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

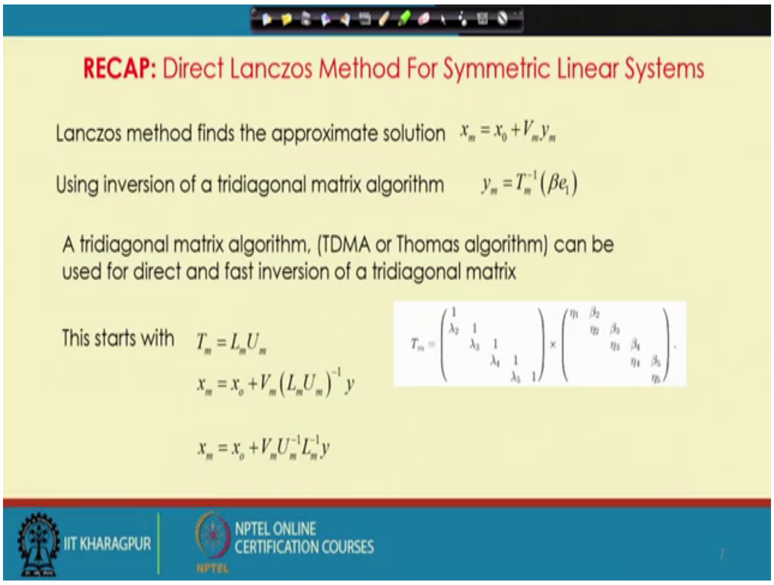
Lecture – 47
Conjugate Gradient Methods (Contd.)

Welcome. So, we are discussing about Conjugate Gradient Method and actually trying to develop a method based on our previous knowledge of full orthogonal methods and direct Lanczos algorithm for tridiagonal method in Krylov subspace method.

So, started with giving a initial given an initial background on Krylov subspaces, and then looked into Arnoldi modified gram Schmidt and full orthogonal method algorithm. And obtained the relationship between the residual and the basis of residual at mth step and the m plus oneth basis of Krylov subspace and observe that how it can be obtained for a tridiagonal for a symmetric matrix a where the Heisenberg matrix h is that of Tridiagonal form.

This is what we obtained in the previous lecture. So, you can I will quickly look into the symmetric linear system solver in Krylov subspace which we have discussed earlier direct Lanczos method.

(Refer Slide Time: 01:24)

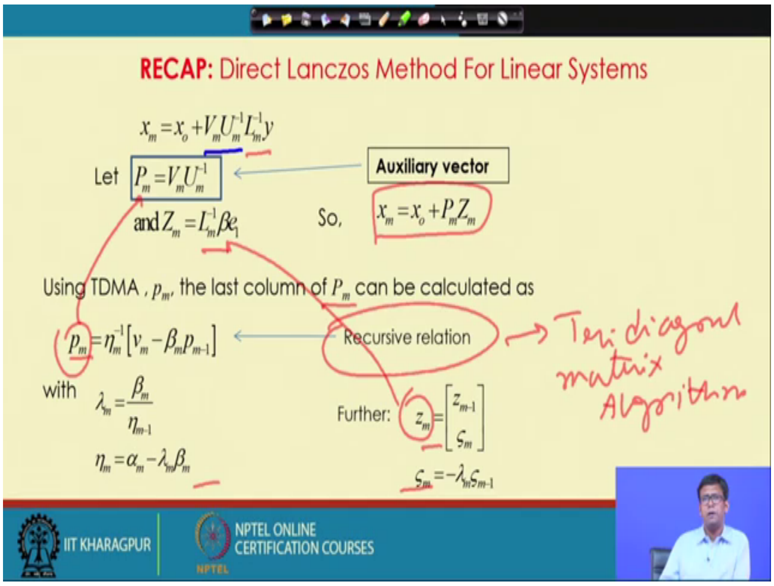


The Lanczos method finds an approximate solution of x m is equal to x 0 plus V m Y m which is a Krylov subspace method. And this x m is equal to x 0 plus k m A comma r 0. Or let me write it down; this is same as x m is equal to x 0 plus k of a r 0 k m. And this is a Krylov subspace method. And it uses inversion of tridiagonal matrix T m where Arnoldis method will give us H m. And H m is equal to T m for symmetric matrix H m is T ma tridiagonal matrix and this comes from the full orthogonal method.

