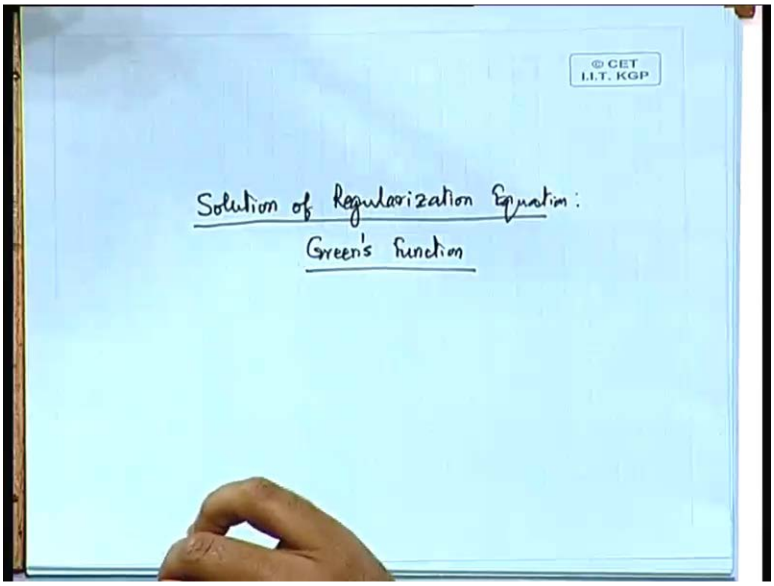
Neural Network and Applications
Prof. S. Sengupta
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 27
Solution of Regularization Equation: Green's Function

Our topic for today is Solution of Regularization Equation.

(Refer Slide Time: 01:07)



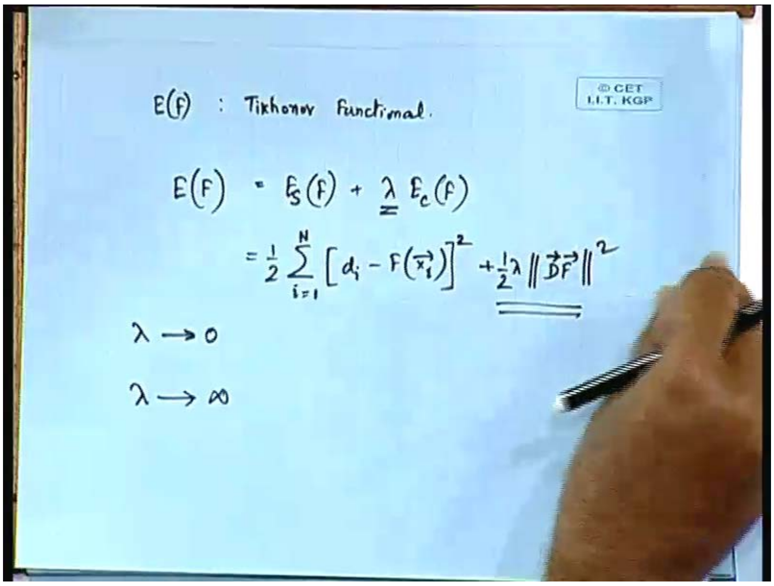
Of course, the solution is quite involved, because unlike the earlier cases where the minimization of the cost functional meant only the minimization of the square error term. In this case it is the square error term plus the smoothness or the regularizing term which also comes in to play. So, I do not think that the solution of regularization equation should be possible within just 1 lecture and we will be taking a couple of lectures for that.

So, today in connection with the solution of regularization equation, we will be talking about Green's functions, which we will use extensively for the solution of this regularization problem. Now, in the last lecture you have seen, that we had modeled the radial basis function, as network as a as an imposed surface reconstruction problem.

So, basically what we had shown in the last class is that, the problem of the surface reconstruction which we have in our hand is not well posed. It can violate the basic properties of well posed problems, that is to say the uniqueness, the continuity and the so what is the third one, existence. So, this three we showed that there are not being fulfilled by the functions, they are not being fulfilled in practical cases, when we are feeding the test patterns to the radial basis function network.

So, for that matter we also said that we will be minimizing a regularization term. And in fact, the kind of equation that we had formed in the process was like this.

(Refer Slide Time: 03:22)



That we got the quantity to be minimized if we write it as E of F, where F is the function which we are approximating. So, E of F that is to say the cost function associated with this F is equal to E S of F plus lambda times E c of F where E S of F is the standard error term. And E c of F is the regularization term and this lambda what we had is the regularization parameter and E F, there is a name for this E F which we had said that it is called as the Tikhonov's of functional.

