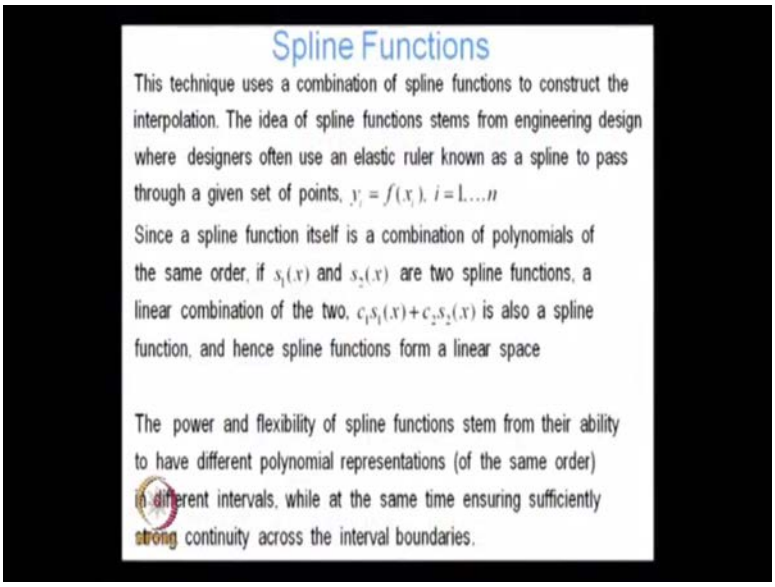
Numerical Methods in Civil Engineering
Prof. Arghya Deb
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture - 36
Orthogonal Basis Functions for Solving PDE's

In lecture 36 of our series on numerical methods in civil engineering we will end up our discussion on spline functions, which we started last time and then move on to a discussion in orthogonal basis functions for solving partial differential equations.

(Refer Slide Time: 00:38)



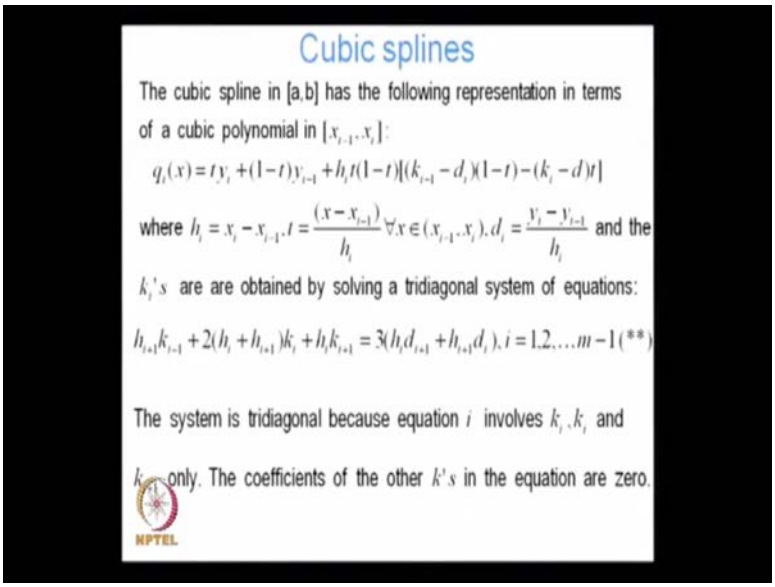
Let us recall that spline functions are another set of basis functions, which can be used for constructing an interpolation. This technique uses a combination of spline functions of combination of functions, and the idea of spline functions we found stems from engineering design, were designers often used in elastic rule as a known as a spline to pass through a given set of points and people found that the spline was nothing but a series of cubic's defined over the intervals of interest.

If I have several grid points and I know the function values at those grid points. The final fit was obtained by considering cubic functions, cubic polynomials over each of those intervals and then ensuring continuity in the function values, as well as in the derivatives across the element boundaries not only the first derivative, but also the second derivative.

In case of cubic splines and these basis functions are independent. So, the different cubic polynomials, different basis functions in different intervals only difference with the type of basis functions at we were considering earlier. These basis functions have local support. So, they are defined over small intervals not over the entire domain.

And the spline functions are combination of these polynomials in the power and flexibility of spline functions stem from their ability to have different polynomial representations in different intervals while at the same. So that is why you can basically modify a complicated function with lot of variation using different functions in different intervals, you do not have to use a very high order. The other alternative is to use a very high order polynomial throughout the interval but that is no longer required. So, you use a relatively low order polynomial in case of cubic splines a cubic polynomial over these intervals and they are different over each interval.

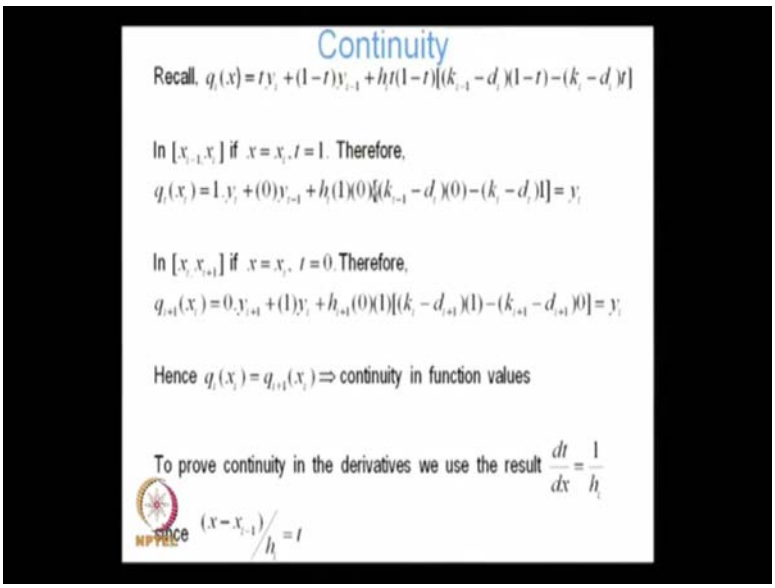
(Refer Slide Time: 02:48)



And we also found that, the cubic spline has this representation. So, if I define the spline over the interval x I minus 1 x i and denoted by q y, I denoting this interval i minus 1 i, then that cubic spline has the following representation, where t is nothing but x minus x i minus 1 by h i. So, it is the parameter over that interval x i minus 1 x I for all x belonging to that interval h i is just the size of the interval d i is the slope slope y i minus y i minus 1 by h i and the k i's the k i's in this equation. In this expression for the cubic polynomial are obtained by solving this tri diagonal system of equations.

And towards the end of the last class, I told you that this k i's are nothing but the second derivatives of this q I and this system is to. So, what we find at the end of last class was that given this representation of q given this representation of the cubic polynomial. It satisfies continuity in the function values and continuity in the first derivatives by construction the way it is constructed. It always satisfies continuity at the ends of the interval both in the function values as well as in the derivatives but in order to ensure that the second derivative is also continuous across the intervals. The k i's have to satisfies equations of these sort and you can see that this system. So, there is m minus 1 equation but there are m plus 1 m plus 1 k i's. So, because of that we need two other constraint equations, which we are going to talk about later.

(Refer Slide Time: 04:44)

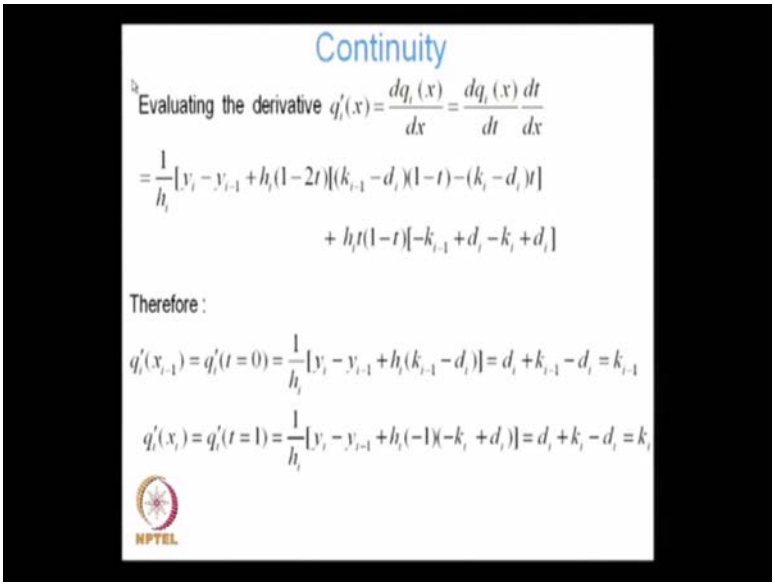


And from as I said that given the definition of y continuity in the function value is satisfied by construction. So, since in x i minus 1 x i, if x is equal to x i, this means t is equal to 1. So, t goes from 0 to 1 throughout the interval. So, if I am looking at the hand point at x is equal to x i at this point, when t is equal to 1. If I substitute t equal to 1 in this equation I can find that, this gives me q i x i is equal to y i. Similarly, if I am looking at the interval x i, x i plus 1 the adjacent interval to that interval if x is equal to x i t must be equal to 0, because it is at the starting point of the interval the left hand starting point.

