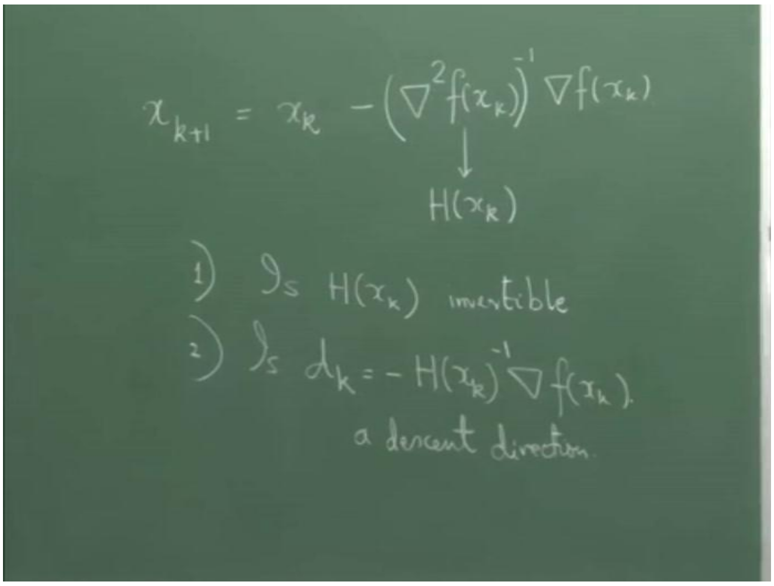
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 9

Welcome once again to this course on foundations of optimization. And we had just learned about Newton's method its properties of very fast convergence to the solution, if you are starting from a point very near the solution which is a very difficult thing because of course, you want to find a solution you do not know the solution. So, Newton method has some sort of this Gesentas business, because you start from a point in which you could be lucky that you are near the solution and in which you could be unlucky that you far from the solution and you go further away from the solution. The other drawback for the Newton methods which is for example, this let me tell you what could be the possible draw back.

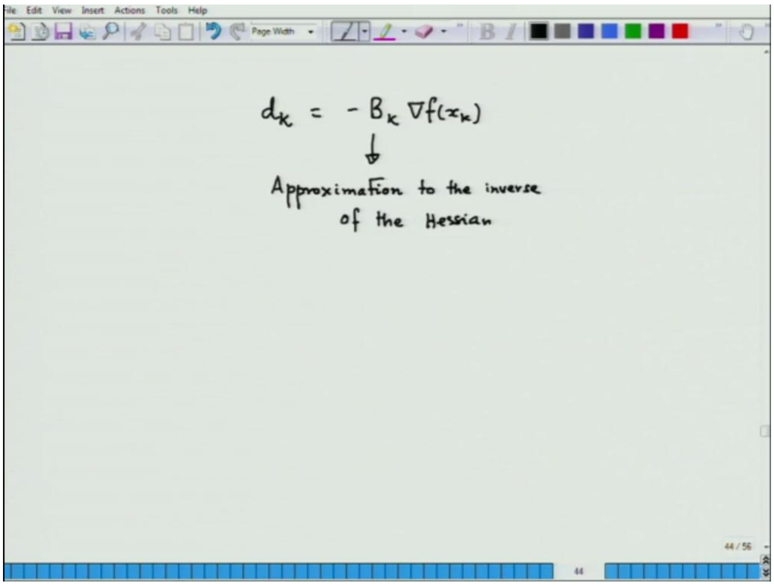
(Refer Slide Time: 01:04)



So, the Newton iteration is of this form k plus oneth iteration is related to kth iteration in the following way this is what is called the Newton iteration which you are completely aware by now. Now, the major issue here is as follows that this, this hessian matrix whose inverse you want to take. Suppose so if I denote this by H of x k which is quite the standard practice in optimization literature. So, question arise is, is follows number 1 is H x invertible, can you get that for functions which are not strongly convex if a function is strongly convex, then it is alright.

