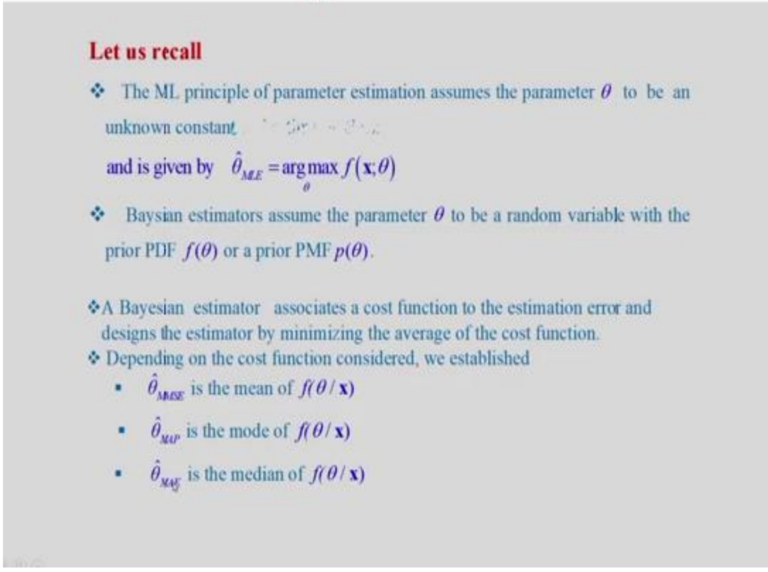
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics and Electrical Engineering
Indian Institute of Technology - Guwahati

Lecture – 17
Optimal Linear Filters: Wiener Filter

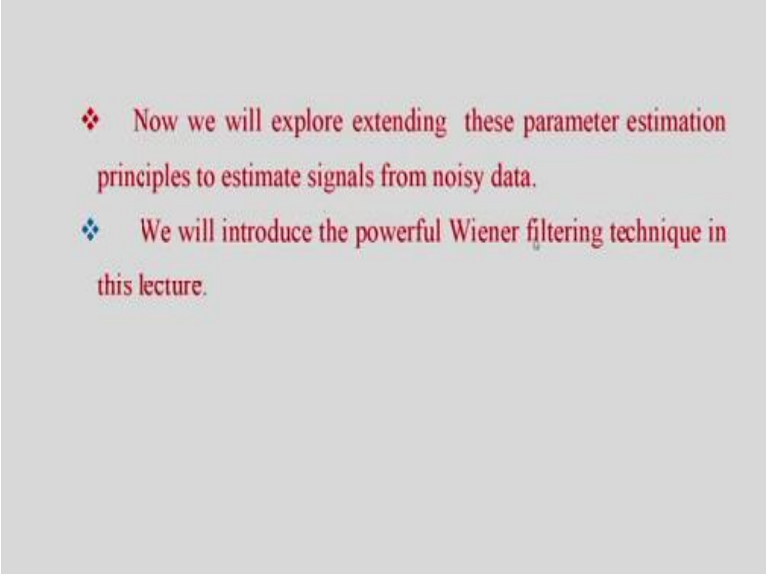
(Refer Slide Time: 00:43)



Hello students, welcome to the lecture on optimal linear filters; let us recall the ML principle maximum likelihood principle of parameter estimation assumes the parameter theta to be an unknown constant and the estimator is given by theta at MLE equal to arg max theta of fx, theta, this is the likelihood function. Bayesian estimators assume the parameter theta to be a random variable with the prior PDF f theta or a prior PMF p theta.

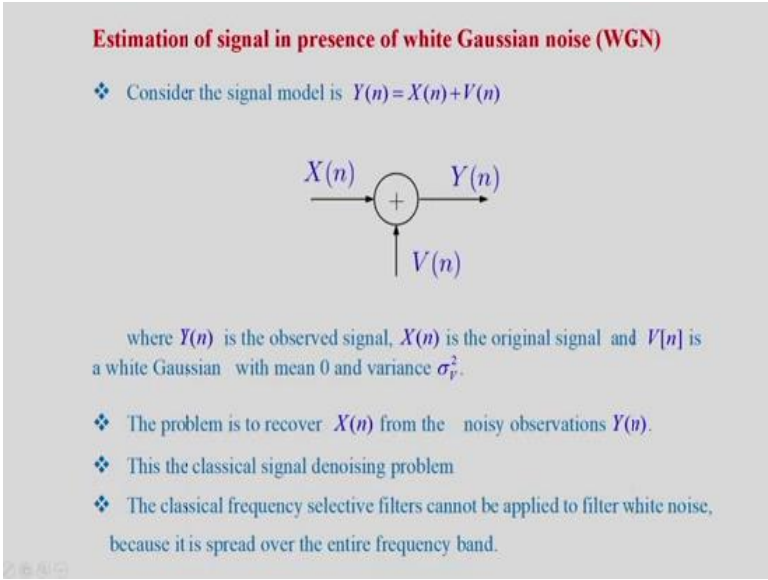
A Baysian estimator associates a cost function to the estimation error and designs the estimator by minimizing the average of the cost function, depending on the cost function considered we established theta hat MMSE that is the mean of the posterior PDF f theta given x, theta hat MAP, maximum a posteriori probability, theta hat MAP is the mode of the posterior PDF f theta given x, theta hat MAE is the median of the posterior PDF f theta given x, MAE here stands for minimum absolute error.

(Refer Slide Time: 02:11)



Now, we will explore extending these parameter estimation principles to estimate signals from noisy data, we will introduce the powerful Wiener filtering technique in this lecture.

