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

Lecture – 39
Introduction to General Projection Methods

Welcome, in last session we are discussing about Projection Methods for iterative solution of Ax is equal to B and we started our discussion considering the steepest descent algorithm. However, what we observed the steepest descent algorithm is only applicable for symmetric positive definite matrices and we ended the discussion with a note that if the matrix is not symmetric neither positive definite for or any for any general matrix what can be an equivalent algorithm.

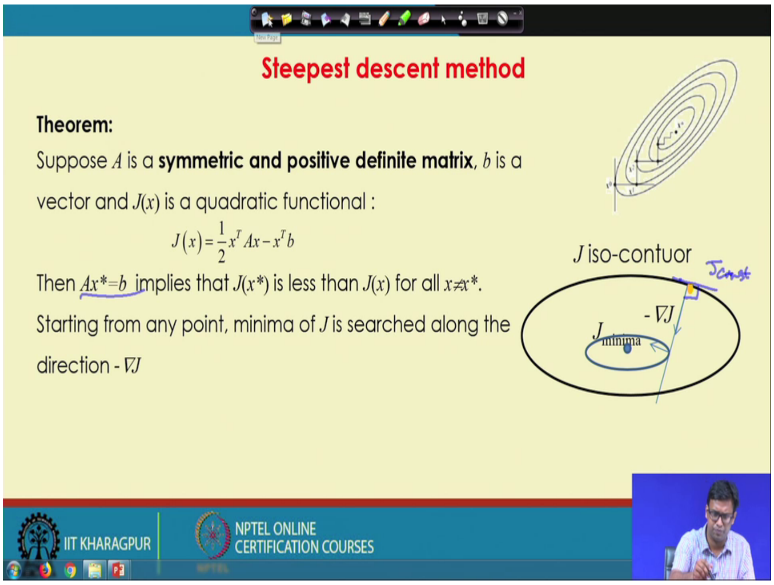
So, before coming into projection methods or a steepest descent algorithm or gradient search algorithm, which we discussed in last few classes; we discussed about another class of iterative solvers which are Jacobi and Gauss Seidel or SOR solvers which are called direct iterative solvers. What we have seen that they are also limited for certain class of matrices which are diagonally dominant or irreducibly diagonally dominant matrices.

So, till now we are yet to find out a method which is applicable for any non any type of non singular matrix. However, when we have looked into direct solvers we have found number of methods which are applicable for any general non singular matrices like Gauss elimination, LU decomposition; not finding solution of normal equation from few methods like this. And, Gauss Jordan method is also one of that, Cramer's rule you can also find out solution for any non singular matrix.

But, we also found a one method called tri diagonal matrix algorithm, which is a very first method. The number of computational steps are extremely small which is only applicable for matrices which are tri diagonal matrices. So, but for general direct solvers, we have seen that they are robust in a sense they can take care of any type of matrices. However, the general direct solvers are limited in a sense that they need lot of computational steps. So, we looked into iterative solvers and as we just said before iterative solvers the number of solvers will be before are restricted either for diagonally dominant matrices which are Jacobi Gauss Seidel or SOR matrices or symmetric positive definite matrix which is the method is steepest descent algorithm.

So, now our question is to find out method for any general solver and we also will like to get faster solvers; that means, in number of less number of iterations with doing less number of computational steps you should be able to reach the solution. So, we will look for a broader class of solvers. This class of solver which you are looking at this stage will borrow their idea from steepest descent algorithm and we will try to generalize it more for general type of matrices which are any non singular matrix. So, we start our discussion on general projection methods which are projection based iterative solvers.