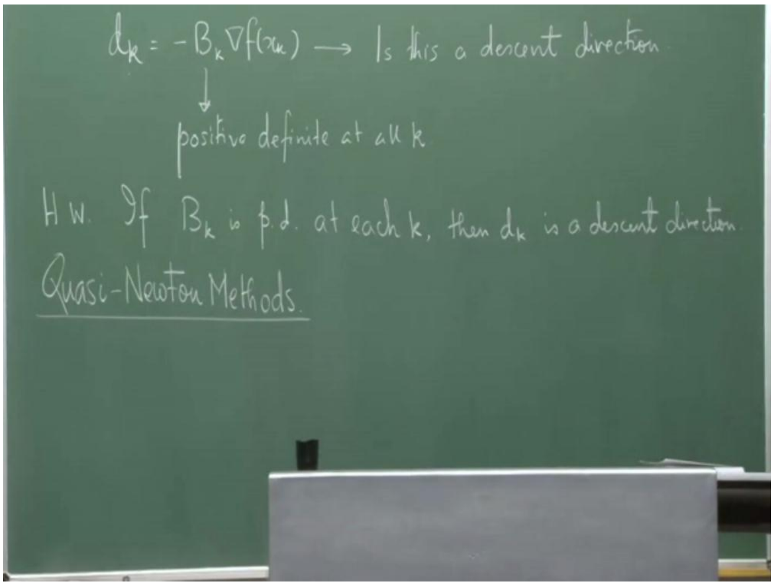
Then for Newton method is possibly very good for strongly convex function and it has unique minima so is much better is. So, this is one question. Another question is even if this is invertible, assume that it is invertible does it mean that I will always have this as a descent direction shall I have this as a descent direction that is a very important question. So, now the question is if I these are my draw backs of the Newton method, how do I go about trying to remedy this method. One approach of remedying is possibly the following there instead of having matrix b instead of having the matrix H x k.

(Refer Slide Time: 03:33)



I considered the matrix I consider the decent direction given by some matrix B k. So, instead of having the inverse of the hessian matrix I consider some matrix of this form. So, this is approximation to the inverse of the hessian, we you can consider it like this when you replaced the hessian inverse, approximation to the inverse of the hessian.

(Refer Slide Time: 04:28)

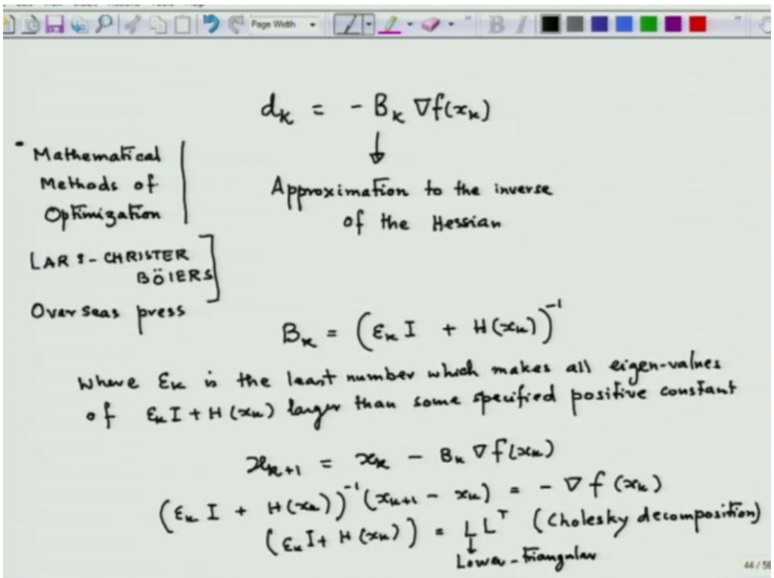


Now, you have chosen direction of this for so so you have solved this question by telling that I will not taken a matrix I do not I would not have to invert the matrix I will a matrix which will approximate the hessian itself the inverse of the hessian itself not hessian inverse of the hessian. So, I do not have to bother about inverting the matrix, but I have to really bother that if I want to write this the direction as this by replacing this as d k possibly I have to check if Is this a decent direction. Now, how do I know this, this would be a descent direction, one guarantee of getting this as a descent direction, this is to design B k at every step k in such a way such that B k is positive definite at all k. So, this is what you would want, because that is that will give you the decent direction. So, if I take d k as this as positive definite at every k prove that d k is a decent direction you can take as an home work.

So, if B k is positive definite which I am writing in short p d at each k, then d k is a descent direction. Now, once I know this let me, the question is how would I construct that B k? Who, who how do I know how to construct? One, so this whole gamete of the story of constructing B k of course, you can start with the initial one could be the identity matrix itself, because that is easy. But then at every step we have to update and get a new B k. And the then new B k should always be positive definite. This thing gives rise to a whole gamete of methods call the Quasi Newton methods or Quasi Newton method, which are quite powerful and actually used in software’s.

