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

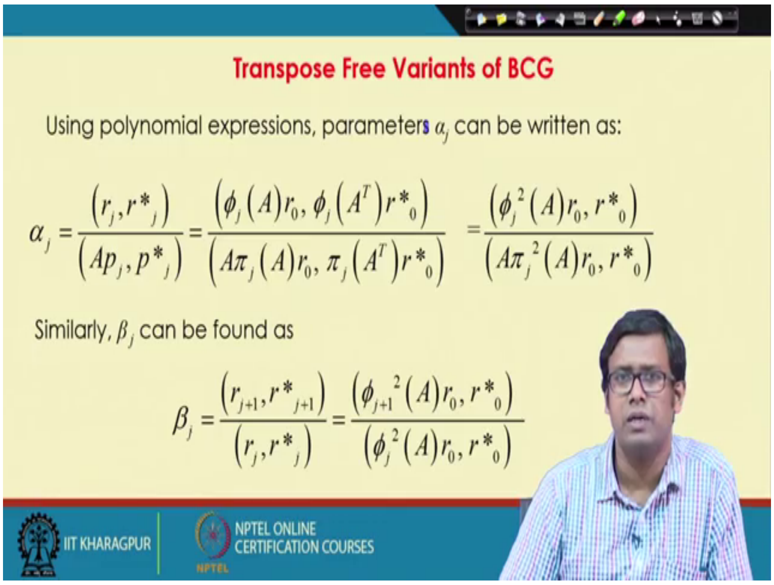
Lecture – 52
Conjugate Gradient Squared and Biconjugate Gradient Stabilized

Welcome. We are discussing about Biconjugate Gradient Method and we are trying to demonstrate the developments over Biconjugate Gradient Method, where the number of computational steps in terms of matrix vector multiplication could be reduced. And also looking into the possibilities of reducing the matrix eliminating the matrix transpose and vector multiplication step.

So, on developing that method in a on a step towards that we discussed that the residual and the auxiliary vector in both the Krylov subspace of A as well as Krylov subspace of A transpose can be expressed as 2 different polynomial functional of A and A transpose. The functional phi is defined for just you have been did it in the last class as the polynomial functional phi is for residual and polynomial function pi is defined for the auxiliary factor..

And we are trying to develop transpose V is variant of Biconjugate Gradient Method..

(Refer Slide Time: 01:26)



So, now is the parameter alpha is written as alpha is equal to r dot r r j dot r j start A p j dot p j star and we write try to write all the polynomial functional instead in cases of r j and p j phi j A r 0 which is polynomial form of r j and phi j A transpose r r 0 star.

Similarly A A into p j is pi j A r 0 and p star j is pi j A transpose r 0. Now r 0 is a vector which is multiplied with the polynomial function. This is this will be supposed to be a suppose it be a polynomial of a matrix. So, when we do this product, we get phi j star. So, this vector will come out and this is this is the dot product which is between these 2 vectors.

So, this will be phi j star A r 0 dot r 0 star and we will get A pi j star a similarly r 0 dot r 0 star. So, we in order to find out alpha instead of doing A transpose etcetera, if we can evaluate the matrix the polynomials we have to find out the polynomial product of phi j square A and A pi j square A also you need to get r 0 r r dot r r and multiply it with that. Similarly in the other parameter beta can be found as r j dot r j r star r j star plus 1 divided by r j dot r j star which is similarly if we replace the values of r j plus 1 r j etcetera we will get phi js plus 1 square A r 0 phi j square A r 0 r 0 star.

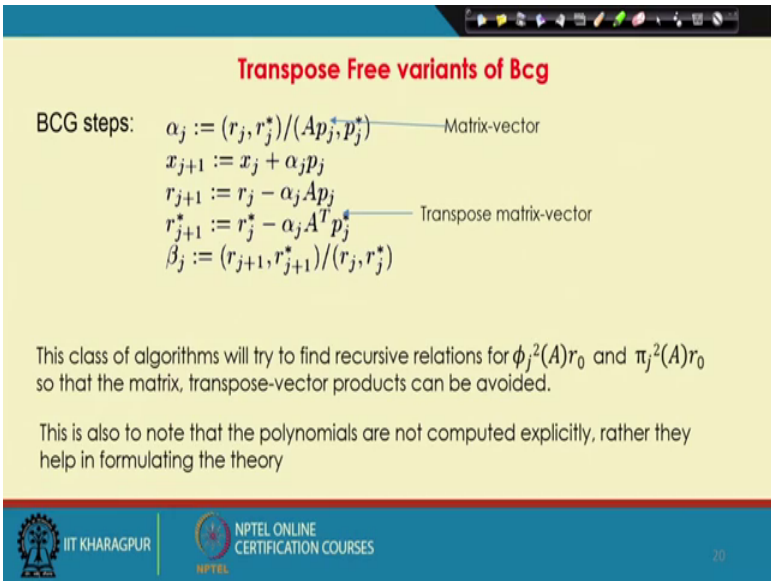
Now, the issue is that that are we actually reducing our steps, if we assume some polynomial form of the residue and the auxiliary vector. If we try to look it as apparently as it is coming that we are not reducing the steps because doing this, finding this polynomial and doing the square of this polynomial will be itself very cumbersome step. So, what will try to do is to develop a recursive relation here.

So, that we really do not have to do the all the calculations at you every step doing square finding doing lot of operations on a matrix rather we can use some simple recursive relation and trying this out. And now we will look into few variant sub that of this method few of the transpose p variants of the method.

First is the conjugate gradients square method, where r 0 the residue norm r will be assumed to of A to behave in a certain polynomial form and. So, that the now the residue norm will try to strict it to here; r 0 star is the initial residue norm of the left space or A transpose space for the Krylov subspace of A r 0 or A while we are solving x is equal to b for the Krylov subspace related to A we will try to focus obtain the residue norm from this part.

So, let us see the steps.

