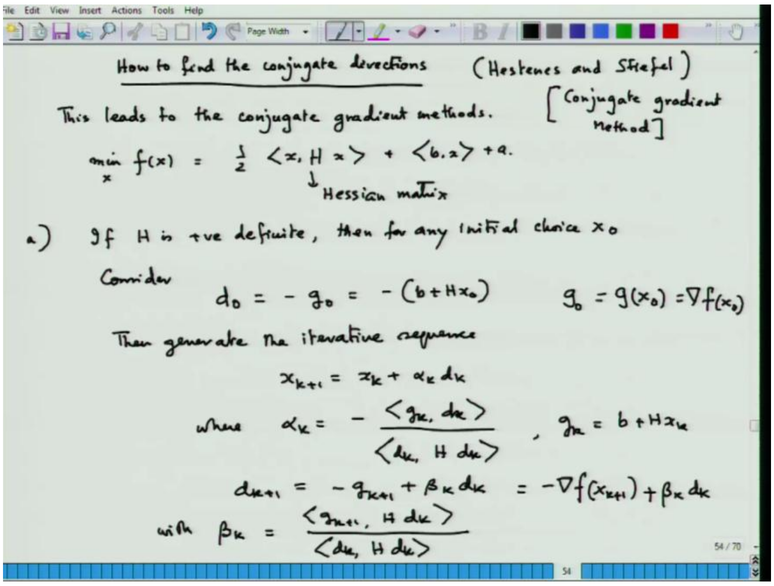
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture – 11

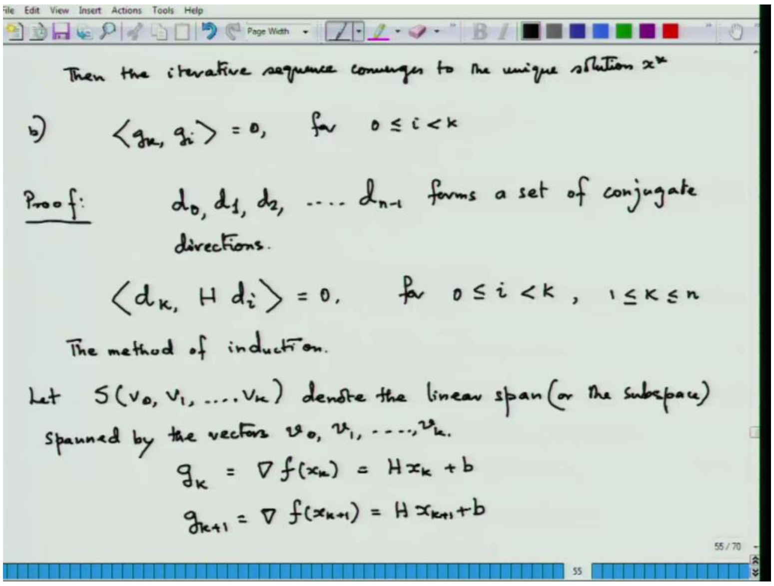
(Refer Slide Time: 00:22)



Last lecture on conjugate direction methods, our primary concern was how to find the conjugate direction because we had already seen that if I know about the conjugate directions solving a convex optimization problem. A quantity convex problem with positive definite hessian is a matter of n steps, but the question clearly lies at how do I find such a conjugate direction. And this method which we had written down in the last lecture was largely this was a method due to Hestenes and Stiefel Hestenses was responsible for a brilliant school on optimization and control at optimum, control calculus variations at Chicago. And here we call this particular approach is called the conjugate gradient method; it is called a conjugate gradient method, because my starting conjugate direction is minus g 0. So, you see what a what we generate here at least in the convex case that convex quadratic case that if I slightly tamper the idea of Stiefel Hestenses method then I can get a algorithm which is much faster than the Stiefel Hestenses method at least in the convex quadratic case.