Now, we are not going to discuss right away Quasi Newton method, we are going to show you some way of constructing B k using the hessian itself. And we are going to then go into some simple and beautiful method called the conjugate gradient method which possibly is one of the not, not only beautiful very powerful method for solving at least convex optimization problems and even some non convex optimization problems of course, differentiable once. So, now one way of constructing B k as given in the book by Lars Christer Boiers I would like you to take down this example. And this book is available as an Indian print it is called mathematical methods of optimization by Lars Christer Boiers.

(Refer Slide Time: 08:50)

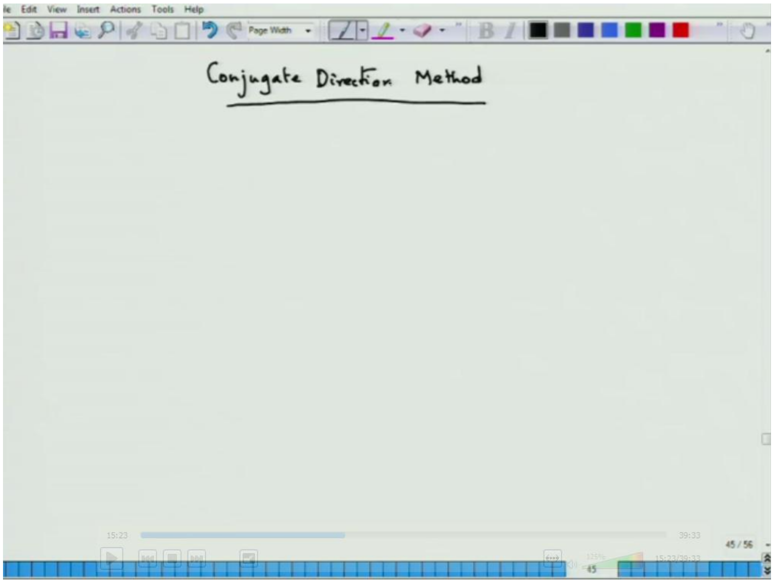


So, one of the books we will be using while discussing is the following mathematical is published in India by overseas press, the author is Lars Christer Boiers of Sweden if I am not wrong I guess he is from Sweden from Sweden from (( )). And it is published by overseas press in India is what Christer, Chriter Boier suggest, Lars suggest to us. He said you could possibly choose B k as follows I is the identity matrix plus the hessian. So, he is not inverting the hessian he is adding an multiple of the identity matrix. And the new matrix He is inverting and calling that matrix B k where epsilon k is the least number which makes all Eigen values of epsilon k I plus H x k larger than some specified positive constant. So, what does this mean?

This mean that I choose epsilon k in such a way that I can make the Eigen value of this matrix lager than a fix, fix some specified positive constant which means that for some k for some epsilon k value. This matrix would be positive definite, because all the Eigen values would become positive. And hence is inverse would remain positive definite, and hence then B k would be positive and then you could positively choose. But the interesting fact that you really do not go to invert matrixes, because matrix inversion is a very very costly operation in a computer if you want to invert a matrix it takes a lot of memory space and all sort of things which computation of expert should call cost of computing. So, people want to avoid matrix inversion.

So, in order to do that one can actually write the this whole thing as follows I can write this as this is absolutely simply manipulation writing B k and then B k as this and then doing this. So, I can write this so this could be mine modified Newton equation, and then what one can do is break up this matrix, because this is a positive definite matrix I can break up this in to lower triangular matrixes here I can decompose in to lower triangular matrixes. So, this is called the cholesky computation and that is how people actually solve it I am not going in to the details. You can see in any book or linear numerical linear algebra. So, this matrix is lower triangular matrix, lower triangular matrix I am not going to the details of how you arrive at such a result. So, that is not our job here. So, this is one of the ways to do it.