(Refer Slide Time: 05:08)



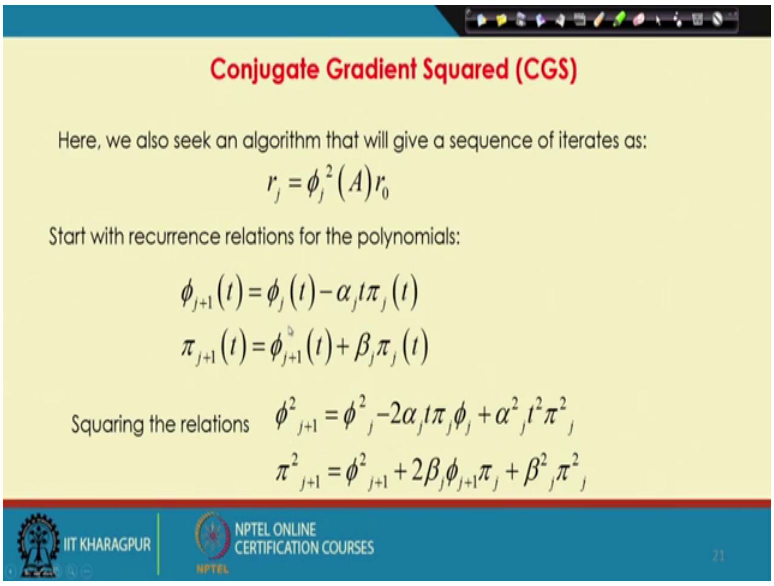
The BICG step is alpha j is equal to r j dot r j start by A p j p j star and we update x j plus 1 using alpha we update r j plus u 1 also using alpha and then, we get r j plus 1 start also using alpha and A transpose p. This step here basically trying to eliminate and then, we get beta j and based on beta j we update the p j. This is the matrix vector product; this is because same matrix vector product is used here. So, once we do this we can store this is the transpose matrix vector product. We are specifically trying to eliminate this particular step. The class of algorithms, will try to find will use recursive relations for phi j square A and pi j square A.

So, that in the alpha calculation, I required phi j square A. So, if I can have some recursive relation there, we can probably avoid the matrix as well as transpose matrix vector products. You see the steps. This is also to note that polynomials, these polynomials are not computed explicitly rather they help in formulating the theory. So, none of this polynomials phi j and pi j, I am writing down the b polynomial expression and trying to work on them do lot of matrix to the power j calculations using polynomial such terms.

We are not doing that we are not even writing the polynomials explicitly, we are assuming there is a polynomial form and it behaves like a continuous function of a higher order in this in the space and we will use this polynomial form to develop a recursive relation and finally, get a theory.

So, the idea of both the methods we look Conjugate Gradients Squared Method as well as Biconjugate Stabilized Method, the idea of this methods is using some recursive relations to make this calculation simple and for that we are using a polynomial expression of residue vector and the auxiliary vector..

(Refer Slide Time: 07:20)

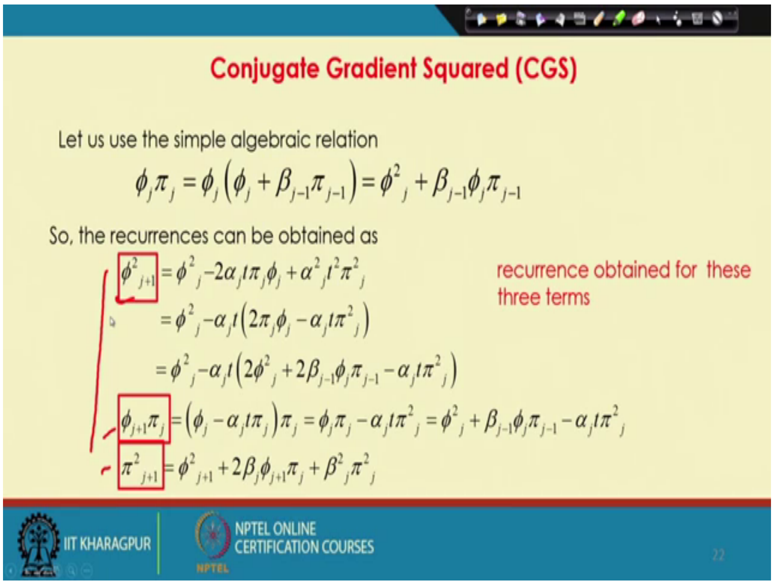


So, we see the conjugate gradient square method, we see can algorithm where the norm of the residual at j-th iteration is obtained as r j is equal to phi j square A r 0 and we start with recurrence relations for the polynomials phi j plus 1 that is the next iteration how phi will be updated is equal to phi j minus alpha j t pi j, there is a recursive relation. These relations actually came out of a algebraic manipulations so that at end we get a nice form, and nice way to replace the Biconjugate Gradient Algorithm, keeping the main principle same that the r and p vectors should be or truth follow the orthogonality relations that we started with.

Similarly for pi we get another recurrence relation. This is these are the recurrence relations we are assuming to start with. Now using this relation recursive relations phi j plus 1 square, t I am dropping here because all these are functions of t. So, the function functional of t, I am dropping here. So, we are try trying to develop the relation for the polynomial, keeping it as a function of a single variable t, but same relations will of course, all for polynomials when will a same polynomial when applied over the matrix A.

So, phi j plus 1 square is phi j square minus 2 alpha j (Refer Time: 08:54) terms squaring this alpha square j square t square pi square. Similarly pi j plus 1 square is phi j plus 1 square plus 2 beta phi j plus 1 pi j plus beta j square pi square j. So, phi j plus 1 square is related with all the phi j pi j square terms alpha j square beta j square which will be evaluated in last step if we start with some guess value.

(Refer Slide Time: 09:23)



Apart from that there is a product term pi j phi j and there is a product of phi j plus 1 pi j. So, you have to see the relations from them. Then, using simple algebraic relation phi j pi j is phi j into pi j is again we wrote pi j plus 1 is beta j plus phi j plus 1 plus beta j pi j. So, pi j is pi j plus 1 is phi j plus 1 beta j pi j. Therefore, pi j will be phi j beta j minus 1 pi j minus 1. So, phi j pi j is equal to phi j plus phi j plus beta j minus 1 pi j minus 1 which will give us phi j pi js phi j square plus beta j minus 1 phi j pi j minus 1. So, in case we know already know pi j, we can find out; you already know phi j and we know the last iterations value we should be able to find out phi j pi j.

