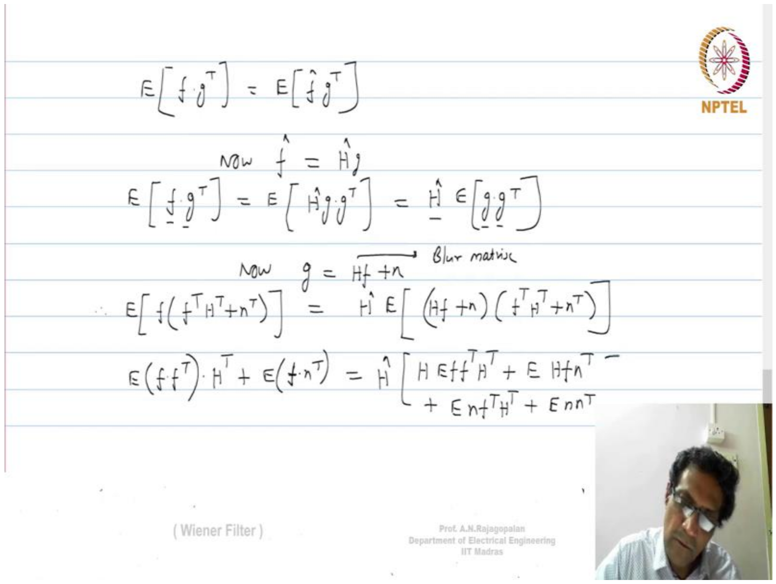
So, what this means is that, if you were to look at f hat 0 comma 0, which is the first intensity in the focused image, that will be the first row of H hat that multiplies g. So, it is like I am observing all of g and I am taking linear combination of all the obser, of all the observed intensities in g in order to be able to arrive at this f hat.

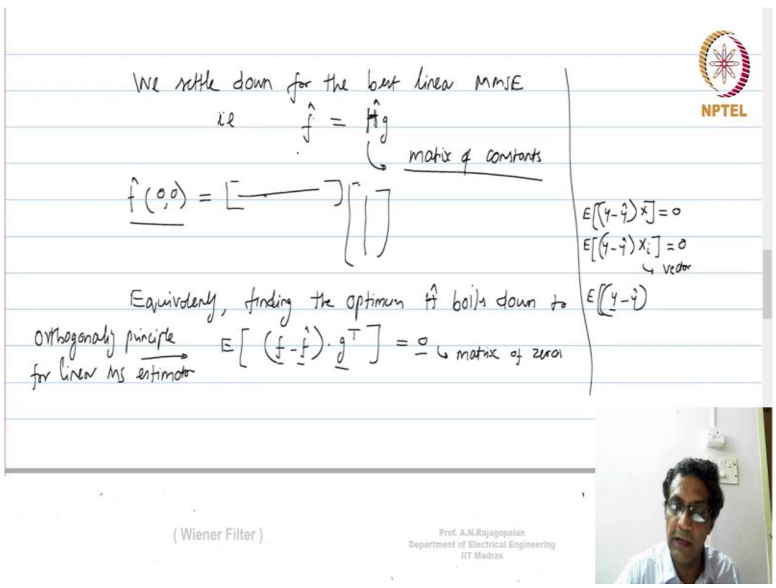
And then these coefficients will had to be found out such that they kind of give you, so their expectation f minus f hat, f hat square is as small as possible. And therefore, and we know that the equivalent, equivalently what this means is finding the optimum H hat, equivalently finding the optimum H hat boils down to the condition expectation f minus f hat, f minus f hat g transpose is equal to 0.

Remember f is a vector, f hat is a vector, g is a vector. Now this is a matrix now, this is a matrix. Now, matrix is 0. Now, why is this suddenly coming up? Because of the fact I mean if you had, now as an (())(16:33), think about it, if you had, X minus Y and, if you have Y to be a scalar and if your observation was also a scalar, then we had just one equation.

Then we extend it, this to the case when Y minus Y hat was still Y was still a scalar but then we observed really a vector and that we said Xi is equal to 0 and therefore this meant, that this whole thing when you work it out for every Xi, then this becomes a vector because then you have a vector of 0’s. Now, instead of this, now what we have is Y itself is now a vector and therefore at this condition of this orthogonality, so this is nothing but the orthogonality principle of this is based upon the orthogonality principle for linear mean square estimator, based upon that principle it will turn out that y f minus f hat into g transpose gives you the 0 matrix.