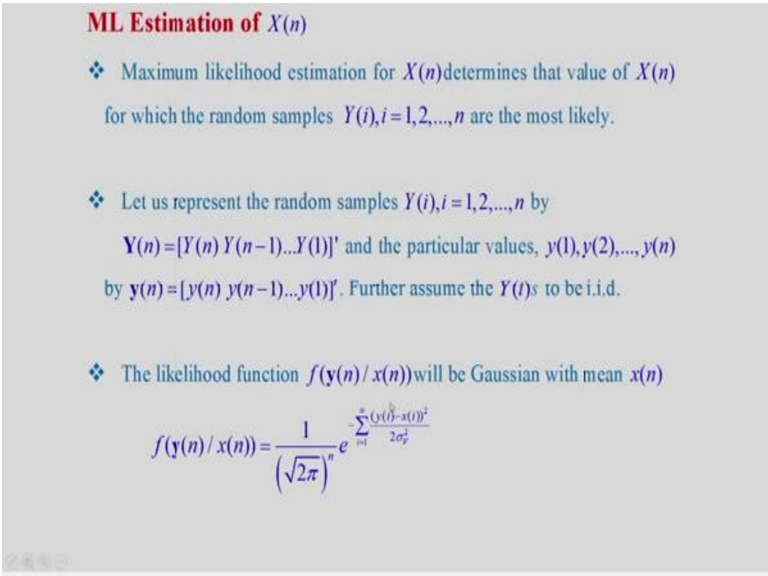
(Refer Slide Time: 02:26)



Let us start with the problem of estimation of signal in the presence of white Gaussian, this is abbreviated as WGN. Consider the signal model Yn = Xn + Vn, this is the addition Xn plus Vn, we are getting Yn, where Yn is the observed signal, Xn is the original signal and Vn is a white Gaussian noise with mean 0 and variance sigma v square. Now, our problem is to recover Xn from the noisy observation Yn, this is noisy observation.

From noisy observation, we want to recover Xn, so this is the classical signal denoising problem because this signal is noisy; we want to remove the noise. The classical frequency selective filters cannot be applied to filter white noise because it is spread over the entire frequency band because we know that white noise has uniform power spectral density over the entire frequency band, indicate some discrete time white noise, the power spectrum is spread from minus pi to pi.

(Refer Slide Time: 04:00)

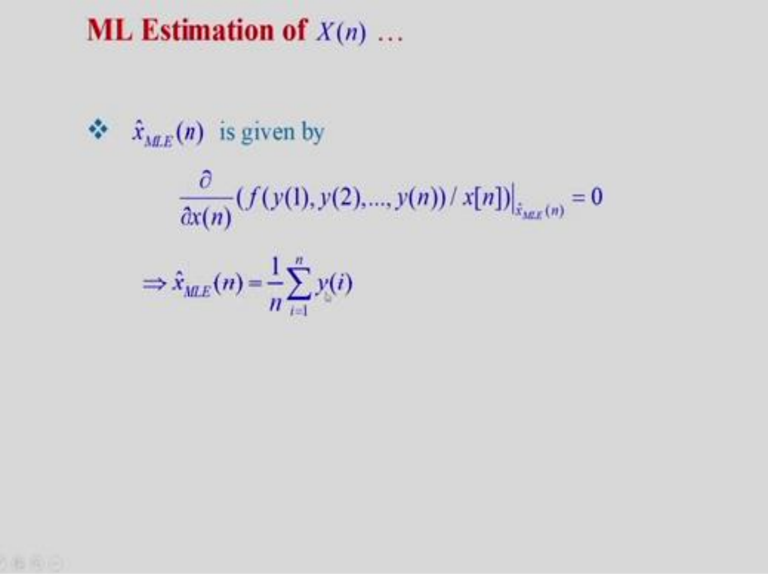


Therefore, the classical frequency selective filters cannot be applied to filter white noise, let us start with ML estimation of Xn, maximum likelihood estimation for Xn determines that value of Xn for which random samples Yi; i going from 1 to n are the most likely. Let us represent a random samples Yi; i going from 1 to n, we will represent it by this observed data vector, Yn vector is equal to Yn, Yn - 1 up to Y1 transpose.

This is the way we represent data Yn vector and the particular values of this random vector are given by y1, y2 up to yn and it is denoted by yn vector that is the vector comprising of yn, yn - 1 up to y1 transpose. Further, we assume that Yi is to be iid, to apply the ML estimation criterion, we assumed the Yi's to be iid; independent and identically distributed. Now, the likelihood function f yn given xn will be Gaussian with mean xn.

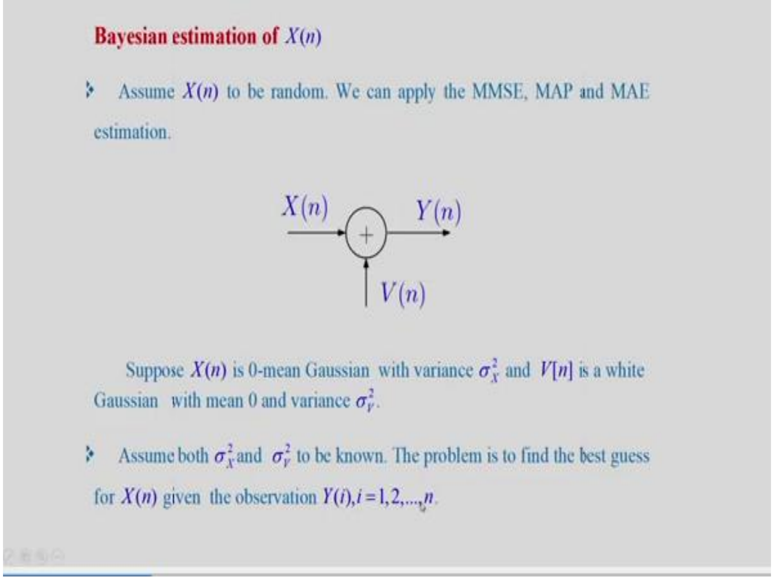
So, we can write f of yn given xn is equal to 1 over 2 pi to the power n into e to the power minus summation yi - xi whole square divided by 2 sigma v square, i going from 1 to n, so this we get assuming xi is a iid, therefore each is Gaussian distributed, therefore they will be jointly Gaussian distributed like this.

