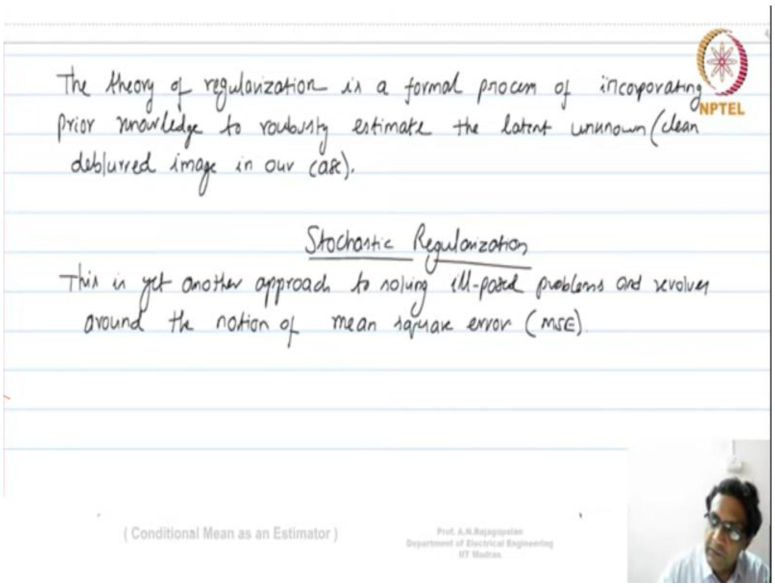
Image Signal Processing
Professor A.N. Rajagopalan
Department of Electrical Engineering, Madras
Lecture 78
Conditional Mean as an Estimator

As I earlier, regularization as a theory is something that actually allows you to robustly estimate an unknown when there is noise and when probably there are there are issues with existence, uniqueness and so on. So, in essence, if you want to summarize this theory of regularization in a few sentences, you can effectively is not say the following, the theory of the realization here let me just write it down.

(Refer Slide Time: 00:45)



The theory of regularization is the formal process of incorporating prior knowledge and we saw this already that not throwing in as much knowledge as you have about the unknowns helps in estimating them robustly, process of incorporating prior knowledge to robustly estimate the latent unknown, the latent or the hidden or whatever the actual quantity that is of interest, the latent unknown which in our cases is the clean image. In our case, clean deblurred image in our case. For us, for the image deburring problem. it turns out that what we are trying to solve for us is clean deblurred image in our case.

We saw for example, how a deterministic regularization theory can actually be employed for this purpose and we saw a constrained least square solution where you could throw in a prior and that that is a prior could be in terms of the Laplacian of the image, should be as small as possible for or for example, if we talk in terms of norm of the gradient of the major long x gradient, major long y and so on.

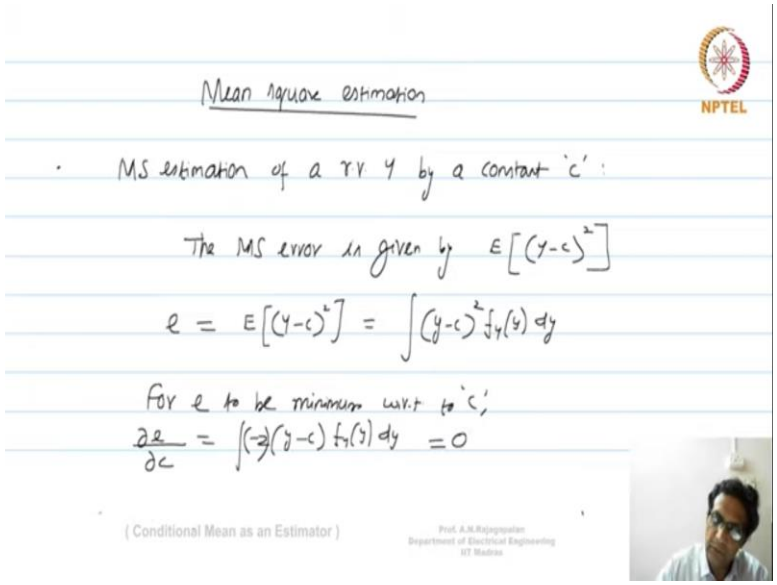
Now, we will actually turn our attention to stochastic ways of actually solving these kinds of ill-posed problems and of course it is specifically for us the image deblurring problem. We will like to see how to solve them using a stochastic framework using what I would say, stochastic regularization route and we will see that the same kind of things which we had earlier then we did a deterministic regularization in the sense that the same kind of things that we had like smoothness, prior and so on or whatever prior which we could employee.