(Refer Slide Time: 17:39)





Or we can simplify this as expectation or expectation f hat, what is it, we have expectation f g transpose, is equal to expectation f hat g transpose. Now, what we have for f hat, now f hat by itself is H hat into g. and therefore what we have as expectation f into g transpose is equal to expectation f hat, this is H hat g, g transpose or we could write H hat expectation because H hat is simply a constant matrix, matrix of constants, so here g g transpose.

Now, each is a vector here, is a vector, so f g transpose is a matrix, this is a matrix, H hat is a matrix and so on. Now, we will actually bring into the observation model because I until now you have observed, we have done exactly what we did for LMSE and so on, we have not brought in anything regarding the relation between g and f.

Now, we know that g is equal to H f plus n, this is something that we know and this H is the blur matrix, this is different from H hat, H hat is our estimator, this is a blur matrix, our blur matrix is H. And therefore, what we can do is, we can write this as expectation f into g transpose, so g is this because g transpose f transpose H transpose plus n transpose is equal to H hat into expectation g, g transpose is H f plus n into g transpose is f transpose H transpose plus n transpose.

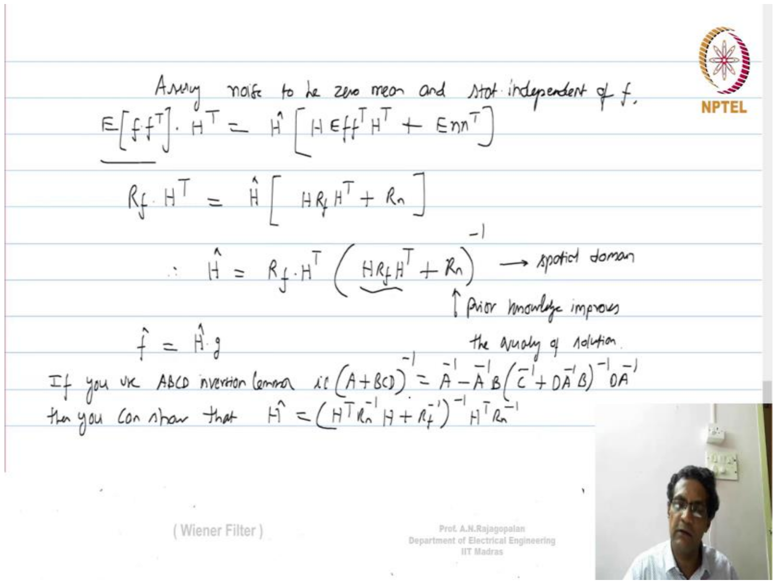
So, if you simply expand this then and then if you try to push, push your expectation inside then you will get expectation f f transpose into H transpose plus expectation f into n transpose, each is a matrix, so please observe, is equal to H hat into expectation, now it will H f f transpose H transpose, which we can simplify as, which we can write as H expectation f f transpose H transpose plus expectation H f n transpose plus expectation n f transpose H transpose plus expectation n n transpose.

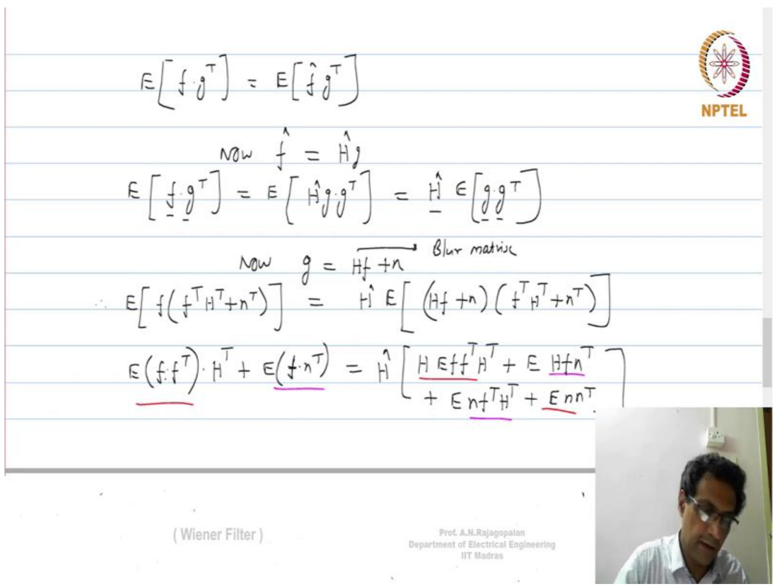
Now, let us assume that, now one additional thing that I should mention out here is that even though we said that we are going for simply a linear estimator which means that we are probably doing some optimal. But remember that if f and g are jointly Gaussian then this will also be the conditional mean, even that is linear in g.

And typically, in our kind of observation and those observation models where noise is independent of the signal, everything is Gaussian, the noise is Gaussian, the signal is modelled as jointly Gaussian and the observation model is linear, it can be shown that f and g turn out to be jointly Gaussian.