So, at is equal to 0 if I substitute t is equal to 0 at that interval in this equation and evaluated at x i i instead of q i. I will use q i plus 1 and evaluated at x i, I am also going

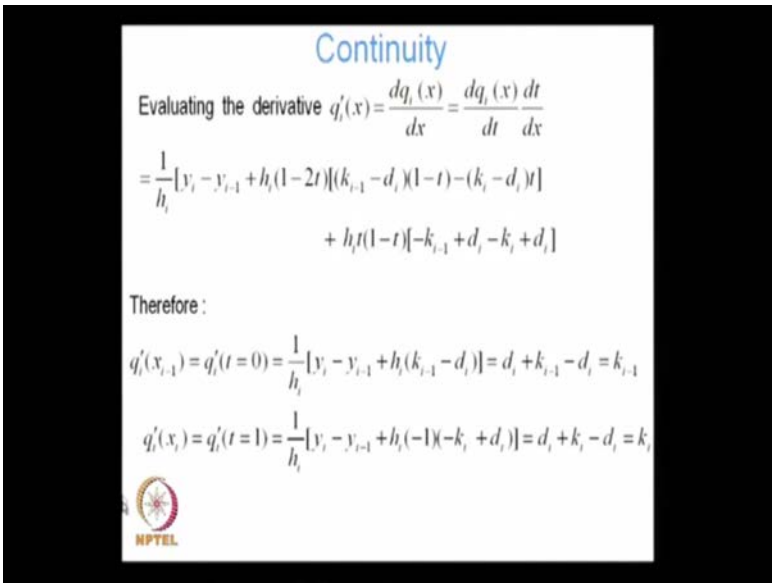
to get y i that means, what the function values q i x i is equal to q i plus 1 x i, that means, it is continuous at the interval of the function values of are continuous at the interval boundaries to prove continuity in the derivatives. We use the following result d t d x is equal to 1 by h i and try to find the derivative of this expression. I try to find the derivative of that expression.

(Refer Slide Time: 06:23)



So, I find q i prime x, while prime x instead of now t is the independent variable. So, d q i d x i want to find but then I have used this intermediate parameter t. So, I use d q i d t, d t d x. d t d x is equal to from this expression d t d x is equal to 1 by h i. So, I have 1 by h i times d q d x, d d t of this expression. I have d d t of that expression. So, I get that and then if I want to evaluate q i prime. So, I look at the interval x i minus 1 x i and I want to evaluate q i prime at x i minus 1. Our q i prime means the derivative, if this function is belongs function cubic function with respect to this interval x i minus 1 x i.

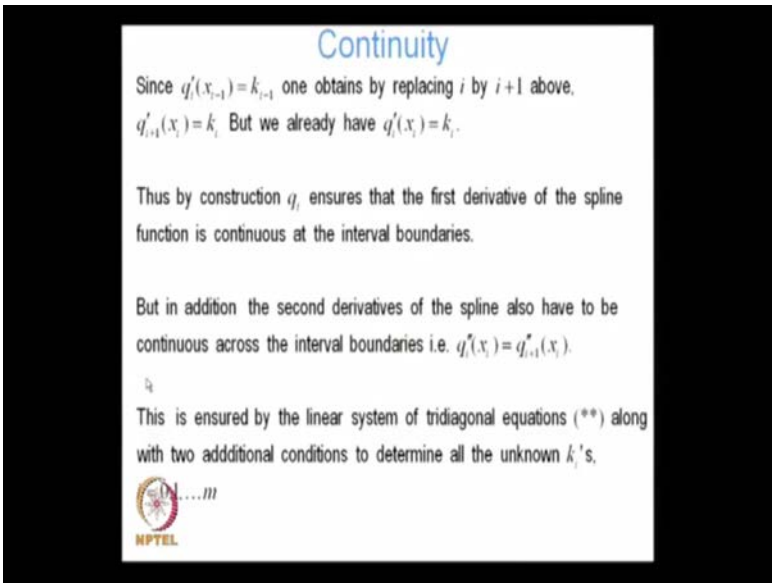
(Refer Slide Time: 07:24)



And I am going to evaluate it at x i minus 1. So, that means I evaluate it at t is equal to 0 because the interval is x i minus 1 x i t is equal to 0 at x i minus 1 t is equal to 1 at x i. So, I evaluate this at t is equal to 0 if I do that I get k i minus 1. So, i just substitute t is equal to 0 in this expression then if I want then I want to evaluate q i prime at x i.

So, I am using the same interval x i minus 1 x i and I am evaluating the derivative at x i that means I am evaluating the derivative at t is equal to 1, if I do that I substitute t equal to 1 in this expression I get is equal to q i that equal to k i.

(Refer Slide Time: 08:15)

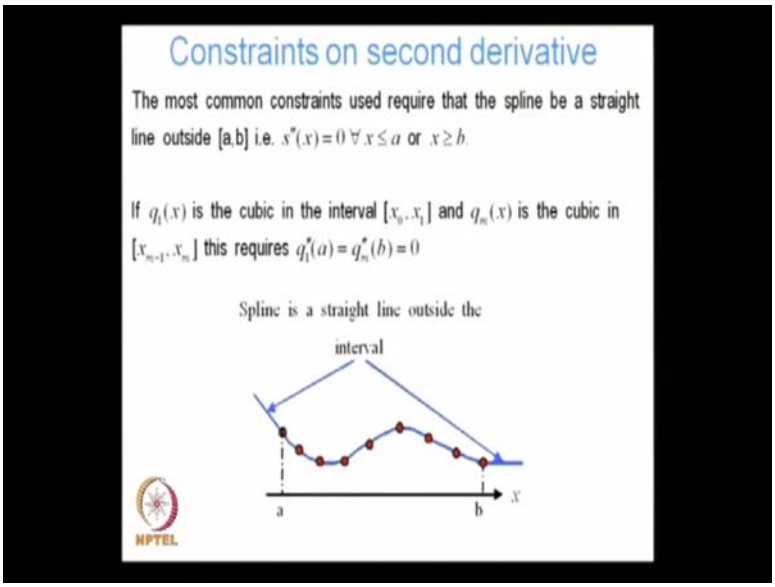


Then, I use the fact that since q i prime q i prime at x minus 1 is equal to k i minus 1 if I want to evaluate if I replace i by i plus 1 I can get q i prime q prime at i plus 1 just replacing i by i plus 1 in this expression x i is equal to k i. But, we have already found that q prime at q i prime at x i is equal to k i. So, that means for the interval x i minus 1 x i, if I evaluate the derivative at the right hand point, I get k i. For the interval x i x i plus 1, if I devaluate the derivative with at the left hand n, I get k I that means, that the derivatives are equal at the boundaries and therefore, we have continuity in the derivative as well.

Thus by construction q i is the construction, q i has continuous function values and derivatives of the spline function at the interval boundaries but in addition the second derivatives of the spline also have to be continuous across the interval boundaries, that means, this condition has to be satisfied at the end of interval x i minus 1 x i. The second derivative must be equal to the second derivative at the beginning of the interval x i x i plus 1, which is this condition. And then if I impose that condition on this equation.