There we will show that there is a parallel, even here when we adopt a stochastic regularization route. Moreover, the one specific filter what is called the affiner filter, which we are going to see for image deburring which is a popular filter. We will see that is a filter that incorporates the regularization notion in an implicit manner while of course, it will have stochastic estimate in an explicit manners something like a, let us say, a map estimator but then in this course should we limit ourselves to the to the affiner filter and we will see how this regularization theory that we seen earlier but then when you adopt a stochastic road, for example, when you arrive at a affiner filter, you will be able to see as to how the prior comes in and how it helps to regularize your solution and so on.

So, the idea is to embark on stochastic estimators. So, from now on instead of talking about least square, we will talk about mean square. So instead of saying least square error, until now whenever we did a deterministic kind of approach, it was always about least square error, now we will talk about mean square error. So we will talk to the mean square estimation, we will talk about estimators is that are constrained to be linear, we will talk about estimators that are non-linear in nature and so on and when a non-estimator can also turn out to be linear at what conditions and so on and how all of this leads you to the affiner filter. That is the idea.

So, let me write this as stochastic regularization. This is yet another approach to solving ill-posed problems. Problems, specifically for us deblurring problem and revolves around the notion of mean square error as opposed to least square error which we had seen earlier, means square error, MSE as it is called.

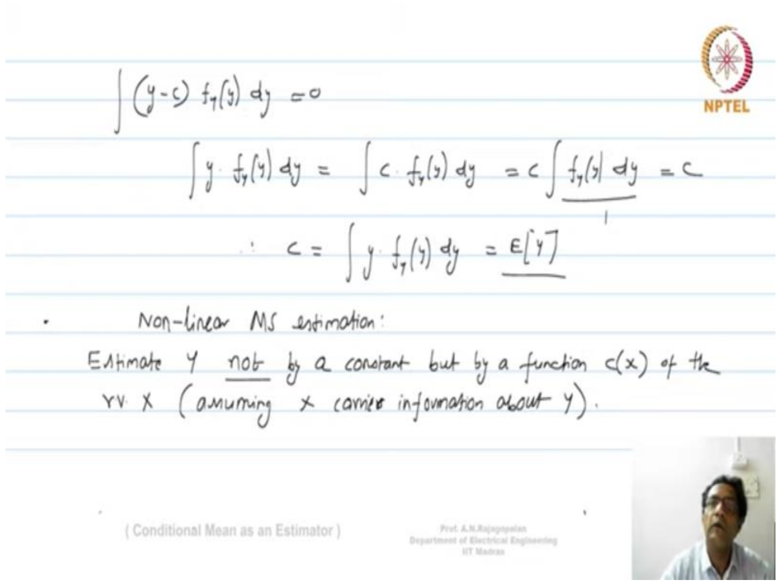
(Refer Slide Time: 05:47)



Let us first let us look at what we mean by mean square estimation. What do we exactly mean by that? Suppose we start by saying that we would like to do mean square estimation, which I would write as MS, mean square estimation. Suppose, I am interested in doing mean square estimation of a random variable y that is we are interested in an estimate random variable y but suppose by constant c this may not be the most obvious thing that you would want to do but let us say. We start with the simplest of cases, which is to say that I want to make an estimate of the random variable y by a constant c. So, the mean square error rate is in fact, given by expectation because now we are dealing with random quantities. Y is random although c is a constant expectation y minus c square.

We would like to be like to find the C such that this expectation y minus c square is as small as possible. So, suppose the error rate is indicated as e then e is equal to y minus c square which in turn is y minus c the whole square f y of y d y and for e to be minimum with respect to C, which is a constant in our case with respect to c. Let us do doh e by doh c. It will be equal to or whatever d by dc, if you wish will be equal to 2 times y minus c f y of y d y valid input minus 2.