So, here because we our initial direction, initial conjugate direction is found by taking the negative of the gradient at the starting point x naught, g 0 of course in this case you might be wondering g 0 is actually nothing but g of x naught. Now what Hestenses Stiefel shows that if you can construct a new vector which is the gradient at f x k plus 1 which you can also write it as like that. So, once you know d k you know x k and you find the alpha k which is of this form and then you get the x k Plus 1. Now, your d k plus 1 to go from x k plus 1 to x k plus 2 is found from d k in this sort of manner. Now the remaining question the main major part of the proof lies in the fact that this d 0, d 1, dot dot dot sorry d n this d n minus 1 d series of vectors are actually conjugate direction.

(Refer Slide Time: 03:20)

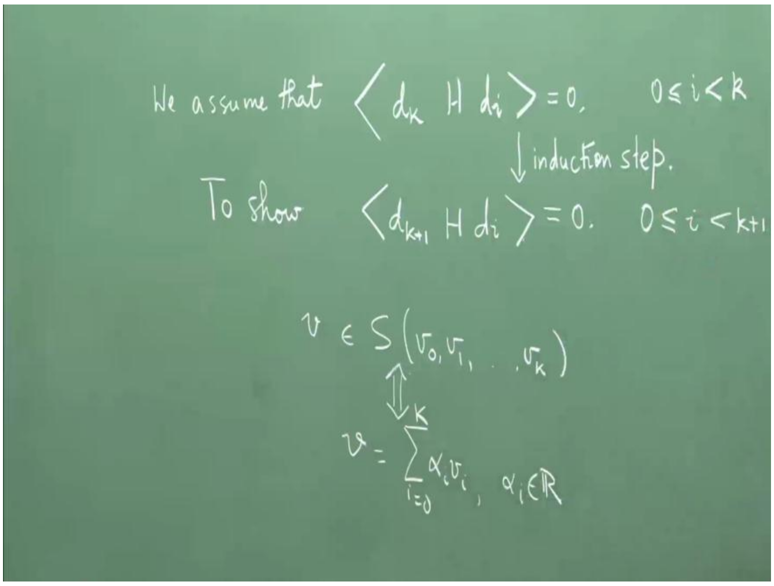


That is we will start to prove this thing, a proof methodology would first show we have to show essentially that is what we need to show at d 0, d 1, d 2 and so on up to d n minus 1 forms a set of conjugate directions. And once I know that these are conjugate directions then it is as same as what we had discussed before. And how do I prove that and what I need to prove. So, what I want to proof is that, so what I have to prove that d i H d j always a conjugate directions with respect to H because the original problem as the hessian matrix has H, and of course you have to consider this as positive definite which I have already written. So, H is my hessian matrix if you want to stress that at every point x.

So, if you prove d i H d j or d k H d i is equal to 0 and k is not equal to I or another way of showing is that this is equal to 0 for all i. So, you fix the k and you change the i's for all i, you just strictly and then k this is equal to 0 and this is true for, so if you can show these then you basically show that d i in a product H d j is equal to 0 whenever i is not equal to j. So, if k is one then you put d 1 H d 0; if k is 2 then you have 0 and 1 k is 2 then i is 0 and possible values of i's are 0 and 1. So, it will be d 2 H d 0 is 0 d 2 H d 1 is 0. So, in this way you basically show that d i H d j is equal to 0 when i is not equal to j. So, what would be our proof approach.