(Refer Slide Time: 06:01)



And now except MLE is given by the partial derivative of the likelihood function with respect to Xn at X hat MLE is equal to 0 and from this equation, we get X hat MLE and is equal to 1 by n into summation yi; i going from 1 to n, so this way we can estimate the signal value from a set of observed data.

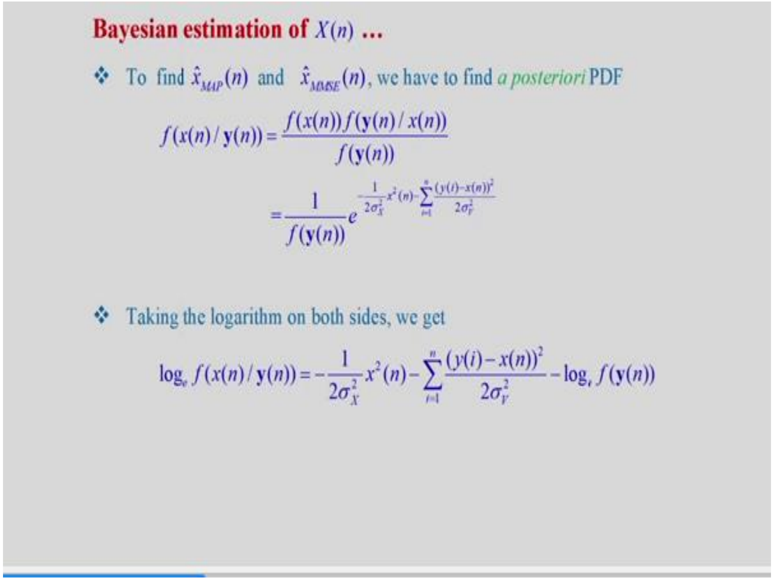
(Refer Slide Time: 06:33)



Next, we consider the Bayesian estimation of Xn, here we will assume Xn to be random, we can apply MMSE minimum mean square error estimation MAP; maximum a posteriori probability, MAE minimum absolute error estimation techniques, so this is the model as earlier; Yn is Xn + Vn, this Vn is white Gaussian noise and Xn has some prior probability distribution, suppose Xn is 0 mean Gaussian with variance sigma x square.

And Vn is a white Gaussian noise with mean 0 and variance sigma v square, so we assume Xn to be 0 mean Gaussian with variance sigma square and Vn to be Gaussian white noise with mean 0 and variance sigma v square. Assume both sigma x square and sigma v square to be known, so we assume that both are known samples. The problem is to find the best guess for Xn given the observation YI; i going from 1 to n.

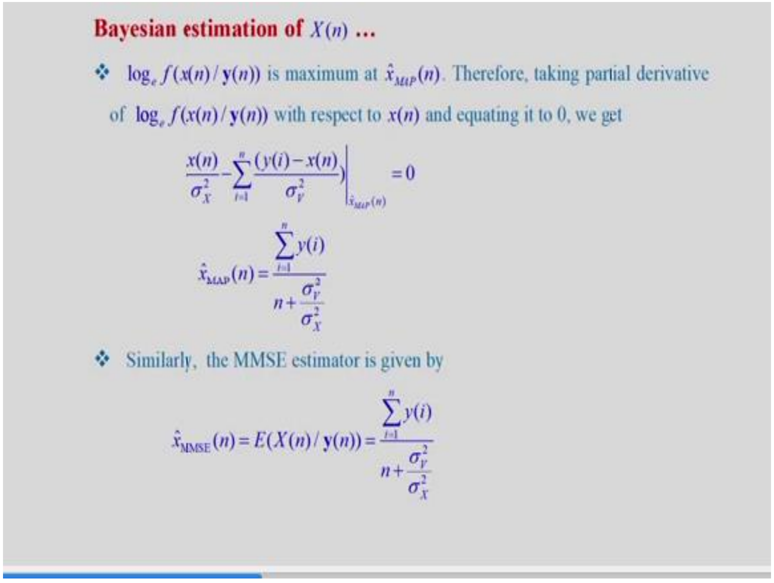
(Refer Slide Time: 07:56)



So, this is the Bayesian estimation problem, to find x hat MAP and x hat MMSE n, we have to find out the a posteriori PDF that is f of xn, given yn, yn is a vector that is f of xn into f of yn given xn divided by f of , so this we can write as; because this part is e to the power minus 1 by 2 sigma x square into x square n and this part is given by e to the power minus summation yi - xn whole square divided by 2 sigma v square, i going from 1 to n.

So, we can write the posterior PDF by this relationship where f yn does not involve xn, now we will take the logarithm on both sides, so log of the a posterior PDF that is equal to minus 1 by 2 sigma x square into x square n minus summation, i going from 1 to n of yi – xn whole square divided by 2 sigma v square minus log of fyn. So, now we can apply the condition for MAP and MMSE.