(Refer Slide Time: 03:21)



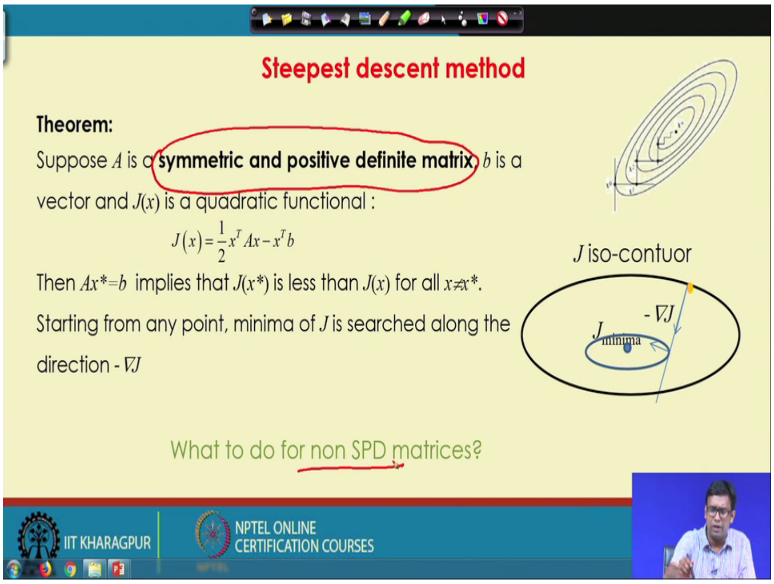
So, you quickly see what we have done in steepest descent method. Suppose A is a symmetric and positive definite matrix and b is a vector, we are solving Ax is equal to b and Jx is a quadratic functional and Jx is equal to half x transpose Ax minus x transpose b. Then Ax star is equal to b implies that Jx star is less than Jx for all x is equal to x is not equal to x star for all x. So, finding solution of Ax is equal to b is replaced by finding minima of the functional Jx and in a sense we can get find this minima iteratively.

So, the we know that is why you do not called it direct iterative solvers because we are not iterating for solution of Ax star is equal to b rather we are solving iterating for finding minima of Jx. So, what we do we get a J iso-contour that is the value of J is constant along this particular surface and then we can see that the value of J reduces fastest in the direction of minus grad J from here.

So, one thing quickly we can say is that J this if I draw a tangent here this is J constant line, if I draw a tangent here this is J constant line. So, if I draw a line perpendicular to tangent or minus grad J this is the direction where J is changing in a fastest way in this line J is J is constant. So, J is not changing at all along the tangent, along the normal, or along minus grad J J is changing fastest.

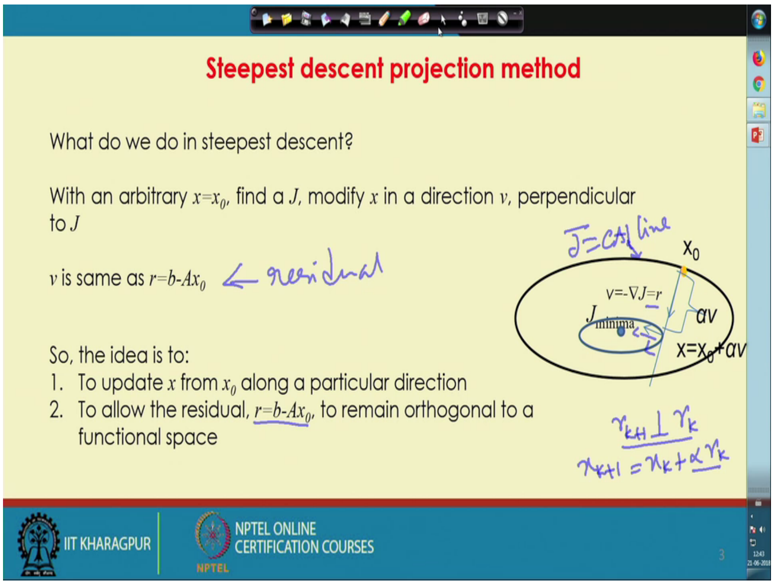
So, you move along minus grad J that is J is radio, we move our x vector along minus grad J. So, J reduces in the fastest direction and reach somewhere where it becomes tangent to another iso contour and then change the direction and that way we can keep on changing the search direction and we will finally, reach J minima. And, at this location the value of x is the solution of Ax star is equal to b that was the idea of steepest descent method. We have discussed in detail in last few classes.

(Refer Slide Time: 05:43)



And the question remains that steepest descent is said to be designed for symmetric positive definite matrix. So, what to do for non symmetric positive definite matrix?