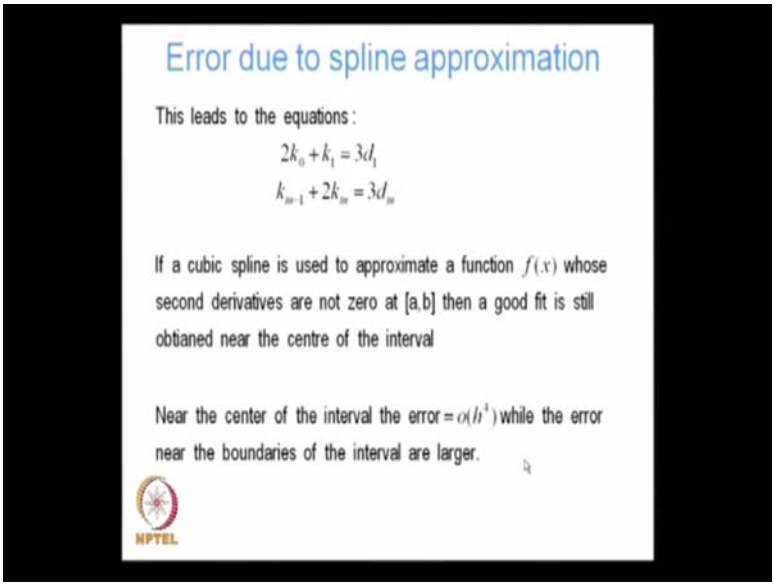
If I use this equation, I put I impose that condition. That is the second derivatives are continuous at the interval boundaries on this equation. I get these equations here, these diagonal system and then I need to solve that tridiagonal system. But, as I said we need two additional equations but because there are fewer equations than the number of unknowns. So, we need to solve that system with constraints.

(Refer Slide Time: 10:22)



And the constraints we use require that the spline be a straight line outside the interval. If it is a straight line, that means it is curvature got to be 0, curvature got to be 0 that means that the second derivative got to be 0. So, as double prime x is equal to 0 for x lesser than or equal to a and x lesser than or equal to b. So that means we are looking at the first interval and the last interval. So, if q 1 x is the cubic in the first interval and q m x is the cubic in the last interval, this means that q 1 double prime at a must be equal to q 1, q q m double prime at b because q m refers to the last interval and both must be equal to 0 both must. you can see these are spline's straight line outside that interval. What happens?

(Refer Slide Time: 11:21)



So, these conditions give me, these two additional equations you can see that, the first equation involves k 0 and k 1, because it repels to the first interval the last equation involves k minus 1 k m, because it repels to the last interval and these two are the additional constraints. Now, I have m plus 1 equation for my m plus 1 constraint. I can solve them, I get my k values and once I know my k values, I can split my I can fit my spline throughout the interval.

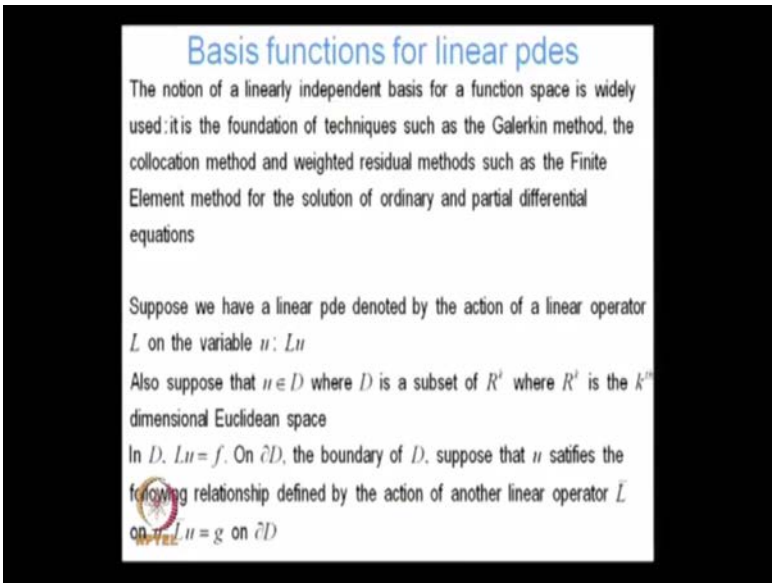
So, it is not expensive because the system as we have seen earlier is a tridiagonal system solving a tridiagonal system is not that hard. Now, we have said that we are going to use a cubic spline to exactly fit the function values but suppose in that interval but suppose my actual function does not have 0 curvature, then my spline has got 0 curvature, then

what happens if my actual function does not have 0 curvature, then in that case there is some loss of accuracy the fit is not as good. So, the good fit is still obtained at the center of the interval but towards the end of the interval the errors are somewhat larger. So, near the center of the interval the error is of the order of h to the power 4. So, you can see the error is really, it is the order of accuracy is really high. So, h to the power 4 if I as you reduce the size the error becomes goes down like (( )) where the error near the boundaries are somewhat larger, if the real function does not have 0 curvature at the ends of the interval.

So, with that I want to draw close to my discussion of spline functions. So, these are very useful functions, because they have this. They give me very good accuracy recall, if I have a polynomial with a lot of with a maxima and minima in order. If I have a function with a lot of maxima and minima in order to have a polynomial maxima and minima in a certain interval lots of stationery points within an interval. If I have to use a polynomial to fit that function, it has to be a very high order polynomial and as we have seen with a high order polynomial, if we have equidistant grid we are going to get erratic behavior ill-conditioning and we. So, poor fit at the ends of the interval.

So, alternative to do that, we found that were one alternative was to use chebyshev interpolation, other alternative is to use splines, where you use a relatively lower order polynomial do not have to use a very high order polynomial use a relatively lower order polynomial fit it across that interval which assumes different forms within each subinterval same order but different form and then you can get a good fit even using a equidistant grid.

(Refer Slide Time: 14:37)

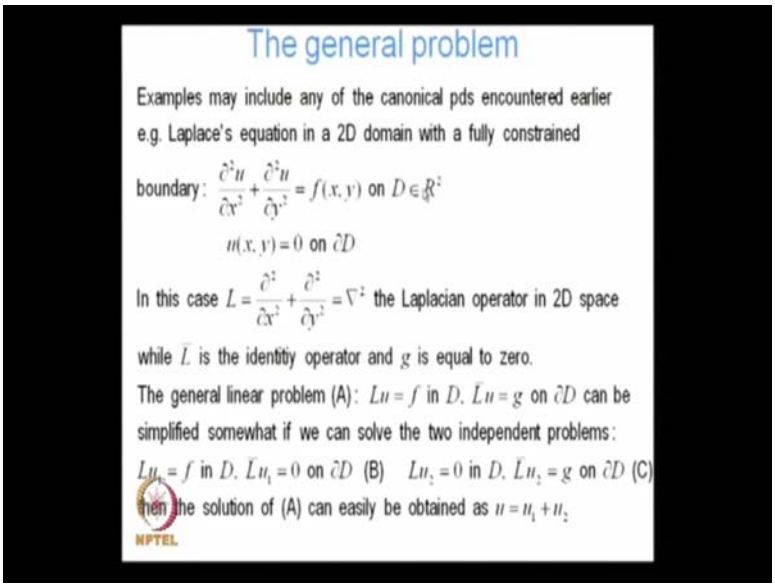


So, with that we are ready, we have talked enough about orthogonal basis functions. So, we want to use these basis functions for solving our partial differential equations and we can say, we will see that, this idea of orthogonal basis functions is the foundation of most of almost all widely used methods used for the numerical solution of partial differential equations including the finite element method. So, let us do that.

So, notion of linearly independent basis functions for a function space is widely used. It is the foundation of techniques such as the galerkin method, the collocation method the weight and weighted residual methods such as the finite element method for the solution of ordinary and partial differential equations. Suppose we have a linear partial differential equation denoted by the action of a linear operator L which is have encountered earlier on the variable u. So, L u is also, suppose that u belongs to the domain D, where D is a subset of R k, R k is the k'th dimensional Euclidean space, so two dimensional Euclidean space or 3 dimensional Euclidean space.

In D L u is equal to f. So, this linear operator operating on u is equal to f. a given function f suppose and on the boundary of D so on. Del D which is the boundary of D let us suppose, that u satisfies another relationship again defined by another linear operator in this time, in this case L bar. So, L bar u is equal to g on del D. So, within the domain D L u is equal to f on the boundaries of that domain L bar u is equal to g.

