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

Lecture – 50
Lanczos Biorthogonalization and BCG Algorithm

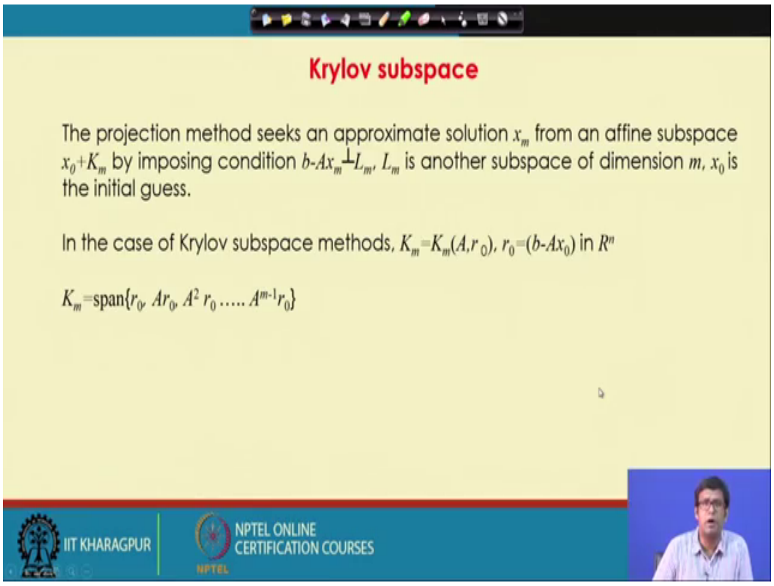
Welcome. So, we are looking into Krylov subspace based method and we have looked into a number of methods involving Arnoldi's, method full orthogonal method, generalized minimum residual method. And, then we looked into Krylov subspace methods which is same as Arnoldi's or full orthogonalization method, but specifically for symmetric matrices. And we saw that there are a number of interesting things happen for symmetric matrices especially we get a conjugacy of the auxiliary vector and the auxiliary vector comes due to the fact that the h matrix becomes a triangular matrix or symmetric matrix. And from there we have looked into Lanczos algorithm which has been modified as direct Lanczos and further into conjugate gradient method, which is a very efficient and very simple method for solving symmetric matrix problems.

Now, the goal of the present lecture will be and of course, this will be continued in subsequent lectures that whether conjugate gradient type simple and efficient solver can be designed for general any general matrix not only for the symmetric matrices. Earlier you have seen that most of the matrix solvers has certain limitations. In a sense that Jacobi gauss Seidel or (Refer Time: 01:45) are applicable only for diagonally dominant or irreducibly diagonally dominant systems. And conjugate gradient steepest descent are applicable for symmetric matrices.

If we look into symmetric positive definite matrices especially steepest descent; if we look into GMRES that is probably the only method which have looked into right now is applicable for more wider classes of matrix, even the matrices asymmetric GMRES can be applied; if any it is not strictly diagonally domain and GMRES can be applied. However, so our present target will be looking into general purpose matrix as the matrix may not be diagonally dominant, may not be symmetric, but the matrix has to be nonsingular for having a solution.

So, for any nonsingular matrix can we develop a method and we will explore Krylov space subspace methods more to look into that part. So, what we will look in this particular class is Lanczos bi orthogonalization analyzation which is a variant of Krylov subspace method and then how a conjugate gradient type of methods can be developed for general matrices for non symmetric matrices using Lanczos by orthogonalization.

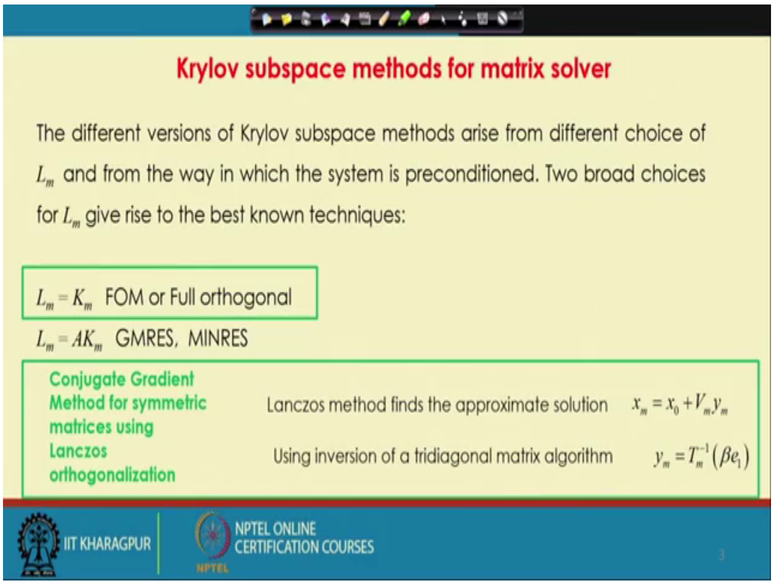
(Refer Slide Time: 03:11)



We have earlier done with Krylov subspace methods and the projection method for solving A x is equal to b using Krylov subspace method, is a method that seeks an approximate solution x m from an affine subspace x 0 plus K m by imposing the condition that the residual b minus A x m is orthogonal to L m.

L m is another subspace of dimension m, x 0 is initial guess, K m is also a subspace of dimension n in case of Krylov subspace method K m is the Krylov subspace of a and r 0, a is the matrix and r 0 is the initial residual minus x 0 in R n. And K m the Krylov subspace is defined as span of r 0 A r 0 A square r 0 to a to the power m minus 1 r 0 that is A m dimensional space in r m.