(Refer Slide Time: 06:01)



So, we go to the next slide that, what do we actually do in a steepest descent method? We choose with an arbitrary X 0 and then find out a direction a vector v which is equal to minus grad J. This is a we find the value of J here, find the value of J here. So, we write this as J is equal to constant line, this is J is equal to constant line or J iso contour. Find minus a grad J which is v and move x along that direction by a certain value alpha v, we move a certain distance alpha along the alpha into v in this direction and then we will again change the direction.

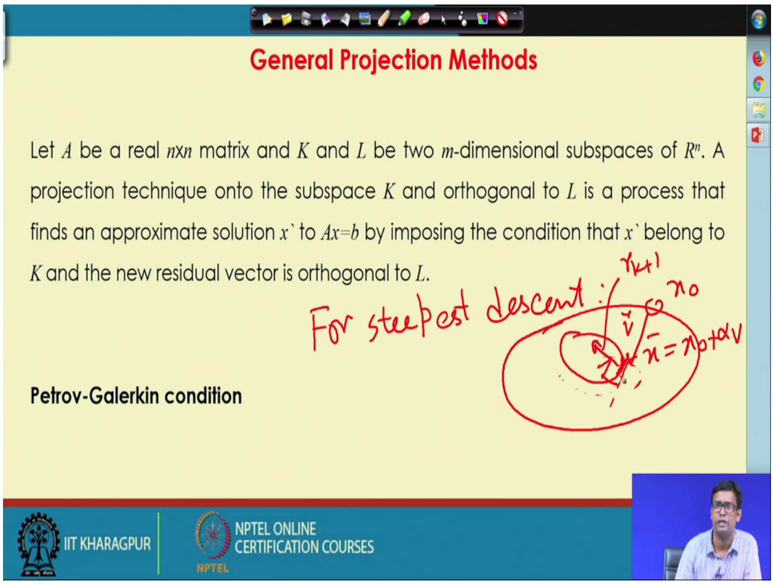
With an arbitrary x is equal to x 0 find a J modify x in a direction v which is perpendicular to J. So, start with an a arbitrary value modify x, in a particular direction and while modifying x in the in the direction we have chosen the direction which is perpendicular to J, v is same as r is equal to so v is same as r, r is equal to b minus Ax 0 what we call to be residual. So, the idea is update x from x 0 start with an arbitrary value x 0 update x from there along a particular direction.

Also in that case allow the residual r which is b minus Ax 0 to remain orthogonal to a functional space. So, x is updated along certain direction here this direction is r d and also when updating x allow this value r to remain orthogonal to certain functional space. Here the space is J constant line it can be a different space in different application also. So, these two are the basic steps in the philosophy of steepest descent projection method, that update x in a in along particular direction in during one particular iteration and then allow the residual to remain orthogonal to a functional space.

What will happen in the next step? We will allow, we will update x in this particular direction and r will remain orthogonal to the previous r or we will be now we be orthogonal to the new J contour iso contour which is perpendicular to the previous r direction. So, we will get r k plus 1 which is orthogonal to r k and x k plus 1 is equal to x k plus alpha into r k. So, update x along one particular direction and the residual is orthogonal to or some functional space this alpha k or this is orthogonal to J k also.

So, with this idea we will try to generalize an iterative method that x is updated along a particular direction and the residual is orthogonal to another to one particular functional. They may be same here they are probably they are same they may not be same and this will give us a different class of method.