So, essentially it shows that this is an orthogonal method. Where Y m is found out by inverse of the Heisenberg matrix which is a off tridiagonal matrix now. And a TDMA type of algorithm tridiagonal matrix algorithm can be used for direct and first inversion of the tridiagonal matrix. And this starts with the step that T m the tridiagonal matrix is product of a lower triangular matrix. So, you can write T m is equal to L m into U m, and then upper triangular matrix. And we can substitute it x m is equal to x 0 plus V m Y m is T m inverse b y.

So, this is T m inverse y Y is Y m here. And further one step gives us x m is equal to x 0 plus V m V m inverse L m L m inverse y this is this is what we have done in D-Lanczos algorithm earlier.

(Refer Slide Time: 03:34)



And x m is equal to x 0 plus V m U m inverse L m inverse y. We can now say that this V m U m inverse this will put this is what we do in a TDMA type of algorithm. We have divided this as U m inverse L m inverse and right two different forms of the matrices P m is equal to V m U m inverse and Z m is equal to L m B m inverse.

So, you can write x m is equal to x 0 plus p m Z m. and then we find an iterative relation for calculating P m which is the which is the last column of P m based on; and we will change the values of m recursively and get different columns of the matrix P m. And using this recursive relation we can calculate what is P m from there we get V m u m inverse. And similarly Z m is also calculated using a recursive relation.

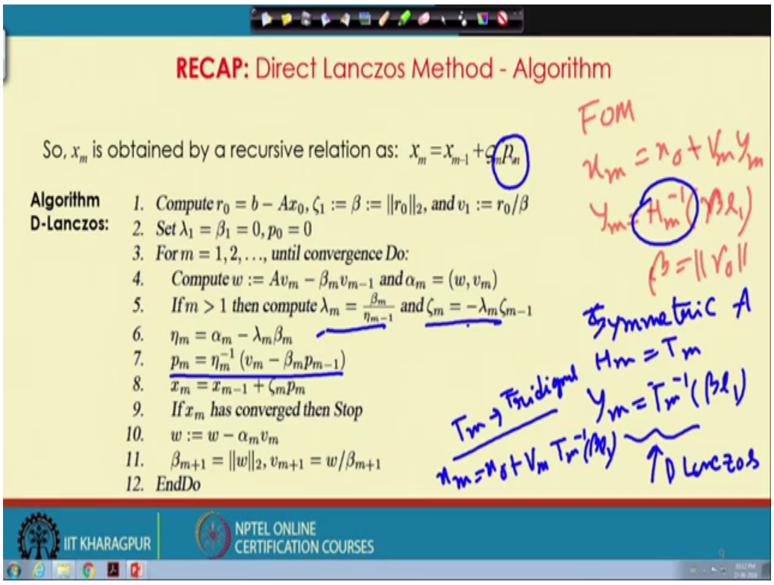
So, this recursive relation is heart of a tridiagonal matrix algorithm. And we have done it in very detail when discussing about direct solvers for tridiagonal matrices which is tridiagonal matrix algorithm or Thomas algorithm. Remember, when we did looked into finite difference approximation of one dimensional heat conduction equation or one dimensional Laplace equation. And got a tridiagonal matrix and then observed that instead of doing gauss elimination a TDMA using recursive relation we can write a td write an TDMA algorithm which you will give us very first solution. So, this can be utilized here.

So, this is this is how the direct Lanczos method works that using this recursive relationship you find Z m and P m and then you substitute P m here and Z m here. And then you write x m is equal to x 0 plus b m Z m; one important vector that we construct in this method is the auxiliary vector P m we are using a recursive relationship for finding out P m. And we can do it in a more elegant way we can probably get a very much better solution here.

The problem with recursive there is one very significant much significant problem with recursive relation that we need in today's talk about parallel computing etcetera recursive relations do not work in parallel computing. So, if this recursive relation can be substituted by some other method we will probably get a better solution algorithm.

So, we focus on looking into the auxiliary vector P m. And interestingly some further characteristics of auxiliary vector will come out in a while.