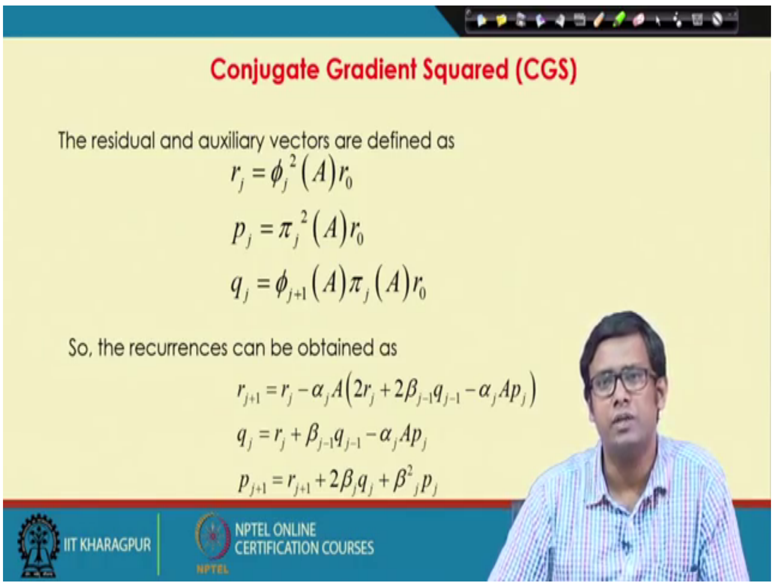
So, the recurrences can be obtained as phi j plus 1 square is which was phi j square minus 2 alpha j t pi j phi j plus alpha j square t square pi j square is equal to phi j square minus alpha j t common 2 pi j phi j minus alpha j t pi j square, which is again phi j square minus alpha j t j now pi j phi j can be replaced by this 2 phi j square plus 2 beta j minus 1 phi j pi j minus 1 minus alpha j t pi j square. So, phi j plus 1 square can be evaluated through this recursive relation and if phi j square phi j pi j pi j phi j and pi j are known and also beta j minus 1 alpha j on this terms are also. These are right recursive relation, if we know the last iterations value, we can calculate it. Similarly phi j plus 1 pi j, we substitute phi j plus 1 which is phi j minus alpha j t pi j into pi j and we get a relation phi j square plus beta j minus 1 phi j pi j minus 1 minus alpha j t pi j square.

So, if I know again phi j pi j and pi j minus 1 beta j minus 1 alpha j, we can find out phi j plus 1 pi j similarly pi j plus 1 square is also can be found out by phi j plus 1 square which is already evaluated plus 2 beta j pi j plus 1 phi j pi j plus 1 phi j is also evaluated plus beta j square pi j square.

So, we get a set of recursive relations of phi j plus 1 square phi j plus 1 pi j and pi j plus 1 square here. So, recurrence is obtained for these 3 terms and now we will see that how we will this help us remember for r and p we have a expressed in terms of phi and pi and we obtained the alpha in terms of phi and pi instead of directly writing it in terms of matrix vector and vector-vector products.

With this recurrence how we how things can be helps? So, this recurrences available is applicable for phi of a single valued function t as well as phi of a phi applied over a matrix A. So, if I know phi applied over the matrix A phi 0 phi 1. So, apply define a matrix a we can phi phi 2, phi 3, phi 4 etcetera; similarly, other terms..

(Refer Slide Time: 13:13)



The auxiliary vectors are defined as r j, now in conjugate gradient square method auxiliary vectors are defined as residual and auxiliary vector r j is equal to phi j square A r 0; p j is equal to pi j square A r 0 and q j, I define another vector q j which is phi j plus 1 A pi j A r 0. So, the recurrences can be obtained in terms of r p and q phi square pi square and phi r obtained and we can write nice recurrence relations in terms of r q and p.

So, we develop the recurrence relations for the polynomials; express the residue and auxiliary vectors in terms of polynomials and now we obtain the set of recurrence relations for r j plus 1 q j and p j plus 1. Now, we see why matrix vector and vector-vector products were matrix transpose vector products are required; required to find out alpha and beta.

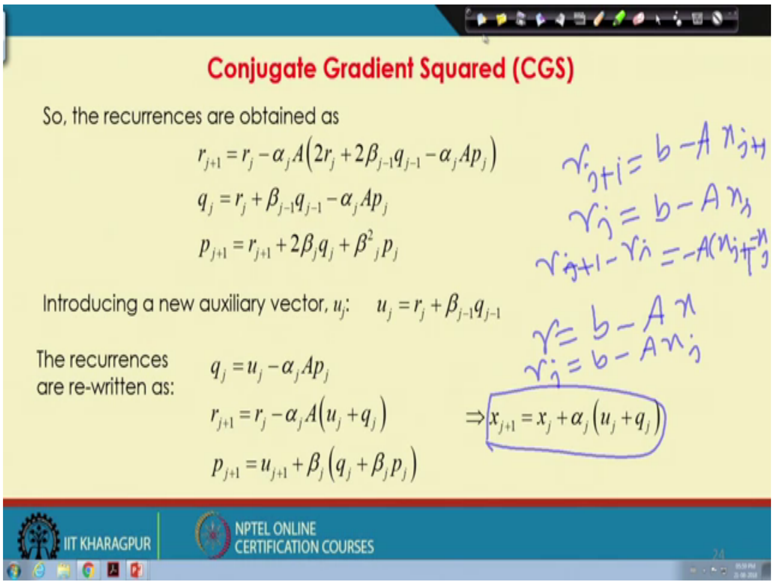
Matrix vector product was required to find out alpha which is A p; matrix transpose vector product if we look into the older BICG algorithm, main BICG algorithm matrix vector product was required to find out r j plus 1 star which will be again required to find out beta j; matrix transpose vector product. This is the matrix transpose vector product which is required to find out r j plus 1 star.

Now, if I already get a recursive relation for r j plus 1 star without using matrix transpose vector that problem is solved this step is eliminated. So, we get the recurrence relationship r j plus 1 is equal to r j minus just substitute the relationships in terms of we substitute that we got this recurrence relations and we substitute it that phi j square plus 1 r 0 is r j.

Substitute these relations. That r j plus 1 r j minus alpha j A 2 r j beta j minus 1 q j minus 1 minus alpha j A p j. So, one matrix vector product is still staying there because from the scalar form of the polynomial form, you know came to the vector form. So, one matrix vector is still staying there. The same matrix vector product is being utilized here. However, we did not need to till now we did not need to think of any r transpose A transpose the x product or A transpose p product; all this is A p product.