(Refer Slide Time: 09:27)

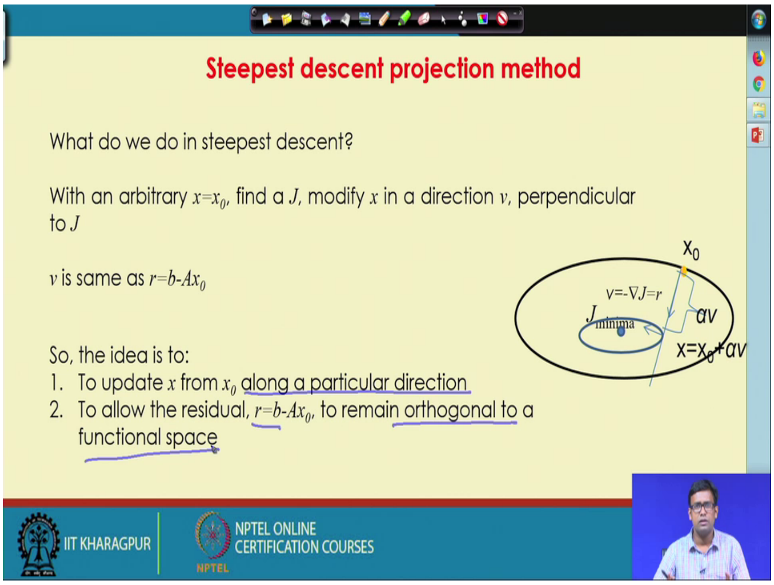


And this is called a general projection method the theory of general projection method is that. Let A be a real n into n matrix and K and L be two m dimensional subspaces of R n. So, they may not be a complete a one d space may not be a complete vector space of R n their subspaces of R n and their m dimension let us assume their m dimensional subspaces of R n.

A projection technique onto the subspace K and orthogonal to L is a process that finds an approximate solution x prime of Ax is equal to b by imposing the condition that x prime belongs to K. The new solution, the approximate solution the solution we obtain after an iteration belongs to certain space because we are updating it along particular direction. So, it belongs to a particular vector space that is the idea. So, x prime belongs to a certain space and their new residual vector r k plus 1 which is Ax prime minus b is perpendicular to a orthogonal to a functional space L.

So, there are two spaces x prime belongs to one space and the residual is perpendicular to another space. So, why because finally, we will when the solution will converge we will see that the residual is 0 ok. So, then we will take dot product of residual and it will be anyway you should give a 0 at any point of time. So, we approximately we take a one space, one sub space in which residual is orthogonal to that subspace and the new solution is updated along another particular subspace. So, there are two subspaces with which we can generalize the projection idea. Again I will go back to the previous slide.

(Refer Slide Time: 11:30)



So, this comes from that that x is updated along a particular direction; that means, x belong to the update of x belong to one particular subspace and the residual r is orthogonal to a particular subspace. So, there are two subspaces involved in steepest descent type of algorithm.

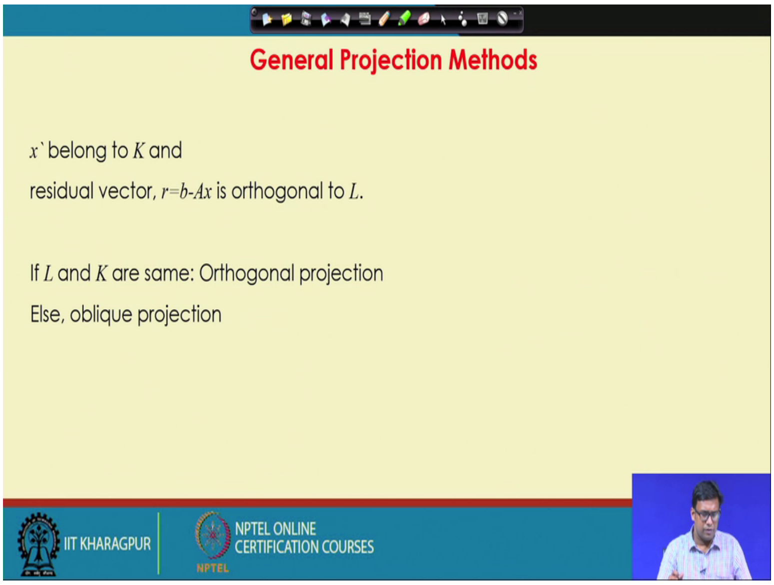
In steepest descent we can see that that x is the update of x is along one particular direction. While I am doing that the residual is remaining orthogonal because residual will be my new search direction. So, residual is orthogonal to another particular direction here which is the iso contour of J or the in the previous tangent to the iso contour of J or previous residual direction.