(Refer Slide Time: 16:34)



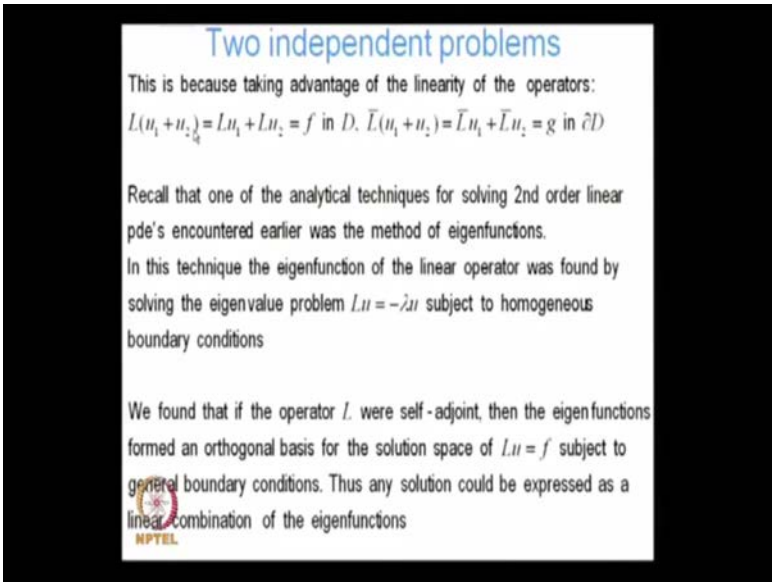
So, examples can include any of the canonical partial differential equations, we encountered earlier for example, Laplace's equation in a 2 D domain. So, del 2 u del x square del 2 u del y square is equal to f of x y on D, which is the subset of the 2 dimensional Euclidean space and u is equal to 0 on del D.

In this case, L obviously is del square x square by del square plus del square y square laplacian operator in 2 D space, while L bar is the identity operator, because L bar u is equal to u. So, L bar is the identity operator and g in these cases equal to 0. So, the general linear problem A, if we define the general linear problem A to be L u is equal to f in D and L bar u is equal to g on del D can be simplified somewhat if we can solve two independent problems what are the 2 independent problems 2 independent problems are L u 1 is equal to f in D and L bar u 1 is equal to 0 on D. Now, instead of solving this problem we solve. So, the instead of solving this problem, we solve the homogeneous problem, First the homogeneous boundary condition.

So, L u 1 is equal to f in D but L bar u 1 is equal to 0 on del D, same problem with homogeneous boundary condition, that the first part, the second problem is L u 2 is equal to 0. So, within the domain L u 2 is equal to 0. The boundaries L bar u 2 is equal to g. So, it is the complimentary problem. So, this is the problem b and it is complementary problem c, if we solve B and C then we can always obtain a solution for A as u is equal

to u 1 plus u 2 as the sum of these two solutions, why? Because this operator L is a linear operator.

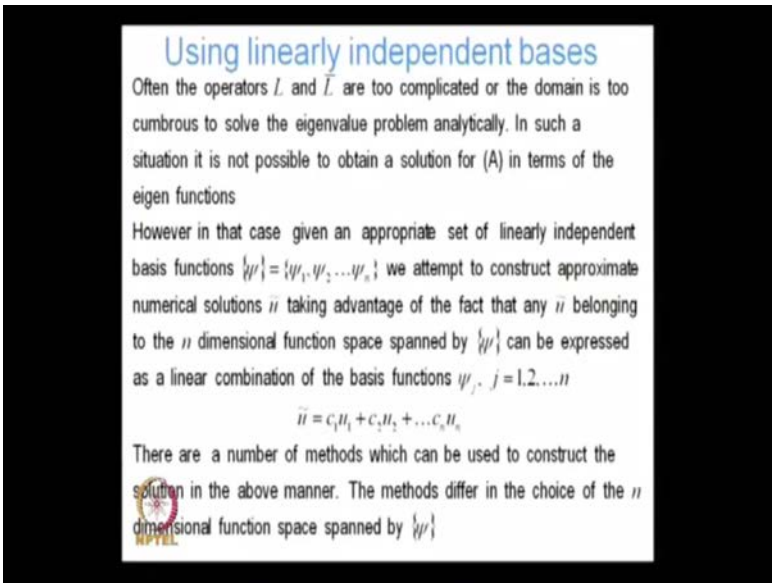
(Refer Slide Time: 18:40)



So, L u 1 plus L 2 is equal to L u 1 plus L 2 which is going to be f in D, because L u 2 is equal to 0 in t L u 1 is equal to f in D. So, L u 1 plus L 2 is going to be f in D and L bar u 1 plus L 2 plus u 2 again taking advantage of linearity it is going to L bar u 1 plus L bar u 2 this is equal to 0, that is equal to g. So, this whole thing u 1 plus u 2 satisfies my original problem. So, if I can solve the problems B on and C then I always construct solutions to my original problem A, I adding those two solutions together taking advantage of the linearity of the operator.

Recall that one of the analytical techniques we looked at for solving second order linear partial differential equations was the method of Eigen functions. In this method we found the, we solve the Eigen value problem. So, initially we solved this problem L u is equal to minus lambda u subject to homogeneous boundary conditions and we found that, the solution if this operator was self–adjoint, if this operator was self –adjoint then the Eigen functions formed an orthogonal basis for my solution space of this function. So, it formed an orthogonal basis for the solution space. So, any solution could be constructed by taking a linear combination of my Eigen functions that we have seen in great detail earlier we have used that technique for solving the wave equation and the diffusion equation and so on.

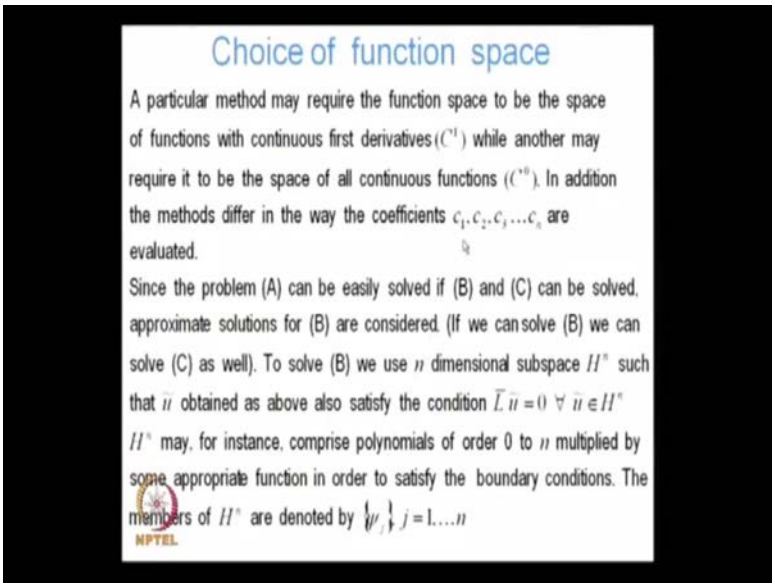
(Refer Slide Time: 20:27)



However the operators L and L bar are often too complicated. They never if they are not as simple as my Laplacian, if they are complicated or if my problem domain geometry is really horrendous, it is very complicated. Then it is not possible to solve this problem analogically to get the Eigen functions. So, then my analytical approach is break stop, I can no longer solve my Eigen functions and I can no longer construct my solution using those Eigen functions. Then what do I do?

Well, in such a situation it is not possible to obtain a solution for in terms of Eigen functions. However, in that case given an appropriate set of linearly independent basis functions we attempt to construct approximate numerical solutions u tilde. Taking advantage of the fact that any u tilde belonging to the n dimensional function space spanned by psi. These basis functions can be represented as a linear combination of the basis functions. Now, there are number of methods, which can be used to construct the solutions in this manner. These methods differ first on the choice of the basis functions, first on the choice no not just the basis, it depends on the choice of the space, choice of the function space, n dimensional function space spanned by psi.

(Refer Slide Time: 21:57)



Particular method may require the function space to be the space of functions with continuous first derivatives. Another method might require the function space to be a space of continuous functions only. It only requires continuity in the function values not in the derivatives. In addition, the methods may differ in the way the coefficients are evaluated. The coefficient c 0, c 1, c 1, c 2, c 3, and c n the way I form my coefficients.

So, it depends on the function space on which function space. I choose to construct my approximate solution. So, once I know my functions space probably the basis functions are known. So, then I depends on the basis functions and then finally, depends on the coefficients. I depend on the coefficients since the problem A can be easily solved, if B and C can be solved approximates.