(Refer Slide Time: 06:50)



So, in the direct Lanczos algorithm the x m is obtained by a recursive relationship as x m is equal to x m minus one plus xi m P m and P m and xi m are found using recursive relations.

Now, with this background we will probably try to look into more detail on the properties of the auxiliary vector P m, because till now till coming into Lanczos algorithm we were only interested about the residual r and new residual r m initial residual r 0 and V m what are the bases of Krylov subspace and the Heisenberg matrix H m H m and V m comes out of the Arnoldi modified gram Schmidt method. And based on my initial guess I can find out what is the initial residual r 0.

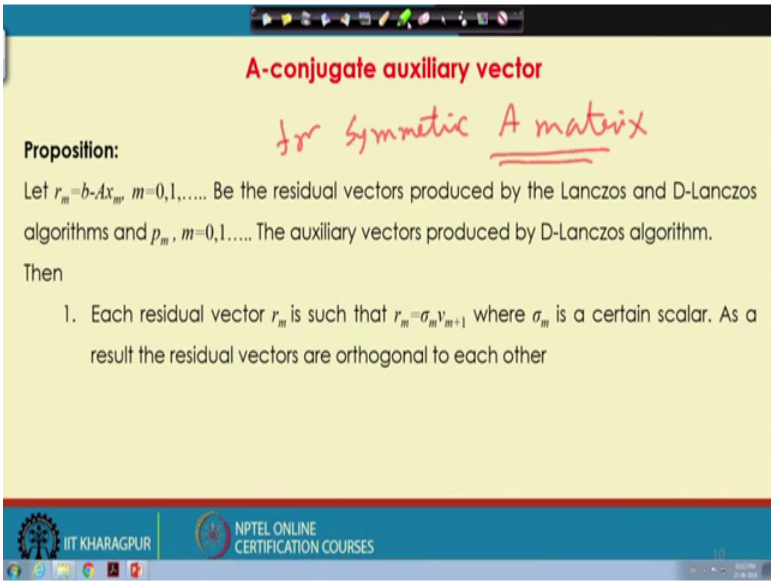
If we have r 0 H m and V m we full orthogonal method tells that we should be able to find out f m is equal to x m is equal to x 0 plus V m y m right. So, full orthogonal method tells us x m FOM full orthogonal method is equal to x 0 plus V m y m where y m is equal to H m inverse of basically beta e one where beta is equal to mod of r 0.

The problem in the full orthonormal method that we have to find out inverse of a upper triangular matrix which is not as probably not as costly as a general matrix A because a can be of any shape and this is an upper triangular matrix. However, this will involve certain number of steps. At least the steps for required for back substitution of a gauss elimination process. So, instead now for try for symmetric A H m is equal to T m and y get we get y is equal to T min verse beta e one which is now the d Lanczos method which we are discussing here. And this is this T m is tridiagonal.

So, inverting this T m needs the recursive relationships and recursive steps like find out xi m find out lambda m find out P m all these regressions are related for inverting T m efficiently. Now this particular P m, we have introduced for while we are looking into inversion of T min versions instead of writing a writing it x m; x m is equal to x 0 plus V m T m inverse beta e 1 I write x m is equal to x m minus one plus xi m p m.

So, this is a new vector or an auxiliary vector which we are introducing at this stage. Let us look into the auxiliary vector in to little more detail.

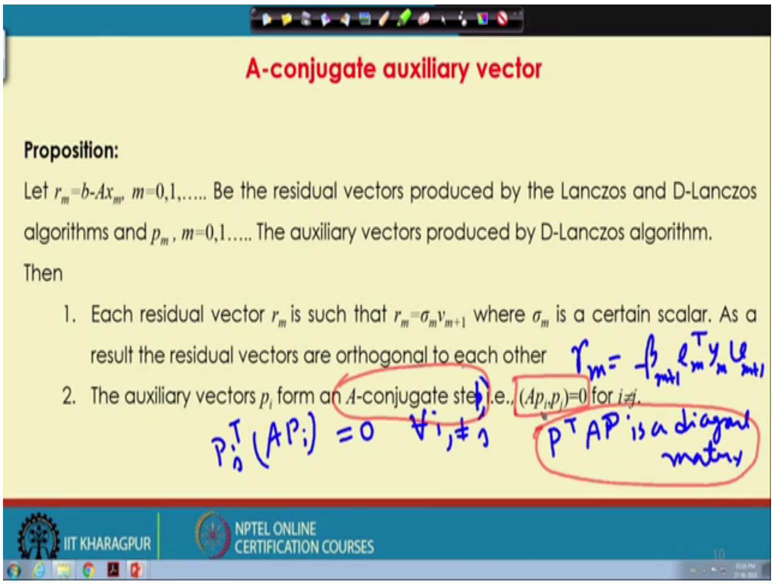
(Refer Slide Time: 10:30)