(Refer Slide Time: 07:47)



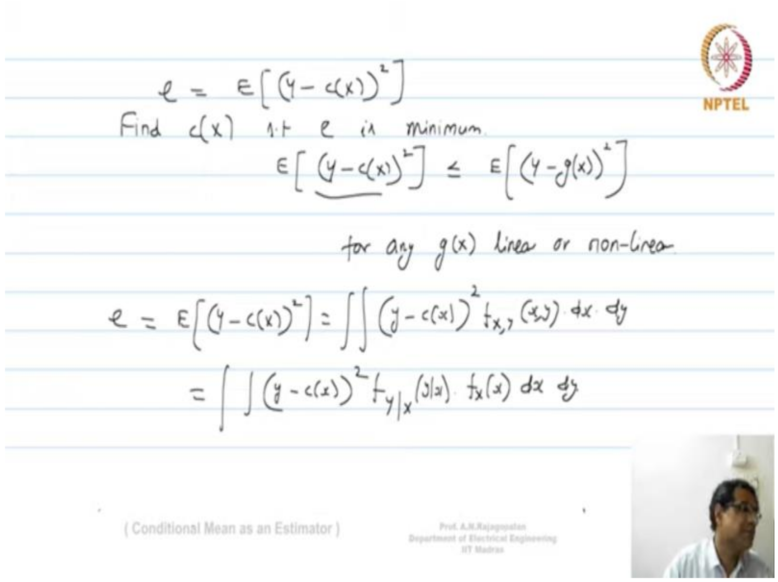
This should be equal to 0 which in turn means that integrals y minus 2 can just be thrown away and then we will have y minus c f y of y, the y is equal to 0 and this comes from here because minus 2 goes away or in other words, we will have integral y f y of y d y is equal to integral c into f y of y d y. However, because c is a constant we can pull c out and we get integral f y of y d y, but this is 1 because this area under pdf. Therefore, this is equal to c. Therefore, we get c to be equal to integral y into f y of y d y, which is p expectation of y.

So, that actually amounts to saying that in the absence of any other information, if I had to make a best means square, a minimum mean square estimator for y in the absence of any other information, but I do not observe anything about y, there is nothing else that I know then the best estimate that I can make is, the best minimum mean square error estimate that I can make of y is simply the mean of y, the mean value of y, expectation of y.

Of course, it is not really very interesting because you would not typically want to do that. What you would instead want to do is the second situation wherein let us say that we asked for a non-linear mean square estimate. Here, of course, this estimator is simply a constant. Now, let us look at non-linear mean square estimator.

We call it non-linear because it turns out that the estimation is generally non-linear. So here what we are going to do is estimate. So the problem statement changes is estimate y is not by a constant, but by a function by a function, let us say c of x of a random variable x, assuming x carries information about y because the whole idea is that we are not able to observe y directly but then through x which is another random variable and suppose x would carry information about y, then we would like to find out the optimum c effects such that expectation y minus e f x square is as small as possible. Assuming x carries information about y. So, suppose we know that x carries information about y then we would like to find c of x, which is optimal.

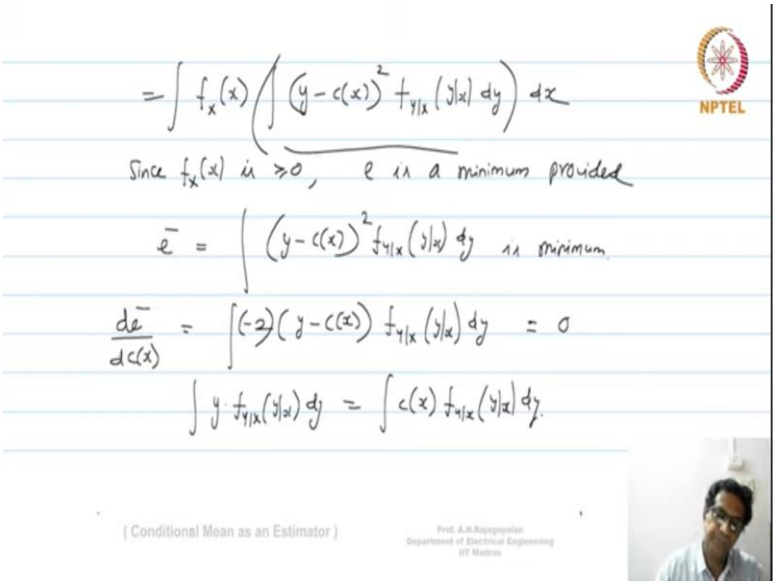
(Refer Slide Time: 10:44)