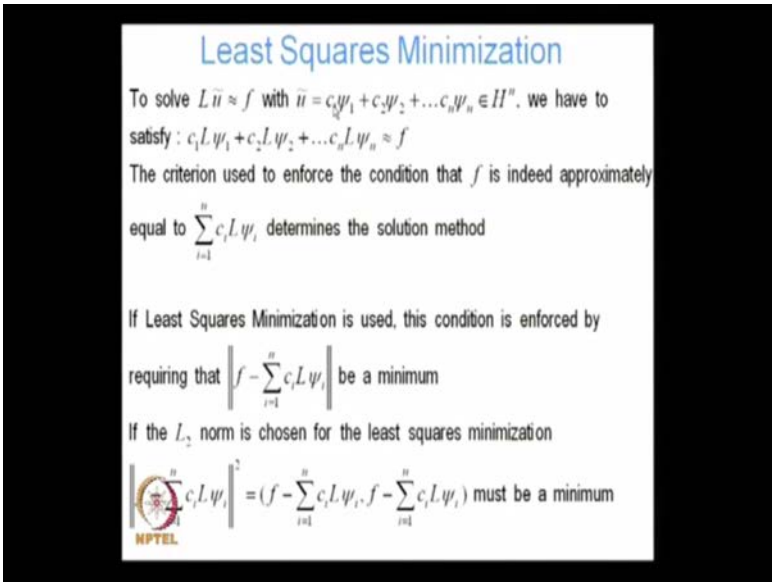
We will consider approximate solutions to the problem B to this problem. We will consider approximate solutions to the problem B. So, if we can solve B, we can easily solve C as well. So, we will only consider solutions of B. The same methods can be used to solve C and once we have solved B and C we can construct solutions to A to solve B. We can use the n dimensional function space H n such that, u tilde obtained as above also satisfies.

So, u tilde obtained as above I mean u tilde obtained like this. So, that is a typo. So, it is c 1 psi 1, c 2 psi 2 and c n psi n. So, if I construct u tilde like this, I construct it. But, u

tilde also has to satisfy this condition. So, u tilde belongs to the function space is constructed by taking a linear combination of the basis functions spanning.

The function space H n which I am denoting as psi 0, psi 1 through psi n. But, u tilde in addition it satisfies the boundary condition for my problem B which is L bar u tilde is equal to 0 for all u tilde belonging to H n. So, H n may for instance comprised polynomials of order 0 to n multiplied by some appropriate function in order to satisfy the boundary conditions. The members of H n are which we have already seen are denoted by psi j j is equal to 1 to n.

(Refer Slide Time: 24:43)

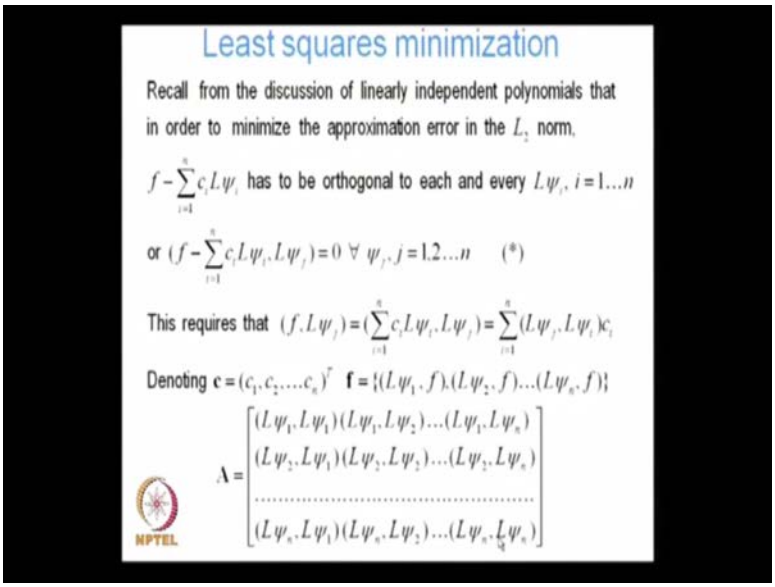


To solve the L u tilde as approximately equal to f with u tilde is equal to c 1 psi 1, c 2 psi 2 through c n psi n, we have to satisfy this equation c 1 L psi 1 plus c 2 L psi 2 plus c n L psi n is approximately equal to f. The criterion is very important, the criterion used to enforce the condition, that f is indeed approximately equal to sigma c i L psi i determines the nature of the solution method as we will see least squares approximation is different from galerkin, why because the criteria used to enforce this condition is different.

So, the criteria defines the method, the criteria used to enforce this condition, that L u tilde is indeed approximately equal to f determines the approximate solution method, if least squares minimization is used. This condition is required, is enforced by requiring norm of f minus sigma i equal to 1 to n c i L psi i be a minimum only then it is the best solution it is the minimum.

If the L L 2 norm is chosen for the least squares minimization that means, this square which is in the L 2 norm. This is equal to the inner product norm square is equal to the inner product. So, this is equal to f minus sigma i equal to 1 to n c i L psi i f minus. This inner product must be a minimum.

(Refer Slide Time: 26:29)



And we have seen from our discussion of linearly independent polynomials, that in order to minimize the approximation error in the L 2 norm. This has to be the residual, the residual f minus sigma i equal to 1, c i L psi i has got to be orthogonal to each L, each L psi i. Because I am taking the linear combination of the L psi i's to construct my approximate solution.

My approximate solution is equal to c 1 L psi 1 plus c 2 L psi 2 this must be equal to f. So, f minus L this 1 must be equal to these building blocks, L psi 1 my approximate solution is actually c 1 psi 1 and so on and so forth, but this is my approximately equal to linear combination of these and we have seen that in the least squares method the residual must be equal to these quantity must be orthogonal to these quantities L psi 1, L psi 1 L psi 2 L psi n.

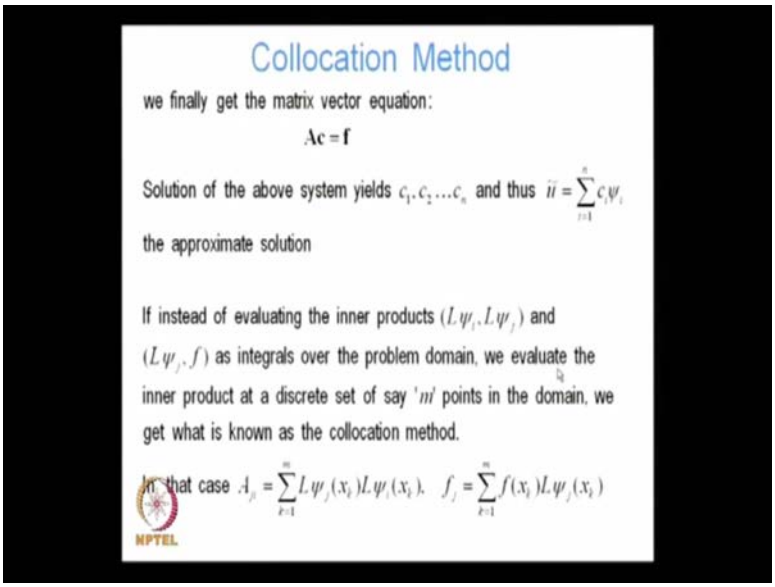
It must be orthogonal to the building blocks of my of what is approximately equal to f? The quantity, which is approximately equal to f is that is clear. So, earlier we were looking at we were not looking at differential equations. So, there was no L, there we are just trying to fit a function but now we are solving a partial differential equation. Now,

there is operator L. So, the difference must be orthogonal to L psi 1, L psi 2, L psi 3 and L psi n.

So, this must be orthogonal to that. For all psi j, j is equal to 1 through n. And this requires if I just expand this out expand this inner product out that means, f L psi j must be equal to sigma c i L psi j, L psi j, psi i, psi j sigma c i L psi i, L psi j which again if I pull out the constant I get that. So, denoting these c i's as the vector c vector f f L psi 1, f L psi 2 with the scalar components of f being L psi 1 inner product f L psi 2 inner product f and so on. So, I get a vector here, I denote that as a vector and here I get a matrix whose components are these dot products.

So, the first component is L psi 1 inner not a dot product, it is inner product. So, L psi 1 L psi 1 second component is L psi 1, L psi two-third components are L psi 1, L psi 3 and so on. So, then I get a system which is equal to A, c is equal to f A, c is equal to f where i this is my A. So, if I know my basis functions, I can evaluate the components of A, because this is nothing but L psi 1, L psi 1 inner product of L psi 1 with itself. This is L psi 1 inner product L psi 2. So, I know the components of A, I know the right hand side, because my function f is known my L psi 1, L psi 2, L psi 3 are known. So, my hand side is also known. So, then I can solve for my c's and I get my solution.

