Matrix Solvers
Prof. Somnath Roy
Department of Mechanical Engineering
Indian Institute of Technology, Kharagpur

Lecture – 49
GMRES (Contd.)

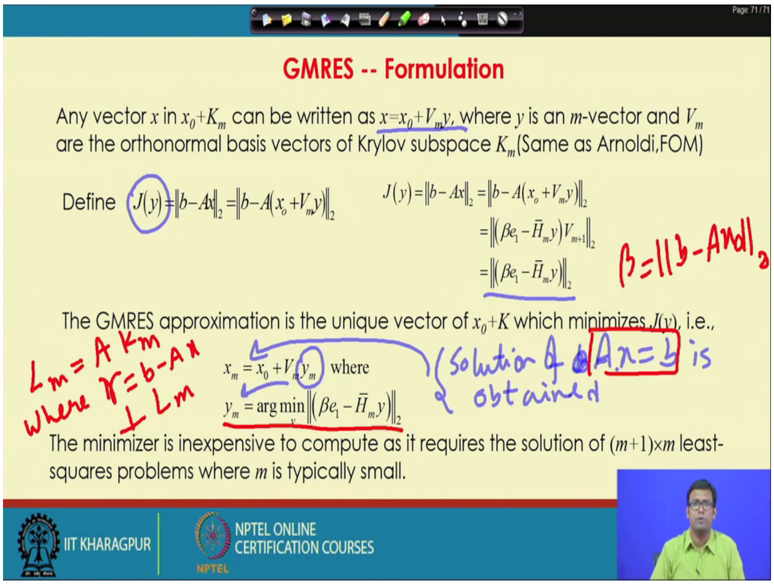
Welcome we are continuing our discussion on Generalized Minimum Residual Method; which is one of the Krylov subspace methods as we discussed which is applicable for any nonsingular matrix.

Earlier we have discussed a number of Krylov subspace methods and one of them which is a very first method called conjugate gradient; we have seen that it is only applicable for symmetric positive definite matrices. Now, we are trying to look into methods which are fast as well as applicable for any type of matrices not specific to symmetric matrix or not specific to diagonal dominant matrix like that.

So this is generalized minimum residual method which we said is applicable for any matrix and is a first method. And one important thing with general Generalized Minimum Residual Method or GMRES; we popularly call it one important point with GMRES is that; this is an oblique projection method; that means, the spaces k m and L m are not same.

So, let us start the discussion on GMRES formulation which we covered at the last session I will quickly recap on the formulation part also before going into algorithm in this particular session.

(Refer Slide Time: 01:47)



Any vector x in the affine space x 0 plus K m can be written as x is equal to x 0 plus V m y, where y is an m vector; y has m components a vector in R m and V m are the orthonormal basis vectors of Krylov subspace K m.

This is the basic idea of Krylov subspace method and up to this any method any Krylov subspace method is same with Arnoldis method or with a full orthogonal method that we generate the bases vectors; V m orthogonal orthogonal bases vectors V m of Krylov subspace and say that the updated value of x will be the approximated value of x plus vector in the Krylov subspace or a linear combination of members of V m.

Now, the GMRES statement looks into a functional J y which is L 2 norm of the residual b minus A x and so, if we start with any general value of x we replace x is equal to x 0 plus V m y; what we obtained here. We can sorry x 0 plus V m y we can substitute it here and with little further analysis we have seen it in last session that this L 2 norm is same as beta e 1 minus H m bar y was L 2 norm what H m bar is a Heisenberg matrix that we obtain from Arnoldis modified Gram Schmidt method and beta is the magnitude of the initial residual.

So, we can also write that beta is equal to mod of b minus A x 0 or maybe second norm of that. So, beta e 1 minus H m y. This is a functional J y which is second norm of the L 2 vectors; we get that J y has this particular value. And now the GMRES approximation tells that there is a unique vector x 0 plus K which will minimize J y and there is the solution exact solution x.

So, if we can find out a vector x 0 plus K which minimizes J y that will be the solution of b equation b minus A x. So, my final x will be x 0 plus K which minimizes this now minimizing J y is equivalent to minimizing this particular function beta e 1 minus H m bar y is L 2 norm.

So, we have to find out a y which minimizes this particular norm and then we multiply this y vector with V m y with the basis vectors of Krylov subspace; V m and this x m is the solution of the equation. So, solution of b minus of A x is equal to b is obtained and this solution is nothing, but this x m where this particular value of y m which comes from the minimization of beta e 1 minus H m bar y. That is GMRES statement or GMRES approximation for solving x is equal to b.

So, one important point is that it is right if we think of conjugate gradient or if we think of steepest descent method it was minimizing a different functional and that functional was J x is equal to x transpose A x half of x transpose x A x minus x transpose V.