And there comes a proposition that r m is equal to b minus A x m for a different values of m be the residual vectors produced by Lanczos or D-Lanczos algorithm. That is for symmetric a matrices right, when we write Lanczos and D-Lanczos we have ensured that for symmetric A matrix. This is important, its only for symmetric matrix. And p m is at different values of m 0 2 0 1 up to larger numbers. P m are the auxiliary vectors produced by D-Lanczos algorithm.

And then each residual vector r m is such that r m is equal to sigma m V m plus 1; where sigma m is certain scalar. As a result the residual vectors are orthogonal to each other. Exactly this are this we have discussed of in the last session that r m gets the form.

(Refer Slide Time: 11:46)

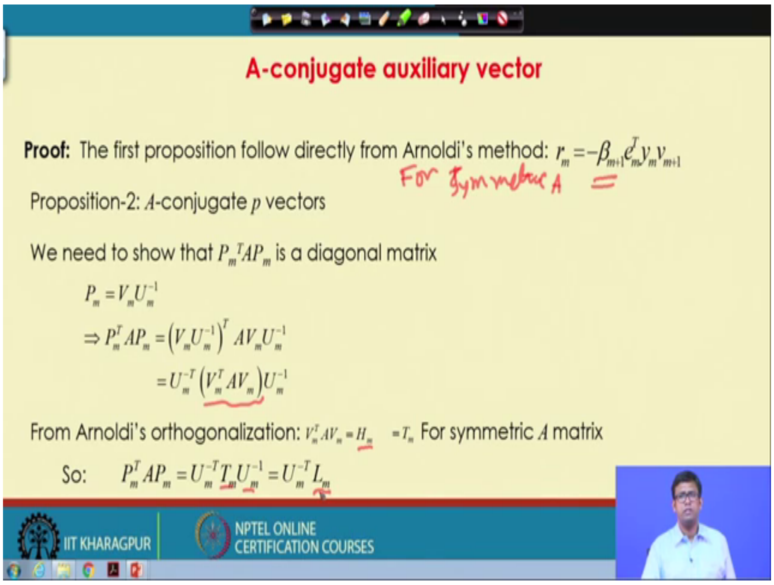


Let me write it down for you: r m r m is equal to minus beta m plus 1 e to the power m T Y m V m plus 1; this one. And there is another proposition which is more important the auxiliary vectors p i for an a conjugated step; from an a conjugate step. That is A p i p j is equal to 0 for i not is equal to 0.

So, A P i and if we take it is dot product with P j this is equal to 0 for all i not is equal to j. Or we can write p transpose A P is a diagonal matrix. That is only when P i transpose A P i that row will come that will give a diagonal one value. And for all other combination of i and j it will give us zeros. So, this is called a conjugate step here. This is the use a different pen this is a conjugacy that A P i dot p j is equal to zero. Now, we will show that it is actually coming to 0..

Or p transpose A P p transpose A P is a diagonal matrix. So, p is are being utilized for deriving the recursive relation. Now what we can see and they are also following a recursive relation what you can see is that before we instead of going into the recursion probably if I can have 1 p i can get idea about the other p s simply by the fact that each p is orthogonal to the other a A p multiplied with A p transpose A p is equal to 0 in case of i is not equal to j or p transpose A p the capital p transpose A p matrix p is the columns of small p and the columns of capital p p transpose a p is a diagonal matrix.