So, now this statement does not tell about any iterative technique it says that one approximate solution or x prime can be found out that is not the right solution. So, we have to keep on modifying the approximate solution which will be which will take us to the best approximated solution where the difference between the actual sol exact solution and approximate solution is minutely smaller it will it should converge.

So, for an iterative method and this condition is called a Petrov Galerkin condition that you update find an approximate solution of Ax is equal to b imposing the condition that x prime the approximate solution x prime belong to a particular subspace k and the new residual vector is orthogonal to L. So, if we again go back that we can write the for steepest decent method. What we are doing? I am updating so I have started from some value x 0 here, I have updated x is equal to x 0 plus alpha V right. I have updated x along this particular direction v.

And sorry this there is some issue here I think I should wipe it out. Up to when I have updated till the this particular line becomes tangent to the new functional space and then the new search direction will be orthogonal to heat or this the new search direction will be the new residual space r k plus 1 which is orthogonal to this particular direction.

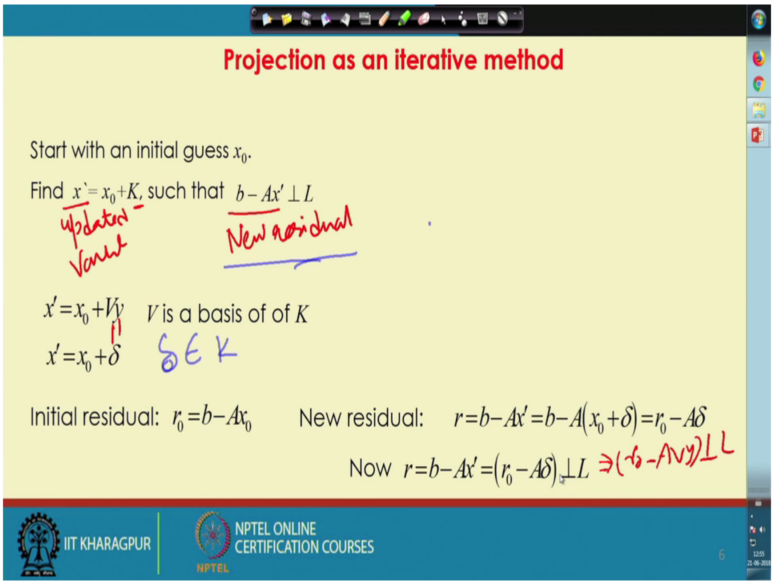
(Refer Slide Time: 14:53)



So, x is updated along one direction and R k plus 1 is orthogonal to another direction there is there is the idea here; x prime belong to a subspace K and residual vector r is equal to b minus Ax is orthogonal to L. Here the subspace and K and L are same means steepest descent algorithm x y belongs to or the it is not exactly x prime in the steepest descent we will see the iterative application, but if we assume that we start with the gaze value x is equal to 0. So, x prime will be along be my along the vector b, b minus Ax Ax is equal to 0. So, along b x prime will be along one particular direction and the updated residual will be perpendicular to that direction only in steepest descent.

But in a general case x prime belong to one particular a space K and the residual vector is perpendicular to another particular space or L and L and K may be same or may not be same that is a different coefficient. If L and K are same which happened in the steepest descent method, they are called Orthogonal Projection. If they are not same they may be L may be transformed where transformation of K, L may be completely unrelated with K. There are several methods we will discuss soon that is called an Oblique Projection. This steepest descent we can say that L and K are same and both are basically R K or minus grad J and whatever the way we want to present it.