Here it is and when it was minimizing this that functional it, we observed that the residual vector is orthogonal to the space L m which is same as K m. Here it is minimizing a different functional therefore, the residual vector b minus A x is not orthogonal to K m rather it is orthogonal to another space L m and if we try to relate it with the discussion we had in minimum residual method not generalized minimum residual method in one dimensional projection process when we discussed about minimum residual method.

We have seen that the space on which residual vector is orthogonal is A V where V is the space of the in which x is updated. And therefore, here we using the same discussion we write that L m is equal to A K m where r is equal to b minus A x is orthogonal to L m.

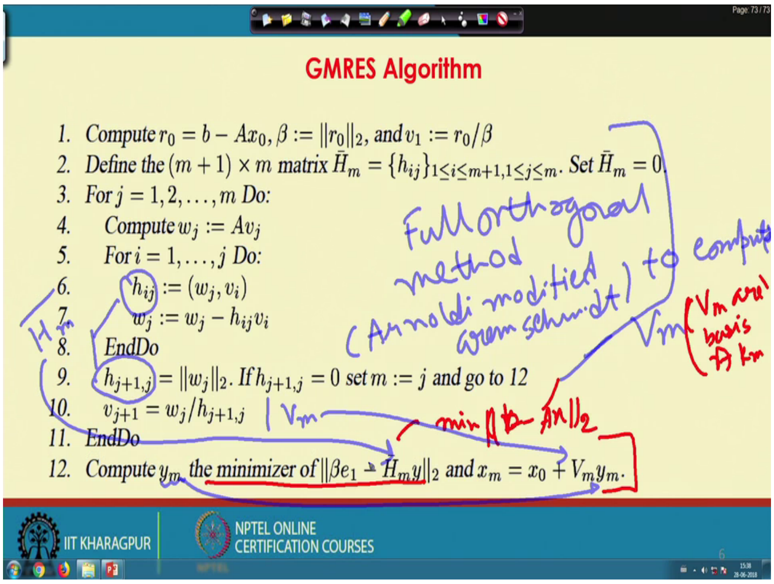
Residual vector is orthogonal to a special m and L m is equal to A K m and what is K m? K m is the Krylov subspace. And the formulation takes us to the fact that we have to find out first Krylov subspace bases vectors V m and then we have to find out and y m which minimizes this function; beta e 1 minus H bar m y and this is equivalent to minimizing the L 2 norm of the residual. And with this y m we multiply this y m with the bases vectors V of V and we will get the solution.

So, essentially solving A x is equal to b boils down to find minimize find out the minimizer of b eta E 1 minus H bar m y L 2 term. So, instead of solving A x is equal to b what we are trying to do? We are trying to find out a minimizer of beta e 1 minus H bar m y is L 2 norm.

So, now the minimizer is an expensive to compute as it requires a solution of m plus what into m least square problems. And m is typically smaller m is the maximum dimension of Krylov subspace or the dimension of Krylov subspace up to which the solutions converge and it is seeing the name is typically smaller than the actual matrix size.

So, it is solving a smaller number of least squared equations than the actual number of equation and therefore, the solutions become faster.

(Refer Slide Time: 08:38)



If we look into the algorithm quickly; the GMRES algorithm it, so what again if we go back once before coming into the algorithm sorry; for that that what does it mean need? It needs computing of V m and it needs computing of y m.

V m is computed exactly same as V m at the V; V m are the members of the Krylov subspace bases of the Krylov subspace. So, V m is computed exactly same in similar way as we have done in full orthogonal method or Arnoldi modified Gram Schmidt.

And once we compute V m then there is a mixed rather smaller tape step which is computing the minimizer of beta e 1 minus H bar m y. So, if we look into the algorithm the first part it starts with an initial guess value x 0 and computes the initial residual finds out the value of beta; which is magnitude of the initial residual vector.

And then the later steps is right you start you take v 1 as the unit vector along r 0 or the initial residual. And then take multiply this with A and subtract which is a new bases for of Krylov subspace.

But as you want to find that orthogonal bases vector you take the new bases and consider the old bases which is v 1 and take their dot products, subtract the dot product multiply with the old bases from the new bases. So, you get a new orthogonal bases and get orthogonal bases vectors and in that way you also calculate the Heisenberg matrix and do till the Heisenberg matrix gives a j plus 1 jth term 0.

That means all the independent vectors in Krylov subspace all the independent orthogonal vectors for Krylov subspace are found out or the entire bases of Krylov subspace is found out do up to this part of m. And this part is very same as full orthogonal method or it uses Arnoldi modified Gram Schmidt; to compute V m and V m V m are basis of K m.

K m is same for all the cases it is the Krylov subspace which is same for any of the Krylov subspace that full orthogonal method arnoldi c g K must be same. So, this computation computation of V m is exactly same as other Krylov subspace (Refer Time: 11:48) methods.

And once you have finish this part then this is what is specific in GMRES is the minimizer of beta e 1 minus H bar m 2. So, here you are actually finding out minimizer of r minus b minus A x L 2 norm; this is a new part in GMRES.