(Refer Slide Time: 14:32)

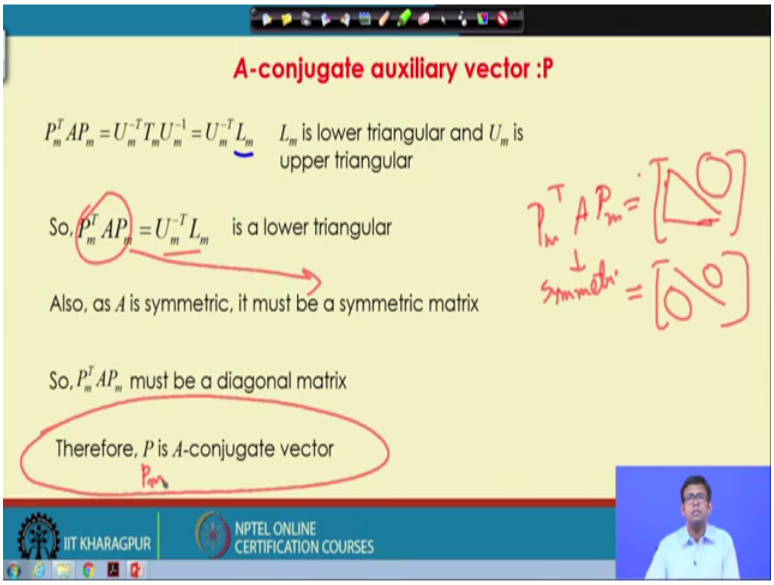


The first proposition followed directly from Arnoldis method which we have seen earlier. And Arnoldis method for everything we should write for tri for symmetry A where you can so write beta right. And the proposition 2 A conjugate three vectors. We need to show that p m transpose A P m is a diagonal matrix; P m is V m U m inverse that is how the auxiliary matrix auxiliary vector matrices defined p m transpose P m is V m U m inverse transpose a V m. So, a b transpose is b transpose a transpose that will be U n inverse transpose into V m transpose A V m.

Now, from Arnoldis method what we have seen is that V m transpose A V m is equal to the Heisenberg matrix H m. And which is same as the tridiagonal matrix T m for symmetric matrix. So, what we get? P m transpose A P m is equal to U m inverse transpose T m U m inverse T m is a tridiagonal matrix; U m is a our triangular matrix.

So, T m U m inverse gives us a lower triangular matrix L m this can be verified very easily by looking into the properties of triangular matrix. And that inverses and transposes also.

(Refer Slide Time: 16:19)



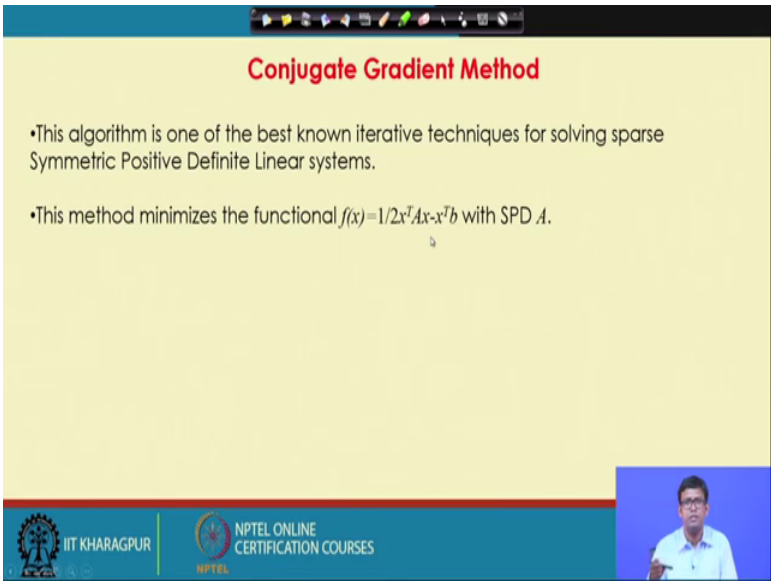
So, P m transpose A P m is equal to U m inverse transpose U m inverse. And T m U m inverse is because U m is an upper triangular matrix. So, T m U m inverse will be a lower triangular matrix L m. So, P m transpose A P m is equal to U m inverse L m; inverse transpose L m. L m is a lower triangular matrix U m is the upper triangular matrix. So, U m transpose inverse multiplied by L m that also gives us a lower triangular matrix.

