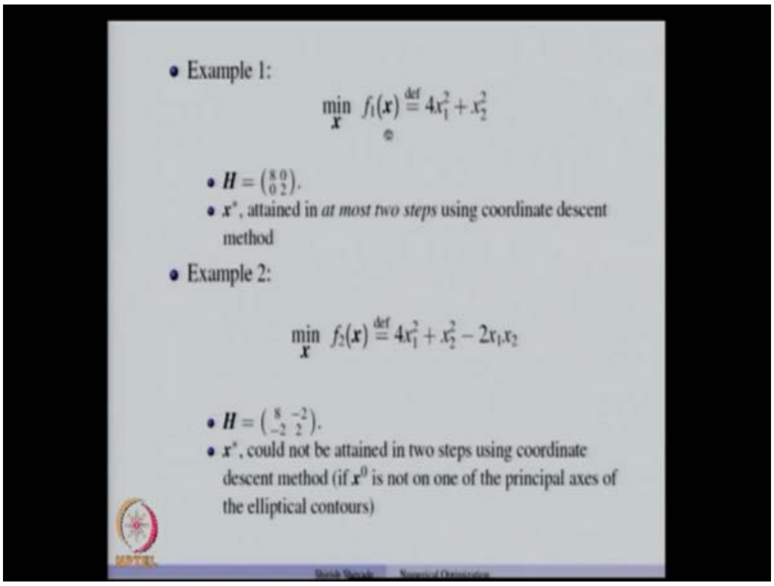
Numerical Optimization
Prof. Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Lecture - 18
Conjugate Directions

Hello, welcome back. In the last class we started discussing about coordinate descent method and we considered a couple of examples.

(Refer Slide Time: 00:29)

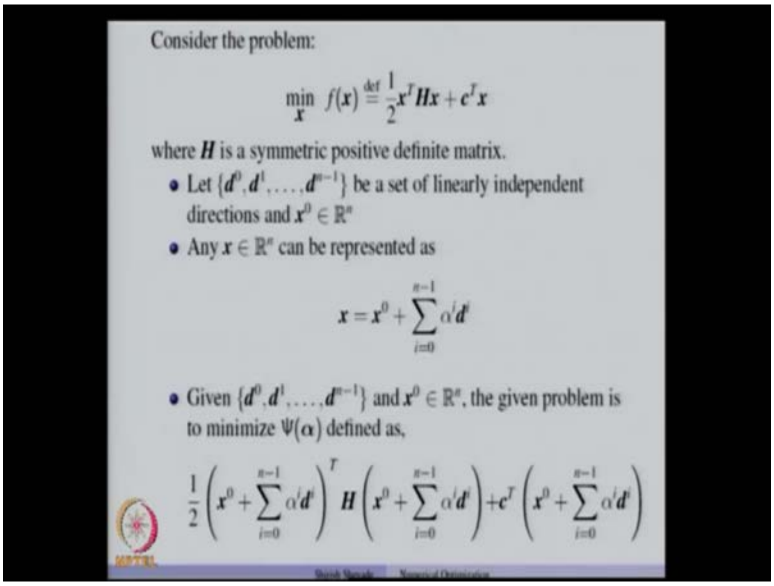


So, one of the examples that we considered was to minimize f 1 x, where f 1 x is defined as 4 x 1 square plus x 2 square and other example was to minimize f 2 x, where f 2 x was f 1 x minus 2 x 1 x 2. So, the difference between the 2 functions; f 1 and f 2 is that in f 1 the terms are separable in terms of x 1 and x 2. And in f 2 the terms are not separable, because there is a term which involves x 1 and x 2. So that objective function here in this problem is not separable in terms of x 1 and x 2, and that is reflected in the hessian matrix of that objective functions.

So, if you look at the hessian matrix of this objective function it is a diagonal matrix, while here if you look at the hessian matrix it is a positive definite, but not a diagonal matrix. So, both the hessian matrices are positive definite. But in one case it is a diagonal while in other case it is not a diagonal matrix, and when we used coordinate descent method to solve this problem.