(Refer Slide Time: 16:21)



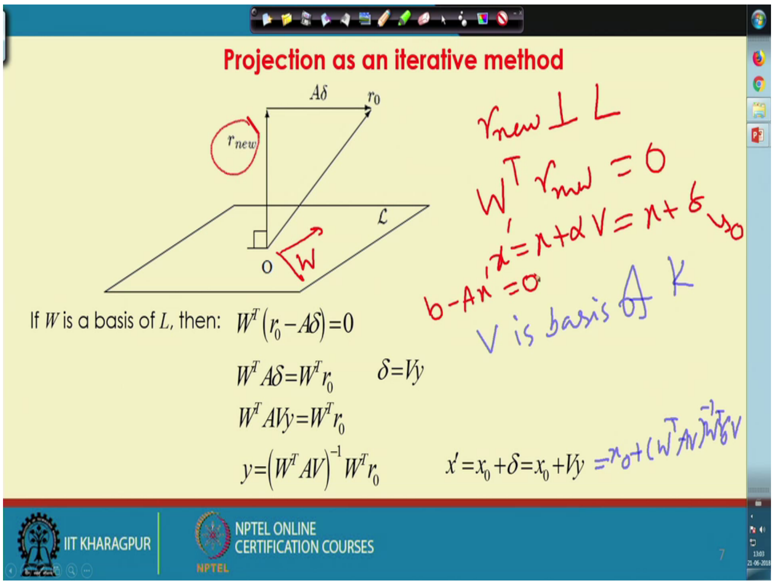
Now, if we redefine this method for an iterative to design an iterative method. We will start with an initial guess x 0 we will find an approximate solution on the new iterate iterative value x tilde x prime which is x prime is equal to x 0 plus k the update belongs to a certain space k such that the new residual b minus Ax prime is equal to is perpendicular to L. So, this is the updated value and this update is in along the vector space L and this is the new residual. And this is remember that this is during each iteration, so in the next iteration K and L might change in one particular iteration there we can see there is 1 space K along which I am updating x. So, I will start with x 0, I will give some K takes some vector in the vector space K add that with x 0 and find out the updated vector x t prime.

Now, the residual b minus Ax prime is orthogonal to one particular vector space L. K and L if they are same we will call that to be orthogonal projection. If K and L are not same we call them to be oblique projection. So, x tilde x prime is equal to x 0 plus V y where V is a basis for K. So, what is y? Y is basically combination of scalars which is multiplied with the basis vector or we can write x prime is equal to x 0 plus delta. Delta is the amount of change finally, the amount of change which is done to x added to x 0 to get x prime. So, here we can write that delta belongs to the vector space K, this thing belongs to the vector space K. And the initial residual R 0 is equal to b minus Ax 0. New residual R is b minus Ax prime. When we found the new residual we really we do not bother about the initial residual because our condition is that that new residual is, sorry our condition is that that new residual is orthogonal to L.

So, r 0 minus A lambda must be orthogonal to L and we find r is equal to b minus Ax J prime which is r 0 minus A lambda and this is orthogonal to L. So, now what we will try to do we will try to find out the basis vectors for L and when we will replace this basis vectors for L in this equation and take the dot product with the basis vectors of L with r 0 minus b lambda and get it is to be 0. And by that way, we should be able to get some relationship because if I look into lambda this lambda is nothing, but V y lambda is nothing, but V y. So, we can probably the next step will be r 0 minus A V y is perpendicular to L. So, if I get basis vectors of L and take a dot product that should be 0 that will give a give me some idea about y based on the how V and L are related.

If a L and K are same it will be probably much straightforward to find out y otherwise also there are some certain ways. So, let us go for one more step here, that is what we find out that the new residual is r 0 minus A lambda which is orthogonal to a space L; we have not said anything about L. We know that there is some spaces L forget the steepest descent where L is equal to R or minus grad J. In general projection method x is updated along a particular vector space K. Now, this updated x is giving me a residual b minus Ax updated or b minus r 0 minus a lambda r 0 was the previous residual. This r 0 minus A lambda is perpendicular to some vector space L.

(Refer Slide Time: 21:17)



And if we look into the projection method, we had an old r 0, what was the last step? R 0 sorry r 0 minus A lambda is perpendicular to L. So, if I subtract a lambda from r 0 we get the new residual r new which is perpendicular to L. So, you can write r new is perpendicular to L. Now, if we think W is the basis of L. So, there are a set of vectors which define W which is the basis of L we can write W transpose r 0 minus a lambda is equal to 0 because r new is or. So, we can also write W transpose sorry write that W transpose r new is equal to 0 because r new is orthogonal to L. And L is spanned by W vectors right this set is called probably W which are the independent linearly independent vectors spanning the entire set L.