(Refer Slide Time: 04:05)

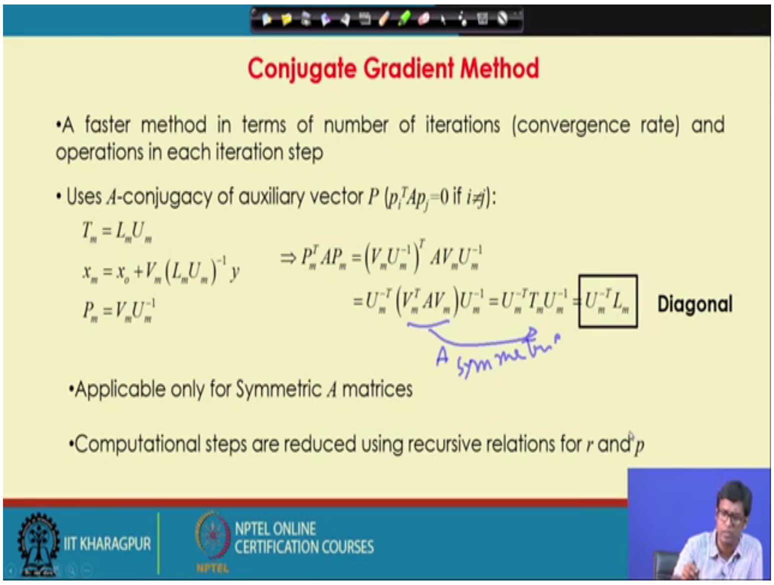


The different versions of Krylov subspace methods arise from different choice of L m and the way the system is preconditioned. Preconditioning will look in look later we have earlier discussed that this will come later. There are true broad two broad choices for L m which you have seen till now in the earlier examples of Krylov subspace methods; one is L m is equal to K m which is full orthogonal or method or FOM another is L m is equal to AK m which is GMRES or minimum residual method.

This is a complete orthogonal projection and the second one is an oblique projection. Now if we look into orthogonal projection in detail then we will see that for symmetric matrix we can get Lanczos orthogonalization and can get conjugant method, gradient method from full orthogonal method. Lanczos method will find the approximate solution as x m is equal to x 0 plus V m y m where V m is the basis of Krylov subspace. And y m can be obtained by inverting the tri diagonal matrix T m which comes from Lanczos orthogonalization of the Heisenberg matrix upper part of the Heisenberg matrix basically assumes a tri diagonal form we have looked into that.

So, y m is equal to T m inverse beta e 1. And we have seen that inversion of a tri diagonal matrix is simple and as well as TDMA type of algorithms can be used for direct inversion; there can be there are recursive relations which come and this makes the problem simpler.

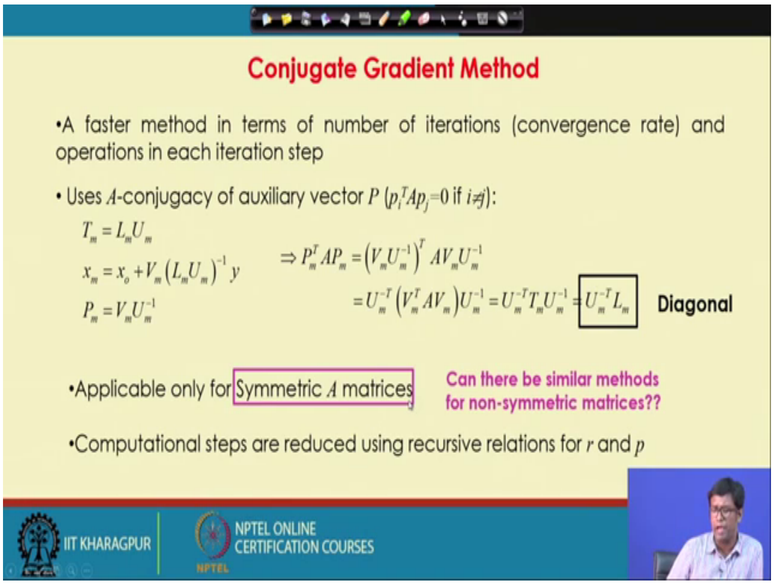
(Refer Slide Time: 05:45)



Conjugate gradient method is a faster method in terms of number of iterations or convergence rate it's the convergence rate is a function of root over of condition numbers. So, the convergence is faster and also the number of operations in each iteration step is smaller, because we can use something like a recursive relation for tri diagonal matrices here. A conjugacy of auxiliary vector P p i transpose A p j is equal to 0 is illustrated we will quickly see what is an auxiliary vector.

So, tri diagonal matrix T m can be decomposed as a lower and upper triangular matrix and the V m U m inverse this becomes an orthogonal this becomes the auxiliary matrix auxiliary vector matrix P m. And all the columns of P are mutually A conjugate to each other that is p i transpose A p j is equal to 0 if I not is equal to j. So, we can get P m transpose AP m is equal to U m inverse transpose L m and we can show that this is a diagonal matrix and that is called the A conjugacy of the auxiliary vectors. This is however, this A conjugacy arises only when A is A is a symmetric matrix asymmetric only for asymmetric matrices.