So, we had shown in the last class that E S of F will be expressed in the standard way of half times summation of i equal to 1 to N d i minus F of x i, given that we have N number of test patterns ranging from x 1 to x N. So, i varying from 1 to N that is it and d i being the corresponding desired output. So, it is d i minus F of xi whole square which is the standard error term which is there and we summit up for all i equal to 1 to N.

So, this the standard term plus we had the lambda times E c of F, which is written as of half of lambda times D of F, the squared norm of this. Where D happens to be D also in fact, D is the differential operator. So, essentially what we have is that the regularization is to be performed, there is a smoothness term that we have introduced over here. So, if we combinedly minimized this E F that will mean, that we want satisfy the minimization of error to the best extent possible.

Also imparting the smoothness to it to the best extent possible, so we have to solve these two together, because these term added to this actually forms our functional to minimized. Now, up till here there we discussed in the last class itself, but there are one or two points that is worth noting. You see that in this equation if we make lambda tending to 0, that means what that if I make lambda tending to 0 make this lambda very small, in that case E F can be approximated by E S of F.

In fact, that is what we have been seeing all throughout, for the case of the single layer perceptron as well as for the multi-layer perceptron. Whenever we were taking the cost functional we were always taking the standard error term square. So, lambda tending to 0, obviously puts the restriction on to the data space that we have got. And with lambda tending to 0, it does not give any importance, any significant importance to the smoothness condition.

It said that whatever data is supplied, I do not care whether it is smooth or not, whether the data contains noise or not, you have to pass a surface, which should be such that, the standard error term is minimized. Whereas, the other extreme could be, that we make this term lambda very high, let us say theoretically we make lambda tending to infinite. If we do that, then we can see from this equation, by that process we make the importance of the first term, that is the standard error terms less.

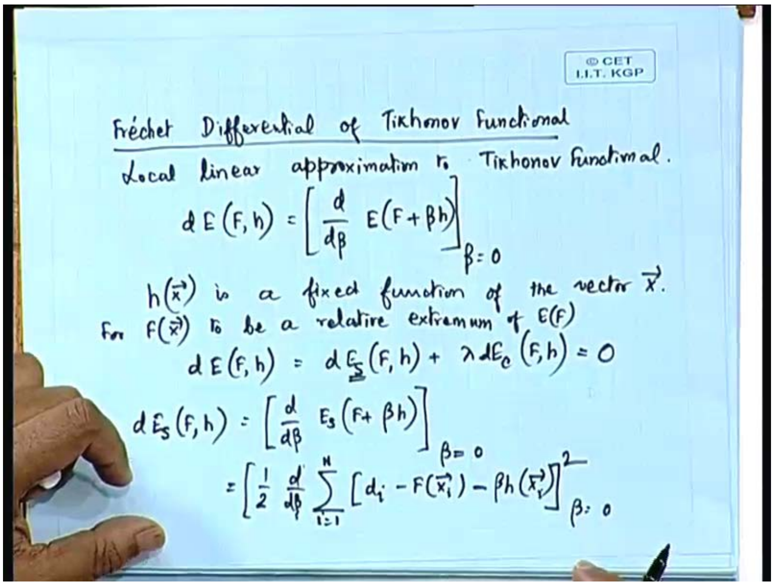
And we are putting are too much of importance to the second one, that is to say to the regularization term, which means to say that as if we are trying to say, that the data said that is given to you is unreliable. So, what you simply do is that, you simply interpolate a functional where the smoothness criteria is absolutely fulfilled, where the smoothness is there and where the minimization of this differential term takes term.

So, these two are the extremes that other way put too much of importance to the given data or we do not put any importance to the given data at all. So, in practice this lambda that we have got should be somewhere in between this two, it should not be very low. Just to make the regularization term insignificant it should be very high just to make the given data set unreliable, so we have to make a balance of these two.

Anywhere given that this kind of a functional exists with us, that is to say given that a Tikhonov functional is the one that we are trying to solve. Naturally the thing that we have to do is to determine that, what is the best functional F, our problem at hand is to determine the best functional F, which minimizes this Tikhonov functional E of F that is what our aim should be. So, the question is that if we have to do that, then surely we have to find out that, what is the differential of these E F terms and we have to minimize that differential.

So, what we do is that, in order to solve this equation, we first formulate this problem in terms of differential which we are calling as the Frechet differential.

(Refer Slide Time: 09:46)



So, we are going to write in the form of Frechet differential of Tikhonov functional. And let me explain what this Frechet differential essentially means, now just let us go back to the theory of the elementary calculus. Where you have seen that, if we have got a curve and if we want to determine the local value of the curve at a given point, what do we do. Then we consider the tangent to the curve at that particular point, where we would like to find out the value of that differential.