(Refer Slide Time: 30:12)



If I only get the matrix vector equation a c is equal to f solving the above system a c 1 c 2 through c n and once, I get c 1 and c 2 through c n and i get u tilde is equal to sigma c i

psi i. So, that is my least squares method, then if instead of evaluating these inner products. So, these are inner products that means. It involves evaluating an integral it is continuous assuming. That the functions are continuous and I am evaluating that inner product over my domain of interest. So, if instead of evaluating that integral, that is treating it as a continuous basis function. This psi i's as continuous basis functions.

If I treat it as a discrete problem and I start evaluating the inner products at discrete points suppose m points in the domain, we have what is known as the collocation method. So, in that case my a j I, I am going to evaluate as the sum and that sum is going to involve these sort of L operating on psi j evaluated at x k. So, I am again, I am evaluating my basis functions at discrete points at x 1, x 2, x 3 to x m. So, L psi j evaluated at x k times L psi I evaluated at x k and then I sum it over k is equal to 1 to m, I have think like that, I have made a mistake here, there is no sum here.

So, a j i is equal to L psi j L psi i x k there is no sum here, there is no sum. So, each of those components are evaluated there or is there can we see, no there is a sum I apologize again, there is a sum ,because this involves I am just evaluating that integral but I am evaluating that integral at a discrete set of points So, there is a sum I am sorry, for the confusion.

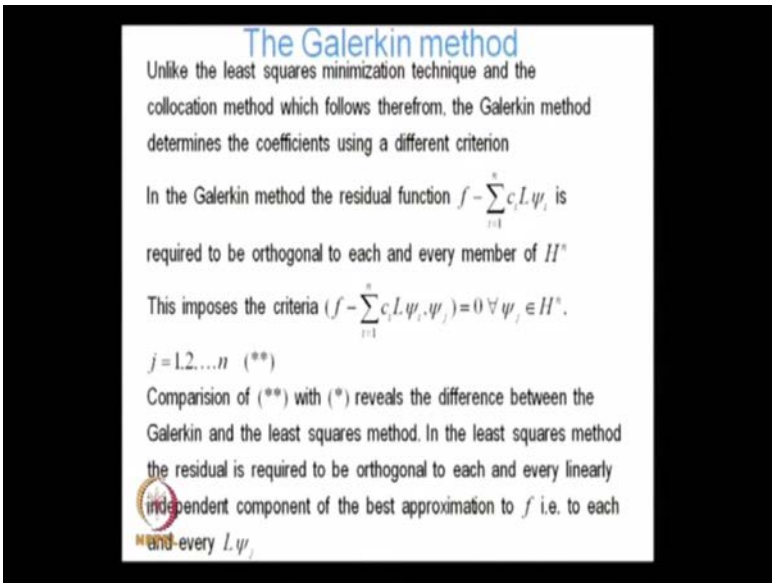
So, the only difference is that instead of evaluating that inner product as an integral. I am evaluating it as a sum. So, I am evaluating these basis functions at each of my grid points and then summing them up to get. So, to get A j I, I am evaluating psi j at k is equal to 1 psi i at k x x 1 then L psi j at x 1 L psi i at x 1 then I am adding to it L psi j at x 2 times L psi i at x 2 and so on. So, it is the sum I am sorry for the confusion and then f j to evaluate the f j i similarly, I convert it into a sum.

So, again i evaluate L psi j at x 1 times f evaluated at x 1 add to that L psi j evaluated at x 2 times f evaluated at x 2 and. So, on and. So, forth and I get my f j the contribution to f of sort of taking the projection with psi j. So, this allows me to evaluate these evaluate my coefficient matrix as well as my hand side in a discrete fashion without performing the integration.

This gives me, what is known as the collocation method, because as we will see in case in the collocation method typically, these psi j's are delta functions. So, basically they are the delta functions. So that means, that each of these basis functions are 1 suppose psi

j is 1 at x 1 but 0 everywhere else psi 2 is 1 at x 2 but 0 everywhere else and so on. So, if these are delta functions, then in this summation only. So, L L, it is becomes a lot simpler, if once they are delta functions is that clear. So, the number of contributions they comes much smaller in that summation. And actually, if they are delta functions then we are going to get whole thing becomes uncoupled. So, my coefficient matrix is actually going to become a diagonal matrix see.

(Refer Slide Time: 35:11)



So, then we let us talk about the galerkin method, which is slightly different from the least squares method, unlike the least squares minimization technique and the collocation method which follows there from which we have seen earlier, where the collocation method is just the discrete version of the least squares method. It is just the discrete version of the least squares method. The galerkin method determines the coefficients using different criteria.

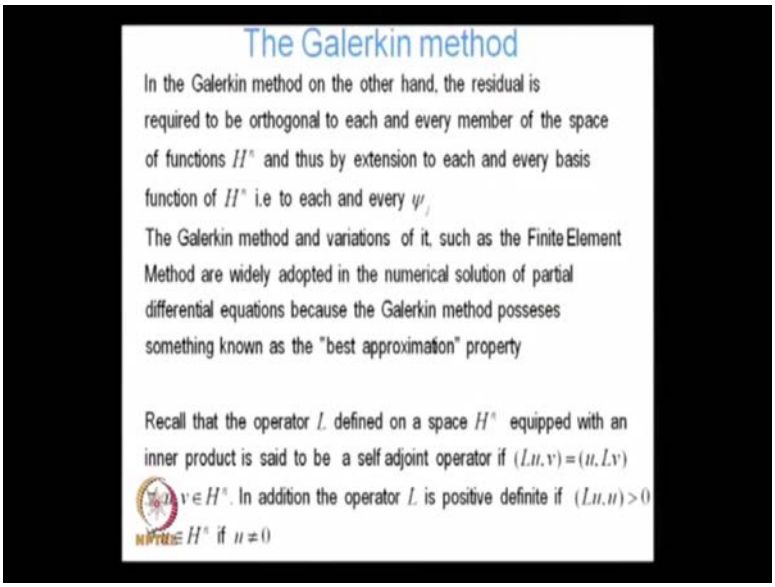
In the galerkin method, the residual function which is f minus sigma i equal to 1 to n c i L psi i remember that, we are constructing the best approximation to f using the basis functions psi i. So, L psi i linear combination of that is approximately equal to f is required to be orthogonal to each and every member of H n. This is different, because the least squares method what did we see, this had to be orthogonal, this residual was orthogonal to each of the L psi i. So, it is orthogonal to the linear operator operating on each of the psi i, L psi one, L psi two, L psi 3 and L psi n. So, residual is orthogonal to L

psi i's but in this case, the residual has to be orthogonal to the basis functions themselves. So, it is orthogonal to each and every member of H n this imposes the restriction that this imposes the criterion that f minus sigma i equal to 1 to n c i L psi i comma psi j is equal to 0 for all psi j belonging to 0 to n.

So, if I compare this with let us see with that, you can see the crucial differences are orthogonal to L psi j, that one in the galerkin method it is orthogonal to psi j. In the least squares method, the residual is required to be orthogonal to each and every linearly independent component of the best approximation to f. It is and the best approximation to f is considered to be this.

The residual must be orthogonal to each and every linearly independent component of the best approximation to f linearly component independent component of best approximation to f are L psi 1, L psi 2, and L psi 3. So, this is by the definition of the least squares method f must be orthogonal to each and every linearly independent component of the best approximation to f, that is why we have this criteria, while in the galerkin method this residual is required to be orthogonal to my basis functions for that function space. So, that is a very important point of difference. So, that is the basic difference between the galerkin method and least squares minimization.