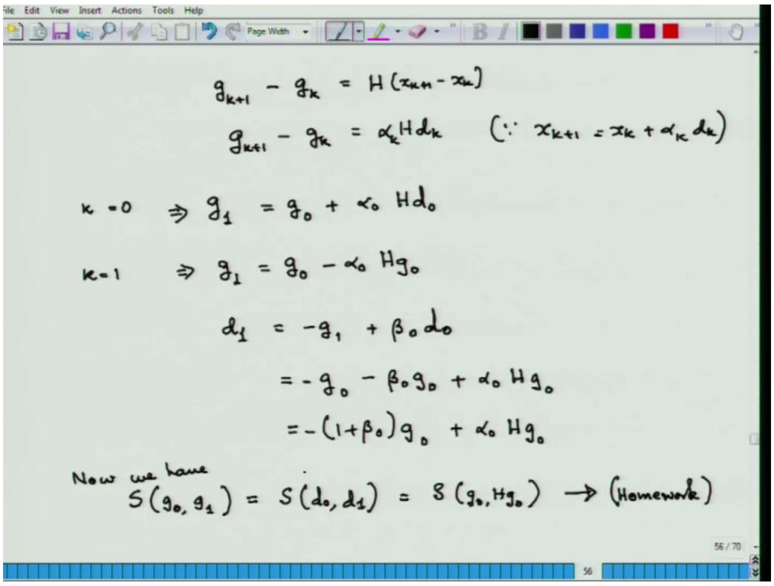
In this for these when you have these sort of things, when you have to prove something about a large number of elements, you use the method of induction you prove for the two, two of them and you prove for three and then you prove for k's and you prove for k plus 1 and that is what we are going to use here the method of induction. So by the method of induction, what are we going to show?

(Refer Slide Time: 06:59)



We are going to show the following we are going to say that let I will assume that let this is equal to 0. So, we assume for every i which is bigger than or equal to 0, and strictly less than k. So, once I know this what I have to show that, so this is what I have to show. And that is exactly your induction step. So, this is your induction step. So, once this is known then we try to use it now. Let S, v 0, v 1, v k denote the linear span or the subspace whatever linear span is a subspace or the subspace spanned by the vectors v 0, v 1, v k. So, what do I mean by this term spanned by the vectors v 0, v 1, v k, so you say that any vector say v is element of the span or e sometimes they write l s linear span or just l span of this. So, v is this if an only if this is a symbolic of only both ways v can be written as summation alpha i v i i is equal to 0 to k; where alpha i is in odd that is any element in this space can be written as a linear combination of the elements v 0, v 1, v k. So, that is the meaning of linear space. Now g k is grad of f x k which is H x k plus b. So, g k plus 1 which is grad of f x k plus 1 is H of x k plus 1 plus b.

(Refer Slide Time: 10:38)

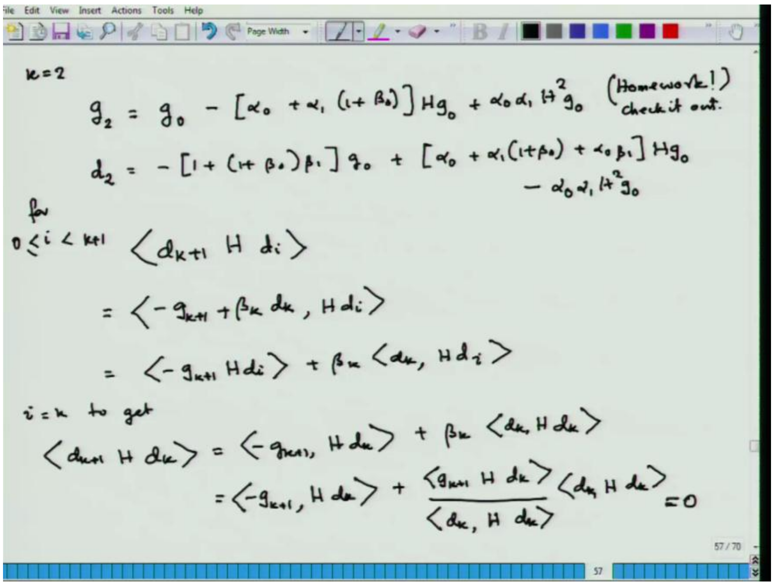


Naturally g k plus 1 minus g k is H of x k plus 1 minus x k. Now you know that if I go back to this setup x k plus 1 is written like this. So, x k plus 1 minus x k is alpha k d k. So, it is g k plus 1 minus g k is alpha times H of d k, where x k since x k plus 1 is x k plus alpha k d k sorry. Here I should put alpha k. So, this is what you have actually. So, this is the basic stuff we have now let us see what does this show. So, if I put k equal to 0 then I will have g 1 for k equal to 0; it implies that g 1 is g 0 plus alpha 0 H d 0 g 1 is a linear combination g 0 and H.