Or to have an infinitely small linear segment of the curve there, so the local value of the curve at that point will be determined by the tangent at that point. An likewise the local linear approximation to the Tikhonov functional will be given by, we can write the Frechet differential's definition as, the differential written as d of E. And in fact, the arguments to this we are writing not only F, but another term another function which we are writing as h and what is this h will be clear to you very shortly.

So, we are going to write this differential Frechet differential as d d beta, where we introduce new variable beta. And we take the differential or the derivative of the functional E of F plus beta h, now E of F we know. So, what we are simply doing is that, we are taking E of F beta h and we are taking a derivative of this with respect to beta. And this derivative we have to compute at beta is equal to 0 and if we compute that, then this is the local linearly approximation.

So, this is what I have written just know is the local linear approximation to Tikhonov functional. So, here actually this h of x, what we have written this h function that we have introduced over here is a fixed function fixed function of the vector h, of the vector x. So, what we have as the necessary condition for the functional F of x to be a relative extremum of E F, because after all what we are going to do is that, we are going to find an extremum of this E F functional.

So, the necessary condition for that is, that is the differential E expressed as arguments F comma h should be equal to the differential d E s of F comma h plus lambda times d E c F comma h, you understand that this is very clear. That we are writing the old equation only, that is to say E F is equal to E s F plus lambda E c F, this equation we are simply going to rewrite in terms of it is local differentials. And the local differentials are d E F for this one, d E s for this one and d E c for this one, so it is d E F is equal to d E s plus lambda times d E c.

And this differential, since we have to find out the F x which has to be an extremum of the functional F x, extremum of the functional E F. So, for we can say that for F of x to be a relative extremum of the functional E F, then the condition that we have to impart is then, this d E F is equal to expressed as the summation of this. That should be equal to 0, because we want it to reach the extremum of E F when we want to minimize this.

So, now what we can do is that, now that we have got two terms E S or rather the differential d E S and the differential d E c, we can attempt to now solve for this d E S and d E c separately. We can write down the corresponding differential expression for that, because E S we know very well that E S is expressible by this equation, that is to say half times summation i is equal to 1 to N d i minus F of x i this whole square. So, this is the expression for E S of F.

Now, of course what we have got is E S of F comma h, that means to say that we have got to find out the derivative of E S of F plus beta h. So, just to write it down d E S F h by applying the same definition, Frechet differential definition if we write down now for E S, that is to say for this if we write down Frechet differential expression. Then that will be as before that can be written as d d beta of E S F plus beta h for beta equal to 0.
Now, E S F plus beta h, now E S is what Es is d i minus F x i, so if write E S F plus beta h, in that case it will be d i minus F x i minus beta h x i sure means instead of F x i it will be F x i plus beta h x i written in this form. So, we can simply write down this as half, half obviously coming from here this half term to say half in to d d beta, so the half can be taken out of this whole expression half d d beta.

And now introducing the summation that we have got for i equal to 1 to N and then, the error functional. Now, the error functional will be in this case d i minus F x i and not only F x i, because now the argument is F plus beta h, so that is why it is F x i minus beta h x i, this thing square which we have to calculate at beta equal to 0. H is non function, h is a fixed function of the vector x we have all ready said that.

Student: ((Refer Time: 18:15))

Square will be inside the bracket

Student: ((Refer Time: 18:19))

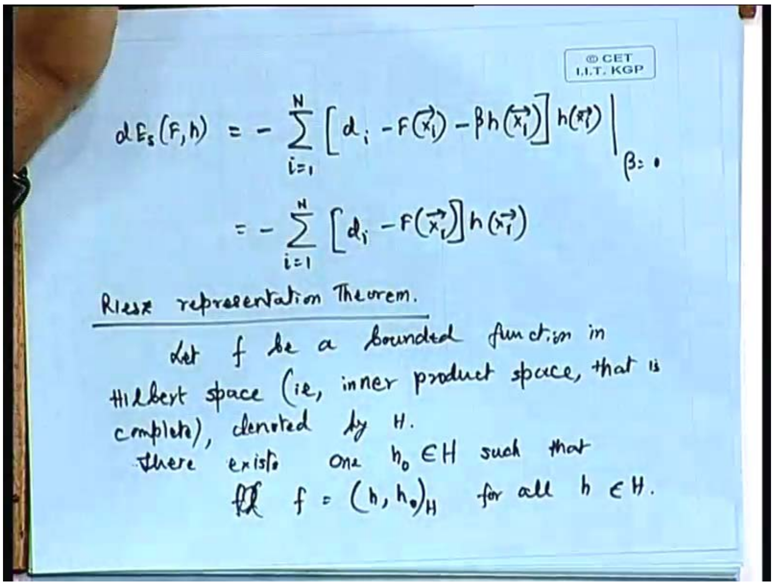
There is one bracket yes, you are very correct there is one more bracket, because this square bracket was corresponding to the summation terms, this terms argument that is the error terms, square bracket ended here we are taking the square of that. And then, this square bracket is for the overall term and there we have to compute it for beta equal to 0.