(Refer Slide Time: 38:26)

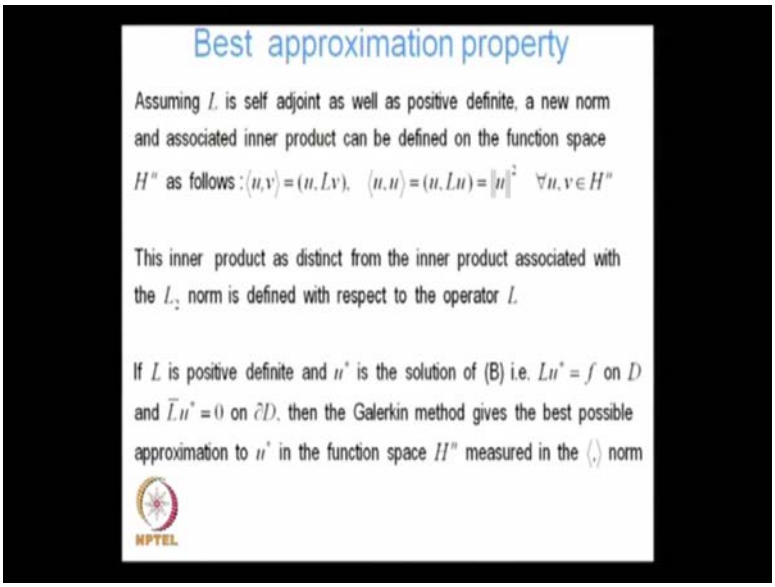


In the galerkin method on the other hand, the residual is required to be orthogonal to each and every member of the space of functions H n and thus by extension to each and

every basis function of H n, it is required to be orthogonal to each and every psi j. The galerkin method and variations of it such as the finite element methods are widely adopted in the numerical solution of partial differential equations why because the galerkin method possess something known as the best approximation property and by extension since the finite element method is also a galerkin method there are variations but let us assume that. So, it is a galerkin method it also possess the best approximation property.

Recall that the operator L defined on a space H n equipped with the inner product is said this. We have discussed when we talking about partial differential equations operator L is self adjoint if L u v is equal to u L v for all u v belonging to H n in addition. The operator L is positive definite if L u comma u is greater than 0 for all belong A, u belonging to H n if u is not equal to 0. So, that it is positive definite that is true.

(Refer Slide Time: 40:00)



Assuming L is self –adjoint as well as positive definite we define a new norm new norm associated with the operator L and what is that new norm the norm of u v is we an associated inner product the norm depends on the inner product. So, the inner product u v is equal to u comma L v it is no longer equal to my original inner product which I defined earlier u v. Now, it is the inner product is with respect to this linear operator. So, this is a new inner product I define and using this inner product I get this norm. So, u u using this new definition of the inner product is u comma L u and this norm is equal to

the norm square is equal to the inner product. So, this norm square is equal to that because the norm square we said is equal to the inner product, so in this new definition of the norm. So, every a function space has a certain inner product a function space with an inner product and the norm is always related to the inner product the norm arises from that inner product.

So, in this case the inner product is defined by that and this gives raise to this norm u. So, norm of u is a member of that function space. If I am interested in the magnitude of u in the norm of u, how am I going to evaluate the magnitude of the norm of u? I am going to take the inner product of u with itself and now my inner product is different. It is no longer u dot u it is u dot L u and whatever I get there is the square of the norm of u is the square of the norm of u. So, there is a one to one relationship between the norm and the inner product norm vector space in a norm function space, the norm is related to the inner product in this way.

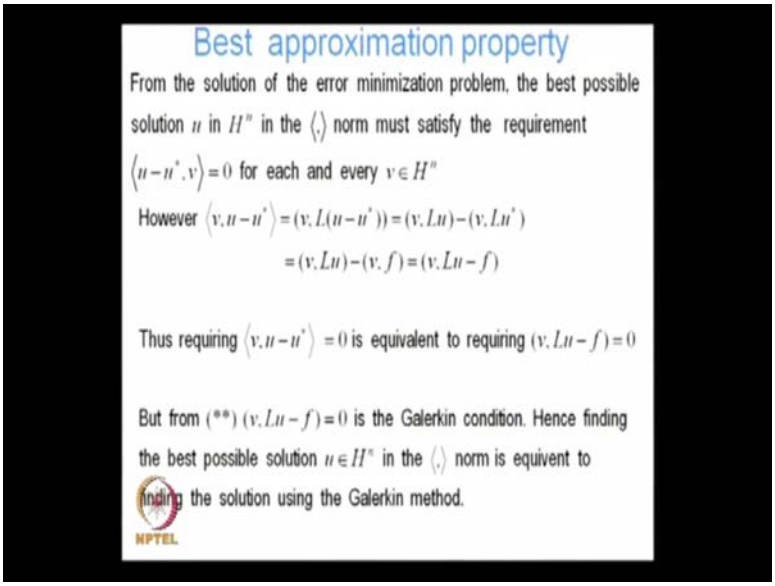
So, the inner product is distinct from the inner product associated with the L 2 norm inner product and it is defined with respect to the operator L. So, what is the best approximation property is this? If my operator L is positive. Definite meaning that it satisfies this property, satisfies that property. If my operator L is positive definite and u star is the solution of my problem my problem B which was L u star is equal to f on D L bar u star is equal to 0 on del D and u star is the solution is the true solution of the problem. Then the galerkin method always gives the best possible approximation to u star in the function space measured in this norm.

Best approximation as soon as we say it is a best approximation, we have to show in what norm it is the best approximation. How do you say it is the best approximation? We say it is the best approximation we if your solution obtained in the galerkin method. That differs from the best solution we calculate the difference between the solution obtained using the galerkin's method and the true solution in some norm which norm is that, this norm which we have defined that gives me the smallest possible norm.

So, it is closest to the true solution in this norm. It is closest to the true solution in this norm. So that is the property. So, if we have a norm defined like this and we get the solution of a problem using galerkin's method. That solution to the problem obtained using galerkin's method differs from the exact solution but the difference is minimum in

that norm. So, in that if you obtained the solution by some other method for instance by the least squares method and you calculate the difference in this norm that difference will be larger than the solution we have got obtained using galerkin's method that difference will be larger than the solution, you obtained using galerkin, this is the property, this is the best approximation property of the galerkin method, which passes on the finite element method and that is why the one reason why the finite element method is so widely adopted.

(Refer Slide Time: 45:06)



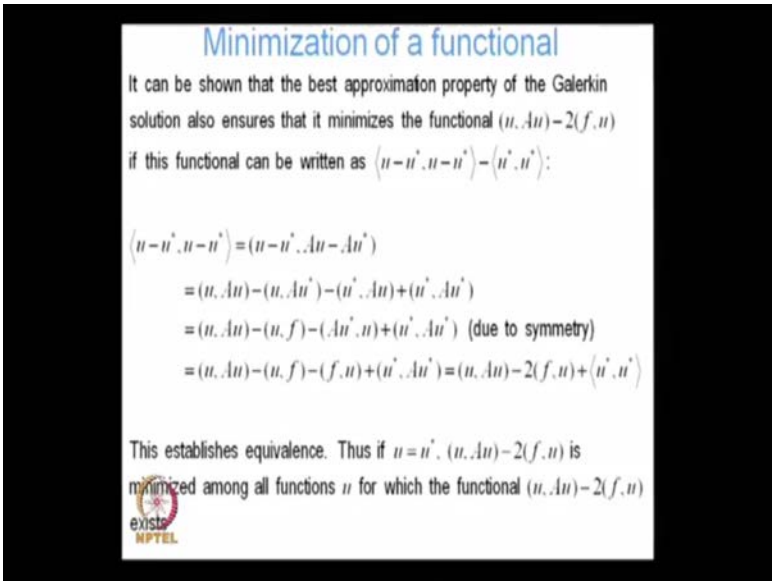
From the solution of the error minimization problem, the best possible solution u in H n u must belong to H n. It is the best possible solution that can be constructed from the basis functions used in H n must satisfy the requirement u minus u star comma v is equal to A 0 for each and every v belong. That is we know solution of the error minimization problem, we know that the residual inner product v must be equal to 0 for each and every v belonging t H n.

However v, if I evaluate that, u minus u star comma v I take advantage of symmetry write it like that and that can be written as v L u minus u star, because this inner product means v comma L u minus u star my usual inner product my usual inner product and that is equal to v comma L u minus v comma L u star again taking advantage of the linearity of the operator. So, I get B comma L u minus v comma f, because L u star is the exact solution u star is the exact solution L u star has got to be equal to f. So, v comma L u

minus v comma f which I can write as v comma L u minus f thus requiring v comma u minus u star is equal to 0 is equal equivalent to requiring v comma L u minus f is equal to 0 but v comma L u minus f is equal to 0 is nothing but the galerkin condition which we have seen v comma L u minus f is equal to 0. That was my galerkin condition. That was my galerkin condition if psi is v and I think of psi as v and this is f minus L u that is my galerkin condition.