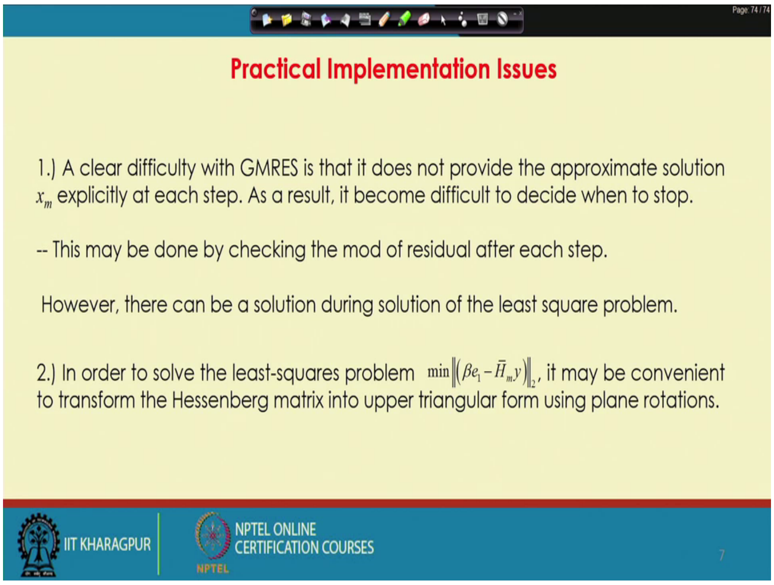
And once you find this minimize that because there is a smaller matrix; the solving least squared equation is like solving normal equation which is much simpler than solving the large A x is equal to V. When you find do this; so, if found all the m's and then you in that process you have generated the Heisenberg matrix elements; these elements will go for the Heisenberg matrix H m by H bar m.

And you find out find out the minimizer and this is the GMRES algorithm. So, the algorithm has couple of parts the first part is find out V m same as full orthogonal method in that part you also compute H bar m or the Heisenberg matrix. And once you have the Heisenberg matrix you find out the minimizer of beta e 1 minus H bar m y; and now go ahead with that y; y m is the minimizer multiply this with V m and we take 0.

So, you find out y m here and then this y m you take here; similarly you have found out V m here and this V m is put here and that that and the H m you found out here this H m is used in this particular step and that gives you the GMRES solution.

So, instead of solving an equation large equation system we are trying to find out minimizer of smaller number of equation of a smaller functional; we are trying to find out small number of equations least square equations.

(Refer Slide Time: 14:04)



However there are few implementation issues; first is that that in GMRES it does not provide the approximate solution x m at each step rather because x m is found out when you have found out y m which minimizes the function.

So, we like an iterative method where we are every time looking into the solution and checking whether the solution is satisfying the actual equation; here we are not finding out x m explicitly till we find out the right y m which minimizes j. So, there is a issue of convergence because the numerical solution you find out is never the very exact solution; there are round of errors or numerical errors each in iterative method the numerical solution converges to the right solution.

So, what should be the convergence criteria? Convergence criteria is something that the error is of the order of round off error, error is very small. The now if we do not find out x m at each step it is find out difficult to find out the error. So, there is a problem here because we are using an iterative method, but through that iterative method we are not finding out the solution; rather we are you doing an iteration for finding out minimum value of a particular function.

So, when should we stop? That is difficult to may make a call because again like if we know x m we can see that how x m is different than x m minus 1 and if the values are small we will say that it has converged and I will stop, but if we do not know x m we are looking only into y m. So, based on what value of y m can we stop and this is a little involved question.

However, there are simpler solution like check the residual with the y m you find out what is the value of b minus A x m or what is the value of beta e minus H bar m y though; it is the minimum value what is the value. If the value is not changing much; that means, you have already close to the minimal value or if you see residual r 0; if this is very small then you can also say that b minus A x is almost 0; so, we can stop that is that can be 1 remedy.

However there can be another solution in the way we propose the least find out the equation of least square solution; we will discuss it later that instead of looking into the value of residual at each step; we can probably directly find out where to stop. The other thing is that when we are solving this least square equation m plus 1 into m equation of a Heisenberg matrix H bar m is an Heisenberg matrix.