Student: ((Refer Time: 18:53))

Fixed function, you see that is to say this is from the theory of calculus itself that what we are doing is that we are finding at this point F. We are finding the value of F plus beta h, that means to say that we are going to an incremental step of beta and h is a functional there h is a fixed function there. So, what we are doing is that, just like the way whenever you are trying to find out the functional value of a curve at a point beta.

What you do F of x you are going to find out you are going to find out F of x plus h beta, very similar way, in this case you are not doing it for the curve, you are doing it in the vector space. So, that is the purpose of this fixed function and now it is just written in this form, so I think up to this we do not have any problem. So, simply it can be simplified further as a matter of fact that, if we take the d d beta of this, if we differentiate this with respect to beta, then things will turn out to be pretty simply.

(Refer Slide Time: 20:20)



What we will be getting there is that, if I write down then you can verify that later on, this becomes equal to minus of i equal to 1 to N. This half terms obviously goes out, because once we are differentiating with respect to beta, because there is a square term over here, this half term gets canceled in the process. And this becomes a form like this summation or d i minus F of x i minus beta h x i.

And this we have to multiply by h of x i, this has to be multiplied by h of x i and this we have to compute at beta equal to 0. Now, this is very clear that why h of x i will come because, you see that there will be in fact, a term which is associated with this h. So, there is beta h terms, so obviously, what we have to do is that it is a square of this and differentiated with respect to beta. So, this obviously, will mean that there is an h x i term which also comes.

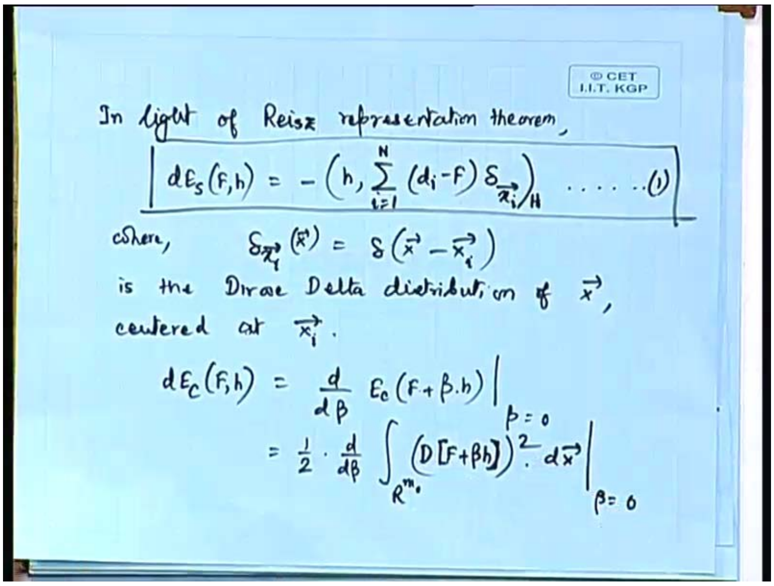
And now computing this for beta is equal to 0 is pretty simple what simply we do is that in this expression if we put beta is equal to 0. Then what results is simply minus summation of i equal to 1 to N d i minus F of x i multiply by h of x i, so this is a good form that we have got that is to say the computation of d E S F h. Now, this in fact, if you look at it this is going to be what, is this a vector or a scalar, this incidentally happens to be a scalar quantity.

Because, d i is a scalar quantity F, the function is a scalar ultimately and h of x i, so here this can be represented simply, this product that we have got it can be represented as inner product of two quantities. And to express it that way, we are making use of a theorem which is called as the Riesz representation theorem. And what that means, is that let f be a bounded linear functional in Hilbert space. And what is Hilbert space, Hilbert space means that for which the inner product space that is complete.

So, the inner product space that is complete, so and to we are denoting it by H, the Hilbert space is denoted by H. Then it says that there exists one h 0 belonging to this H space such that, f can be represented as h comma h 0 in Hilbert space H for all h belonging to Hilbert space H. So, that means to say let us not get very much confused with this terminology this term h comma h 0, what is simply written is the inner product in the Hilbert space.
So, what it means to say is that, this function f that we have got is expressible as an inner product of these two functions h and h 0 which exists in the Hilbert space. So, we are just going to write the functional as an inner product in it is Hilbert space h comma h 0. So, if as the function we take this differential d E S in this expression, if we go back to this expression and try to apply this Riesz representation theorem. Then this d E S for us should be expressible as what as an inner product of h and this summation term.