So, P m transpose A P m is a lower triangular matrix. Now we will see that A is a symmetric matrix. So, if we take p look into P m transpose A P m this is also symmetric matrix. A symmetric matrix is of lower triangular form. So, A is symmetric therefore, P m transpose A P m must be a symmetric matrix. P m transpose A P m a symmetric matrix this is P m transpose A P m which is said to be of a lower triangular form. So, these are zeros here all the upper above diagonal terms are 0 here. This is also symmetric. So, the numbers beyond that below the diagonal must also be 0, so this should have the form of the diagonal 0 and 0.

The symmetric matrix is also a both upper and lower triangle upper triangular as well as lower triangle. So, lower triangular matrix can only be symmetric same for an upper triangular matrix it can only be a symmetry when it is a diagonal matrix. Therefore, P m transpose A P m must be a diagonal matrix. And we get the fact that P or small p is the A conjugate vector or I will write small instead of p all the columns P m are A conjugate vector.

So, now how can we utilize this for deriving a new solution algorithm?

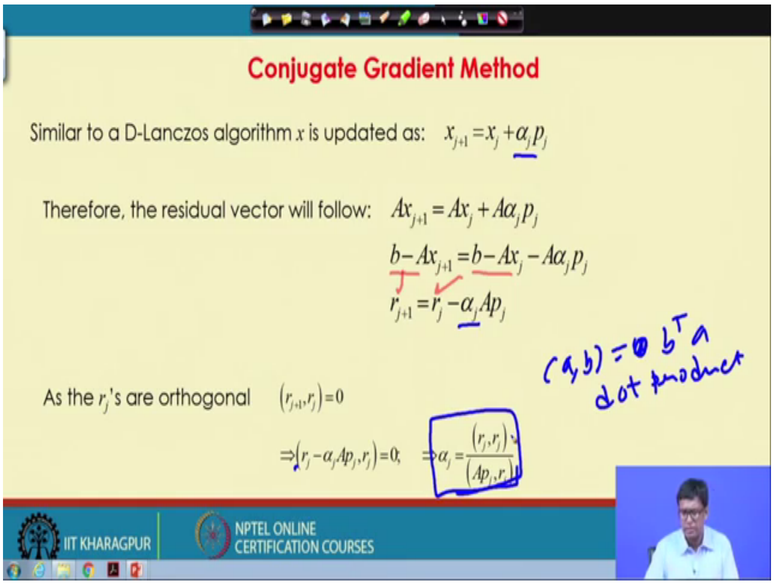
(Refer Slide Time: 18:59)



This is called the Conjugate Gradient Method. The algorithm is one of the best known iterative techniques for solving sparse system positive definite linear; sparse take a for sparse lot of symmetric positive definite linear matrices. Symmetric positive definite matrix which are sparse can be solved by this method and this is one of the best known methods we will show it in a while.

This method minimizes the function f x is equal to half of x transpose a x minus x transpose b for symmetric positive definite matrix A. There is (Refer Time: 19:39) general projection method. So, if the idea comes from the steepest descent type of algorithm that it should minimize a function along a particular direction. And minimize direction of f x is we can show it very easily is the solution of x is equal to b.

(Refer Slide Time: 19:53)