Now, since d 0 is minus g 0, I can write g 1 as g 0 plus alpha naught H. Now you have to put minus here, because d 0 is minus g 0 that is the basic assumption that is how you take the starting one. This is exactly this assumption and which we will now try to write down here that is what you have at this point. Once you have this I can keep on doing the stuff, now if I look at d 1 the direction one, it is minus g 1 plus beta naught d naught where d naught is again minus g 0. So, g 1 can be written as g 0 and d 0 can be written as minus g 0, so b 0 g 0. So, g 1 is g 0 minus alpha naught H g 0, so it is 1. So, y minus g 1 so minus g 0 plus. So, I will have a plus and a minus g 1 is minus g 0 plus minus g 0. So, that is what we will have.

So, now, I can write minus 1 plus beta naught, beta naught is the beta naught here. Beta naught is the beta naught that you already know here. This is beta naught, this is actually scaling up and generating the descent directions the conjugate directions. I can write this as this into g naught plus alpha naught H g naught. So, d 1 is the linear combination of g naught and H g naught, so that is for k equal to 2. So, what I have what I have concluded here, so g 1 and d 1 are the linear combinations of this; d 1 and g 1 are linear combinations g 0 and H g 0. Now I will leave it to as a homework to prove. Now what we will have here is S, now this of course, what we are trying to show that the linear span or the subspace generated by these two are same as these two are same as these two. So, this is a simple exercise in linear algebra which I pose as homework for your course. So, you have to show this as a homework.

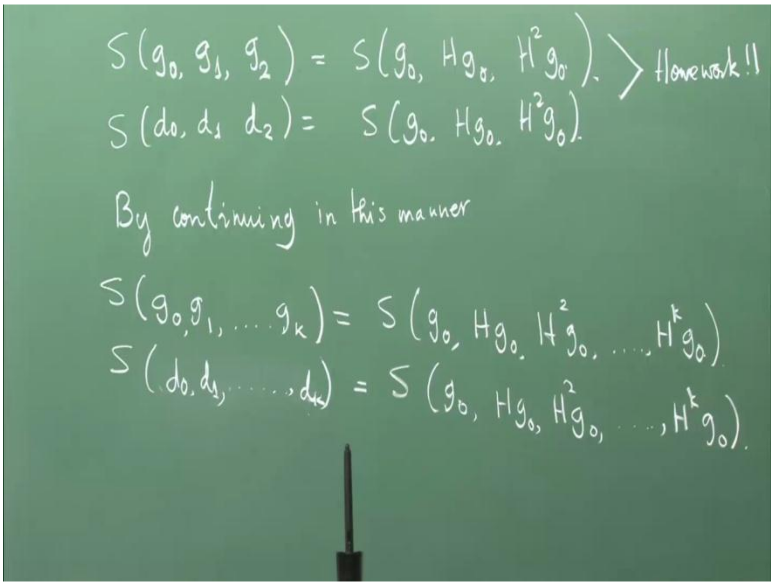
(Refer Slide Time: 15:42)



Let us look at k is equal to 2. So, for k equal to 2, we can again write g 2 as g 0 minus alpha naught plus alpha 1 into 1 plus beta naught 1 plus beta naught into H g 0 plus alpha naught alpha 1 H square g 0. How to compute these g 2 is again by putting k equal to 2 in the expression of g k plus 1 is equal to g k plus alpha k H d k and then successively suppose g 2 is g 1 plus alpha k H d k H d 1. So, you put what is d 1 and g 1 in their place you will get this expression. So, I will ask you to check it out as homework check it out.