In fact, in terms of this Riesz representation theorem, the Frechet differential of this d E S term can be rewritten as follows.

(Refer Slide Time: 25:51)



So, in light of Riesz representation, it is possible for us to write d E S F comma h is equal to minus, here minus is there obviously and this is a product of these two, so we are just expressing it as a dot product. So, this can be in the Hilbert space written as h comma summation i is equal to 1 to N d i minus F, we are not writing F x i again and again. So, it is d i minus F and what we have to do is that we have to write it as this times delta x i, where delta x i is the Dirac delta.

In fact, this inner product is written in the H space in the Hilbert space and here, this delta x i what we have got of this x, this is equal to delta of x minus x i, this is the Direct delta distribution of x vector centered at x i. So, this means to say what that only when this x vector is equal to x i vector, in that case the Dirac delta is equal to 1 and otherwise, when x i vector is not equal to x vector, then that is equal to 0.

So, it is the Dirac delta distribution definition directly, so this term of the Frechet differential of d E S is represented as an inner product of these two in the H space that is the Hilbert space. And now let us us go over to the second term of this, because this was the way we represented d E S and now we have to represent the next term that is d E c. So, to represent d E c we should follow a similar approach, in fact d E s by applying the definition of the Frechet differential directly, where d E F is equal to d d beta of this.

If we have to compute d E c of F comma h is has to be d d beta of E c of F plus beta h commutated at beta equal to 0. And then, we will make use of the expression for E c that we have got, in fact expression for E c is this term half of norm of D F square, this is the term that we have got as a definition for this E c F that is to say the regularization term. So, now to find out the Frechet differential of the regularization term, we proceed as follows.

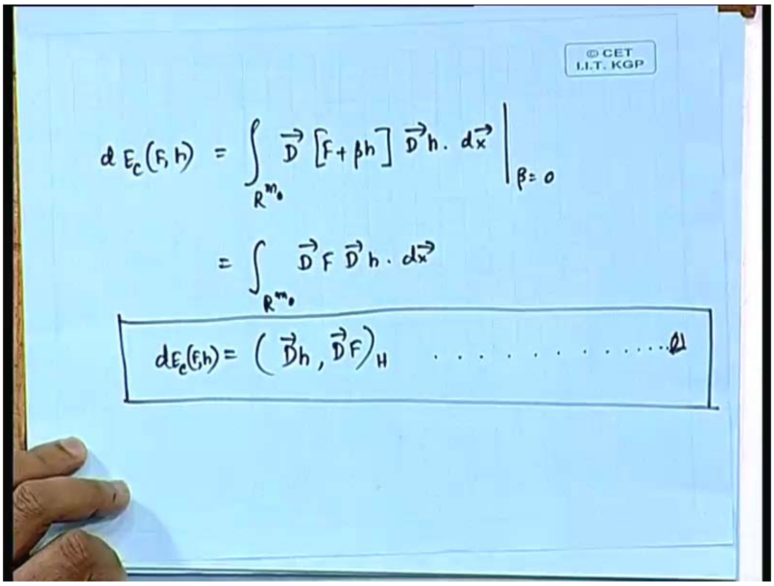
So, the evaluation of Frechet differential for the regularization term gives us d E c of F comma h should be equal to d d beta of E c F plus beta h. And this is to be evaluated at beta equal to 0, which means to say that this is equal to half of d d beta. Now, we apply the definition of this E c and what is the definition of the E c that leads to, what we have to do is simply take the, this is we said that it is d F, the differential of this F, in fact this differential is in the multidimensional space.

So, this is the multidimensional differential operator that we are taking over this F. So, this is this is wrong I should not say D F vector this is D is a vector quantity no doubt, because it is multi dimensional differential operator. So, here in fact, for this thing d d beta of E c this could be written as half of d d beta instead of E c, we can express it as an integral that is defined in the m 0 dimensional real space. The input there are m 0 number of input, so naturally the input space itself is m 0 directional.

So, what we have got following the definition of E c itself, that it can be written as the differential of F plus beta h F plus beta h this argument F plus beta h. And we have to say this argument square of this differential of this differential of this F plus beta h whole square, this is from the definition of E c itself that we have got. And we have to integrate this in the multidimensional x space R m 0 space.

So, this times d x and we have to add it over the, we have to integrate it over R m 0 space and of course, everything to be computed at beta equal to 0. Again we have a half term over here and within this we have got a differential with a square term. So, in effect these ultimately cancel out and what we get as a simplification to this expression this Frechet differential definition of this regularization term.