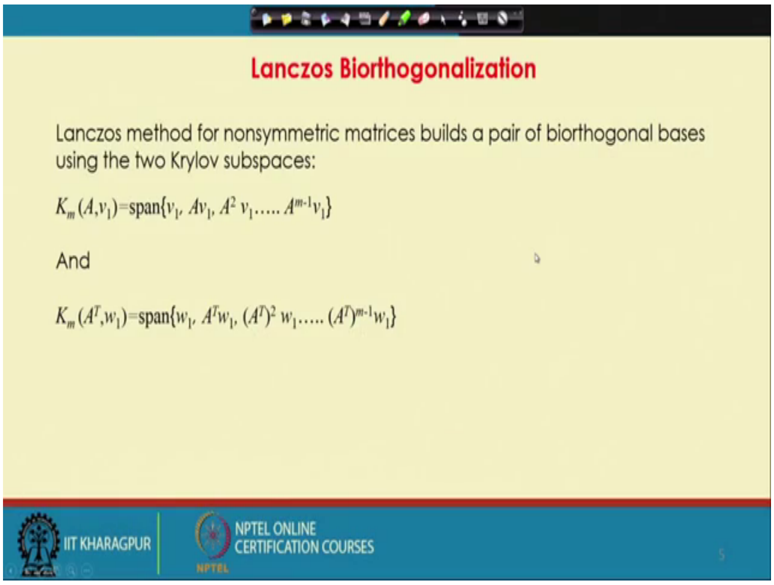
(Refer Slide Time: 07:39)



So, with all its advantages conjugate gradient is only applicable for symmetric matrices and the computational steps are reduced using recursive relations for r and p for residual as well as for the auxiliary vector we can use the recursive relations. So, it is only applicable for symmetric a matrices; and the question is that can there be similar methods for non-symmetric matrices also.

And for that we explore the other variants of Krylov subspace method in which we can some way handle the asymmetric method. And one idea is that that if we have a in the space we use the Krylov space of a for K m, can you use the Krylov space of A transpose as a L m so, that a plus A transpose is a symmetric matrix. So, that the asymmetricity of a is some way taken care of by the asymmetricity of A transpose a plus A transpose a symmetric matrix. So, can we can we pose it like the a problem like this and we will we will explore it.

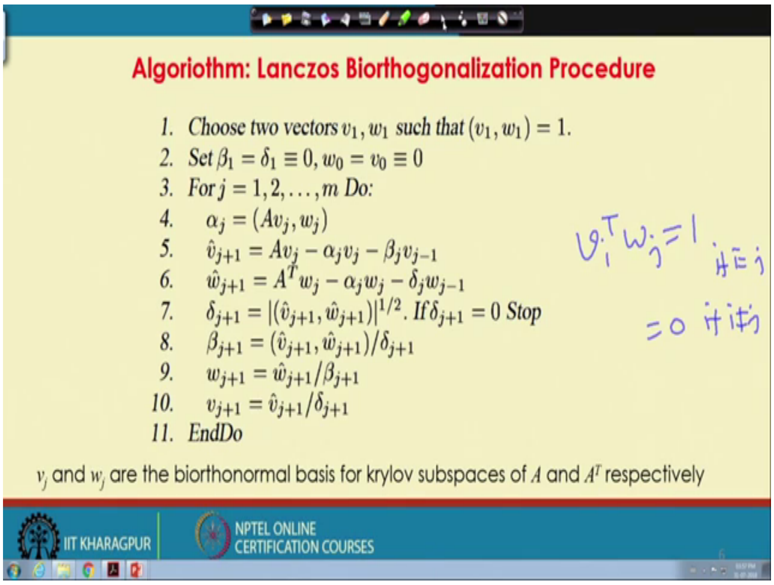
(Refer Slide Time: 08:33)



Lanczos by orthogonalization exactly looks into the Krylov space to Krylov subspaces; one Krylov subspace of one is of A another is of A transpose. So, it builds a pair of biorthogonal basis using the two Krylov subspaces came A v 1 which is span of v 1 A v 1 A square v 1 a to the power m minus 1, v 1 and K m of A transpose w 1 which is span of w 1 A transpose w 1 is square w 1 to a A transpose m minus 1 w 1.

Now, if by Lanczos biorthogonalization method, we get bases of k and K m a v 1 and K m A transpose v 1 Krylov subspace of a and A transpose and these bases are biorthogonal to each other. In that that is a sense that we take one basis of this v v i and we take one bases of this w w; v i transpose w will be 0 if v i or w j v i transpose w j will be 0 if i is not equal to 0. So, one v is conjugate to all other ws except that particular element of w all vector of w all other ws; there is a biconjugacy between v and w.

(Refer Slide Time: 09:51)

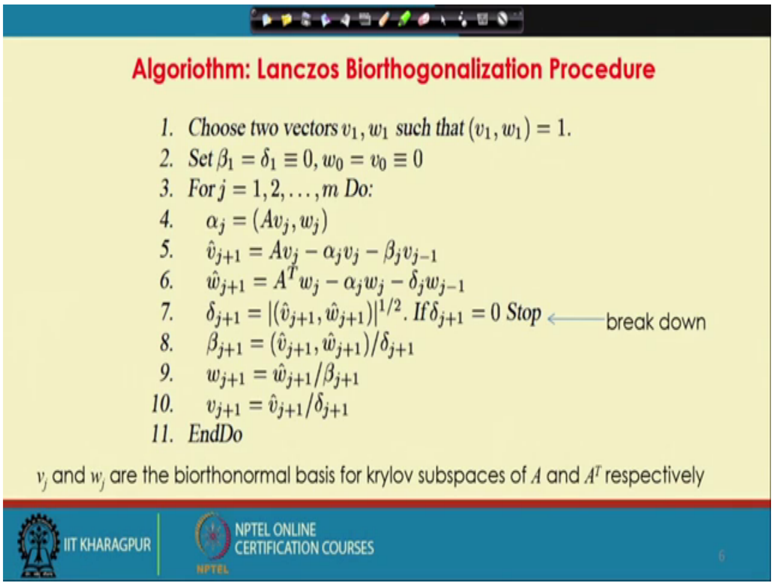


And there is an algorithm for that by which we can get it which is called Lanczos biorthogonalization algorithm it initially starts with any guess vectors v 1 and w 1 with their dot product being 1. So, their orthonormal in that sense the dot product is normal. And sets beta one is equal to delta 1 is equal to 0 and w 0 is equal to v 0 is equal to 0. And then uses something like a recursive relationship for v j plus 1 and w j plus 1 and gets the delta j plus 1 which is a dot product between v j plus 1 and w j plus 1 root of that, and then divides then beta j plus 1 which is this dot product divided by delta j plus 1 and divide w by that beta and v by delta and if delta is equal to 0 this j plus this stop. So, if the dot product between v and w is 0 this algorithm stops.