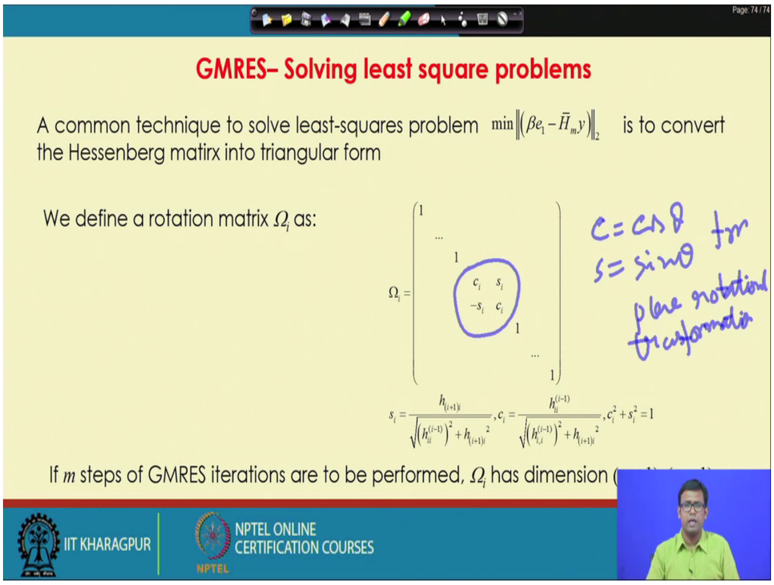
So, instead if we can go on transform the Heisenberg matrix to a triangular form Heisenberg matrix is that a triangular matrix plus 1 row below the diagonal one sub diagonal row. Now if we can transform the Heisenberg matrix to triangular form it is much simpler to find out the solution of least squared equations. And conversion of matrix is possible through plane rotations; so, we will apply a transformation; a rotational transformation on the Heisenberg matrix and convert it to a triangular matrix, upper triangular matrix.

And we will see that the issue of finding out minima of beta e minus H bar m y will be much simple. Because although it is a less smaller number of equations which are solving in least square equations still we are solving a kind of a matrix equation kind of a transpose A x is equal to x transpose b things like that.

So, A is A transpose A x is equal to A transpose b a equation like that which is a least square equation. So, there are some computational cost involved here, but now if we can make it much simpler or by you doing some transformation on the Heisenberg matrix; so, that we come up with a upper triangular matrix the computations will be further simpler.

And these also will tell us give us a very good nice way to decide that what is the value of x m directly so that we can stop the calculations once x m has converged to the right solution.

(Refer Slide Time: 18:34)



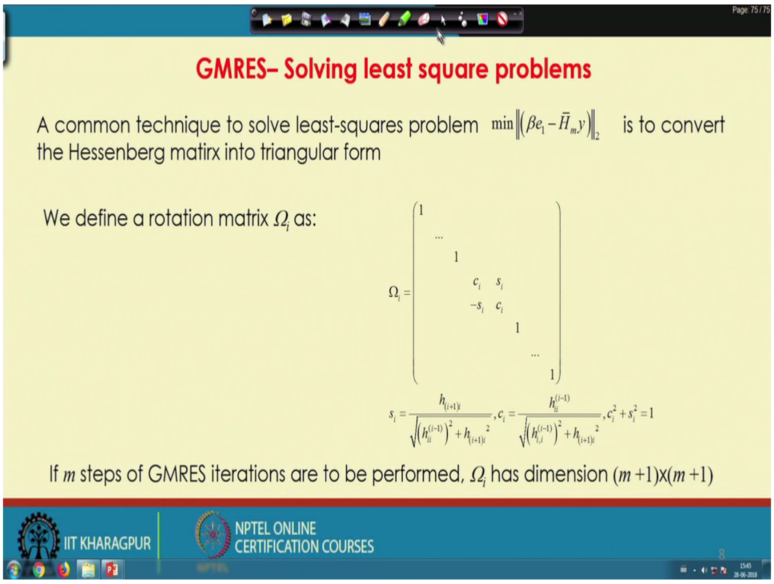
A common technique to solve least square problem is to convert the Heisenberg matrix into upper triangular form and that is done by a rotation.

So, here a rotation matrix chemical omega i is defined as it is a diagonal matrix with only a block of c i s i minus s i c i which are the cosine and sine of the angles which are calculated based on the i plus 1th ith term of; Heisenberg matrix. If we are looking into the ith row only you or if we are trying to take the sub diagonal term from the ith row only; we will look into these values the Heisenberg matrix i plus 1 ith term and of the like this can come after several transformation.

So, what is the value of the diagonal term of ith row after the transformation based on which we can calculate s and c; which is basically c is equal to cos theta, s is equal to sine theta; for plane rotation plane rotational transformation and therefore, c square plus s square is equal to 1.