So, to minimize the error if I minimize the error, if I minimize the error, then I satisfy galerkin's condition to minimize the error I have to satisfy that requirement that I know from earlier to minimize the error. I have to satisfy that requirement and satisfying that requirement is equivalent to satisfying the galerkin condition hence finding the best possible solution u belonging to h n in this norm is equivalent to finding the solution using the galerkin's method. So, instead of finding the best possible solution to u and this norm, if I solve the equation using galerkin's method find the best solution using galerkin's criterion by basically find reduce the make ensure that approximation satisfies this criterion. The galerkin's criterion then I am going to also get my best approximation property.

(Refer Slide Time: 48:28)



It can be shown that the best approximation property of the galerkin's criteria. It is not only minimizes, it not only gives me satisfies that condition it also minimizes a certain functional which if I define that functional like this. It is also minimizes, that functional

what is the advantage of why do we want to emphasis this? The fact, that the galerkin's solution minimizes that functional well we are going to see that later on. Actually it can be shown that, this functional can be written like this u minus u star inner product of u minus u star u minus u star minus inner product of u star u star, why well let us see how let us see.

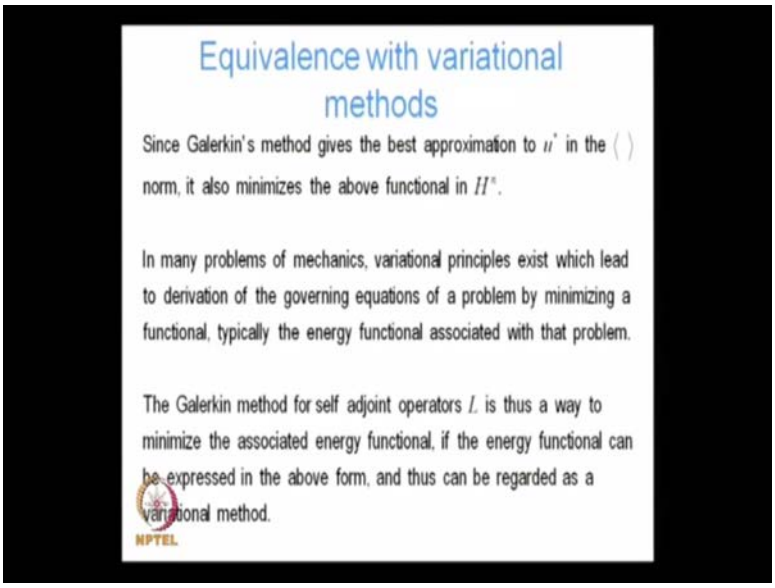
U minus u star inner product u minus u star, I can write it as u minus u star A u minus A u star by definition of the inner product. This I can write as u comma A u minus u comma A u star minus u star comma A u plus u star comma u A A u star taking advantage of linearity, this becomes equal to u comma A u minus u comma f because A u star u star remembers the exact solution.

So, A u star is equal to f u comma f and again I have interchanged u due taking advantage of symmetry. We can write as A u no sorry, I have not interchanged. I have taken advantage of the fact, that my operators self adjoint. So, A u star comma u A u star comma u plus u star comma A u star u star comma A u is equal to A u star comma u, because the operator A is the self adjoint operator.

This I can write as u comma A u minus u comma f minus A u star. I can again replace by f minus f comma u plus u star comma u star A star. So, this is equal to u comma A u and then again taking advantage of symmetry, I can write f u as u f. So, minus 2 f comma u plus u star u star inner product u star comma A u star. I can write as inner product of u star u star is that how I define my inner product.

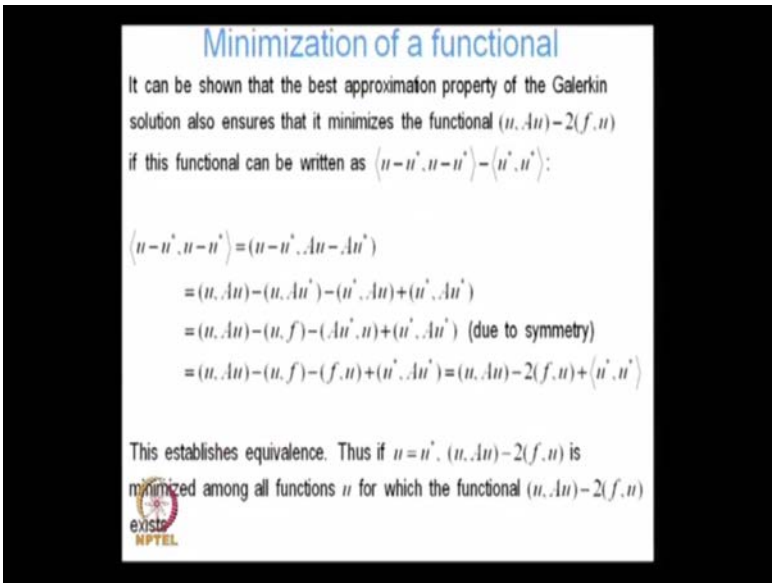
So, you can see that inner product of u minus u star u minus u star minus inner product of u star u star is actually equal to u comma A u minus 2 f comma u if u is equal to u star, that means if u is equal to the true solution, then this thing becomes 0. So, this difference is minimized u comma A u minus twice f comma u is minimized among all functions u for which is the functional u A u minus 2 f u exists, because then it becomes 0.

(Refer Slide Time: 51:54)



Since, galerkin's method gives the best approximation to u star in that norm. It also minimizes the above functional in H n since galerkin's method.

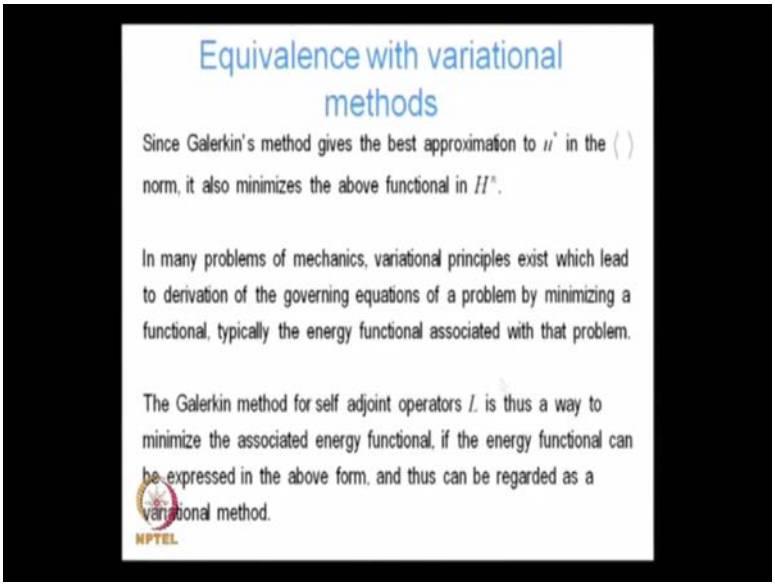
(Refer Slide Time: 52:06)



We have seen that galerkin's method gives me the best approximation to u star since galerkin's method. So, if I calculate u using galerkin's method. I am assured that, this difference is going to be a minimum, because galerkin's method gives me the best approximation to u star. So, this norm is going to be minimum, this norm is going to be minimum, that means that also gives me, it minimizes this functional. It minimizes the

galerkin's method minimizes this functional. Why is that important? Well we are just going to see.

(Refer Slide Time: 52:45)



In many problems of mechanics variation principles exists which lead to derivation of the governing equations of a problem by minimizing a functional typically, the energy functional associated with their problem. So, I am able to write an expression for the energy and by minimize by the variation principles which by minimizing that energy allow me to get the equations of motion.

So, we write an expression for the work. The principle of virtual work for instance principle of virtual work. So, all that depends on this I do not want to get much mechanics but there you can write an expression for the energy minimizing that you can get the equations of motion. So, the galerkin method for self adjoint operators L is thus a way to minimize the associated energy functional, because we have seen that by when we solve the problem using the galerkin's method, when we find the solution using galerkin's method, we are also minimizing that potential, that functional like this. So, when we are solving the problem using galerkin's method we are minimizing that potential. So, if this potential of this functional I am always talking potential, but if this functional is the energy potential it is the energy potential energy associated with that problem. I am also minimizing the energy.

So, galerkin’s method for self-adjoint operators, L is thus a way to minimize the associated energy functional, if the energy functional can be expressed in the above form and thus can be regarded as a variation method. The galerkin method what self –adjoint operators can therefore, be regarded as a variation method, because it leads to the same result. Is that clear? I think we will stop there, and we will talk about weighted residual methods and specialize them to finite element methods in the next class.