And now, we have to see how what are the recursive relations for alpha and beta or how to use these recursive relations to find out alpha and beta. Because alpha and beta are r q and p can be updated q is a new auxiliary vector, we have introduced here. r q and p can be updated through this relations alpha and beta cannot be..

(Refer Slide Time: 16:51)

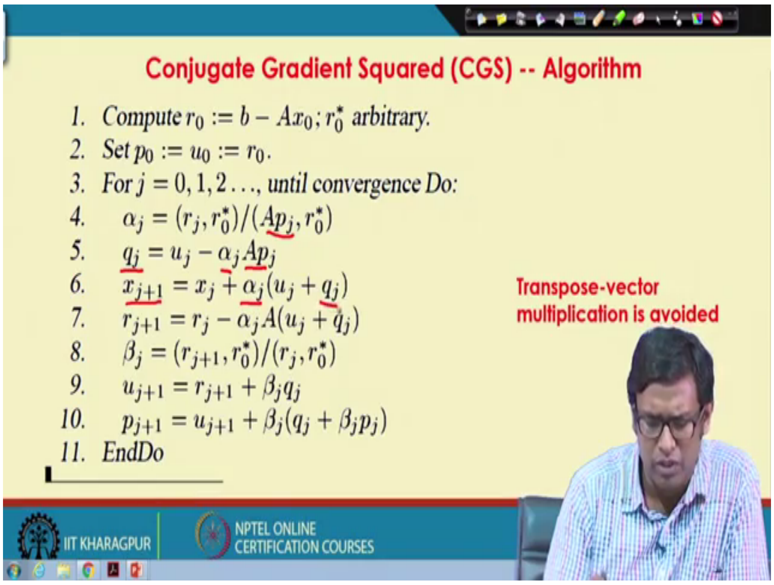


So, the recurrences are obtained in terms of r j plus 1 q and p into and we introduce a new auxiliary vector u j which is r j plus beta j q j minus 1. So, this becomes basically twice of r j and then recurrences can be written as q j is equal to q j is equal to u j minus alpha A p j; r j plus 1 is equal to r j plus 1 a j u j plus p j; u j.

So, q plus u p j plus 1 is u j plus 1 plus beta q j p beta j. So, there is still I can see there are still 2 matrix vector products one is A p j; another is a u j plus q j. However, the matrix vector transpose made transpose matrix and vector product that has been avoided. And once we can get r j, we have the relation that r is equal to b minus A x.

So, r j is equal to b minus A x j. So, when we have a relation for r j plus 1, we can get the relation; update relation for x j. Only the negative sign will be positive here and this will be divided by A. So, r say this is r. I can write r j plus 1 is equal to b minus A x j plus 1 r j is equal to b minus A x j. So, r j plus 1 minus r j is equal to minus A into x j plus 1 minus x j. Using this we can up get the updated relation for x j plus also..

(Refer Slide Time: 18:55)



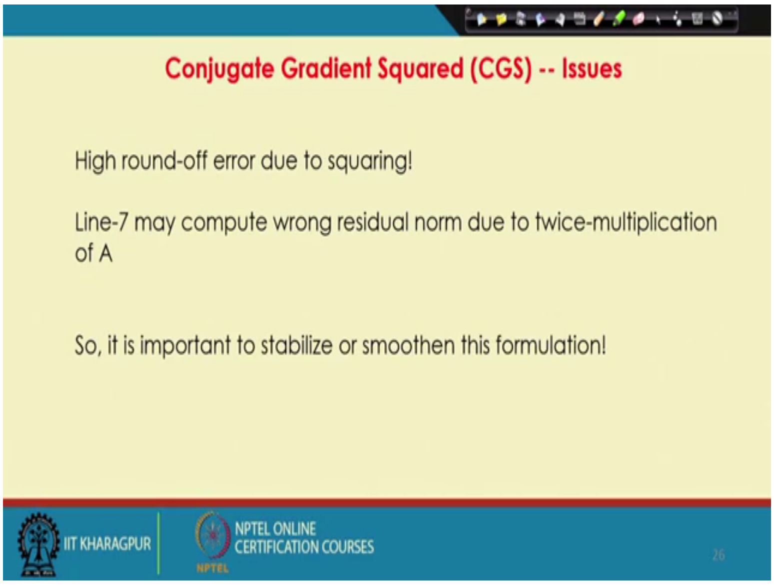
So, you get a complete Conjugate Gradient Squared Algorithm, where alpha and beta are calculated in same way as matrix vector product. For calculation of which needed a matrix vector product of A p into calculate alpha, similarly q also needed the matrix vector product of A p; r j plus 1 instead of using a transpose matrix vector r j r star; star as not required to we do not need to calculate the star here only. r 0 star will be sufficient to complete the entire algorithm because r j plus 1 star has a relationship with r 0 star which you are exploiting here and instead of using A transpose, we are using another matrix vector product A into u plus q j.

So, the transpose matrix into vector product can be completely eliminated in Conjugate Gradient Squared Method and please note that we have used a polynomial based formulation, but never use the polynomial explicitly. There is a beauty of this method. Transpose vector make a vector transpose matrix into vector multiplication is avoided. Still I can look into 1 significant 1 serious issue here; there is while calculating q, I am using alpha into A p. Alpha is calculated using A p..

So, the round of error that I am obtaining from while calculating A p is A multiplied with the; or added with the round off error I am doing with alpha A p. Similarly while doing x j plus 1, I am doing x with alpha which is obtained through the round off error it is still there and then, I am multiplying it with q and getting another sort of round off error.

Multiplication does the actually is reducing the error rather; that means, adding up of several round off error here. So, there is also a chance of numerical instability because round of the error is high. The implementational problem due to A transpose; transpose matrix vector multiplication is avoided, but the numerical instability issue is still there..