So, it is your duty to check it out whether what we have written on the board is actually on the screen is actually correct. And now d 2, so here g 2 is a linear combination of g 0 H 0 and H square g 0. And so d 2 is written as so the coefficients are getting lightly slightly complicated once we introduce the other variables. So, if I know the gradient vector at this and this g at gradient vector at g 0 that is x naught, then I can know I can have a lot of information about the problem. So, here beta 1, beta naught I calculate accordingly to that beta k formula given in the beginning. Now what I want to say is that you observe again that g 2 is a linear combination of g 0 H 0 and H square g 0, while d 2 is also linear combination of g 0 H 0 and H square g 0 which shows that then from there you can have this conclusion.

(Refer Slide Time: 19:00)

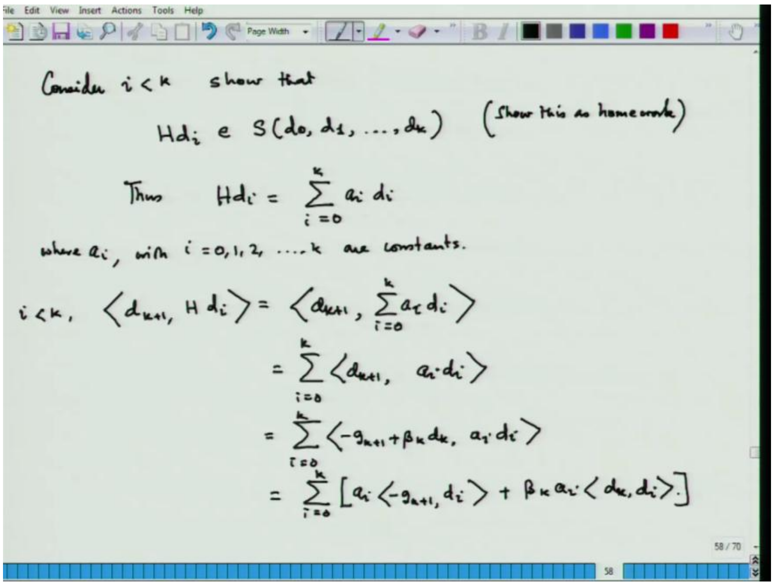


That a span generated by g 0, g 1, g 2 is same as the span generated by these vectors. So, again the span generated by see our whole goal now is we are trying to put that these are conjugate directions you see how much work is involved. It is very easy to state the theorem anybody can understand possibly why is done the bit of optimization, but a problem lies that once you start proving that their conjugacy of the directions then the difficulty starts. Now this is also equal to the span of, so that is what happens. So now, if you continue in this way, for say k steps k minus 1 steps essentially or k steps essentially then you will have the following by continuing. So, we know the result up to k. So, let us one go up to k by continuing in this manner that is you see what we did was d 1 was a linear span of 0 g 0 and H g 0. Then we expressed d 2 in terms of d 1 and press pack the value of d 1 to get g 2 and d 2, and now in that there are three terms now g 0, H g 0 and H square g 0 expressing g 2 and so is d 2.

So, what we found is these two. So, these two you have to again find as homework, it is not very difficult, it is plain simple manipulations. And by continuing in this manner, so I am continuing in a manner same as what we have just done here, and if you do so then I can extend these things. So, our stuff is up to k, we do not know what is there in k plus 1 that is exactly what we are set out to prove. This is true and of course, once you know this, you can of course, extrapolate and say that this is true then not g I made a mistake the directions is same as this. So, the thing in the top part of the board and the bottom part of the board looks similar, but it is enough that if you prove the top part and once you prove the top part you can understand the bottom part it is just an extension of that. So, once I know that, I now start doing the stuff. Now what I can now look at this expression, what I have to compute is for i for any i which is strictly less than k plus 1 what I have to compute is let us see how to compute this expression - d k plus 1 is of course minus g k plus 1 plus beta k d k. So, what I get here is minus g k plus 1 H d i plus beta k d k H d i.