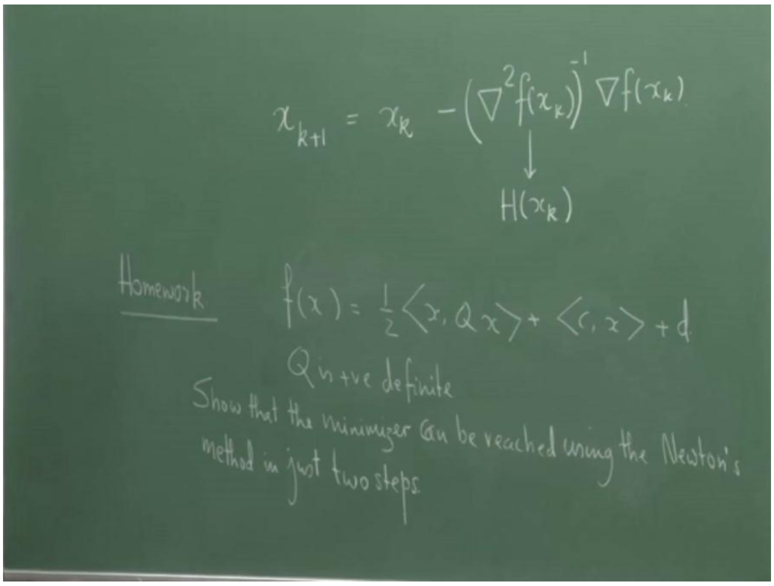
(Refer Slide Time: 13:57)



But now we will go into very interesting approach called the conjugate gradient method, do look its name might look very strange, we are having descent directions. And all those things suddenly we were coming to something called conjugate gradient. See the problem is people knew that it is once you are out of this steepest descent business it is not so easy to generate descent directions. So, there was the whole lot of story for developing Quasi Newton methods or Quasi Newton methods. And there also a story of not bothering so much about directions of descent, but developing certain directions along which you will definitely reach the minimum.

At least you will definitely reach the minimum for the quadratic case in finite, finite number of steps that leads us to what is called the conjugate gradient method or conjugate direction method? So, let me talk about conjugate or rather I would say instead conjugate gradient I would say conjugate direction that would be possibly better here. So, ultimately will come to something up on conjugated gradient method let me see conjugate direction method. Now, before I close of the discussion on Newton, let me just give you a small home work I think you should be able to do it.

(Refer Slide Time: 15:36)

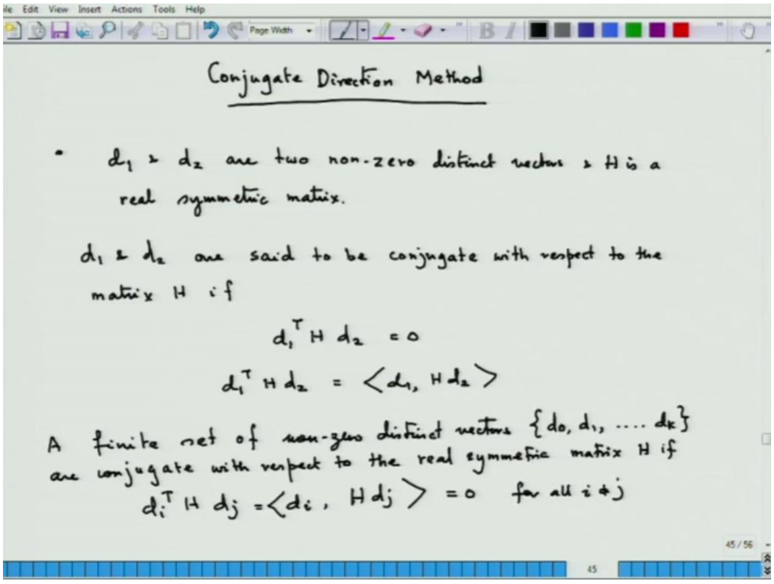


For the home work here as follows consider a function f x is the quadratic function where Q is positive definite. Now, if that is the case show that I can reach the minima of this problem, you know this minima can be very easily computed. So, show that the minima can be reached by using the Newton's method in just 2 steps for that minima minimum or minimizer can be reached using the Newton's methods in just 2 steps. So, please note this home work.

Now, let us speak about the conjugate direction method. So, we will start talking about conjugate directions. We will use the combination of 2 books to discuss about this. And this 2 books are are available in India. So, I am using the 2 books which have Indian additions, because I want to also encourage the, the viewers to buy these books and follow the talks accordingly. So, there will be certain things in the book which might not be very clear to you and it might be very clear to you when you listen to this lectures. So, it is very, very important that I I maintain books which are available in the Indian market that that; that is what I think to do, because numerical optimization is something which is been used by many many users, not only mathematicians they are used by engineers, they are used by many many other people physicist or biologist whoever.