We saw that when we minimize f 1 x, we attend the solution x star in at the most 2 steps while here we required more than 2 steps to attained the solution. So, the main problem in this case is that the hessian matrix is not a diagonal matrix. So, is there any way to transform the problem into some other space or in terms of other variables so that, this function becomes separable in terms of the variables. And if that happens then the hessian matrix will be a diagonal matrix and we would be able to use coordinate descent in that new space and get the solution. And at the most n steps for a n-dimensional problem.

(Refer Slide Time: 02:46)

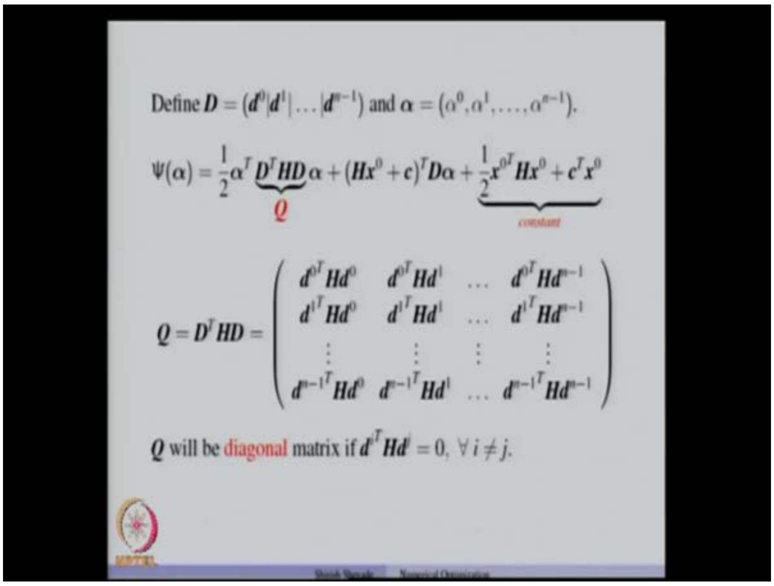


So, let us consider a general quadratic programming problem unconstraint problem, where do not minimize f of x which is defined as half x transpose H x plus c transpose x, where h is a symmetric and positive definite matrix. Now, let us assume that we have set of n directions d 0 to d n minus 1 which formal linear independent set. So, in some sense they form basis for n-dimensional space. So, they can be thought of has a basis for the n-dimensional space. Now, let us consider any point x 0 which is in n-dimensional space x 0 can be any point. Now, we know that any point in the n-dimensional space can be written as a linier combination of d 0 to d n minus 1 and plus x naught. So, this is a representation of x in n-dimensional space with respect to the basis d 0 to d n minus 1 and some initial point x 0.

Now, suppose if we replace this x in this equation into this equation. So, use this right hand side into this equation. So, the objective function then becomes a function of those alphas, alpha 0 to alpha n minus 1. So, given the basis say d 0 to d n minus 1 of n-dimensional space and some point x 0. We can rewrite the given problem in terms of the variables alpha and that is given here.

So, you will see that x here is replaced by the right hand side of this expression and now, we have a problem to minimize psi alpha with respect to alpha. Note that this alpha is the vector and that contains alpha 0, alpha 1 up to alpha n minus 1. And this alphas are real numbers, each of the alphas are real numbers. So, we have representation of x as obtained here using n real valued alphas. Now, once we have this problem we can minimize psi alpha with respect to alpha.

(Refer Slide Time: 05:47)