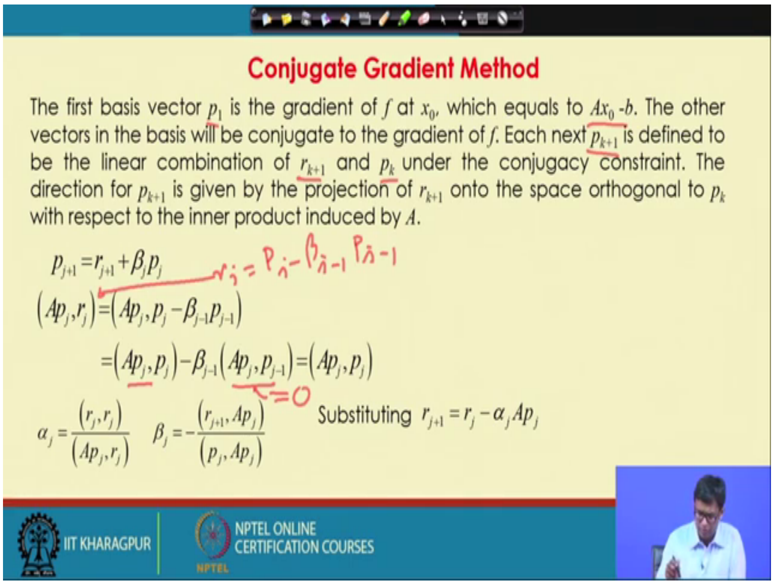


Similar to D-Lanczos algorithm x is updated as x j plus 1 x j plus a p j. So, they have a recursive like relation for exchange you know it is an (Refer Time: 20:08) should not say recursive relation or the iterative relation for x j. The difference is that the recursive relation for p will be substituted by the A conjugacy theorem. Therefore, the residual vector will follow a x j; if I multiply A both side A is j is equal to A x j plus a alpha p j substitute both side from b b minus A x j plus 1 which is the residual r j plus 1 is equal to b minus A x j which is the residual r j minus alpha j A p j.

Now, we have the earlier first proposition that r j's are orthogonal to each other. So, if I take a dot product of with r j in the both side r j plus 1 r j transpose will be 0. And we can get using the values of r j and p j. So, if we can get alpha j. So, if I know about the residual and the auxiliary vector at any j-th level I can find out the coefficient alpha I am sorry I can find out the coefficient alpha through which x will be modified in that level.

So, as r j's are orthogonal means this a b is equal to 0 i think we have probably discussed in somewhere. So, a b is equal to s means b transpose a. So, r j plus one r j is equal to 0 means the dot product or dot product dot product of r j and r j transpose r j plus 1 and r j 0 r j plus 1 is r j minus alpha A P j if we substitute we had a relationship for; A is equal to r j dot r j divided by r r j dot a p j.

(Refer Slide Time: 22:07)



The first basis vector p 1 is a gradient and this is the main idea of conjugate gradient. Choose the first basis vector p 1 is the gradient of f at x 0. The first basis of p auxiliary vector space p is the gradient of gradient f at x 0 which equals to A x 0 minus b. This is the first basis vector.

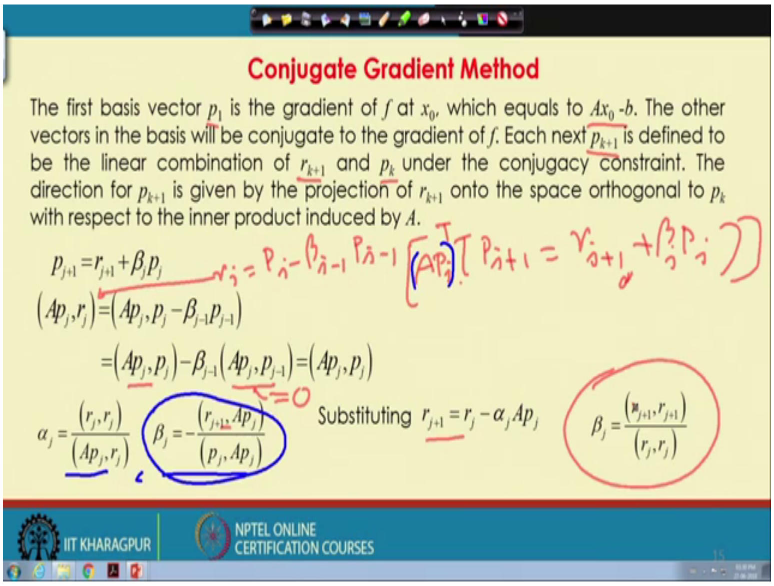
The other vectors in this bases will be conjugate to the gradient of f. Each next p k plus 1 is defined to be a linear combination of r k plus 1 and plus p k under this conjugacy constant. The direction of P k plus 1 is given by the projection of r k plus 1 on to the space orthogonal to p k with respect to the inner product induced by A..