(Refer Slide Time: 32:46)



That leads us to this can be written to that is to say, the d E c of F h F comma h, this can be written as the integral over this R m 0 space of D F plus beta h. And we are going to have in fact, D with the vector sign that we need to do and here, this times D h, this is coming from the differential itself, because we are taking that and this D h d x and this we have to compute and beta is equal to 0.

So, which means to say that this argument which is now F plus beta h, that will become F, so this simplifies further to R m 0 integral D of F times D of h d x. And this integral is one and the same as writing that in Hilbert space, we could express as the inner product D h comma D F in H space. And we can call this as the equation number 1 let us say, so what we did so far again, so that to keep in tune with what we have been doing.

Again we have taken the, we have resorted to the calculation of this d E of F h, the differential of the Tikhonov functional, expressed this as the summation of these two. And individually this d E S and d E c we have expressed and the form which we choose in order to express that is the inner product form. So, E S of F h has been expressed as the inner product form like this. In fact, let us call this as equation number 1, because this equation will be of use to us.

And the second one that is, say d E c what we have got at the end d E c of F comma h this to be equation number 2. So, these two are quite important to us and based on this two very important relation, we will now go further to have a first, we need to have a proper representation of the equation of the regularization equation in the differential form. And then, only we will try to solve that, yes please any questions at this stage.

Student: ((Refer Time: 35:39))

This one

Student: ((Refer Time: 35:45))

Here

Student: ((Refer Time: 36:03))

From here to here

Student: ((Refer Time: 36:08))

Now, beta is 0, anyway this you can verify this is an established result, so I think with a bit of manipulation or with the differential computation we can get it, it is not difficult to get one. So, this is the differential form that we get, this is an established equation, so...

Student: ((Refer Time: 36:47))

1 minute this one

Student: ((Refer Time: 36:57))

You see, what we are doing is that, what is E c's definition, E c's definition says that it is differential of F, now we are computing the differential of that at the point beta equal to 0. So, what we have to do is to consider the function not as F, but as F plus beta h with beta is equal to 0, so we have to consider the function there, so it is the differential of that squared.

Now, this differential that we are computing, this D of this square that we are getting is in the unit input space. So, what we have to do is to take the differential of the input space which is given by this d x and then, integrate it over the input space itself this R m 0 is the input space directly. So, that is how we got this overall integral term. And then, what we are getting is this d d beta term that goes out for us.

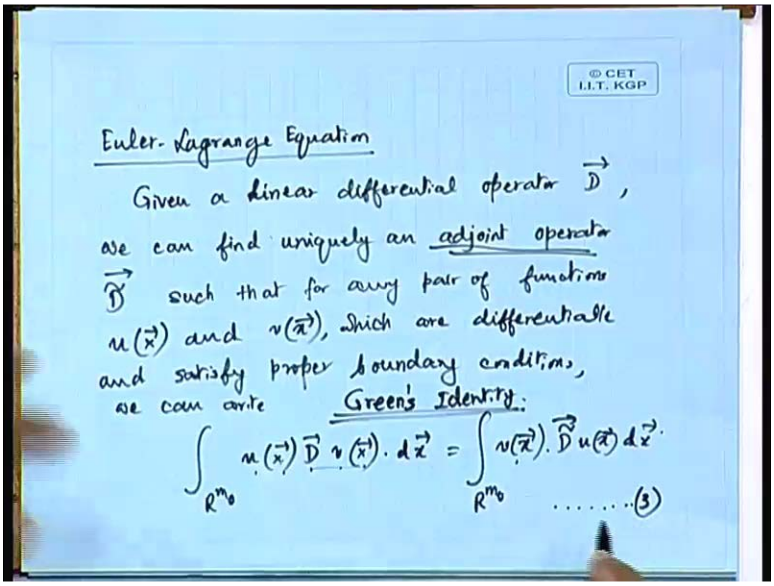
In fact, just a minute why are saying that there will be a beta square term out there I think beta square

Student: ((Refer Time: 38:39))

Beta is 0, so ultimately we are getting this result, so D h D F, so inner product of that, so that is important. So, we have got d E s and d E c both representable in the inner product form. Now, our objective is to write this differential in a manageable form, first of all that we have not been yet able to write it down, our functional minimization that we are going to achieve is not yet written in a manageable way.

So, we are going to write it in a more manageable way and for that, we make use of a popular equation, which is called as the Euler Lagrange equation.