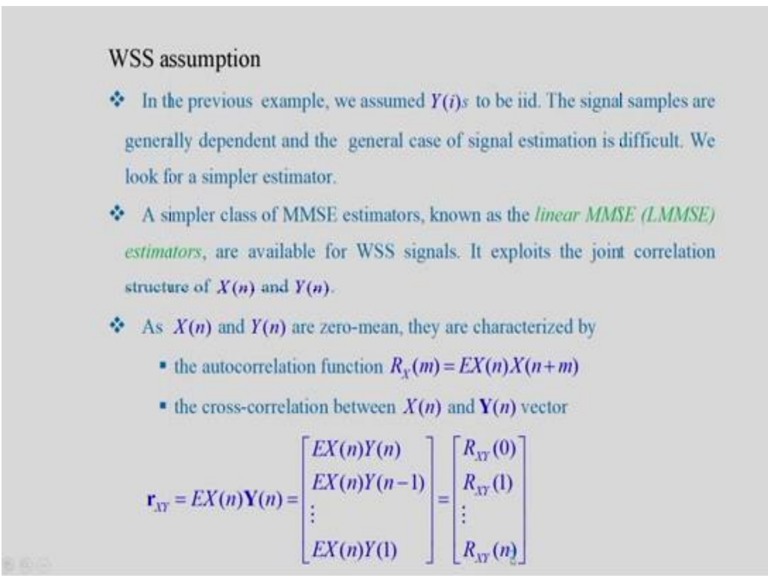
(Refer Slide Time: 09:19)



Now, this log of the a posterior PDF is maximum at x hat MAP n, therefore taking partial derivative of log of fx given n with respect to xn and equating it to 0, we get this condition; xn divided by sigma x square minus summation, i going from 1 to n of yn - xn divided by sigma v square f x hat MAP n is equal to 0, so this is the condition we get by applying partial derivative with respect to xn.

And if we simplify this expression, we will get x hat MAP n that is the maximum a posterior estimate of xn as summation of yi; i going from 1 to n divided by n + sigma v square divided by sigma x square similarly, the MMSE is given by x hat MMSE n is equal to conditional mean of the posterior PDF that is E of Xn given yn, so that is given by summation yi; i going from 1 to n divided by n + sigma v square by sigma x square.

(Refer Slide Time: 10:50)

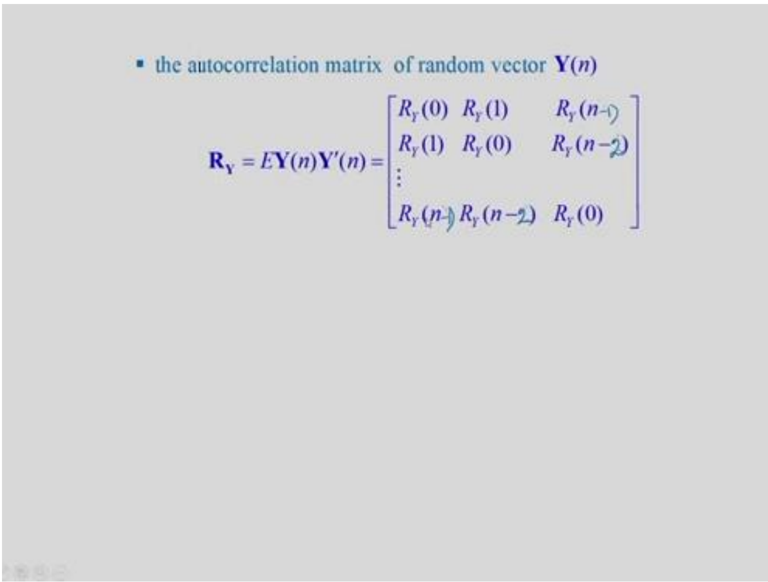


Actually, the same quantity we will get because of the symmetry of the Gaussian PDF, in the previous example we assumed Yi is to be iid, the signal samples are generally dependent, so iid assumption we cannot use and the general case of signal estimation is difficult, we look for a simple estimator. A simple class of MMSE estimators known as the linear MMSE; LMMSE estimators are available for WSS signals.

We have to consider the signals to be white sense stationary, it exploits the joint correlation structure of Xn and Yn in terms of autocorrelation function and cross correlation. We are assuming Xn, Yn to be 0 mean therefore, they are characterized by the autocorrelation function Rx of m that is E of Xn into X of n + m, the cross correlation between Xn and Yn vector is given by small rxy vector, this is the notation and this is equal to E of Xn into Yn vector.

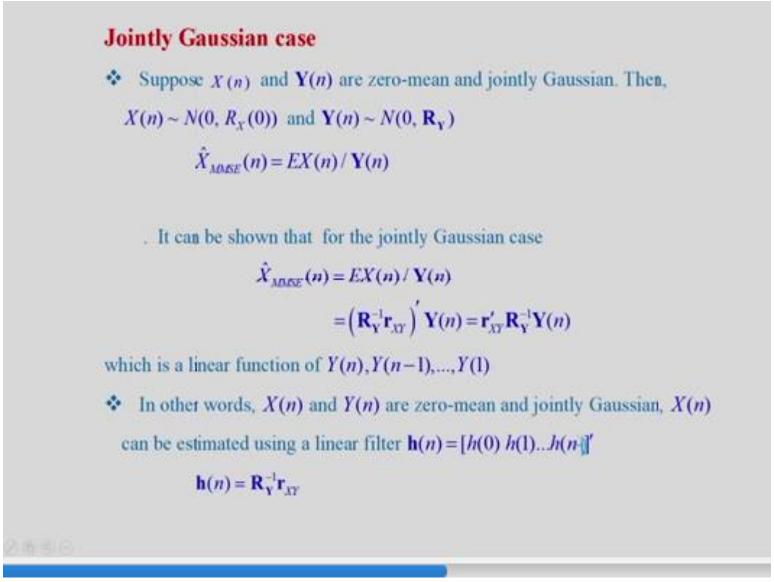
And this is given by this expression E of Xn into Yn, E of Xn into Yn - 1 and so on up to E of Xn into Y1 and using the joint stationarity property, we get this term is equal to rxy of 0, this term is equal to rxy of 1 and this term is equal to rxy of n – 1.