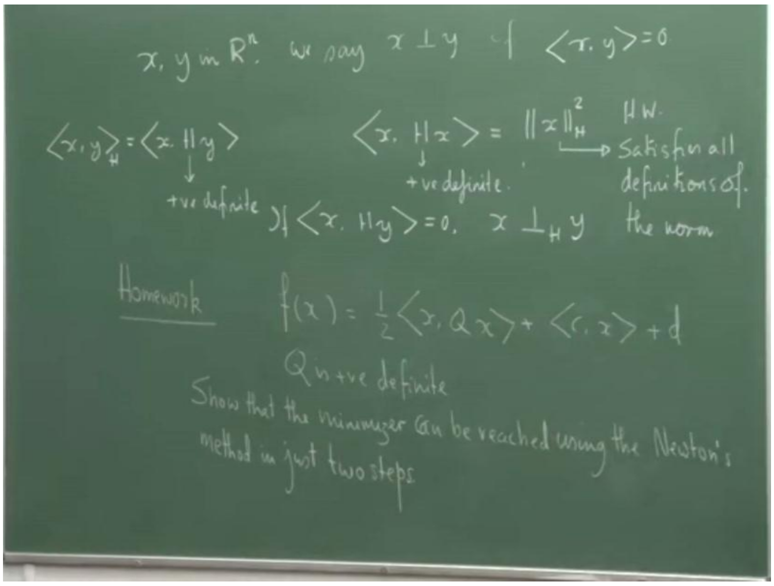
So, this would have a large audiences and books must be available. So, you cannot just learn by listening to lectures, because lectures would skip lectures would give you home works and all sort of things. But books would allow you to tell a bit more in detail and so as a result of which I would like to use books which are used which, which could be obtained in India. Now, another book which I am going to use along with the book of Christer Boiers is the book called practical optimization, algorithms and engineering application by Andrea's Antoniou Wu Shang Lu it is a Springer book. So, why to tell you the publisher Springer? So, this book is quite useful I will also use very soon the book by Roger Fletcher the legendary Roger Flexure, one of the greatest optimization geniuses in the World. So, he is sometimes called optimization resort. So, I would now start discussing conjugate direction method from this particular book.

(Refer Slide Time: 20:24)



Now, so when are two vectors distinct d 1 and d 2 two vectors, say d 1 and d 2 are two non zero distinct vectors in R n in R n, whatever I do not mind anything you can say R m, and H is a real symmetric matrix in course and off course, of. Now, when are these two vectors d 1 and d 2 said to be conjugate with respect to the matrix H. So, d 1 and d 2 are said to be conjugate with respect to the matrix H if d 1 transpose H d 2 is equal to 0. So, of course, I can write d 1 transpose H d 2 has d 1 in a product H d 2. Now, it is very important to know what sort of concept this concept of conjugate direction, this is generalizing did this concept come out of the blue from the air somebody discover it or there is something else which is involved here. So, that would lead us to notion of what I called an an elliptic norm.

(Refer Slide Time: 22:49)



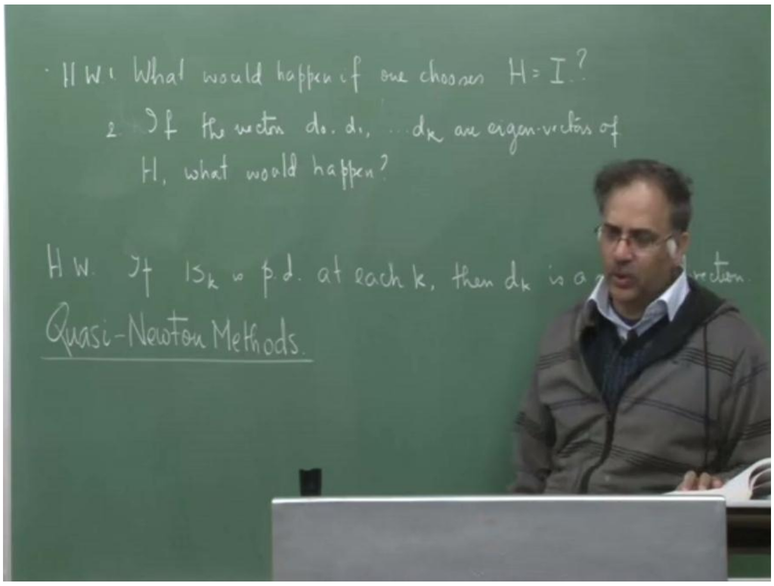
I feel if you have two vectors say x and y in R n, so x y is in R n we say that x is perpendicular to y or y is perpendicular to x, if the inner product of x and y is 0, there is the meaning of perpendicularity of two vectors; that is linear product is 0, because Cos 90 is 0. Now, if you look at it, look at this definition it is some sort of a perpendicularity. Now, what sort of perpendicularity it is? Now, if you look at let me define now consider this sort of inner product where H is positive definite. Now, if I write this, this is what is called see the norm square with respect to the positive definite matrix H it is again positive definite.