So, it starts with one take one particular v 1 w 1 and for next v is obtained as v is equal to A v minus alpha into v alpha is A v dot w minus beta into v j minus 1. So, from v the a a certain amount of j minus 1 of v is v j is subtracted as well as alpha j v j is subtracted where v alpha j comes by dot product of v j and w j and the similarly w j plus 1 comes. So, through this method it is seen that v j and w j are biorthogonal to each other which is v j and w j are bi orthonormals bases of Krylov subspace A and A transpose. So, v j transposed v j dot w j 1 is equal to 1, and v j dot w v j dot w j is equal to one, but v j dot w i is 0 if I is not equal to 0.

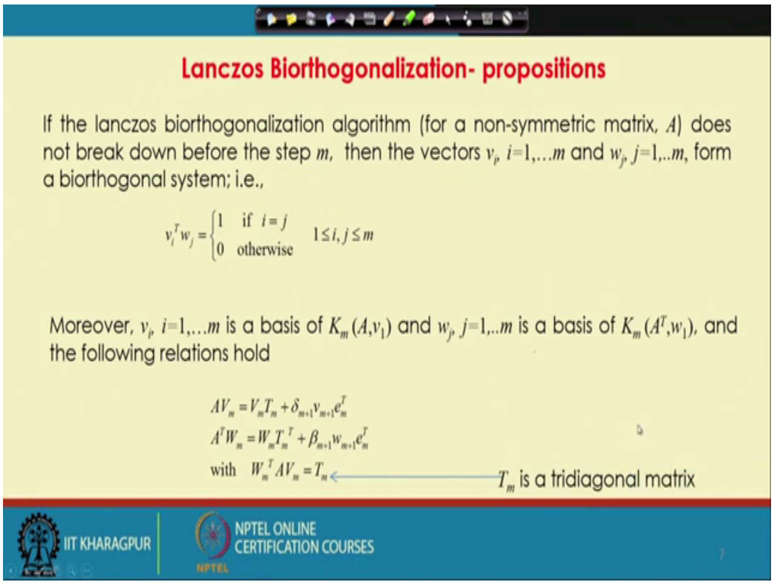
That is the property of biorthogonality or we can write v i transpose w j is equal to one if i is equal to j is equal to 0 if I is not is equal to j and that is hence ascertained by this particular algorithm.

(Refer Slide Time: 12:14)



And the process breaks down once we get the product between v and w for the particular j to be 0 that means, there is no further independent basis of the Krylov subspaces which can be generated all the independent bases vector all the independent vectors have been found out or we have been we have calculated the entire bases of these two spaces A and A transpose. This took the Krylov spaces of A and A transpose.

(Refer Slide Time: 12:45)



Lanczos biorthogonalization is follows this proposition if Lanczos bi biorthogonalization algorithm for a non symmetric matrix A does not break down before the step m that; that means, before the step m V m transpose v i transpose w i is non-zero. Then for vectors the vectors v i, i is equal to 1 to m and w j j is equal to 1 to m form a biorthonormal system, that is v i transpose w j is equal to 1 if i is equal to j and otherwise it is 0, and this is ascertained by the way the Lanczos algorithm has been devised.

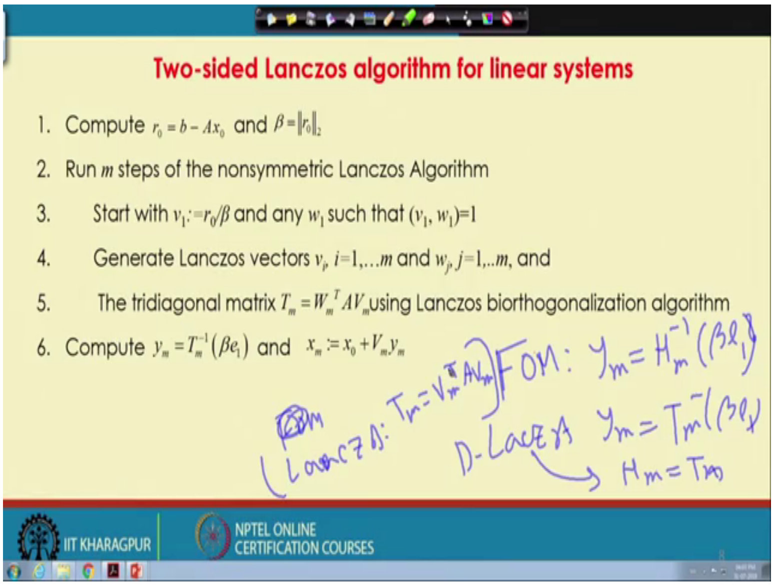
So, it is the algorithm is for finding out basis of A transpose v and A transpose w and Krylov subspace between A transpose and w and a and v and there are these Krylov subspaces are can have any bases any m bases, but the basis our vectors are found using Lanczos by orthogonalization algorithm in a way that, v i transpose w j is equal to 1 or 0 if i is equal to j they are 1 otherwise there is 0 and v and w are biorthonormal bases of this two kyrlov subspaces. This looks little abstract, but we will see that this is a great utility when you will try to define derive an algorithm from here.