So, we can call our error now. We can redefine error as expectation y minus c of x, where x is also a random variable. Now c minus x square. Now such that so find c of x, such that e is minimum or in other words, what you are sort of saying is that expectation y minus c of x square should be less than or equal to expectation y minus any g of x square for g of x linear or non-linear is what you are asking. That means the variance that you have here, that you can compute using c of x should be such that right, this variance is less than or equal to any other function of x which you might want to take: linear or not.

Now again, let us first expand this e. So, we have e to be equal to expectation, let us first look at this as c minus x whole square which we know is integral y minus c of x the whole square and now, we need a joint pdf because of the fact that both x and y are random variables. Let us put a double integral d x d y and this furthermore, we can split this as double integral y minus c of x the whole square and this let us write this down as f of y given x, y given x f x of x d x d y.

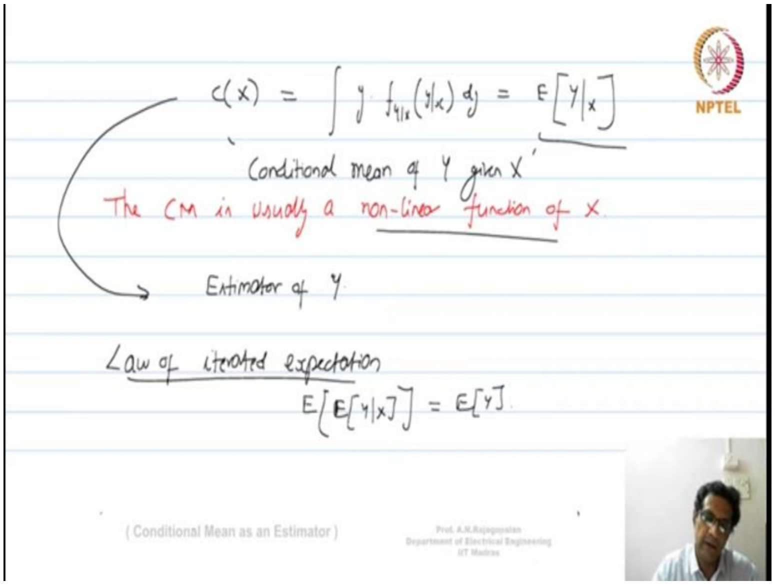
(Refer Slide Time: 12:46)



This can further be written as integral f x of x. Let us just pull this f x of x out from the second integral. Then inside we will continue to have y minus c x the whole square and then f y given x, y given x d y and then d x outside because this entire thing is of course a function of x. Now, since f x of x is always greater than or equal to 0 e is minimum, e will be a minimum provided the term inside equivalently. It means provided e bar, let us denote this as e bar, this is integral y minus c of x square f y given x d y is minimum or in other words, we can look at the e bar by d c x. I have changed e to e bar.

So, d c of x if you do then you get integral or anyway, all these integrals are typically from minus infinity to infinity unless stated otherwise. So, this would mean you will have 2 y minus c f x, minus 2 maybe then f y given x, y given x d y. This whole thing is equal to 0. So, in order to find the optimum estimate, we have equate this to 0 or in other words, we will have integral y f y given x d y is equal to integral c of x f y given x d y. This is equal to c of x.

(Refer Slide Time: 14:56)