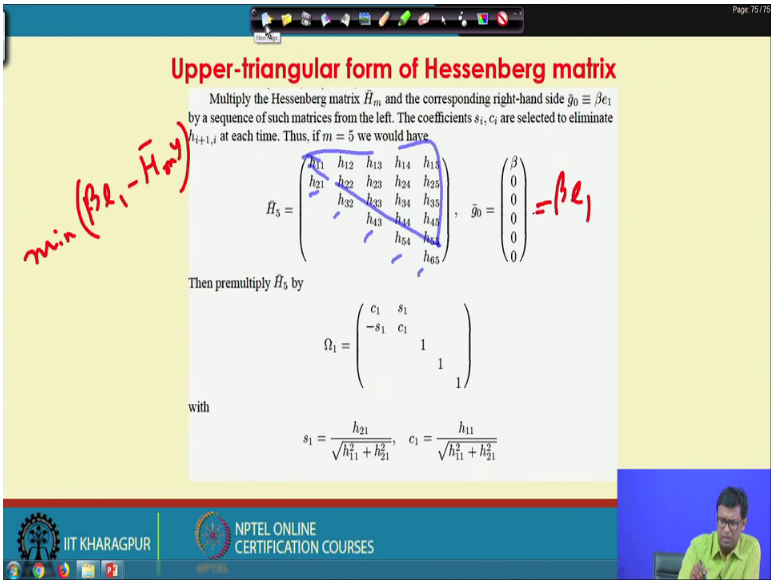
So, if m steps are performed; then sigma will if there the H bar m has m m plus 1 into m plus 1 dimension sigma m; i will also m plus 1 into m plus 1 dimension. So, how does it work we will see on a 5 take a 5 by 5 Heisenberg matrix and look into the example.

(Refer Slide Time: 20:23)



This is an example which is directly copied from (Refer Time: 20:22) book. So, we will see how this matrix can help us in finding out the triangular form.

(Refer Slide Time: 20:27)



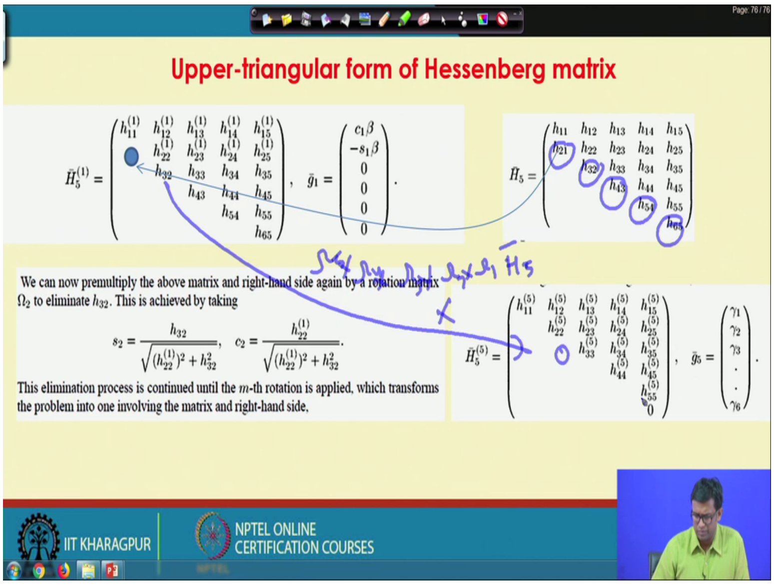
For example we take a Heisenberg matrix which has basically 6 rows and the last term is h 65 and, but this is m plus 1 into m plus 1. So, m is equal to 5 we take this as Heisenberg matrix for m is equal to 5. And the corresponding beta g matrix which is beta as the first component and remaining 0 is there. So, we are solving basically beta e 1 minus.

So, g 0 is this is this is basically beta E 1 beta e 1 minus we are trying to find out the minimal of beta e 1 minus H bar m y. So, if I because I am I have to find out minimum of beta e 1 minus H bar m y. If I do a transformation on H bar m or if I multiply H bar m with sigma i; I should multiply H bar m with c beta with also with sigma i; so, that it carries the meaning.

So, entire thing is being multiplied by the same vector. So, now sigma 1 the pre multiplier we for first row is c 1 minus s 1; s 1 c 1 and this will our target is to give a triangular form here. So, triangular form will be a form like this. So, we have to keep on eliminating these vectors.

So, by multiplying with sigma 1 our goal is that H 2 one will be eliminated from H bar 5. And of course, g z g bar 0 after this multiplication will not be beta e 1 rather; it will change from E 1 something else will come here. So, this is the pre multiplier matrix and the according to the formula given in that slide.

(Refer Slide Time: 22:27)