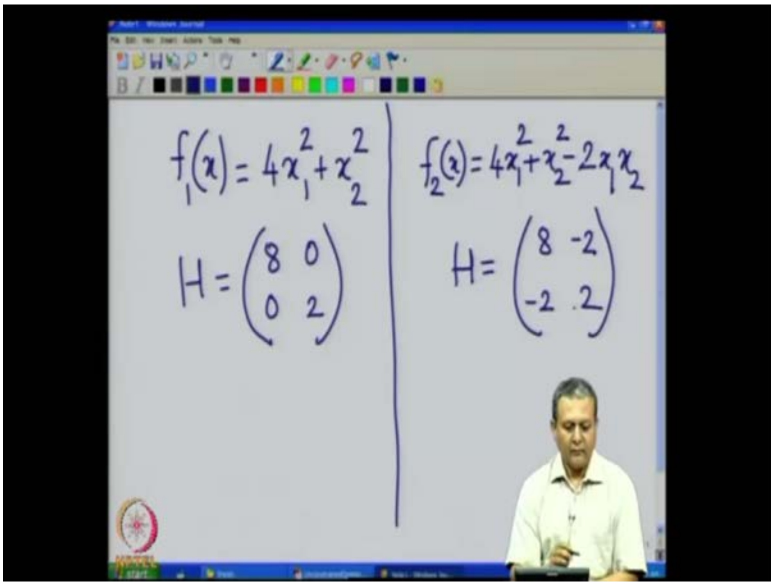
Now, let us use some notations. Let us denote by D, a matrix which is obtained by putting d 0 to d n minus 1 in n columns of that matrix and let us denote alpha to be alpha 0 to alpha n minus 1.

Now, if you rewrite psi alpha in terms of this compact notation, what will see is that. The psi alpha is nothing but half of alpha transpose D transpose H D into alpha plus H x naught plus c transpose D alpha plus half x naught transpose H x naught plus c transpose x naught.

Now, x naught is a given point. Since, it is a quadratic function the hessian matrix is constant and c is also a constant vector. So, this entire term is a constant. So, when we want to minimize psi alpha we can ignore this term because this term does not involve any expression involving alphas. So, we have to concentrate only on these 2 terms when we want to minimize psi alpha with respect to alpha.

Now, the hessian matrix of this quadratic function is D transpose H D and let us call that matrixes Q. So, how will those matrixes Q looks like. So, the matrix Q looks like this where on the diagonal you have d 0 transpose H d 0, d 1 transpose H d 1 and d n minus 1 transpose H d n minus 1. And of diagonal elements are d i transpose H d j is the, i j-th elements of the matrix Q. Now, if you recall our earlier example where we want to minimize a quadratic problem.

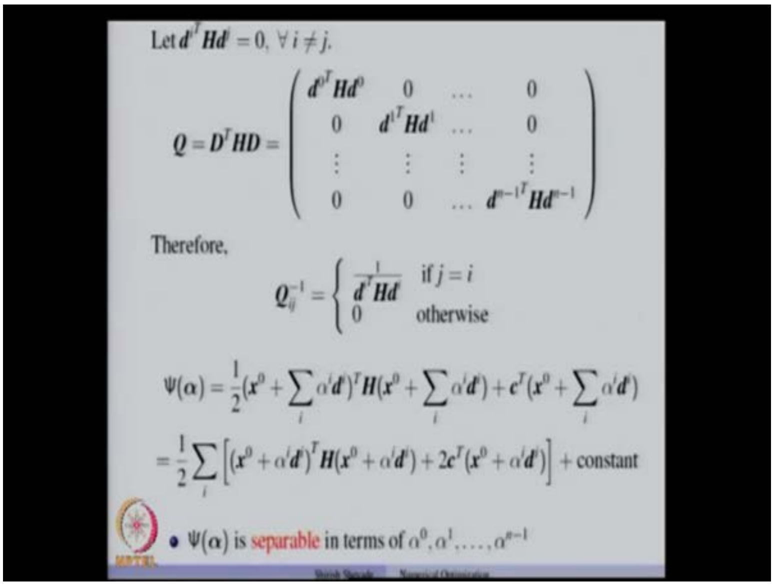
(Refer Slide Time: 07:52)



So, in our first case f 1 x was 4 x 1 square plus x 2 square and in the second case f 2 x was 4 x 1 square plus x 2 square minus 2 x 1 x 2. So, if you recall the hessian matrix, hessian matrix was 8, 2, 0, 0 and in this case the hessian matrix was 8, 2 minus 2 minus 2. So, in this case because the hessian was diagonal we could do the coordinate descent we could apply the coordinate descent method here and get the solution at the most 2 steps and that was not possible here.