So, now, what I do is to restart this thing and look at I am looking for i strictly less than k plus one. So, I can choose i equal to k. So, put i equal to k, to get now what happens to this that is a very important thing to know. Now if I put i expect this should be 0 and let us see what happens. Now I put b k is expression this one. So, now, b k would become, so what happens is that you see this gets cancelled with this and this, so this remains and which is nothing but the negative of this. So, ultimately this becomes 0. So, that is what the choice of beta k is particularly important and that comes out in a very natural way. So, now, which I leave for you to think about how did people conjugate of this expression for b k. It has got linked with what we have discussed in the previous section. Now, here we have this part. So, this part is 0. So, now, we have to look at scenarios where k is strictly less than 1; i is strictly less than k.

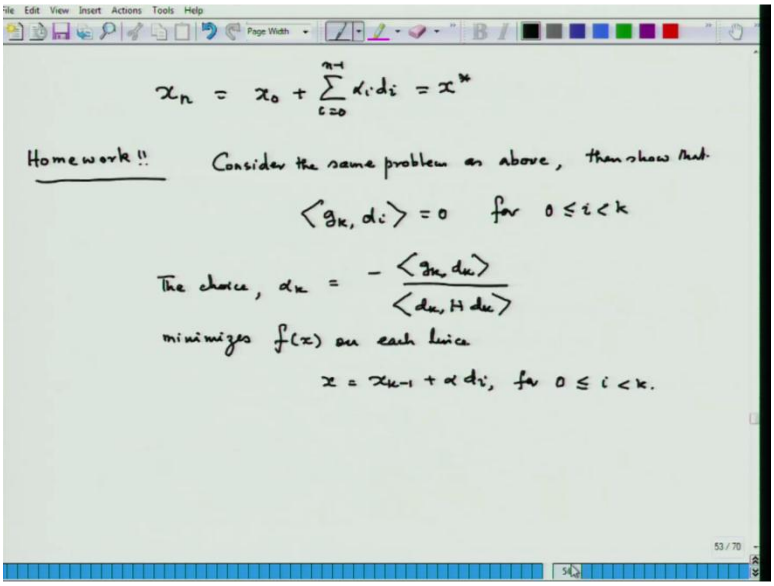
(Refer Slide Time: 28:05)



Now consider, now if I do that if I take now consider i is strictly less than k, now using these facts which I had already written down on the board; using these two facts, I expect you to prove show that this is a simple exercise in linear algebra and. So, I think this fact is important when you learn unconstraint minimum optimization. So, prove this, show this as homework. Now once you can, this is true actually H d i can be shown as a combination of g 0, g 1, g 2, g k and all those things. So, then you can basically show that because this is equal to this. So, you this so which means this is equal to this and this is equal to this, so this is equal to this.

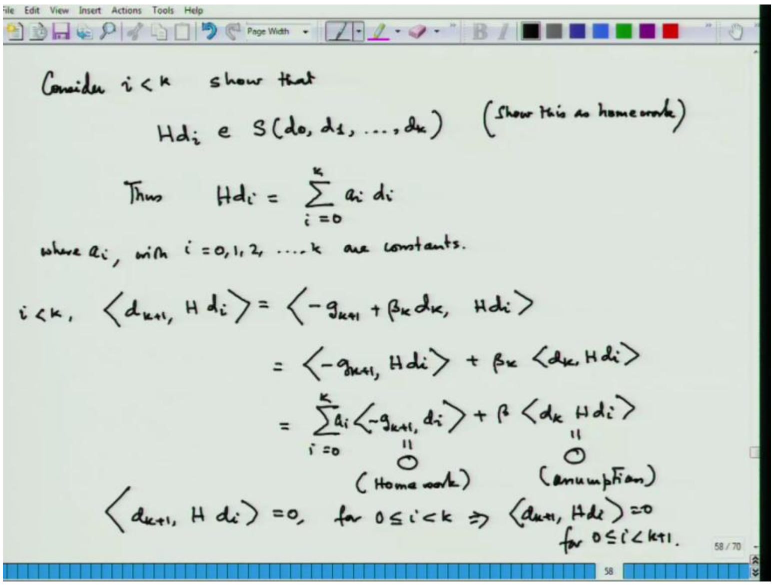
So, H d i becomes an element of that and so hence we can show that H d i thus H d i can be written as H d i yields you have some scalar a i, so i is equal to 1 to k this is how it happens linear combination of this. Where a i with i from these are constants and now once I know this then I again go back to this part, because now I have to look into the case i strictly less than k, because i equal to k, I know that d k plus 1 H d k is 0. So, this can be again written as because of H d i is minus a i d i. Now what I am going to how do how do I prove that I need to prove that this is 0. So, how do I do this. So, I will write down H d i H d i is so this goes the calculation. Now what is d k plus 1; d k plus 1 is nothing but minus g k plus 1 plus beta k d k, and this a i d i. This will give me i is equal to 0 to k; if I break it up inside; this is minus g k plus 1 a i of course, is a scalar which I can write in the front. So, I can take this scalar out and so I can have a i times minus g k plus 1 d i plus d k times a i d k d i. Now for this to prove this 0, we need to go back to the homework that we gave in the last lecture.