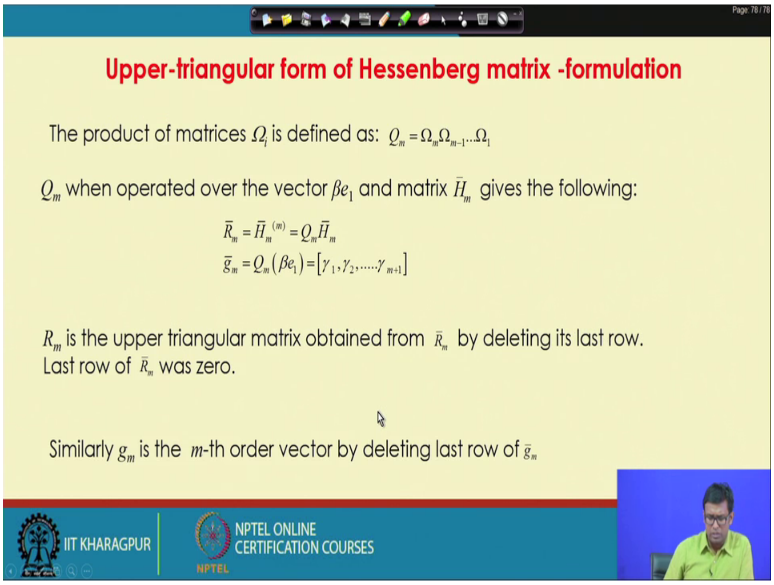


So, when we multiply omega 1 with H bar 5 h 21 is related. So, this term is related and there is a minus s 1 beta which comes here. Now we calculate s 2 c 2 and s 2 to take away h 32 term and when we multiply this the this like from here; we can see the 3 2 term this has been cancelled out. And this there are further more components in g by 5 g 5.

So, once we do these multiplications with h 2 h 3 like we take H bar 1, H bar 5 and then multiply omega 1 into omega 2 into omega 3 into omega 4 into omega 5. So, with each multiplication with multiplication with omega 1; we cancel out this with omega 2 cancel out this omega 3 cancel out, this omega 4 you cancel out this omega 5 you cancel out this.

So, finally, we end up in a upper triangular matrix and the g bar 1 is g, g 1 or the g matrix is converted into a 6 all the terms are populated instead of 0. So, it is a 6 nonzero terms in general (Refer Time: 23:58). So, through a plane rotation we can take any H bar matrix of Heisenberg matrix and convert it to a upper triangular form with the last row having 0 on the.

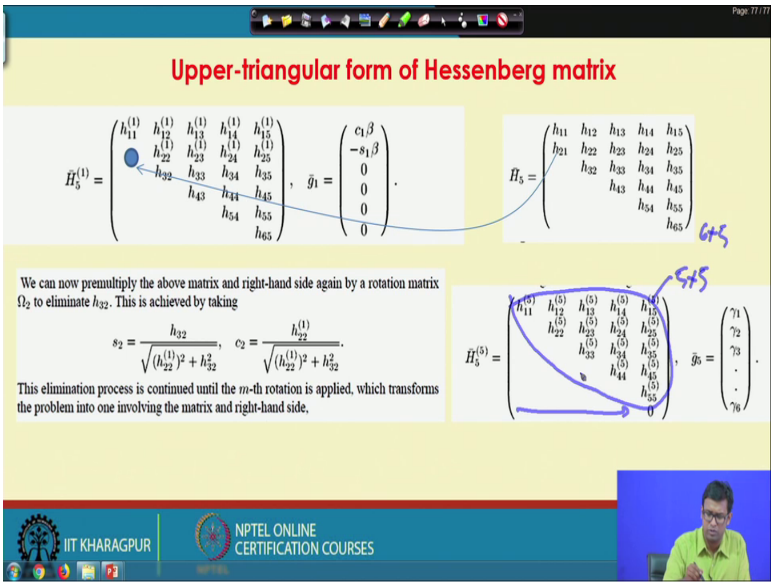
(Refer Slide Time: 24:18)



The product of the matrices omega i is known as Q m omega m omega minus 1 to omega 1. Q m when operated over vector beta 1 and matrix H bar m gives the following; R m bar is the upper triangular form with the lower row everything 0 is H m bar H m Q m H bar m. g m bar is Q m beta e 1 gamma 1 to gamma n plus 1.

The R m is the upper triangular matrix obtained from R m bar by deleting the last row; since the last row is 0. So, if we look into the sorry.

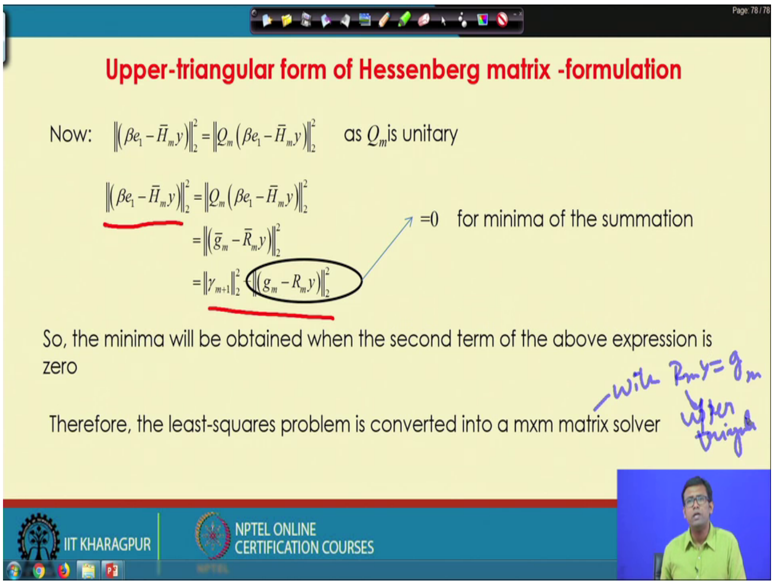
(Refer Slide Time: 25:08)



See if we look into the last row of this is the 6 by 5 matrix, but the last row this is completely 0 in. So, this will give me and 5 by 5 upper triangular form if I delete the last row; that is exactly what we are doing here.