So, in that sense in most situations that you are dealing with perhaps wherein noise is modelled is Gaussian, the signal can be modelled as Gaussian, the noise and signal and then the observation model is linear, and noise is independent of the signal, then in such cases it also, can also be shown that g and f are jointly Gaussian, in which case even though you are only looking and even that looks like you are only doing linear estimation, but then you are actually doing the conditional mean. So, in that sense a Wiener filter is still a powerful filter, so we do not have to underestimate it simply because it is linear.

(Refer Slide Time: 21:28)





Now, assuming, now it is safe to assume that, assuming noise to be 0 mean and statistically independent of the, of your signal f, independent of f, because f is one (())(21:48) estimate. So what will happen is, so you see that the term here will go to 0, the f n transpose be a 0 matrix, this will go to 0, and f transpose, therefore what will remain is, this will go to 0, so the only terms that will remain are this and this and this, which will turn out to be, now suppose we write Rf to be the covariance, f f transpose is this (())(22:15) covariance of f.

And therefore if you write that as Rf or else let me just write this as expectation f f transpose into H transpose is equal to H hat into, what have we got here H Rf H transpose, H Rf H transpose plus expectation, let write this as f f transpose, let us, let me not, let me bring in this later, plus expectation f f transpose H transpose plus expectation n n transpose. And suppose we indicate the covariance of… of your f as Rf, then we get Rf into H transpose is equal to H hat which is your filter into H Rf H transpose plus suppose we indicated noise covariance as Rn, then we have Rn.

And therefore, H hat will then be Rf, so in this case you have to multiply it from the right, H transpose into H Rf H transpose plus Rn the whole inverse, that will be your, that will be the filter, H hat will turn into Rf H transpose H Rf H transpose plus Rn the whole inverse. Now, I would like you to notice that, the regularization part, if you are wondering where is your regularization, the regularization has sneaked in, in a kind of implicit way.

Look at this Rn, because of the fact that we suppose we have some knowledge about statistics of the noise and if you can (())(24:06) that, then the Wiener filter can actually accommodate that in a very nice way. And this you see, this is something that we have seen even when we did constraint least squares, we use to, we would think of some term that gets added to this in order to improve the stability, in order to improve invertibility, in order to improve the condition number, all that is happening here, so this is why I said that there is a nice sort of a parallel between sarcastic regularization and deterministic regularization.

In fact, if you go through a map estimator, a maximum a posteriori kind of estimator, then you can even do this regularization in a very implicit way, sorry explicit way. Now in this case, regularization has happened and you know in implicit way. So, this is your prior knowledge which is gotten in, improves the quality of your solution. So, this quality of solution f of your image, so it is like improving the numerical stability of this inversion process.

And this entire thing is as you can see in the spatial domain and the Wiener filter is very general, in the sense that you can assume that all of this, so even if you have space variant blur and if you knew it, you can throw it in. The only tricky part is knowing Rf because you might wonder who gives me the covariance, so the power spectral density which is a Fourier transform of the auto correlation function, I will talk about it in a minute.

But as of now this spatial equation is you know can actually accommodate any kind of blur, space invariant or space variant and so on. And your f hat, of course eventually will be H hat times g, because g is your observation. So, you have to multiply this filter with this. Now, the spatial domain thing other than the fact that this prior now I just want to also point out that if we use the, if you use the ABCD inversion lemma, if you use ABCD inversion lemma, if you know what the ABCD inversion lemma is because here it looks like noise is a prior.

Now Rf, you may think this Rf really a part prior even you can show that there is a dual role, where if you use the ABCD inversion lemma, that is A plus BCD the whole inverse, if you see this equation here, it has exactly the same form, Rn is A, H is B, Rf is C, H transpose is Z and then this is A inverse minus, this is kind of a nasty thing A inverse B into C inverse plus D A inverse B the whole inverse DA inverse, so you are head might soon, I mean if you look at this inversion lemma, but no then it is not so bad.

So, if you use this inversion lemma into this equation play a small little trick, then you can show that, then you can show that H hat can be equivalently written as, if you just write that down, (())(27:10) equation down, can be written as H transpose Rn inverse, this I just leave it to you to show H plus Rf inverse the whole inverse into H transpose Rn inverse.

So as you can see, now another way to look at it is that the prior, which is in terms of the image information, the covariance, so on which is some kind of statistical information, that can be brought in, so either can come in this form as we have seen here or it can come in this form. (())(27:43) either case it is a prior that is actually entering into the picture in order to be able to improve the estimate of your f.