(Refer Slide Time: 33:02)



So, consider that, so how to prove this 0. See now you observe that. So, if the same problem I have asked you to show that g k d i d i as a conjugate directions is equal to 0. So, up to k, I have conjugate directions. So, g k plus 1 d i, i is strictly less than k. So, it is strictly less than k plus 1, so this is 0. So, this becomes equal to 0, because of this assignment just go back and try to see. So, if each d i is so if we I have kept here g k plus 1 then I have d i, where d i are behaving as a conjugate direction till k then this holds to k, so I go back. So, using that particular think this becomes 0 while I think there is a small little bit of change I will do in the calculation. Let me rub off this to make it much more effective, let me not put a i d i here just I am going to change the calculation.

(Refer Slide Time: 34:26)



So, let me do something here. So, what I do is I first take the d k plus 1 and write it as minus g k plus 1 plus beta k into d k and that I put with H d i, because then I can have minus g k plus 1 H d i plus beta k d k H d i. Now this is 0 is already known that is the assumption we have made. And here I will have by putting H d i is this, I can write this part only as i equal to 0 to k a I into g k plus 1 d i. So, we know that from again I go back to this homework where is that. So, whenever k is bigger than this i. So, k plus 1 is bigger than i, because i is strictly less than k then this is always 0. So, then using that fact we can now conclude that this is 0 and this is anyway 0, because that is the assumption. So, what I have is that this is 0 and what I have is that this is 0, this is 0 from the homework and this is 0 from assumption.

So, ultimately, you have d k plus 1 H d i is equal to 0 or i strictly less than k bigger than or equal to 0 of course, but and with k I have already proved. So, what we finally, proved is that d k plus 1 H d i is so this would imply finally that d k plus 1 H d i is equal to 0 for I strictly less than k plus 1 and hence we have proved this thing. So, added with this result, this result has two parts is this result. In our next lecture, we are going to prove this part and we are also going to complete the homework that I had given. We are also going to complete this homework where is it this homework as well as we are going to complete the proof of this part. Once we do that we will write down the conjugate gradient algorithm, and once we do that we will stop our discussions of conjugate gradient algorithms for the moment, and then go on to study what is called quasi Newton method. You see the proof here rests largely on the fact of the clever use of identifying two types of subspaces generated by very different sort of basis. So, basically you are finding a subspace the same subspace, but with different basis. So, these writings here is of same subspace, but here this is the basis and here this is the basis. So, it this is generating this and this is generating this.

So, we end our talk today and first in the next lecture we are going to first concentrate on proving the b part of what is left and then which is also very important to know and it is just not a very trivial proof. And once we do that then we can write down the conjugate gradient algorithm, but before we will also complete the solution of the homework that we have given. So, with that we finish our talk today. Thank you very much, please go through this calculations very very carefully because conjugate gradient calculations though they are simple they are not so straight forward because really you really have to do this linear algebraic manipulations to show that and I expect that you really take some time to have a look at these things.

Thank you.