So, R m is the upper triangular matrix obtained from R m bar by deleting the last row; since the last row is zero similarly g m is the mth order vector deleting the last row of g m bar.

(Refer Slide Time: 25:42)



Now, beta e 1 minus H bar m y is equal to Q m beta e 1 minus H bar m y. Now as Q m is a unitary matrix multiplying with Q m will not change the norm will not change their length because it is called Q m is product of cos theta sine theta.

So, Q m has a norm of 1. So, beta e 1 can be seen as Q m into beta e 1 minus H bar ms L 2 norm is Q m into beta e 1 minus H bar ms L 2 norm; this is g m bar minus R bar m y L 2 norm; R bar m has the last row 0. So, we can write g m minus R m y is L 2 norm plus the last component of g m which is gamma m plus 1.

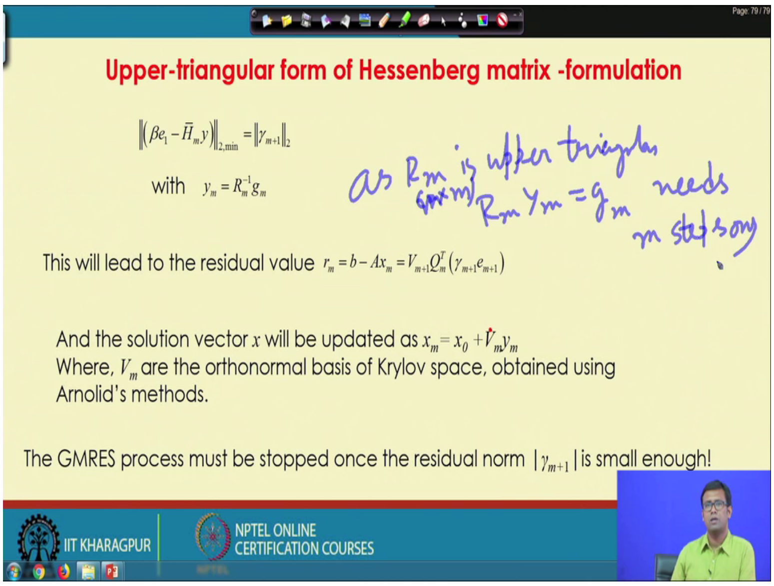
Now, we are trying to find out minimum of this function we are trying to find out its minima or minima of this particular function. And this is a function of y a plus b; b is a function of y; we can change with why when a plus b can be minimum? If b is a is fixed b is a variable if b is 0 a plus b is minimum. So, this function can have a minima; when g m minus r r bar m cell 2 norm is 0 the minima is this is the minima of the summation right.

So, the variable part is 0 and the (Refer Time: 27:10) constant part is still there. So, gamma m plus 1 square is a minimal value of this particular functional and so, plane rotation it comes out very easily. And now more interestingly when for which y this is minima? The y for which g m minus R m y is L 2 norm is 0.

So, now we are in a matrix equation actually this should be a 0 matrix and R m is a upper triangular matrix. R m is an upper triangular matrix; so the matrix equation is you only need n steps to solve the matrix equation. So, this will give you only need m steps exactly; m steps to find out the value of y which will give the minimum value of this function.

So, the minima will be obtained when the second term of the above expression is 0. Therefore, least square problem is can be is converted into a m plus m into m matrix solver with R m y is equal to g m and R m is upper triangular. So, we only need the backward; so, if a Gauss elimination method to solve this equation.

(Refer Slide Time: 28:32)



So, beta e 1 minus H bar m square with a y m is equal to r r inverse m j m; this is what we are trying to find out. This minimize gamma m gamma m m plus once L 2 norm L 2 norm of this and when y m is R m inverse j m.

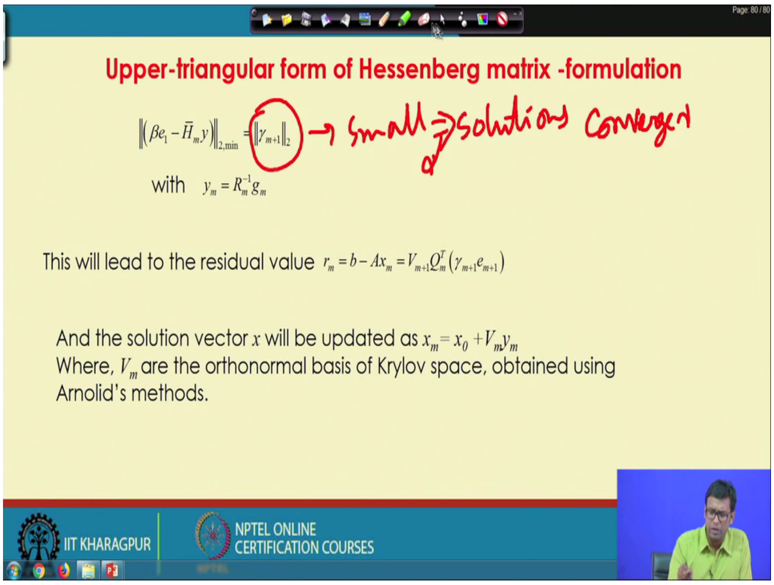
So, this will lead to a residual value b R m is equal to b minus x m is equal to V m plus 1 Q m transpose and this is the minimum value V m plus 1 Q m transpose gamma m plus 1 e m. Why Q m transpose? Because this is when calculating gamma m we have multiplied Q m with the beta vector and the solution will be updated as x 0 is equal to x m plus V m y m.