So, suppose we decide to make this hessian matrix diagonal. So, the way to do that is that to make all this half diagonal entries in this matrix 0 or in other words whenever i is not equal to j make d i transpose h d j to be 0. So, this matrix Q will be diagonal, if d i transpose H d j is equal to 0 for all i not equal to j. Now, if you do that then we get Q to be a diagonal matrix.

(Refer Slide Time: 09:38)



So, let us assume that d i transpose H d j is equal to 0 for all i not equal to j. And then we have Q to be a diagonal matrix where you will see that all of diagonal elements are 0 and the ith entry of the diagonal is d i transpose H d i. So, inverting this matrix becomes very easy because we just have to take the reciprocal of each of the diagonal elements and put them on the diagonal, the rest of the elements are 0. So, they will remain 0 in the inverse matrix.

So, i jth entry of the matrix Q inverse is 1 over d i transpose H d i, if j equal to i and equal to 0 otherwise. Remember that whenever we want to solve or minimize convex quadratic function half x transpose H x plus c transpose x, what we need to is that said assuming that h is symmetric and positive definite matrix. What we do is that, take the derivative or the gradient of the function and set it to 0. So, what we get is h x plus c equal to 0 and in other words x equal to minus h inverse c.

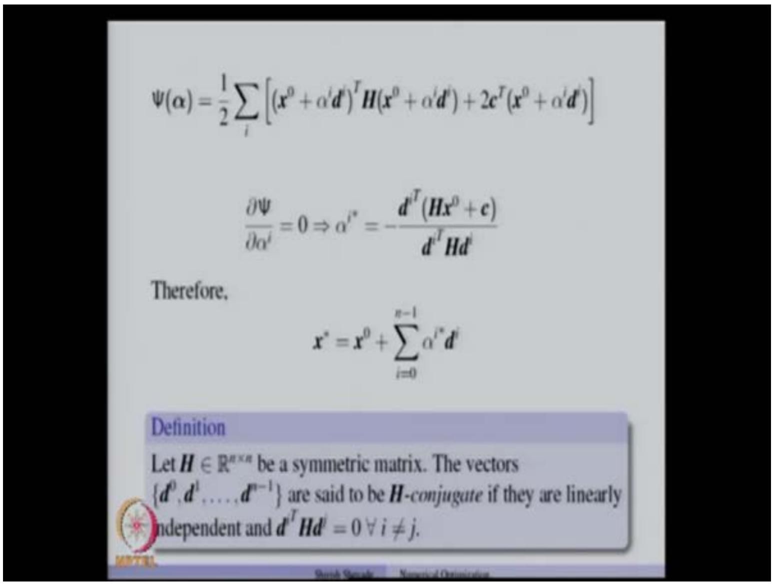
So, that requires inversion of hessian matrix and that inversion is easy, if we have h to be a diagonal matrix. So, Q inverse is like this and now if we look at our psi alpha. So, psi alpha if you recall that we had defined it to be half of x transpose H x plus c transpose x. Where x is represented as x 0 plus sigma alpha i d i, where d i's are i going from 0 to n minus 1 d i's are the basis of that n-dimensional space and x 0 is an initial point.

So, this was our initial objective function and now, with this condition that d i transpose H d j is equal to 0. So, what happens is that the terms involving d i transpose H d j they become 0 in this product and what we get finally, is a terms involving only d i transpose H d i, i going from 0 to n minus 1. So, there is the coupling which was there between d is and d j's in this objective function that vanishes if we assume that and d i transpose h d j is equal to 0 for all i naught equal to j. And therefore, so what we get is that apart from the constant which we saw that, we can ignore because it involves all terms involving x naught H and c. So, we can ignore that part and what remains is the term which is shown here and now, you will see that every term here in this objective function involves only 1 of the alpha i's or in other words now, this becomes a separable problem in terms of alphas.

So, this is a very important observation that psi alpha now becomes separable in terms of alpha 0 to alpha n minus 1. Because, this summation which was there in the earlier term which had this coupling between d i and d j in terms of, d i transpose h d j and that coupling vanishes because of this condition and we get so we have end terms here and each term is depended only 1 of the alphas.