(Refer Slide Time: 39:36)



And what does the Euler Lagrange equation has to tell for us, that is to say that given a linear differential operator D. Operator D we can find uniquely an adjoint operator D we can write it is as D tilda, we call this as adjoint operator D tilda such that, for any pair of functions we take a pair of functions. Let us say the functions that we take are u x vector and v x vector, which are differentiable and satisfy proper boundary conditions.

Then, we can write integral over R m 0 space u x vector D of v x vector, we take two functions pair of functions u x and v x. What property does it have to fulfill, you will see that u x times differential of D v x, in d x is equal to integral over R m 0 space v x D tilda of u x d x. So, you see just note the rhythm of this equation you see here u x is there, and the v x is associated with the differential operator D it is the derivative of v.

And this is equivalently represented as you can write it as v x and take the differential of u x, but not the exact differential, you take the what you call as you can treat it as an inverse differential kind of thing. This is D tilda it is called as the adjoint operator, in fact this is the way of defining the adjoint operator in terms of the differential.

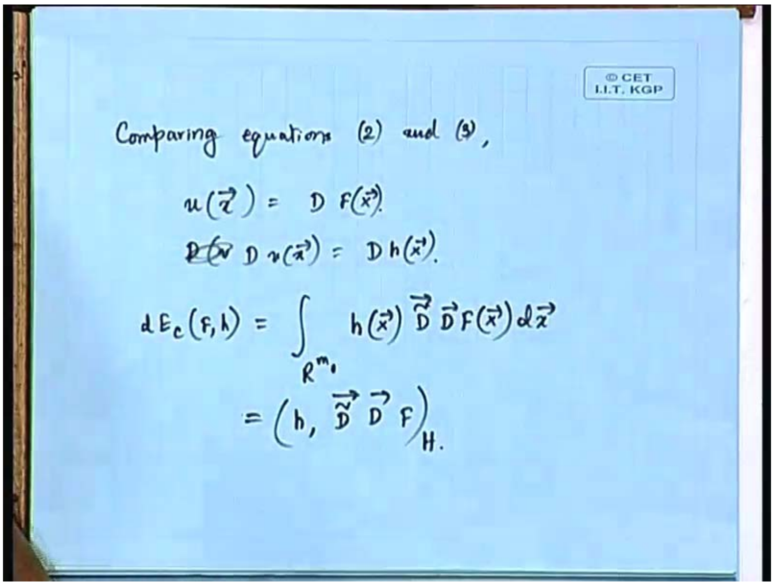
And in fact, it is if you are treating the D, the operator D as a matrix form in the matrix form if you treat the operator D. Then D tilda essentially means the transpose of that matrix D tilda effectively becomes the transpose of that matrix. So, u x D v x written in equivalent form as v x D tilda u x, so it has got matrix analogy it is like this D operator is, this D operator is just like an inverse operator that is happening.

Now, this equation is very popular one, in fact this is called as Green’s identity, so this is known as Green’s identity, so please remember this. So, those who have taken a course on this, on the partial differential equations etcetera, must have come across this Green’s identity, this is a very popular one where it is expressible this way. Now, we can let us see that if we can use the Green’s identity to our best use or not.

Now, let us have a look at the differential that we have got for E c, the differential that we have got for E c simply said, that is the one d E c F h is equal to integral over R m 0 space of course. D F, D h, the D F D h means that if we want to write it in the form of the left hand side of Green’s identity, you see now I am putting this two equation together, the left hand side of Green’s identity u x, D v x, d x.

And now look at this d E c, d E c says D F D h d x as if to say that h here plays the role of v or rather to say D h plays the role of D v and in that case, u x here plays the role of D F. So, now from these two things that means, to say if I call this as equation number 3, Green’s identity if I call as equation number 3, then comparing equation number 2 and equation number 3.

(Refer Slide Time: 45:58)



Comparing equations 2 and 3 what we get u x is equal to D F of x and D v of x is equal to D h of x. So, if I can draw an equivalence from the left hand side of Green's identity, then I can express this in terms of the right hand side of Green's identity. And what does the right hand side of Green's identity say v x and what is v x in my case h of x. So, this can be written as rather to say d E c of F h can be written in terms of Green's identity as R m 0.