Once little complicated way of way the how the mathematical formulations are put into statement, but if look into the step that the direction for P k plus 1 is sorry; the each next P k plus 1 the first P 1 is same as a x 0 minus b as the residual. The next auxiliary vector is the combination of the new residual vector and the old auxiliary vector. This is how we are constructing it fine. So, A p j dot r j is equal to a P j dot P j that means, I can write r j here we can write r j is equal to P j minus beta j minus 1 P j minus 1. And I substitute it here that is that what is the conjugacy what is the r j transpose A P j this is A Pj into p j beta P j minus 1. So, this is A P j in to transpose P j minus beta j minus 1 A P j p j minus 1. And this is equal to 0 from the conjugacy constraint we have already found out.

So, started with the fact that next each auxiliary vector is the previous is a linear combination of previous auxiliary vector and the present residual vector. And then we take the dot product of one auxiliary vector with the residual vector at a conjugate dot product. And it comes out to be A P j P j.

And sorry substituting; so beta j minus 1 becomes 0 right. And now if we take a A P g r j is found out if we take a dot product of missed one step here sorry then probably (Refer Time: 25:46) write like this. So, now from this how to find out beta what should be the linear combination of P that should be added with r j plus 1.

(Refer Slide Time: 25:58)



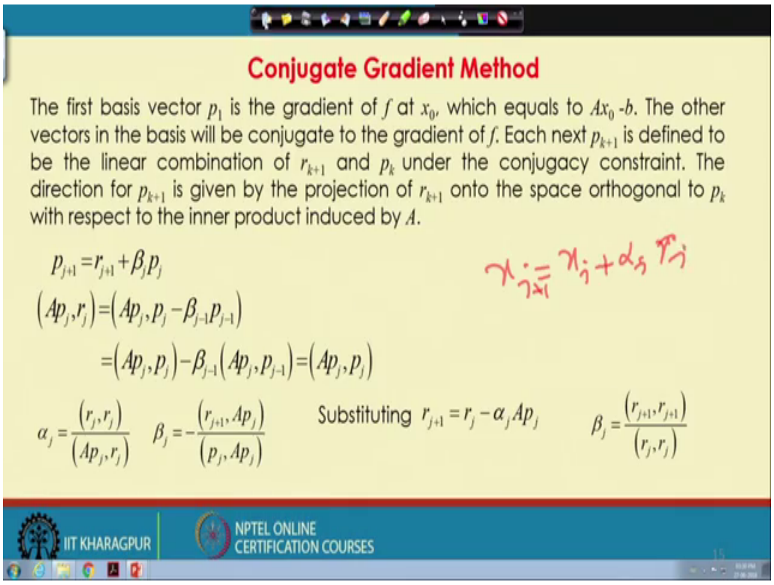
So, I can start working like this that P j plus 1 is equal to r j plus 1 plus beta P j beta j p j. And then P j plus 1 and then take the dot product of the entire thing with a P j take a dot product with A P j of the entire thing. And this will be 0, so this is A P j r j plus 1 plus beta j A P j r j and P j r j we have found out A P j p j.

So, using this, so like I can write this transpose this, if I do take this dot product. From there I can find out what is beta is minus r j plus 1 A P j p by p j p j alpha j is equal to r j p j r j. And this particular derivation I will I will not do it right now here, but this is exactly same as the derivation of alpha j. Which we will follow from here, take a dot product of this expression with A P j. And you will find out you will arrive in this particular expression of beta j.

So, what beta j needs this r j plus 1 A P j divide divided by p j a p j. Now for this is a matrix vector product multiplied with a vector. Matrix vector products are expansionally computationally expensive, though we have we are doing it once here. Now this can be very way well substitute changed by substituting r j plus 1 is equal to r j minus alpha j A P j. And if we substitute this r j plus 1, if we substitute this r j plus 1 here we get beta j is equal to r g plus 1 by r j r j. So, if we start with one value of r j r j transpose r j and at next step we again evaluate r j transpose r j the ratio will give us the beta j.

So, this should be sufficient to update P as well as to update x and r; r also can be updated using alpha X is updated right x is always updated.

(Refer Slide Time: 28:31)



Like x j is equal to x j plus 1 is equal to x j plus alpha j r j.

So, with this updates we can now finish the conjugate gradient algorithm. And we will look into it in the next session.

Thank you.