Moreover v i i is equal to 1 to m is a bases of K m A v plus 1 Krylov subspace of a and v 1 and w j j is equal to one to m is the bases of K m A transpose and w 1 and the following relations hold that A V m is equal to V m T m delta m plus 1 delta is defined to Lanczos algorithm biorthogonalization algorithm v m plus e m transpose e m is the first unique vector.

A transpose w m is w m T m transpose beta m w m plus 1 e m transpose T m is a tri diagonal matrix. So, its transpose is also a triangular matrix with that now the T m is defined such that W m transpose AV m is equal to T m. W m is the matrix containing as its columns all the vectors which are bases of the Krylov subspace of A transpose all the ws.

Similarly, V m is the matrix which contains all the bases of Krylov subspace of a all the v s. And W m transpose A V m is a tri diagonal matrix and this is also this is also a very evidently very easily a apparent because w and v are biorthonrmal w m transpose V m must be a W m transpose A V m must be a tridiagonal matrix. So, T m is a tri diagonal matrix.

(Refer Slide Time: 15:49)



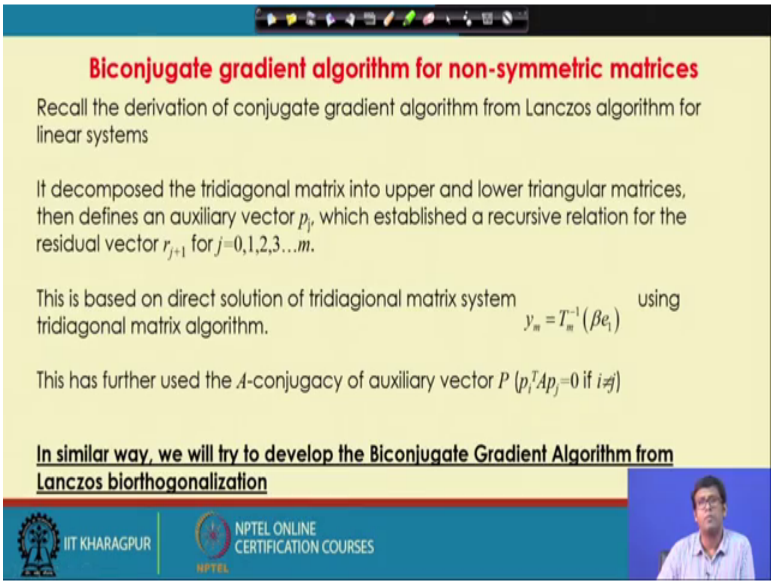
Two sided Lanczos algorithm is found is a device for linear systems or for solving A x is equal to b compute r naught b minus A x 0 and beta is equal to l 2 norm of r naught. Run m steps of non symmetric Lanczos algorithm start with v 1 is equal to r 0 by beta and any beta. So, that v 1 dot w 1 is equal to 0. Generate Lanczos vectors using Lanczos algorithm v v i n w i find the tri diagonal matrix T m and then compute y m is equal to T m inverse beta e 1. So, T m is the again the same as the Heisenberg matrix upper part of the Heisenberg matrix which is a tri diagonal matrix we found out in full orthogonal method.

So, similarly you write T m is equal to W m transpose A V m and compute y m is equal to T m inverse beta e 1 and x m is equal to x 0 plus V m y m. Now if we can remember that h m the relationship for full orthogonal method was Y m is equal to H m inverse beta e 1 and this and in Lanczos say D -Lanczos source or Lanczos for linear systems will say that Y m is equal to T m inverse beta e 1 because H m and T m H m is T a min Lanczos method. The Heisenberg upper part of the Heisenberg matrix is same as that as as is a tri diagonal matrix here how was full orthogonal method coming? It was appearing due to the fact that our K m and L m are same.

Here we are getting a different tridiagonal matrix, but we are getting a similar relationship due to the fact and here T m was if we look into the conjugate gradient method, T m was defined in a different way. But, it was coming from the fact that l K m the Krylov subspace vectors or T m was defined in FOM or in rather not I will write Lanczos, the Lanczos for symmetric matrix T m was defined as V m transpose A V m and this is because V m was orthogonal to the residual vectors. V m is where they x is updated V m is also the subspace V m is also the bases of the L m, V m is also orthogonal to the residual vector.

Here instead we get a w m transpose. So, w m transpose must be orthogonal to the residual vector here; or what we can say or w m must be orthogonal to residue vector here or what we should see say is that that the Krylov subspace of a m transpose r 0 will be the a L m for Lanczos biorthogonalization methods or a variant of that. We will look it into a little more elaborative way.

(Refer Slide Time: 19:10)

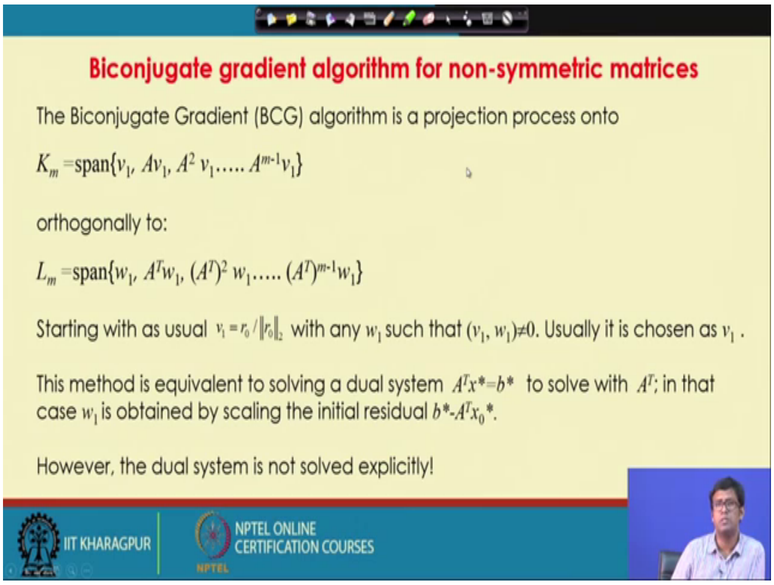


That recall biconjugate gradient algorithm for non symmetric matrices. Let us recall the derivation of conjugate gradient algorithm for Lanczos method. If it decomposed the tri diagonal matrix into upper and lower triangular matrices then defines an auxiliary vector p. So, tri diagonal matrix is decomposed in L m u m or then auxiliary vector p is defined, and a recursive relation is established for the residue vector r j plus 1.