(Refer Slide Time: 12:31)



The autocorrelation matrix of the random vector Yn is given by this Ry matrix that is equal to E of Yn into Yn transpose, this will be a matrix and we will take the expected of the individual elements and assuming stationarity we get Ry0, Ry1 up to Ry n - 1 similarly, second row is Ry1, Ry0 and this way up to Ry n - 2 and so on, last row will be Ry n – 1, Ry - 2 up to Ry0, so this is the autocorrelation matrix.

(Refer Slide Time: 13:14)



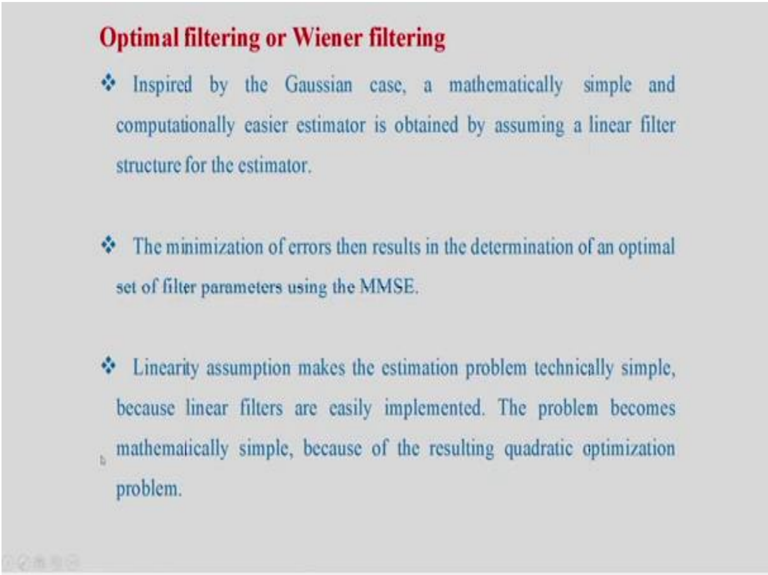
We will consider a simple case that is the jointly Gaussian case suppose, Xn and Yn are 0 mean and jointly Gaussian, then Xn is distributed as normal with mean 0 and variance Rx of 0 because 0 mean therefore, variance will be Rx of 0 only and Yn is distributed as normal with mean 0 vector and covariance Ry matrix, so again this covariance matrix is equal to Ry because of zero mean.

In that case, X hat MMSE at point n will be given by E of Xn given Yn vector, it can be shown that for jointly Gaussian case, this MMSE; X hat MMSE n is given by E of Xn, given Yn that is equal to transpose of RY inverse into rxy into Yn vector. Now, writing this transpose as rxy vector into Ry inverse, we are taking the transpose of the product, so we will get like this.

And Ry is anyway is a symmetric matrix therefore this will remain same, so we will get rxy vector transpose into Ry inverse into Yn clearly, this is a linear combination of Yn, Y n - 1 up to Y1, in other words if Xn and Yn are 0 mean and jointly Gaussian, then Xn can be estimated using a linear filter hn which is given by a vector comprising of h0, h1 up to h of n – 1.

And it is related to the autocorrelation matrix and cross covariance vector by this relationship; hn is equal to Ry inverse into rxy vector.

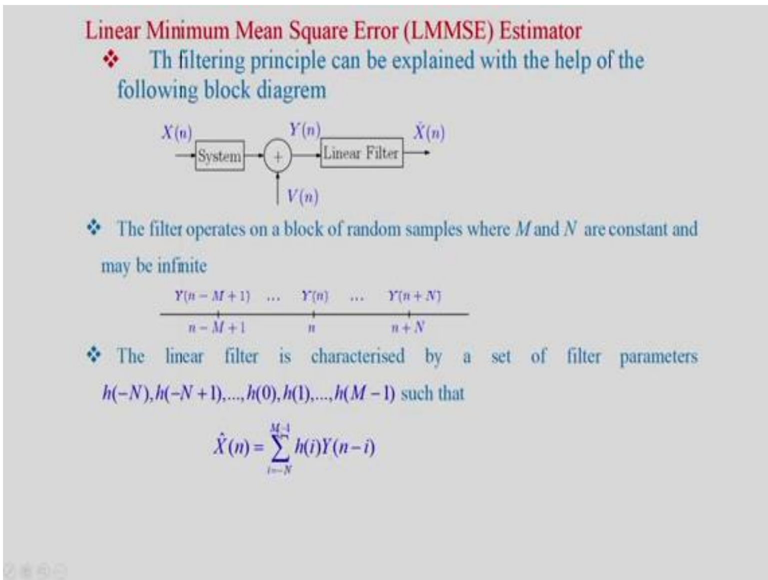
(Refer Slide Time: 15:32)



Inspired by the Gaussian case, a mathematically simple and computationally easier estimator is obtained by assuming a linear filter structure for the estimator. Now, for the estimator we will assume a linear filter structure, the minimization of errors then results the determination of an optimal set of filter parameters using the MMSE principle. Now, why linear assumption; linearity assumption makes the estimation problem technically simple, why?

Because linear filters are easily implemented, it is simply multiplication and summation, the problem becomes mathematically simple because the resulting quadratic optimization problem because of this linear filter structure, the mean square error criterion will be an a quadratic optimization problem.