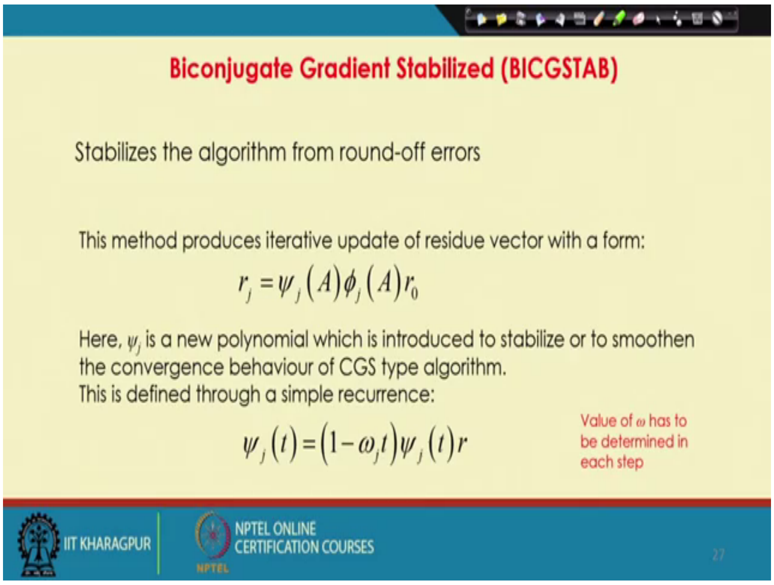
(Refer Slide Time: 21:15)



So, the observations are how the high round of error due to squaring which is possible and line 7, what was line 7? Line 7 is r j plus 1 is equal to r j minus alpha j A u plus q j right and this alpha j is obtained from a matrix vector product. There is some error there. A is again being multiplied with u j plus q j, where q j is also obtained by alpha j A j. So, lot of calculations are involved here; lot of multiplications are involved here and that also shows that round off error can be very high and the residual can be computed wrongly in this step. Line 7 may compute wrong residual.

Now, due to twice multiplication of A, which again can be a serious issue; so it is important to stabilize or smoothen this formulation. This formulation is not also stabilized enough. Though the transpose matrix vector product is that step is eliminated, implementational issue has been eliminated. However, the solution is not stabilized. There can be numerical issue errors also..

(Refer Slide Time: 22:46)

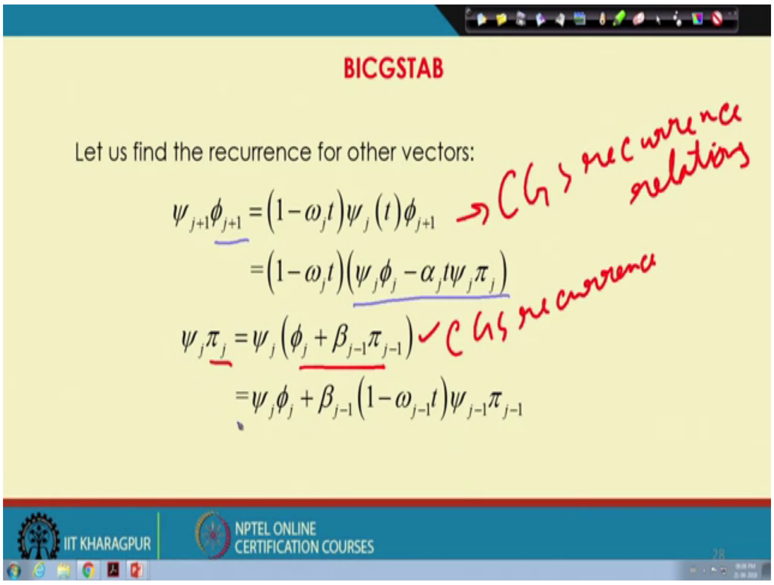


So, we go for a new method which is again using the idea of Conjugate Gradient Square Method and the principle of Biconjugate Gradient Method which stabilizes the round of issues due to round off error which is called Biconjugate gradient stabilized method and very popularly known as BICGSTAB method. The method produces iteratively and update of residue vector of a form r j is equal to phi j A; psi j A phi j A r 0. Earlier it was r j was only determined by phi.

Now, we introduced another component psi. What is the role of psi? The role of psi is to smooth the function; the error is to stabilize the method. Here the psi j is a new polynomial which is introduced to stabilize or to smoothen the convergence behavior of conjugate gradient square type of algorithm and this is defined through the recurrence psi j t is equal to 1 minus omega j psi j there should be psi j minus 1 psi j minus 1 t into r.

So, this in this r is not there, psi j t is equal to 1 minus omega j t psi j minus 1 t. I will I will rectify this equation in when I will upload the slides. So, here psi j is using something it has a recurrence relation; that means, its related with the last times last iterations value and there is a factor omega which is multiplied with t is updating it and based on this omega this will the behavior of psi can be smoothened. Value of omega has to be determined in each step so that it gives the optimum result to smooth the function psi is smooth the error psi.

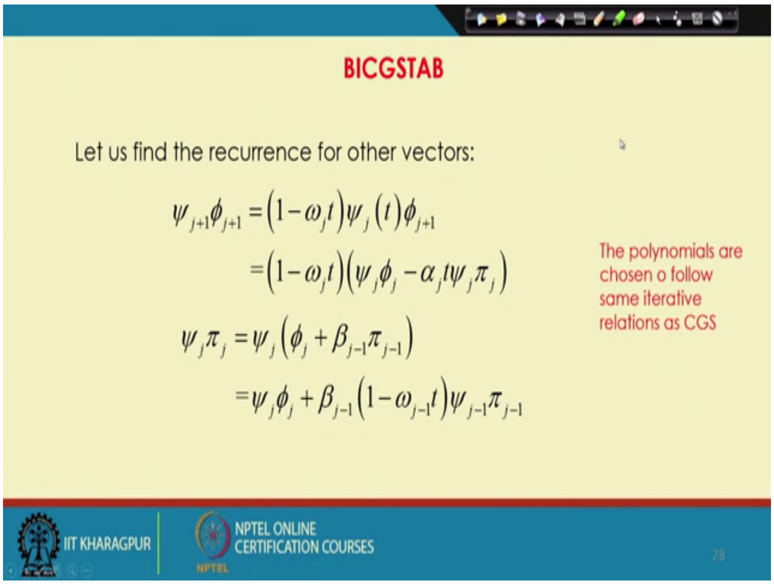
(Refer Slide Time: 24:46)



Let us find recurrence for the other be phi vector the similar way we did it for the last vectors psi j plus 1 phi j plus 1. Now, just write the expression of psi j and into phi j plus 1 and psi j pi j similarly we get a recurrence relation. All the recurrence relation now contain and stay along with alpha contain this; this is the phi j plus 1 and pi j that this recurrence relations are same as the recurrence relations using in CGS method. So, here we are using same relations CGS conjugate gradient CGS recurrence relations..