This is based on the direct solution of tri diagonal matrix system Y m is equal to T m inverse beta e 1 using something like a TDMA type of algorithm which is a direct algorithm. This is further used a conjugacy of orthogonal auxiliary vector p transpose A p is equal to an identity vector if i is not is equal to j p i transpose a p j is equal to 0.

In similar way we will try to now we will now try to develop a bi conjugate gradient algorithm from Lanczos biorthogonal narration. The idea is that y m is calculated in similar way and there is an e conjugacy now there is a conjugacy here there were conjugacy between the bases vectors of v v i transpose v j is equal to 0 if i is not equal to j. Here there is a conjugation between w and v, we will utilize these facts.

(Refer Slide Time: 20:37)



The biconjugated gradient algorithm is a projection process on to the Krylov subspace K m which is a span of v 1, A v 1, A square v 1, A to the power m minus 1 v 1 and it starts with v 1 is the initial residual vector b minus r a a x zeros unit vector on that. Orthogonal to thus (Refer Time: 20:59) the residue vector must be orthogonal to L m which is span of w 1 A transpose w 1 A transpose square w 1 A transpose m minus 1 w 1.

So, now L m is span of Krylov subspace of A transpose and that is a way we are taking any general matrix which is non symmetric matrix, and trying to build an algorithm which was earlier built for a symmetric matrix. For conjugate gradient L m was Krylov subspace of a only, but for biconjugate gradient we will see for non symmetric matrix. So, similar algorithm will come keeping in mind that L m is now not the Krylov subspace of a m a rather it is the Krylov subspace of A transpose.

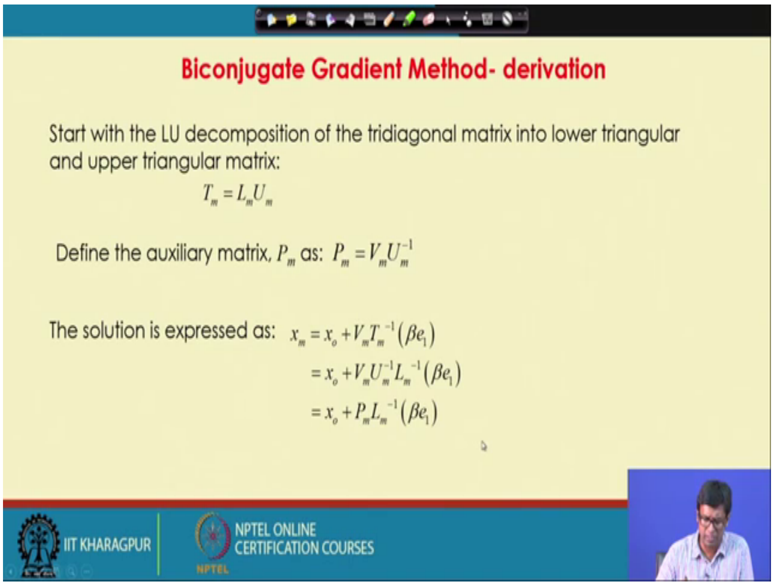
So, we will start with as usual v is equal to r 0 by mod r 0, and that is the unit vector along the first residue direction and take a w such that v 1 transpose w 1 is non-zero. As their orthonormal usually take w is equal to v. The method is equivalent to solving a dual system A transpose x star is equal to b star along with a x is equal to b. And to solve with a A transpose in that case w 1 is a obtained by scaling the initial residual b star minus A transpose x 0 star.

Instead of if we actually have to solve the dual system, we are not solving the dual system, but in the back of this processes this dual system is also being solved. Krylov subspace of a is my K m space here Krylov sub space of A transpose is the L m here. If I take another equation A transpose x is equal to some b star A transpose x star is equal to b star the element K m is just reversed.

So, if I can solve one equation it is identical of solving the other equation also, which is equation for the transpose of a matrix. If we as we can solve that both the equations together, but you we, but as the given problem is a x is equal to b for us we do not solve the transpose equation, which is type of solved in the back of the algorithm.

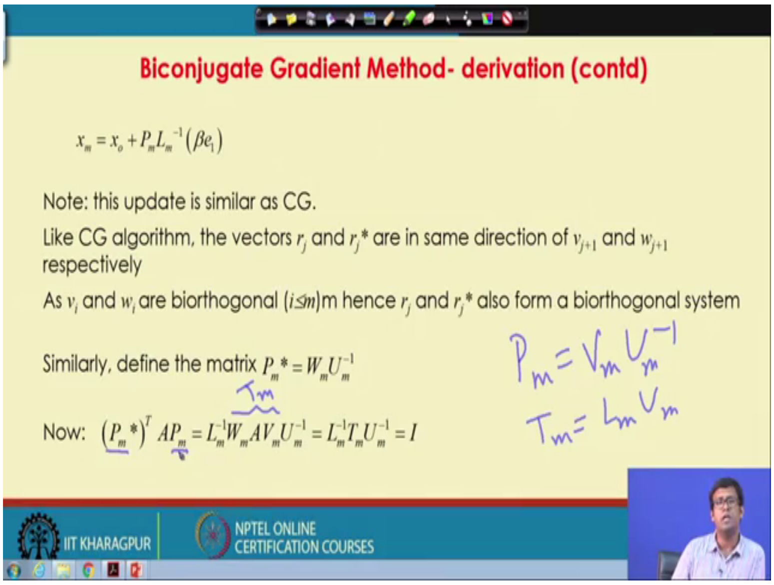
But if we actually have to solve the transpose equation also, able to solve a dual problem a x is equal to b and A transpose x is equal to b star A transpose x star is equal to b star. We have to start with w is equal to b star minus A transpose a x zeros. However, as we are not solving here, it is well we can take any w 1 to start with usually it is chosen as v 1. However the dual system is not solved explicitly.