(Refer Slide Time: 16:36)

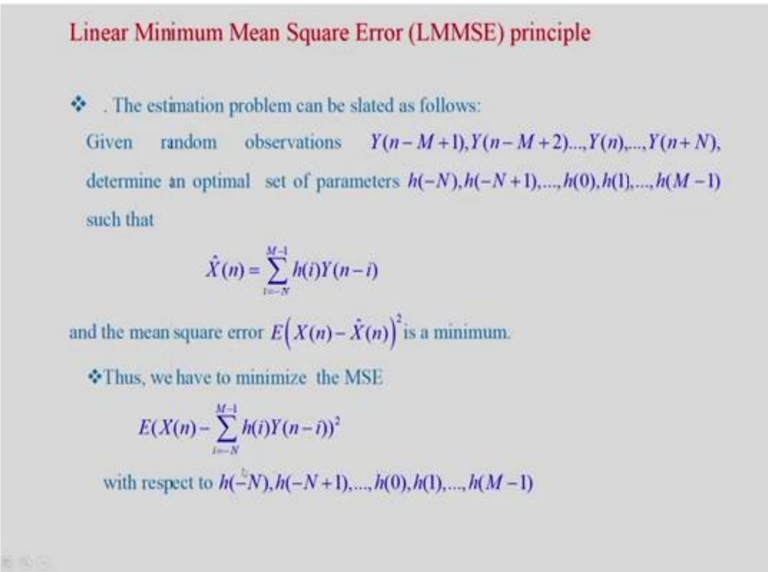


Now, we will discuss the linear minimum mean square error LMMSE estimator principle; the filtering principle can be explained with the help of the block diagram here, so suppose Xn it is pass through some system and the output is also corrupted by noise Vn and this is my observed data Yn and we have to apply some linear filter to get back X hat n, which is an estimate of Xn.

Now, this linear filter operates on a block of random samples where M and N are constant and may be infinite, so this is the block of data, this start from n - M + 1; Big M, this is n and this is n + capital N, we have both past and future data, so it starts at point n - M + 1 and end at n + capital N where M and N may be infinite. The linear filter is characterized by a set of filter parameters corresponding to suppose, this block of data h of – N, h of – N + 1, then up to h0, then h1 up to h of M - 1.

Such that now, filtered output will be given by X hat n that is equal to summation hi into Y n – i; i going from - N to M – 1, so we can get X hat n the estimator for Xn by this relationship that is the convolution operation which is given by summation hi into y n – i; i going from - n to M – 1.

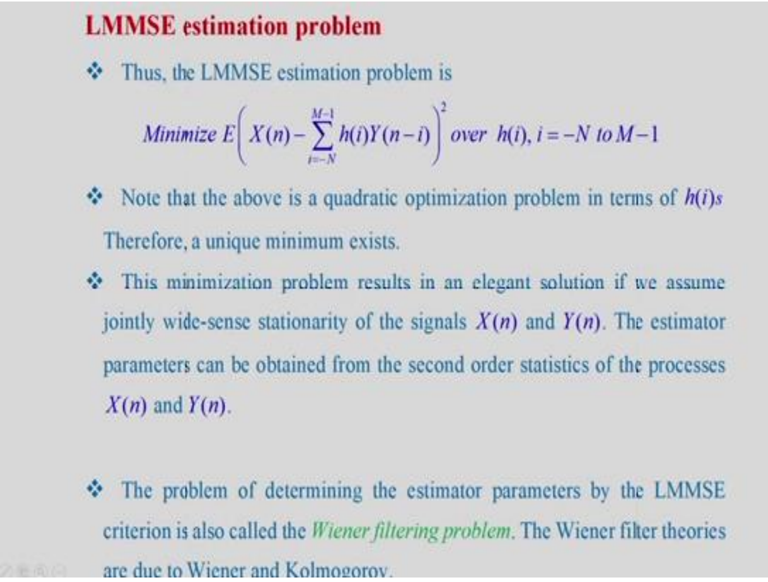
(Refer Slide Time: 18:38)



Now, the estimation problem can be stated as follows; given random observations Y n - M + 1, Yn – M + 2 and so on up to Yn, then up to Yn + capital N, determine an optimal set of parameters, these parameters are h of – N, h of –N + 1 up to h0, h1, up to h of M – 1, such that that estimator is the filter output X hat n is equal to summation hi Y n – i; i going from - N to M - 1 and the mean square error you have Xn – X hat n whole square is a minimum. So, we have the estimator output at the convolution between hn and Yn sequence such that the mean square error E of Xn - X hat n whole square is a minimum.

Thus we have to minimize the MSE mean square error, what is the mean square error; E of Xn minus this is the estimator summation hi Y n – i; i going from – N to M - 1 whole square, so this is the mean square error this we have to minimize, with respect to the filter parameters h of – N, h of – N + 1 up to h0, h1 up to h of M – 1, so we have to minimize this mean square error with respect to these parameters.

(Refer Slide Time: 20:23)