So, the function psi alpha is now, separable in terms of alpha 0 to alpha n minus 1. And once we have this separability, it is easy to optimize this objective function individually in terms of alphas. So, let us see how to do that.

(Refer Slide Time: 14:04)

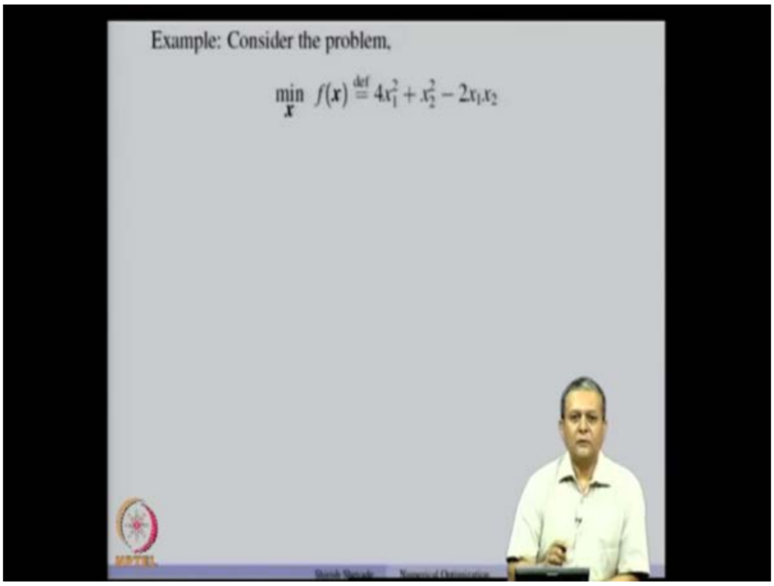


So, we have this separable function in terms of alphas. So, when we want to optimize with respect to alpha, what we can do is that we can take individual alphas at a time and optimize with respect to each of those alphas. So, in the alphas space you can think of it is a coordinate descent method. So, we take 1 alpha at a time and optimize with respect to that and what we get is by setting the derivative of this to 0, we get alpha i star to be minus d i transpose H x naught plus c divided by d i transpose H d i.

So, if we do it for all alphas. So, we will get alpha 0 star to alpha n minus 1 star and then we can get x star by plugging in those alphas star here in this formula to get our x star. So, computation of x star becomes very easy in this way, provided we ensure that d i transpose H d j is equal to 0. So, what are these directions d i transpose H d j is equal to 0? So, suppose we have a symmetric matrix which is n-dimensional then the vectors d 0 to d n minus 1 are said to be H conjugate. If they are linearly independent and d i transpose H d j equal to 0 for all i naught equal to j.

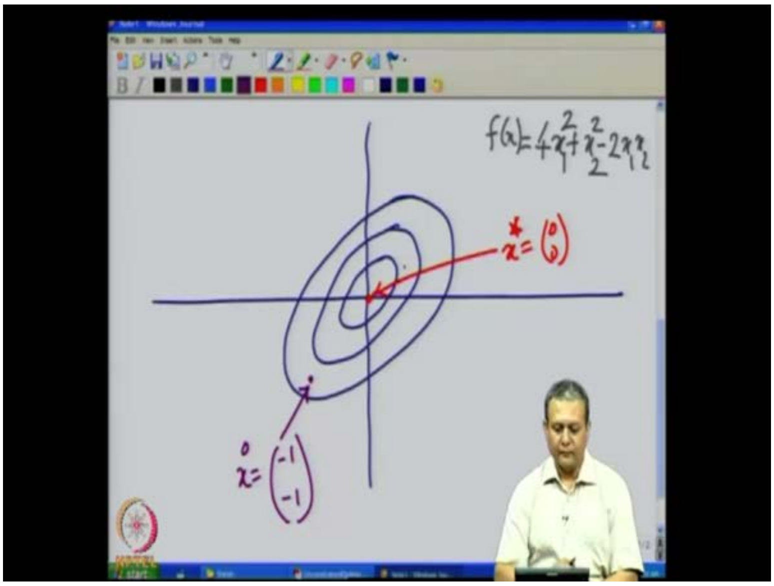
So, in other words, if we are able to get H conjugate vectors for a given quadratic function. Then we can convert it to a separable problem and get this alphas very easily and plug those alphas in the in this formula to get x star. So, what are this H conjugate vectors? These are the vectors which are linearly independent and they satisfy this property that d i transpose H d j is equal to 0 for all i naught equal to j. So, if H is a identity matrix, what we get is the orthogonal vectors. So, when H is a identity and d 0 to d n minus 1 are linearly independent. Then what we have is that d i transpose d j is equal to 0 for all i naught equal to j that means, that the set d 0 to d n minus 1 forms an orthogonal set of vectors. So, that is a special case of H conjugate vectors, when H is identity matrix. Sometimes this is also called H orthonormal set of vectors. So, the vectors are orthonormal with respect to the matrix H which is a symmetric matrix. So, these are call conjugate vectors or H conjugate vectors with respect to the matrix H.