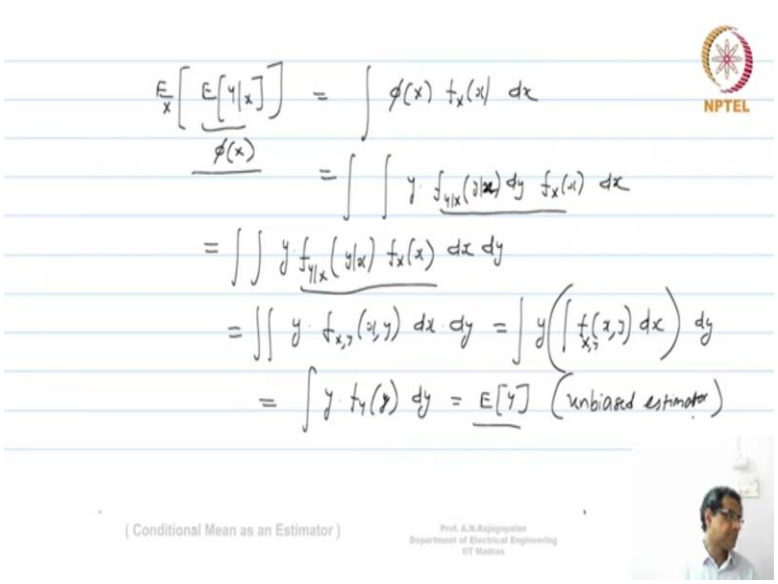


In other words, c of x as you can see is integral of y given x d y or this is also called the conditional mean of y given x. This is the conditional mean or conditional expectation or conditional mean and as it is called mean of y given x. Now, the point is this quantity, this conditional mean is typically the CM. Let us call the conditional mean is usually a non-linear function in x. Non-linear function in a function of x and which is why we call this a non-linear mean square estimation.

Now, c of x is now is estimator of y and if you substitute the value of x in this non-linear function, then you get an estimate of y for that particular value of x. So, c of x is an estimator. So, it is a random estimator and for a specific choice of x, you would get a specific value for y. Now, the conditional mean has a very very interesting property what is called the law of iterated expectation.

There is something called the law of iterated expectation. This is an important property of the conditional mean and what that amounts to saying is that expectation of expectation of y given x is equal to expectation y. We can show this. Now, the main thing that you should kind of bear in mind is that there are 2 expectations here. Remember the outer expectation is with respect to x. The inner expectation is with respect to the conditional mean of y. So, you should actually keep that in mind when you write down the integrals.

(Refer Slide Time: 17:16)

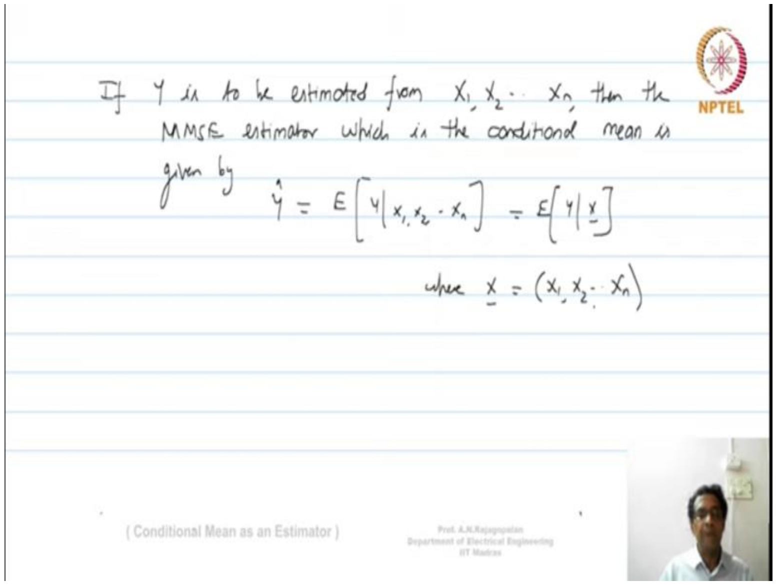


So, suppose we start from the left. We start from the left by looking at expectation y given x. So, suppose you call this as some phi of x, then you know that this is integral phi x f x of x d x since the outer guy is with respect to x and this further is equal to, now this substitute for phi of X, which is nothing but the conditional mean of y given x. It is f y given x d y, and then you have f x of x d x or this in turn becomes integral y. Now, if you were to combine these 2, let us again rewrite this. F y given x given x f of x, d x d y and these 2 will give you the joint pdf, y f x y x y d x d y and this can further be simplified as to get the marginal by integrating this joint pdf.

So, we can write this as integral, you can write is y and then we can integrate f x y with respect to d x and the whole thing d y and this in turn will give integral y and will be left with f y of y. After you integrate x out, f y of y d y that is equal to expectation of y or in other words, if you were able to look at this is called the law of iterated expectation because it is like an expectation over an expectation. The other way to kind of look at it is, the conditional expectation is an unbiased estimator.