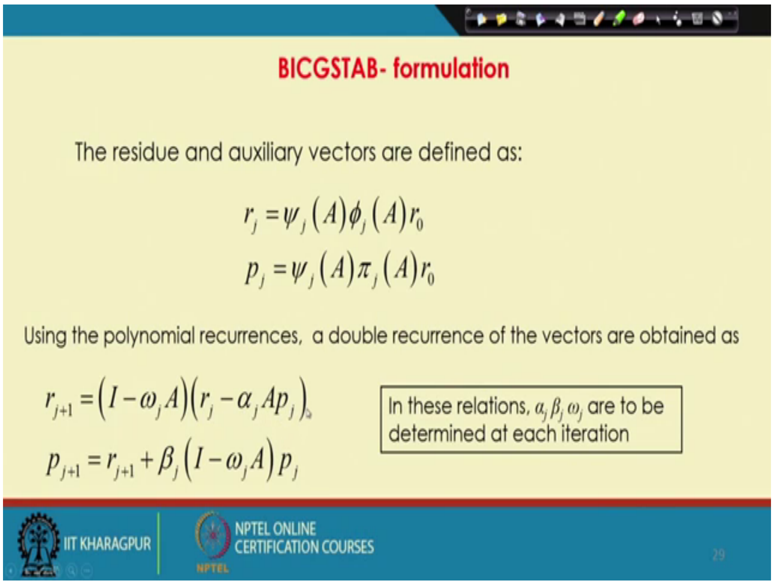
And check the recurrence relations, we have discuss wrote CGS method, same here CGS recurrence. So, the recurrence relations we have developed for CGS we are using for pi j and we are using for phi j plus 1 and we get a recurrence relations for psi j plus 1 phi j plus 1 and psi j phi j; while it be needed to see.

(Refer Slide Time: 26:03)



The polynomials are chosen to follow same iterative relations as CGS.

(Refer Slide Time: 26:06)

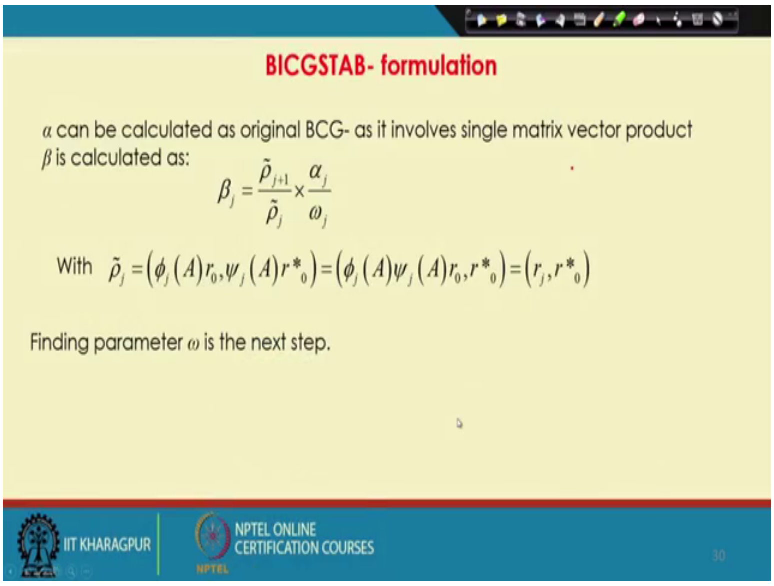


The residue and auxiliary vectors are defined as r j is equal to psi j phi j r 0 you have defined that and p j is equal to psi j a pi j A r 0.

Using the polynomial recurrences or double recurrence of the vectors can be obtained as r j plus 1, the same method we have done for conjugate gradient squared. I minus omega j A r j minus alpha j p j; p j plus 1 is r j plus 1 beta j I minus omega j and if we have to use this recurrence relations, we need to find out alpha, beta, as well as omega.

Once we can find out this, then starting with the first residual and first auxiliary vector, we can find out the next residual and next auxiliary vector. So, the entire problem will be a residue vector finding problem once you can find out omega j beta j and alpha.

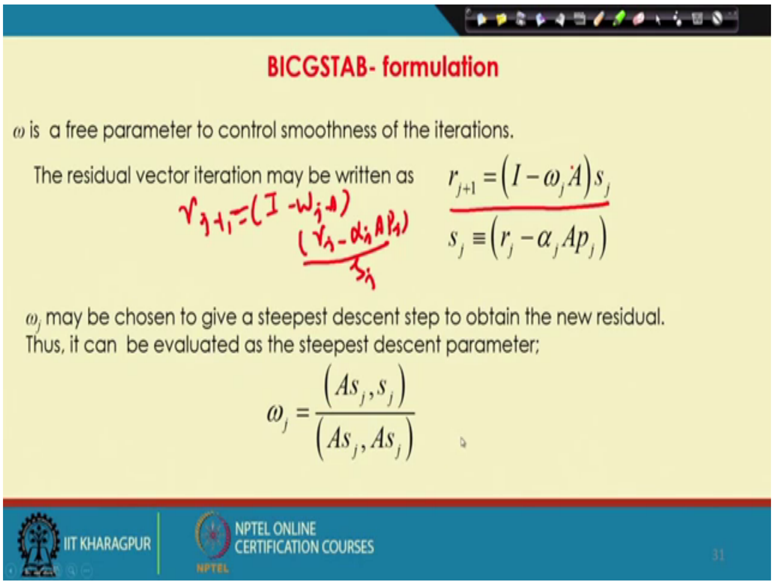
(Refer Slide Time: 27:10)



Alpha can be calculated as original BCG as it in works single matrix vector product. So, the or the original BCG using the orthogonal relation, relation can find out the alpha. Beta can be calculated as rho tilde j plus 1; rho tilde is a new scalar introduced here by rho tilde j into alpha j by omega j, where rho tilde j is phi j psi j A r 0 star which is phi j psi j r 0 r 0 star that is r 0 r j r 0 star, the dot product between r j and r 0 star is rho j.

This is a little involved. I am not discussing it here due to lack of time it should have taken a one more session there, where even look into exercise book here. Finding omega is the next step.