(Refer Slide Time: 17:27)



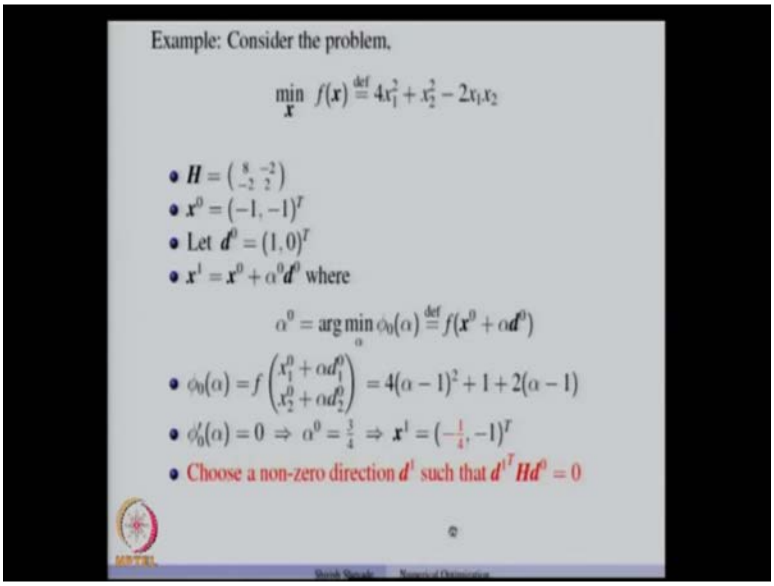
And now, let us consider the same problem formulation that we saw last time. So, we have this problem where we want to minimize 4 x 1 square plus x 2 square minus 2 x 1 x 2.

(Refer Slide Time: 17:50)



So, if you look at the contours. So, this is our function f of x to be x 1 square sorry 4 x 1 square plus x 2 square minus 2 x 1 x 2. Now, we know that this is going to be x star, this is the minimum. And suppose we start from the point x 0 which is minus 1, minus 1. So, this is our initial point and we want to find out some directions d 0 d 1 such that we can reach from x 0 to x star. So, let us see how to do that?

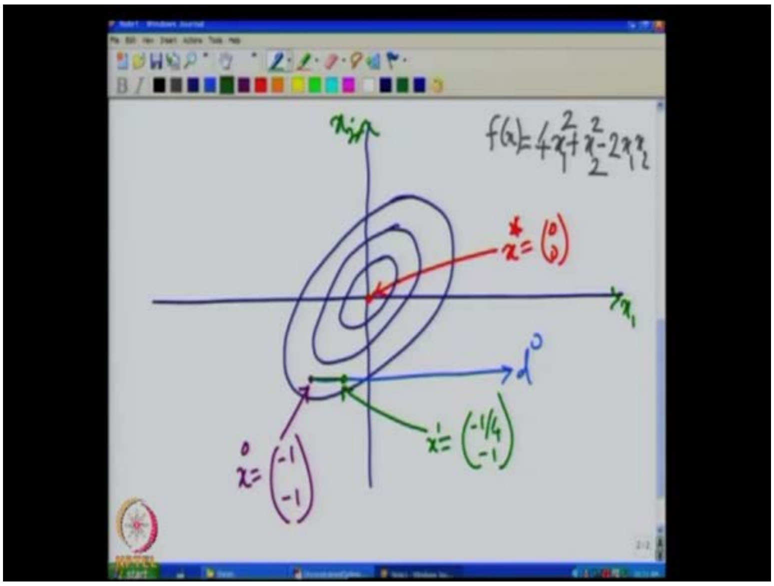
(Refer Slide Time: 19:29)