(Refer Slide Time: 23:42)



So, for the derivation we start with the LU decomposition of the tri diagonal matrix into lower triangular and upper triangular matrix T m is equal to L m U m. We define an auxiliary matrix P m as P m is equal to V m U m inverse. It is very same as it we have defined it in the conjugate gradient case. The solution is expressed as x m is equal to x 0 plus V m T m inverse beta e 1; now T m inverse is u m inverse L m inverse and V m u m inverse is P m. So, x 0 plus P m L m inverse beta u 1 so, this is inversion of a lower triangular matrix multiplied by a and auxiliary vector matrix P.

(Refer Slide Time: 24:30)



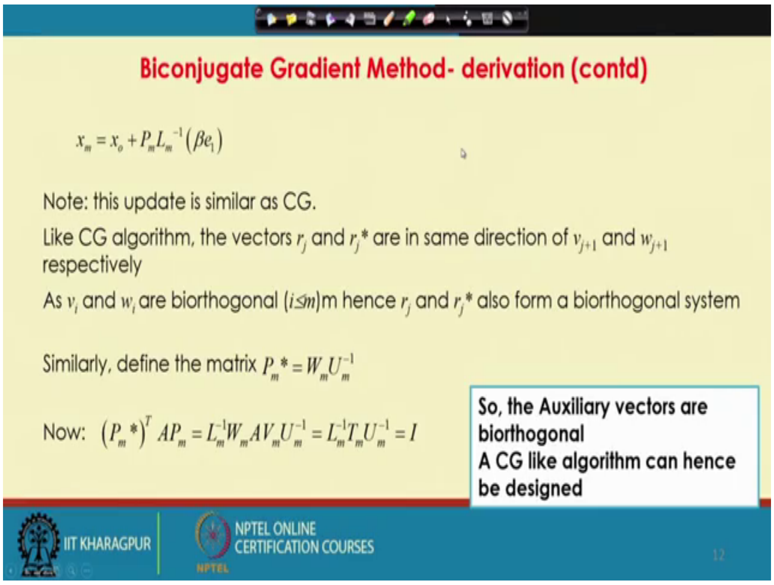
X m is equal to x 0 plus P m L m inverse beta e 1 which is the this one. This update is very similar as conjugate gradient update. Like conjugate gradient algorithm the vector r j and r j star are in same direction of v j minus 1 plus 1 and w j plus 1 right. Because the residue vector is orthogonal to w and what is orthogonal to w? Is v. So, residue vector is along v j plus 1 r j is orthogonal to w j plus 1. So, it should be along v j plus 1 r j star similarly will be along w j plus 1.

So, as v j v i n w are biorthogonal for I is less than m, hence r j r j star should also form a bi orthogonal system. Similarly we can define a matrix P star which is W m U m inverse. P was defined as P m was defined as V m U m inverse. So, the star matrix that is the matrix for the transfer transpose equation part P m star is W m U m inverse. And very interestingly we can show that P m star AP m is L m inverse W m AV m U m inverse and W m A V m W m AV m this is nothing, but the tri diagonal matrix T m.

So, L m inverse T m U m inverse of a lower triangular matrix with a tri diagonal matrix and inverse of the upper triangular matrix and now we can we all we have also seen that T m is equal to L m U m T m is decomposed as L m U m. So, you multiply this we will get the identity matrix. So, P m star and P m the auxiliary vector matrix for A and for A transpose are a conjugate. Earlier for conjugate gradient we have seen that P m is itself a conjugate, but here we are getting P m star and P m there a conjugate.

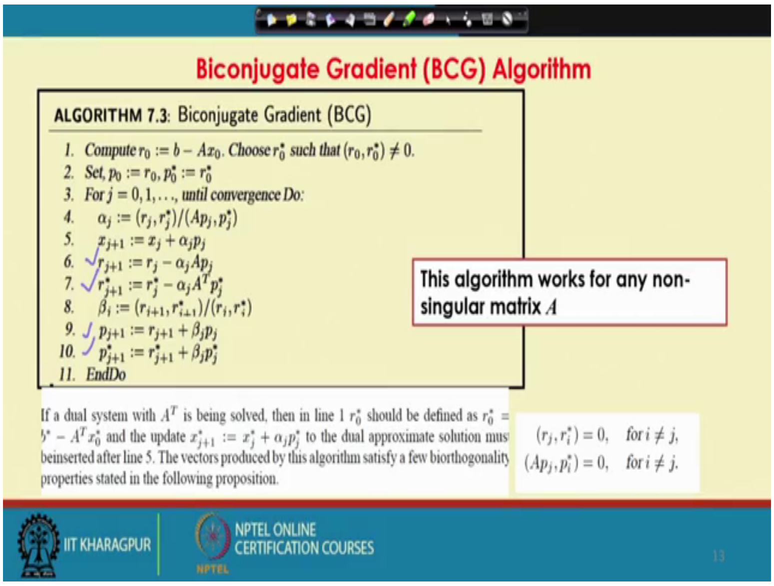
So, we got biconjugacy a (Refer Time: 26:41) a by conjugacy biconjugacy of r j and r j star residue vectors and we got by a conjugacy of auxiliary filters. Now the problem is exactly same as conjugate gradient method we can use similar type of iterate recursive relations and form the algorithm.