(Refer Slide Time: 28:07)



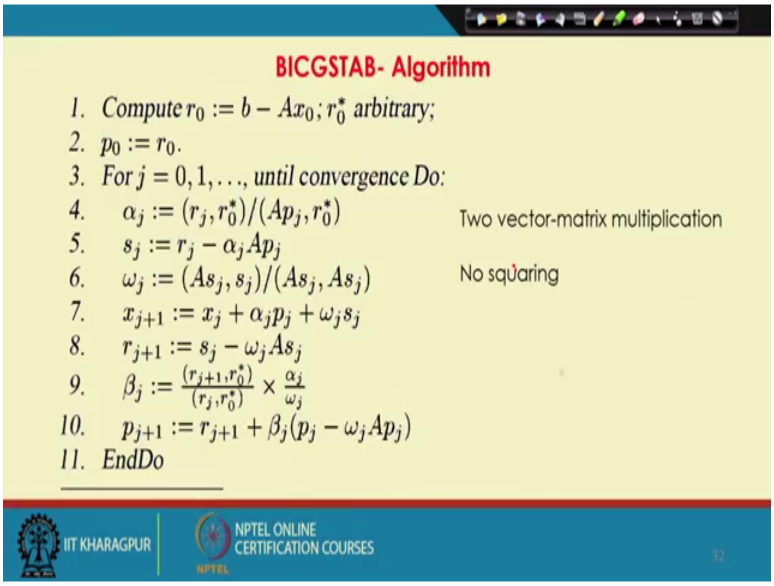
And this is complex. Omega is a free parameter which control smoothness of the iterations. Alpha and beta were related with the Biconjugate gradient properties; the conjugate gradiency of the p j and a p j and p j star in the left and the A space the Krylov subspace and the conjugate gradient between probably between r j and r j star; alpha and beta determined based on this conjugate gradiency property. When is a new parameter which is to control the smoothness itself?

Let us look into this step of omega. The residual vector is written as r j plus 1 I minus omega j A r j A j A j A p j which is we are introducing a new vector s j. So, the residual vector the original relation of omega what is what is the r j plus 1 is equal to I minus omega j A r j minus alpha j A p j. Now I replace that as s j and I write this equation. So, what is this? This is finding out the residue vector from the older residue minus some vector s j. So, I have a older residue victor.

I subtracted something from there and then I try to multiply I minus omega j here to find the new residual vector. This is kind of an optimization problem that finding out the optimal alpha optimal omega, I talked about alpha which is synonymous here in a second I am explaining that; finding out the optimum omega which will give me the first step first step convergence. So, that the new residual becomes 0.

This is in a sense very similar to steepest descent algorithm that find the optimum alpha, find the optimum statistical design parameter based on which the new residual will approach the that is the solutions. The new vector will approach the solution the new residual will go to 0 and exactly similarly in same way omega may be chosen as a steepest descent step to obtain the new residual. At steepest descent step here which gives us omega s is equal to A s j trans dot s j divided by A s dot s j; exactly same as the steepest descent algorithm.

(Refer Slide Time: 30:32)



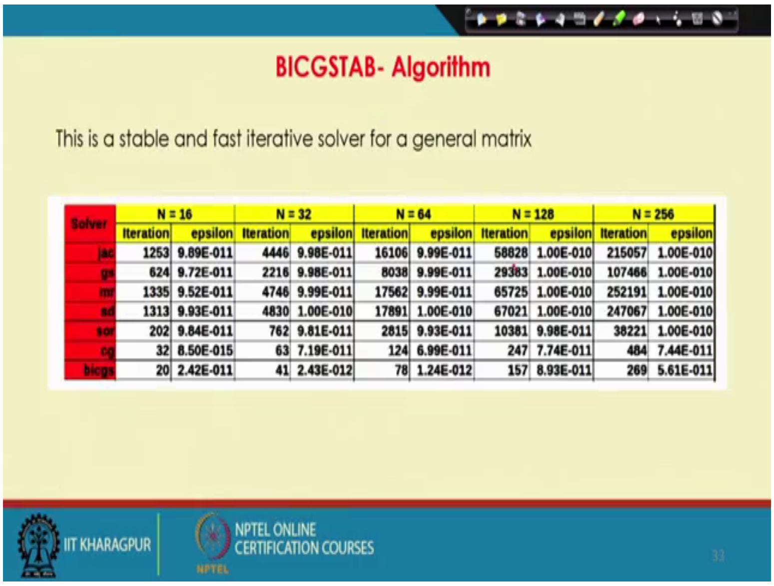
So, we get the Biconjugate Gradient Stabilized Algorithm, where it started with a in any initial value A x 0 and the initial residual r 0 is equal to b minus A x 0; p 0 is equal to r 0. Alpha j is calculated very similar to the Biconjugate gradient original Biconjugate gradient method, but now we using r 0 only. Because r j contains the polynomial of the another a part phi squared a part is contained in r j r j dot r 0 star divided by A p j dot r 0 star.

We introduce a new variable, s j r j minus s A alpha j A p j using s j, we get the optimum omega and x j plus 1 and r j plus 1 are updated using that optimum omega factor. Beta j plus 1 is similarly calculated and we are again using beta j plus 1, we calculate j.

So, what are the matrix vector steps here? This is one matrix vector step here; this is one matrix vector step here and we can also see that there is nothing like a residual which is dependent omega is dependent on s j. So, it is like it omega comes from A s j and then multiplying with A s j, there is no dependence between 2 different matrix vector products.

There are 2 matrix vector products, but they are not coupled. There is nothing like a alpha A s j type of steps here, step present here. Therefore, the numerical errors are also much controlled here. Two vector matrix vector multiplication no squaring, no vector; once matrix vector multiply, again that is not being multiplied with something else no squaring on the matrices.