So, it is sometime called elliptic norm, now it is here a home work to prove that this satisfies this thing satisfies all definitions of the norm. So, consider this as your second home work today. Now, I can say that x and y are perpendicular with respect to H if x H y 0 that is in this particular known it is 0 in this particular set of inner product. So, if this is 0 if x H and y is 0 then we can say that x is perpendicular to y with respect to H possibly, we can invent a symbol like this. Now, including more meaningful to have H positive definite, because we can associate a norm with it, because in a norms of in a products always induce norms. So, here this could be taken as some sort of true inner product so inner product between x and y you.

So, inner product between x and y I can write in terms with respect to H. So, this also is an inner product this satisfies all the properties of inner product this. So, this inner product will only induced the norm if H is positive definite, but in general to make a slightly more general definition we take H 2 be just a symmetric matrix real symmetric matrix define conjugacy. So, d 1 and d 2 are essentially some sort of perpendicular vectors so there so because we can just a perpendicular we use a term conjugate may be this symbol could be used to denote conjugacy.

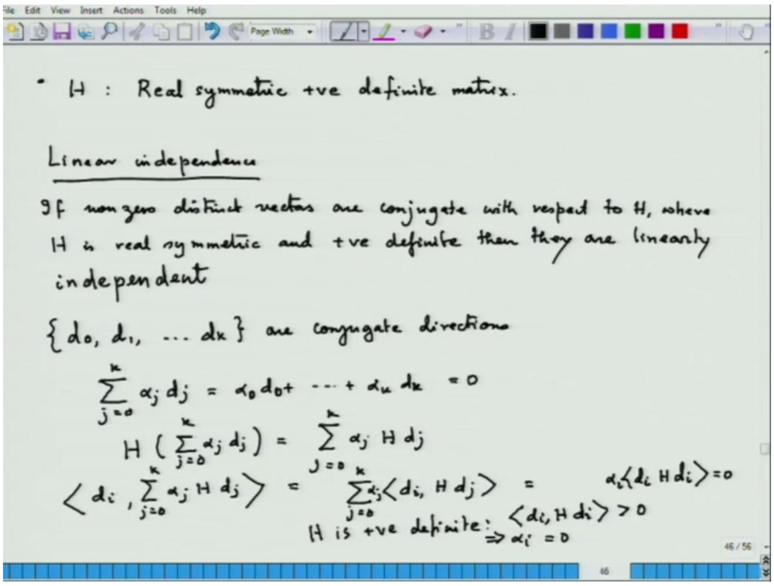
So, now let us see what is the say extend this definition beyond 2 vectors. A finite set of non zero distinct vectors d 0 d 1 d k are said to be conjugate with respect to a real symmetric matrix H if d i H d j is same as d i transpose H d j is equal to 0 for all I not equal to 0. Then I will now ask a few questions as home work again. So, you need to do something on your own also not just listen to the lectures, because that is the way you learn a mathematical subject.

(Refer Slide Time: 28:23)



So, answer is what would happen if one chooses H equal to the identity matrix I? So that is one thing. And the second problem is suppose this d 0 d 0 d 0 d 1 d k are actually Eigen vectors of H, then what is the relation between the vectors? So, if the vectors or Eigen vectors of H, what would happen? So, you have to find out what would happen? Tomorrow I will tell you the answers, but I will not tell you the way to do it, because the the thing is so so simple. Now, we will talk about conjugacy with respect to a positive definite matrix.