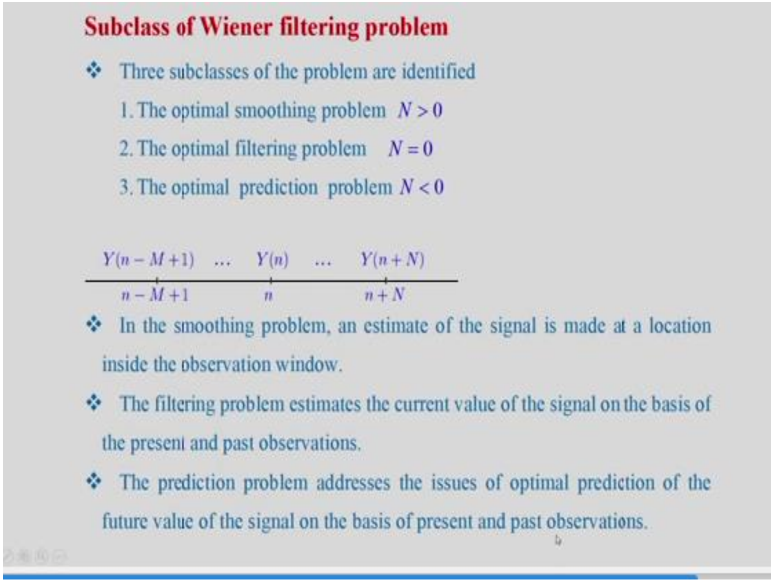


So, we can write the LMMSE estimator some problem as minimize E of Xn - summation hi into y n – i; i going from – N to M - 1 whole square over hi; i going from – N to M – 1. Now, let us see this optimization problem, this is the cost function and this cost function is a quadratic cost function in terms of hi's, therefore the above is a quadratic optimization problem in terms of hi's.

Therefore, a unique minimum exists because it is a quadratic optimization problem, unique optimum exists, the minimization problem results in an elegant solution if we assume jointly wide-sense stationarity of the signals Xn and Yn, this we will assume that signal n observed data are jointly WSS, the estimator parameters can be obtained from the second order statistics that is autocorrelation functions and cross correlation functions of the process Xn and Yn.

The problem of determining the estimator parameters by the LMMSE criterion is also called the Wiener filtering problem. The Wiener filter theories are due to Wiener and Kolmogorov.

(Refer Slide Time: 21:59)

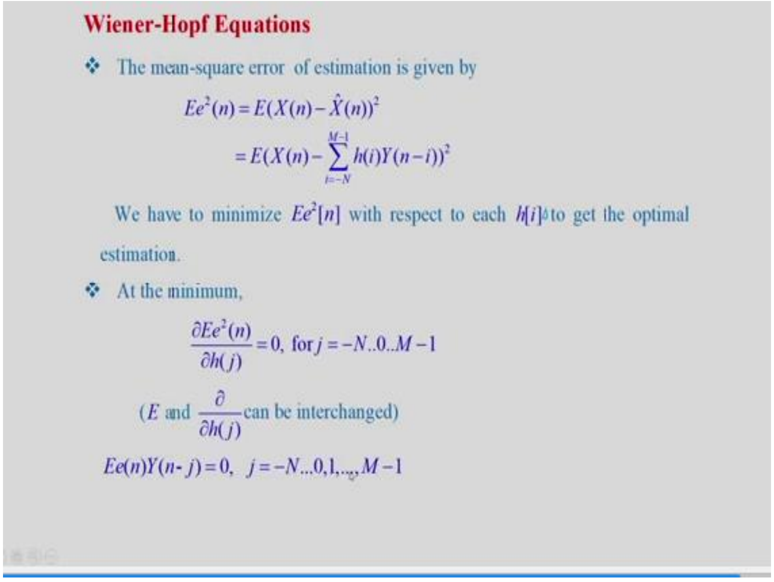


There are three subclasses of the Wiener filtering problem that is optimal smoothing problem, if N is greater than 0; capital N is greater than 0, the optimal filtering problem if N is 0 and the optimal prediction problem if N is less than 0 that means, this N; n + N, suppose if N is greater than 0 then we will consider the future samples also and therefore this will be this smoothing problem.

In smoothing problem, an estimate of the signal is made at the location inside observation window because observation window is up to here and we are making an estimator for this signal at this instant. The filtering problem estimates the current value of the signal on the basis of the present and past observation because if big N is equal to 0, then data will considering up to this point therefore, we will be considering the present and the past observations.

Now, for the prediction problem, N is less than 0 that means, we considered only past data because our begin is this side therefore, we will consider only the past data therefore, the prediction problem addresses the issues of optimal prediction of the future value of the signal on the basis of present and the past observations, these are the 3 subclasses of Wiener filtering problem.

(Refer Slide Time: 23:34)

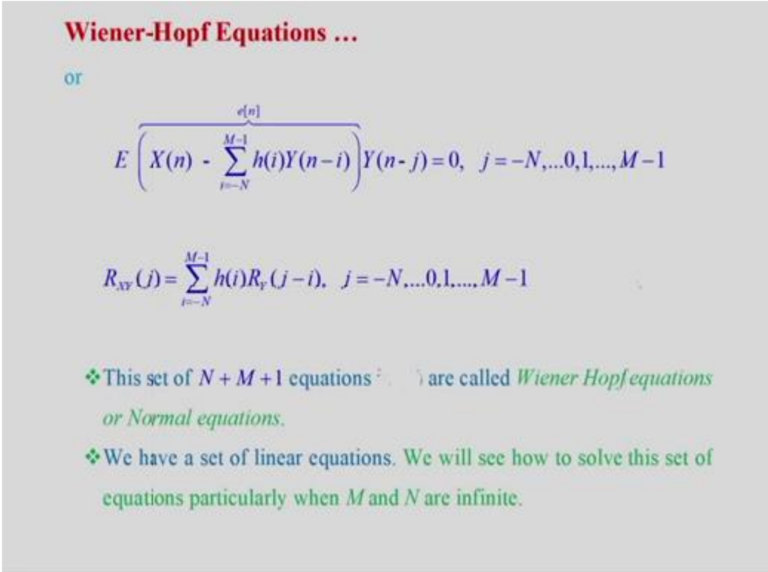


Now, the mean square error of estimation is given by E of e square n that is equal to E of Xn – X hat n whole square and if we write X hat n as the linear combination of hi and Y n – i's, we will get the same expression is equal to E of Xn - hi into Y n – i; i going from - N to M - 1 whole square, so this is the mean square error. We have to minimize E of e square n with respect to hi's to get the optimal estimation.