Hence V m where V m are orthonormal to bases of Krylov subspace obtained using Arnoldis method. And this process must be stopped when the residual is small and when the residual is small means finally, this is gamma m plus 1 L 2 norm that is the residual when this value is small enough; this process should be stopped.

And solving as R m is upper triangular; R m y m is equal to g m needs R m also it has a dimension m into m; m steps only. So, now, the GMRES algorithm can be solved in m number of in order of m number of steps.

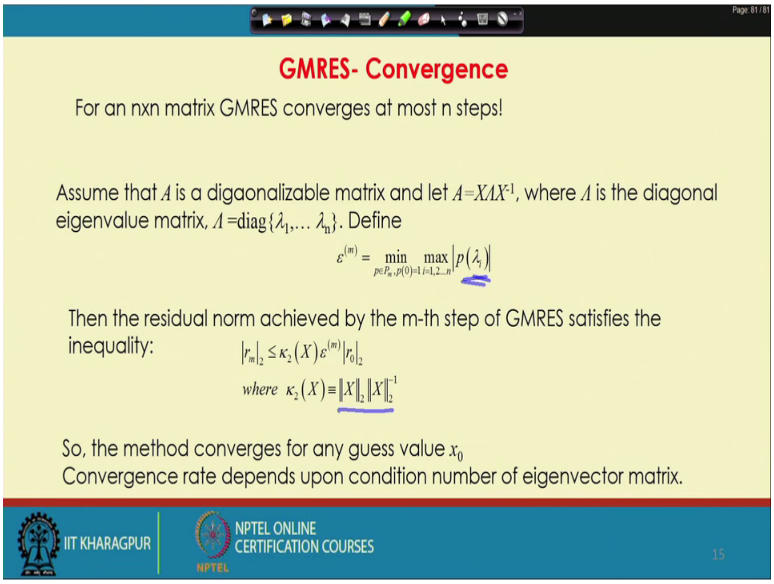
So, if we take a large matrix and we find out the basis vectors of Krylov subspace converging base in which the solution and solution will converge when gamma m is smaller enough. So, that gamma m plus 1 that is the value m; when gamma m plus 1 is smaller (Refer Time: 30:47) of the solution in convergent that we will need only m steps. So, it becomes a much faster method and we.

(Refer Slide Time: 31:01)



So,; so you also got a parameter to stop here and that is when this value gamma m plus 1; when this value is small solutions this will implies their solutions converged. So, we will quickly look into the convergence criteria of GMRES method.

(Refer Slide Time: 31:38)



And n into n matrix GMRES converge at most in n steps m cannot be more than n and n n steps it must converge.

Assume that A is diagonalizable matrix and let X is equal A is equal to X omega X in X lambda X inverse where lambda is the eigen value matrix all the diagonals of lambda has different eigen values of A. And epsilon to the power m is defined as a polynomial of lambda with different values of lambda, lambda 1 and maximum of that for the degree of polynomial will depend on for which this value is minimum.

Then the residual norm achieved by mth step of GMRES satisfies that r m 2 residual norm at mth step is less than the condition number of the eigenvector matrix X into epsilon m r 0 to the power 2. And therefore, the L 2 norm of r 0 is always greater than the L 2 norm of the r m um. So, it should converge for any r 0 and the convergence rate hence depends on the eigen value lambda of A; eigen value of A and condition number of the eigen vectors.

So, the method converges for any guess value x 0; convergence rate depends on condition value of the eigen vector matrix; not the not on the condition value of the matrix rather condition number of the matrix rather condition number of the eigenvector matrix and eigenvalues of A also.

So, this is a very important method because this is applicable for any general type of matrices. Now though this is an important method this is a first method the in implementation wise; it might be little complex if we compare with conjugate gradient where; if we look into the programming we will look soon the programming in conjugate gradient is much simple then the programming efforts from GMRES.

Now we will next few classes, we will explore that if conjugate gradient type of methods can be extended for non symmetric matrices. And we will get a algorithm called by conjugate gradient method in the subsequent classes we will discuss on that.

Thank you.