(Refer Slide Time: 30:16)



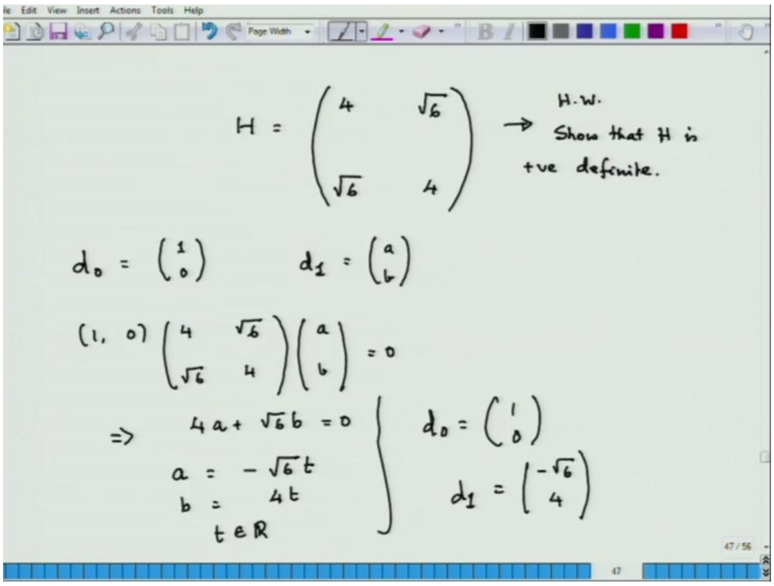
Now in this definition of conjugacy we will consider H to be positive definite matrix H now is a real symmetric positive definite matrix. So, once I have this once I have this I have the first fundamental property of conjugate vectors is linear independence. If non zero distinct vectors are conjugate with respect to H where H is real symmetric, and positive definite I did not write out where are the non zero distinct symmetric and positive definite. Then the set of vectors are linearly independent then they are linear independent.

So, that is what I want to now show you. So, let me show you that if I have a system now d 0 or conjugate. Now, consider this expression some lambda j has some scalar suppose for those scalars I would have this from j equal 0 to k that is alpha 1 d 1, alpha not d not to alpha k d k that is equal to 0 H of summation j equal to 0 to k. So, this is this is by the property of matrices that you really can figure out yourself, this is by the property of matrices. Now, once you have this, now do this; take this inner product.

So, what you would have again by the property of inner products this is not d j this is d i. So, I am taking the inner product with respect to d i now so d i h d j so an alpha j here. Now, for all those i which is not equal to j this is what you would have when this would become 0 by the conjugacy definitions. So, this will become alpha i d i H d i and this is 0. Now, H is a positive definite matrix this is nothing but the norm with respect to H and d i on is non zero. So, you would have summation alpha i, this should be gone. So, it will be just this, because all the other terms are 0.

So, basically because of the positive definiteness of H, because H is positive definite d i and d i is non zero implying alpha i is equal to 0. So, you can take here some other i say d 1. So, you will have this d 2; you will have d d alpha 2 is 0 d 3 alpha 3 is 0 so on so for. And that shows the conjugate direction when H is positive definite is nearly independent. Let me now show you today's class will end by showing an example of how to find a conjugate direction? Say two conjugate direction two directions they are conjugate with respect to each other.

(Refer Slide Time: 35:49)



So, take this matrix H so is a example given in Christer Boier's, Christer Boier's names are quiet difficult. Now, this is a positive definite matrix and now the home work show that H is positive definite. Then let d not be the vector 1 0 and say d 1 which I want to find is say some a b. So, let us see what happens if I do 1 0. So, this is what would happen if this two things are conjugate with respect to this matrix of course, they have to linearly independent. So, this would imply that so I can take a is equal to then this equation will be satisfied for t some real number, t is a parameter actually t is a real number. Now, put t equal to 1 then also it will be satisfy. So, if I take d 0 has 1 0. Now, I can take my d 1 as minus root 6 4. So, these two are conjugate with respect to this matrix Eigen these are of course, you can see not linearly dependent they are linearly independent.

Now, using this idea of conjugate directions we can show that it is very, very simple to minimize a strongly convex quadratic function over R n in the sense that it is n steps see these are not descent directions. I am not telling that conjugate directions are descent directions, but even they are not descent directions for the quadratic case in n steps without any sort of inversion of the positive definite matrix, which defines the quadratic function like this, no no inversion of the q they are on the board. You can show using conjugate direction that in n steps you reach the solution. And that is exactly what we are going to demonstrate in tomorrow's class.

Thank very much.