(Refer Slide Time: 27:03)



So, auxiliary vectors are biorthogonal as CG like algorithm hence can be designed.

(Refer Slide Time: 27:08)



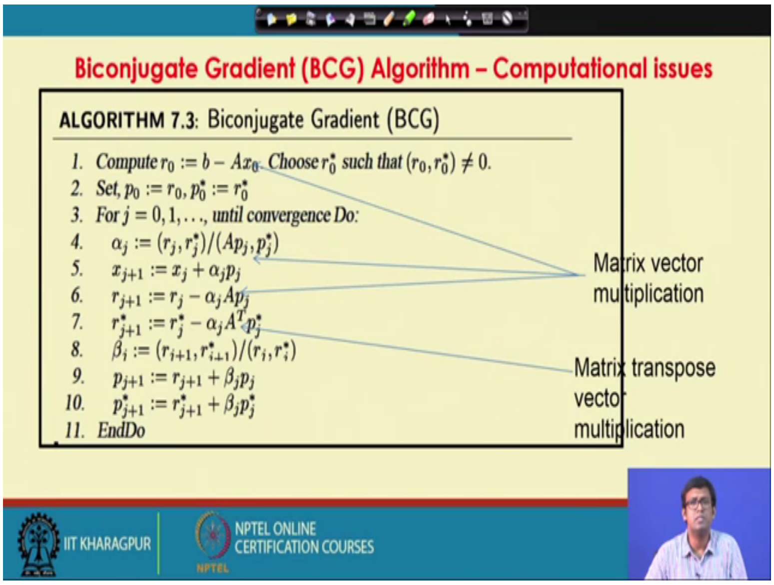
And we get the biconjugate algorithm biconjugate gradient algorithm the initial residual vector r 0 is equal to b minus x 0 we choose r 0 star, such that the dot product for the initial vector r 0 and r 0 star is non zero. Then p 0 is equal to r 0 p 0 star is equal to r zero. So, r 0 and r 0 star are non 0 dot product, but r 0 and r one star will be 0 that that is the idea by conjugacy. So, j is equal to 0 or two convergence find alpha j which is r j transpose r j start by A p transpose p star these are biorthogonal and these are biconjugate update x j plus 1 as x j plus alpha j p j r j plus 1 as r j minus alpha j a p j r j plus 1 star is equal to r j star minus alpha j A transpose p j star get beta i beta j and get p j plus 1 using beta j and p j star plus 1 p star j plus 1 using beta j and r j plus 1.

So, similar recursive relations as conjugate gradient, but now it is for both r star and r p and p star. But for x we have only finding out x j plus 1 because we not interested in finding the solution of the transpose equation you are not finding out x extra j plus 1; and after certain iterations it should converge and we get the converse solution.

The algorithm works for any nonsingular matrix a and the convergence etcetera can be shown as same as the conjugate gradient type of algorithm which; that means, these are first converging algorithms. If the dual system A transpose is being solved, then in the line 1 r 0 star should be defined as r 0 star is equal to b star minus A transpose x star and then x j star plus x j plus 1 star has to be updated from x j star for dual approximate solution after the line 5.

The vectors produced by this algorithm satisfy about five orthogonality properties like r j star r r j dot r i star is equal to 0 if i is not is equal to 0 A p j dot p p i star is equal to 0 if i is not equal to 0, this these are the base of those are proposition based on which this if this method is developed actually. The recursive relations come and then using this we can update the vectors these vectors the updated vectors also must satisfy this particular properties.

(Refer Slide Time: 29:54)



Now, there is a particular issue with biconjugate gradient algorithm, which we will discuss before we finish this lecture and we look into how to solve this issue in the subsequent lectures. That is if we look into the BiCG algorithm, that there are three matrix vector multiplication. If the first one is before the initial step, but during the iterations also there are two matrix vector multiplication Ax 0 is the first one, A p j is one multiplication and then A p j is again used another matrix vector multiplication. On top of that there is one more matrix vector multiplication with A transpose p j and transpose multiplications are difficult in terms of communicating with the memory and the processor.

So, number of at least A transpose p j a p j three matrix vector multiplication then p j is to be multiplied with p star, the number of calculations operational steps are usually more and multiplication with the auxiliary vector and a is also more and we have to also multiply with the transpose vector. And for that the as the number of operational steps are more the in during each iteration, the number of the amount of round off error is also more. And due to this round of error in the convergence of this BiCG or biconjugate gradient method we see there are irregularity.

It does not converge in a smooth way or monotonically there are fluctuations during convergence and sometimes these fluctuations can be large enough. So, that the convergence is disturbed for certain case of the problems they therefore, small perturbations we can see lot of change in the result due to this round off error related fluctuations.

So, to look into more stabilized versions of BiCG method, where this many matrix vector multiplications can be avoided and BiCG stabilized method is one of that method which we will discuss now in that the subsequent classes. And there are few other methods using some polynomial formulation of the auxiliary vector and residual vector polynomial expansion type of formulation of the auxiliary vector we can minimize the matrix vector products, and can get good recursive relations and much simple operations for in during each iteration and better algorithms can be devised.

In the next classes we will look into the much developed algorithms. Keeping in mind that BiCG is the basic algorithm for any non symmetric matrix, which will be now developed using certain polynomial expansions to BiCG stabilized type of algorithms. We will look into it in the next lesson.

Thank you.