(Refer Slide Time: 32:29)

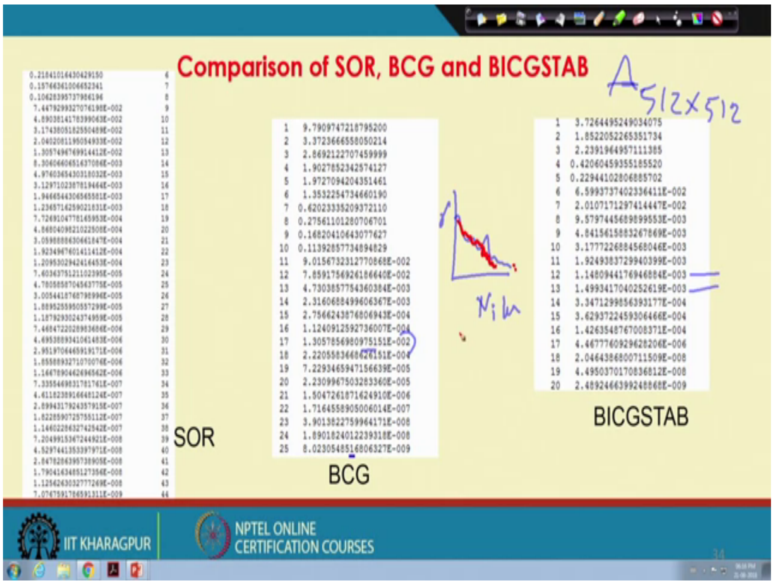


Solver	N = 16		N = 32		N = 64		N = 128		N = 256	
	Iteration	epsilon	Iteration	epsilon	Iteration	epsilon	Iteration	epsilon	Iteration	epsilon
jac	1253	9.89E-011	4446	9.98E-011	16106	9.99E-011	58828	1.00E-010	215057	1.00E-010
gs	624	9.72E-011	2216	9.98E-011	8038	9.99E-011	29383	1.00E-010	107466	1.00E-010
mr	1335	9.52E-011	4746	9.99E-011	17562	9.99E-011	65725	1.00E-010	252191	1.00E-010
sd	1313	9.93E-011	4830	1.00E-010	17891	1.00E-010	67021	1.00E-010	247067	1.00E-010
sor	202	9.84E-011	762	9.81E-011	2815	9.93E-011	10381	9.98E-011	38221	1.00E-010
cg	32	8.50E-015	63	7.19E-011	124	6.99E-011	247	7.74E-011	484	7.44E-011
bicgs	20	2.42E-011	41	2.43E-012	78	1.24E-012	157	8.93E-011	269	5.61E-011

And this is a table and first iterative solver for any general matrix. This is probably or our target to show that we can have a stable and first iterative method for any general matrix, we which can be developed right now algorithm an algorithm, can be developed right now and can be utilized in any in an different purposes of solving equations. And we can see that the number of iterations are much smaller for different size of matrices 16 by 16, 32 by 32, 256 by 256 matrices.

To a very high order of accuracy 256 by 256 matrix, the number of iteration is 269 to 10 to the power minus 11 accuracy.

(Refer Slide Time: 33:08)



So, number of iterations are same as the number of rows of the matrix of same order. So, we finally, obtained a methodology solver which iterates within n steps for an n by n matrix and we can quickly compare 3 methods; Successive Over Relaxation, BCG and BICG. We are Biconjugate gradient and BICGSTAB. For successive over relaxation this is a 512 into 512 matrix. A is 512 into 512 and we are not converging up to retain very high order of accuracy 10 to the power minus 9 we are leaving, so convergence is faster.

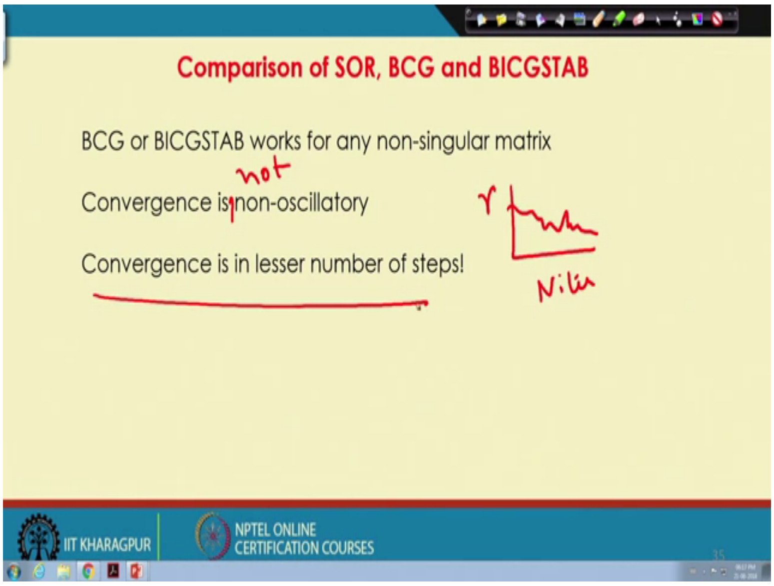
For the if you can see for a SOR with optimum SOR parameter, it took 44 steps. However, the residue is very smoothly falling down; starting from 10 to the power minus 2 the residue goes to 10 to the power minus 9. Below 10 to the power minus 9 and it is monotonically reducing.

If I look into the Biconjugate gradient method, the residue is reducing it converges in 25 steps. However, after say after 13 step, 14 steps, 15 steps, 16 step, 17 steps suddenly residue increases. Again it reduces and again there is some increase in residue somewhere I guess. So, there is a there is an oscillation of residue. There is no increase, but there is an increase, there is a increase in residue here.

So, if I drop r versus number of iteration, it is coming down like this. There is a numerical instability. In that sense, if the numerical if the matrixes much complex and much bigger, this numerical instability may this may disturb the convergence of the method. And same we can see for BICGSTAB, the residue is actually falling down; however, there is some oscillation in the residue.

So, BICGSTAB probably may be showing similar method like this, but there is some oscillation in the residue value 1.144 10 to the power minus 3, 1.49 to the power minus 3; residue increased somewhere, there is some increase in residue donot need no this is not as significant as in BICG. So, this is also an oscillatory scheme and importantly this schemes runs with only high precision machine as well as high precision declare declaration of the variables in the code..

(Refer Slide Time: 35:59)



So, took summarize we got BIC, G Biconjugate gradient or BICGSTAB works for any non singular matrix. Convergence is non-oscillatory. Convergence is not non oscillatory, convergence is sorry convergence is not non-oscillatory. There is a if I drop plot residue verses number of iterations, there is some oscillation in that.

However, the important is that convergence is in lesser number of steps compared to the other methods. So, BICGSTAB is a works for any non singular matrix is a non the not high numerical error method which is a stabilized method and we will give solutions in less number of steps. However, there are some oscillations in the residue while converging.

Thanks.