Like I said earlier when we did image sequence filtering. At the time we had actually talked about this point that if the unknown quantity if itself is random then you have an estimator that is trying to estimate that. The estimate is unbiased if the mean of the estimator is equal to the mean of the unknown that you are trying to estimate. In this case, the unknown is y, it is random and therefore another way to look at it is estimator as phi of x. Therefore, the mean of phi of x is equal to the mean of y. Therefore, this also leads to the conclusion that what you have is unbiased estimator. Another way to interpret this result.

(Refer Slide Time: 20:17)

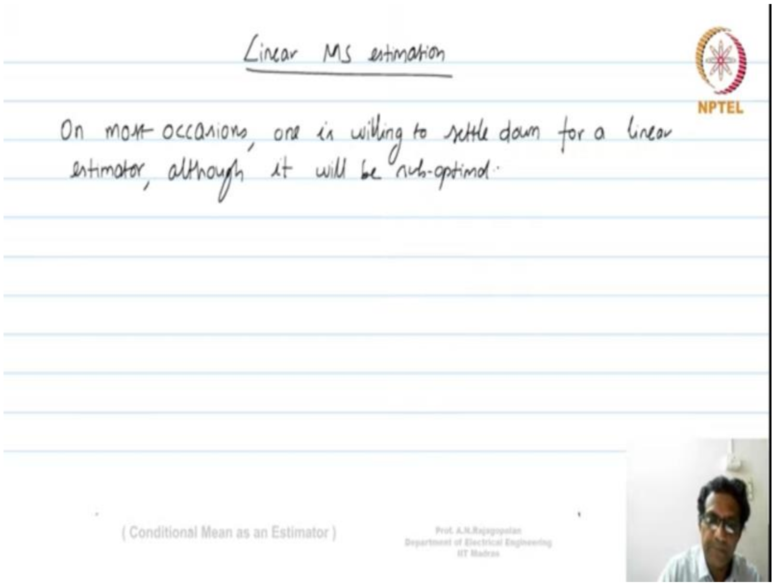


Now, moving forward, this is can also generalize to the case when instead of 1 random variable x, suppose you have the several of them and they still try to estimate a scalar random variable y. Then y is to estimate, this generalises in the following way. If y is to be estimated, from random variables x1, x2 all the way up to xn. Then the minimum mean square error estimator, minimum which is the conditional mean of course, is given by you can write this as y hat which is conditional mean as y given x1, x2 all the way up to xn or you can write this as expectation y given a random vector x where the random vector encompasses x1 to xn, where you can say x equal to x1, x2 all the way xn. All random variables.

Now till now, I have been saying that the conditional mean is in general non-linear function of x. In this case, it will be a non-linear function of x1 to xn, if it affects this depends upon multiple random variables. Then it will be a non-linear function of x1 to xn. Now, it turns out that you could have a very very interesting case where in the conditional mean becomes linear in the sense that in general, a linear estimator would be sub-optimal whereas non-linear estimator such as which is the conditional mean, minimal mean square estimator which is the conditional mean is optimum.

However, if you constrain an estimator to be linear then it is expected to be sub-optimal, but then you can show that situations can arise when under certain conditions on y and x that y and x should satisfy then it can turn out that conditional mean can turn out to be a linear function of ,x in which case a linear estimator is also the best MMS. It is not sub-optimal anymore but in general one should remember that the conditional mean is non-linear in x.

(Refer Slide Time: 22:50)



Now, let us kind of look at linear mean square estimation. Linear MS estimation. Now, as it of course clearly indicates that the linear least square estimation. So linear mean square estimation for most occasions. So, let us write this down. On most occasions, one is willing to settle down or accept for a linear estimator because we know linear estimator is something that gives you a better handle even though it may be sub-optimal but then it is something that we can understand, that we can analyse and theoretically analyse and so on.

Non-linear estimators can be can be really very very tricky to analyse for a linear estimator although, it will be sub-optimal or it will typically be sub-optimal. That is what we mean. Sometimes, in some special cases can be the optimal estimator and we would like to see, examine how to actually arrive at a linear mean square error estimator