Now, at the minimum the partial derivative of E of e square n with respect to all hj is equal to 0, so del of E of e square n del hj is equal to 0 for j is equal to - N up to M - 1 now, this del operation and E operation can be interchanged, so first we will take the partial derivative of E square n that is E of n and then partial derivative of En with respect to h, so that way we will get only Y of n – j.

Because if I take the partial derivative of this error term with respect to hj, there is only one term, hj into Y n – j which involve hj, therefore by taking the partial derivative we will get E of; En into Y n - j is equal to 0 for j is equal to - N up to M – 1.

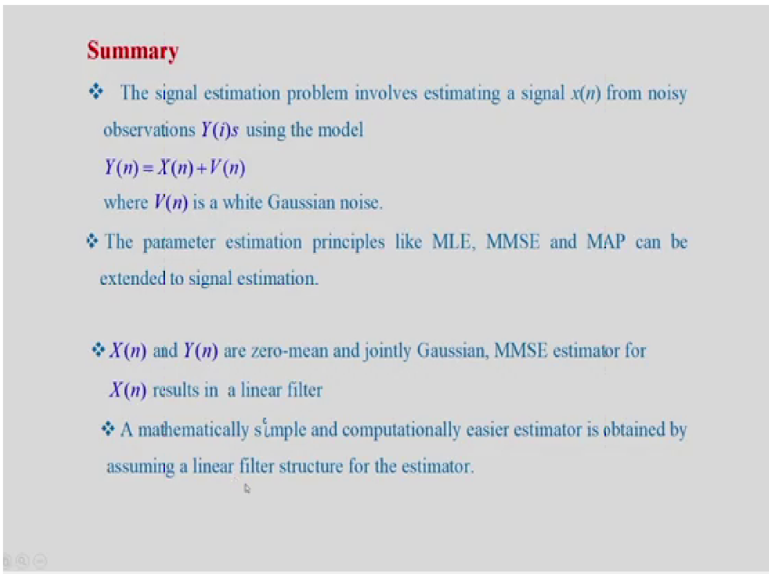
(Refer Slide Time: 25:28)



For N we can substitute this, E of Xn minus summation hi Y n – i; i going from - N to M - 1 into Y n - j that is equal to 0, for j equal to - N to M - 1 now, using the joint stationarity property we will get E of Xn into Y n - j is equal to Rxyj and therefore, we have Rxy of j is equal to summation hi into Ry of j – i; i going from - N to M – 1, where j takes the values; - N, - N + 1 up to M – 1. Now, we get a set of N + M + 1 equations called Wiener Hopf equations or normal equations.

So, this set is the Wiener Hopf equation, we have a set of linear equation we will see how to solve this set of equations particularly, when M and N are infinite, hence such solution of linear equations are easy but here problem is both N and M may become infinite, how to solve the Wiener Hopf equation in those situation is a problematic case.

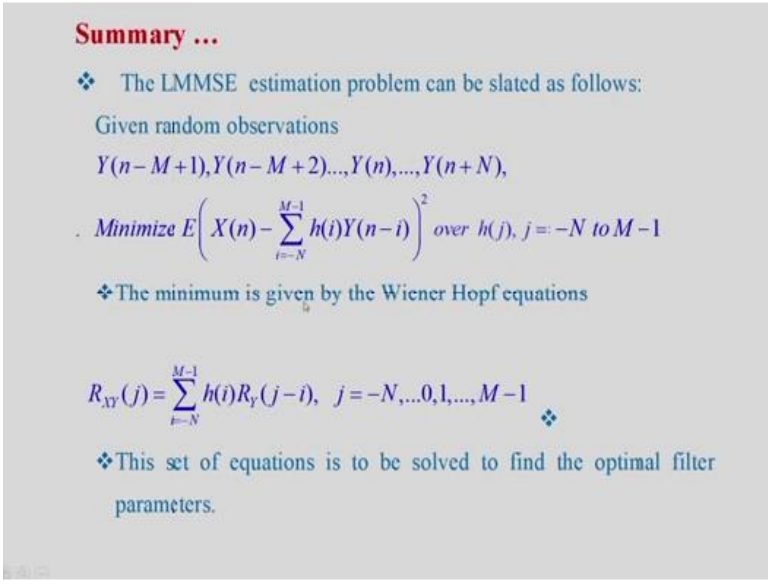
(Refer Slide Time: 26:56)



Let us summarize; the signal estimation problem involves estimating a signal xn from noisy observations Yi's using the model Yn is equal to Xn + Vn, where Vn is a white Gaussian noise. The parameter estimation principles like MLE, MMSE and MAP can be extended to signal estimation case, if Xn and Yn are 0 mean and jointly Gaussian, then MMSE estimator for Xn results in a linear filter.

This is an important observation; if Xn and Yn are zero mean, jointly Gaussian random processes, then MMSE estimator for Xn results in a linear filter, a mathematically simple and computationally easier estimator is obtained by assuming a linear filter structure for the estimator, so we assume a linear filter structure for the estimator.

(Refer Slide Time: 28:01)



The LMMSE estimation problem can be stated as follows; given random observations Y n – M + 1, Y n – M + 2 up to Yn + N, we have to minimize E of Xn - summation hi into Y n – i; i going from - N to M - 1 whole square over hi's; i going from – N to M – 1, so we have to minimize this mean square error with respect to hi's, so the minimum is given by the Wiener Hopf equations that is Rxy of j is equal to summation hi into Ry j – i; i going from - N to M – 1, where j goes from - N to M – 1. This set of equations is to be solved to find the optimal filter parameters which we will discuss in the next lecture, thank you.