So, transpose r 0 minus A lambda is equal to 0 and W transpose a lambda is equal to W transpose r 0.if we started with a guess x 0 r 0 is already known to us what is not known to us is lambda is delta that because delta is a vector which is along K we do not know exact value of delta, but delta is a vector which is along k. So, if we know K; we know that linear combination of the basis of K has given me delta. But what is the exact linear combination that we do not know and that that is what we are trying to find out. Well delta is equal to V y; this V V is basis of the space k.

So, if I if K and L are given we can we know about W and we know about V we do not know about delta or y. So, now, we write W transpose AV y is equal to W transpose r 0 or we get y is equal to W transpose AV inverse by W transpose AV inverse W transpose r 0 or we can get once we have got y we can get x tilde x prime is equal to x 0 plus tilde is equal to x 0 plus Vy x 0 or we can write it down that this is this is equal to x 0 plus W transpose AV inverse W transpose r 0 multiplied by V right ok.

So, if we now try to see what we have actually done what we have actually done is we assumed that x tilde is x prime rather is the new updated guess which is an approximate solution of some equation as Ax is equal to b. Now this solution is obtained iteratively by updating a guess solution. The question is how should we update the guess solution? The idea we got from steepest descent algorithm that we should update the guess solution in a sense we will minimize the functional J; that means, the x will be updated along a direction of gradient of J x will be updated to up to a point till this is tangential to a new iso contour of J.

After that the new update direction of x will be perpendicular to the previous update direction of J and this idea we carried that x is updated along a particular line. This update of x now maybe we know about the line, but how on that line x will be updated, what distance we are remember when we are doing steepest descent we are also trying to find out what is the distance till which you should update x till which we should move along gradient of J x 0 minus gradient of J x 0.

So, the we should move up to a distance while updating it. So, that the new residual is orthogonal to certain plane. So, we have two spaces L and k the basis vectors W and V and we are trying to update x in a way that x is updated along k. So, x is equal to x 0 plus V y.V is the basis of K and this y will come from the fact that the new residual b minus ax is perpendicular to some space W or W transpose b minus ax is equal to 0 though it looks very abstract in next session we will consider some like simplified case like we will take a one particular vector in V and one particular vector in W we will consider them to be one d subspaces.

And we will see that a lot of formulations come out, when we think of that. The only thing is that probably we will take from the idea of general projection method I said like this might look little abstract and the later formulations will be little more abstract. When you will go to kind of space type of methods, but we are updating x in an iterative method in a direction that is given some along certain in that in a direction which is a subspace in Rm.

When we are updating it we are taking care at that b minus Ax is the new after we update what the b minus Ax will get is the orthogonal to another subspace. So, we probably we will have decided about these two subspaces and we will see that this conditions hold and then we will say that my update has been fulfilled that b minus a updater is completed the iteration particulars iteration step is completed. I have updated x such that the new b minus Ax is a orthogonal to a given plane and we will continue this and we will see later we can show that this method actually converges.

So, one very important idea is that when we will go through number of steps in that direction finally, where we will end we will update x by infinite small amount. We will update x almost the K has a almost become the 0 vector we will update x by a very small amount and the new update the new b minus Ax is also 0 vector if x, If x plus x is equal to x plus alpha V right x is equal to x plus alpha V.

So, or a is equal to x plus delta if delta becomes 0 in that case my b minus Ax x prime b minus Ax is also 0 vector. So, when the update has become a 0 vector b minus Ax will also become 0 vectors and then that is on the iterations will converge. We will we will look into it in detail and we will see you and under what conditions for which type of matrices these iterations converge.

In next class we will start from here we will consider a particular subspace along which x is being updated we will consider another particular subspace to which the residual will be orthogonal. And, then we will try to see that that lead gives us a particular iterative scheme. For example, we will get steepest descent algorithm which we discussed last few classes as one by one particular choice of the spaces K and L. And, we will also see that this iterative scheme will converge to the right result that we will look into the next class.

Thank you.