Yes please, please dictate to me, then only I will know that you have understood this h of x yes

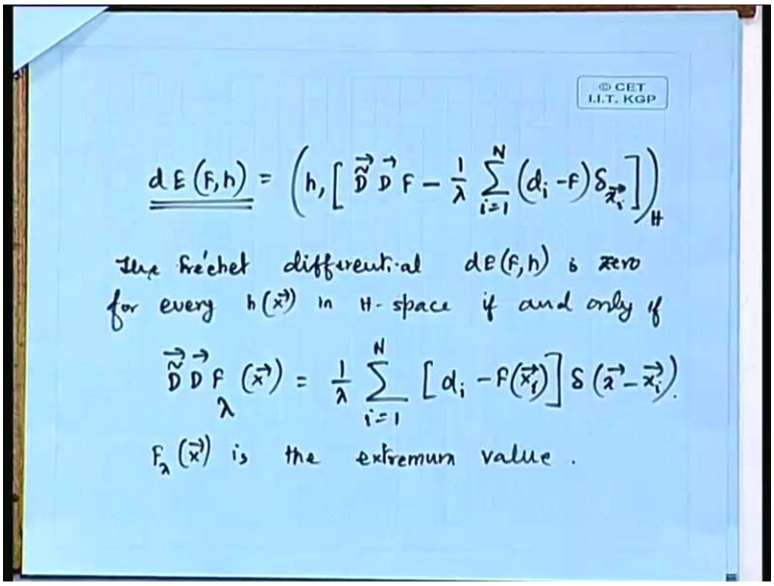
Student: ((Refer Time: 47:15))

D tilda D of F x, why D tilda because it is to be expressed as D tilda u x and what is u x u, x is our case is D F, so it is D tilda D F, so it is h x D tilda D F of x, d x. So, this is equivalently in the inner product form it is expressible as h comma D tilda D F in Hilbert space H, so that is it. And now if we return to the extremum condition, in fact this is nothing but, representation of d E c of F h in different page.

We earlier expressed it as a as the inner product form only, but it was D h comma D F and now we are writing it as h D tilda D F that form in the Hilbert space. Now, since this is the representation of E c, we can equivalently now find out the differentiation. Now, we our aim is to find out the differential of E, not the differential of E S and E c alone, so we have all ready got differential of E S and E c and what we now have to do is to add this E S and E c terms together.

In fact, the addition is of course, through the use of this lambda, it is d E is equal to d E S plus lambda d E c. One and the same as saying that it is if I say 1 upon lambda d E S F h plus d E c F comma h, we can say that lambda is not equal to 0, in that case its possible for us to right it as d E c plus 1 upon lambda d E S. And if we do that then the d E F h can be combinedly represented as the inner product form as follow. So, a combined representation of d E F h in the inner product form is given by this.

(Refer Slide Time: 49:57)



D E F comma h is equal to h comma, we have D tilda D F, now this D tilda D F we can very easily recognize that this D tilda D F is coming, because of our regularization term the regurgitation term here that is the E c term is contributing to D tilda D F. So, h comma D tilda D F, but here we have not only the regularization term, but also the standard error term. And standard error term also has been expressed equivalently as an inner product.

So, what is it that we had got, we had got there as inner product look at the expression h comma this, h comma summation this, so we have to write it as minus 1 upon lambda of this, h comma 1 upon lambda of this. So, effectively it is D D F minus 1 upon lambda summation i equal to 1 to N d i minus F delta x i again delta x i is already defined, it is the Dirac delta. And it is inner product of h with this term and this term takes care of both, this D tilda D F takes care of the E c term and 1 upon delta of this takes care of the standard error term.

And together for both we have got it as inner product with h, so combinedly as inner product of h it is h comma this, so this whole thing should be an inner product in the Hilbert space. Now, what we have to do is that, the Frechet differential d E F h is 0, for every since we are looking for the extremum of this, this term has to be 0, this differential is 0 for every h of x in H space.

So, we can write that the Frechet differential d E of F h is 0, for every h in Hilbert space if and only if, what is the condition if for every h it has to fulfill, what should be the condition to make d E F h equal to 0. This is already written in the inner product form, that means to say that the second term which we have got as the inner product expression has to be equated to 0 simply.

So, simply we have to write it down as D tilda D F is equal to minus this is equal to 0, means as if to say that we have got that equivalently D tilda D F. D tilda D F of x equal to 1 upon lambda summation i equal to 1 to N d i minus F of x i times delta x i means it is delta x minus x i. And in fact, here we are writing it as F and what is this F, this F is the F that we are looking for, that is the F at the extremum, because only at the extremum we are equating d E to 0.

So, at the extremum we are writing it as F lambda, so where F lambda is the extremum value, so F lambda of x is the extremum value. Now, what have we got, we have expressed now the functional in terms of this equation, but we have to now solve for F of lambda which is still not very easy. Because, this involves the derivative terms, differential terms are involved associated with this.

So, what we have to do is to solve using integral transform methods and we do not have time to cover that all though the title of the lecture said that, I wanted to introduce Green’s function ultimately I could not, because of the shortage of time for today. So, in the next class I will be introducing the Green’s function which will be needed for us to solve this regularization equation.

Thank you very much.