Now, we have already seen that the hessian matrix in this case is, this which is not a diagonal matrix. Let us consider some initial point which is minus 1, minus 1 both the coordinates are same. And let us consider the first direction to be the direction 1 0 so that, which we want to optimize only with respect to x 1.

Now, we have already seen this that x 1 is nothing but x 0 plus alpha 0 d 0 where alpha 0 is minimization of phi 0 alpha with respect to alpha. Remember that this alpha belongs to r because we just have a direction d, we do not want to make this alpha to be strictly greater than 0 as we did in earlier cases. Earlier we were working with descent directions so we wanted to move along the descent directions and wanted to have a alpha to be greater than 0. While here the direction that we are choosing, we do not know whether that is a descent direction or negative of this is a descent direction. So, we have to make alpha real valued variable in this case. And alpha 0 is obtained by minimizing phi 0 alpha with respect to alpha where phi alpha 0 is defined as f of x 0 plus alpha d 0. So, we use this last time and founded phi prime alpha equal to 0 implies alpha 0 is equal to 3 by 4 and we move to the point minus 1 by 4 minus 1.

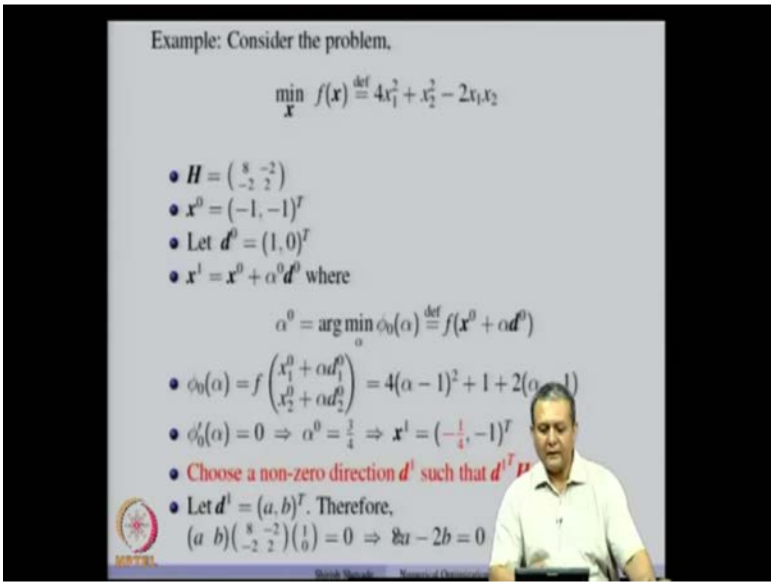
(Refer Slide Time: 21:20)



So, we move to the point, this point is x 1 to be minus 1 by 4 minus 1. So, if you recall in the earlier case we used the second coordinate direction. So, this is our x 1 and this is x 2. So, in the earlier case we use the second coordinate direction and we could not reach the solution in that case. So, let us see what happens if we use a direction which is H conjugate to the previous direction. Remember that we have used this direction as our d 0 and then we got the point in this direction, which is a minimum. Now, let us…

So, after having found x 1 which is minus 1 by 4 minus 1 our next step is to find out the direction d 1. Now, unlike the previous case where we use the direction d 1 to be the direction along the x 2 axis, we will use the direction d 1 which will be H conjugate to d 0 or in other words, we are looking at direction d 1 which is H orthogonal to d 1 or it should satisfy d 1 transpose H d 0 equal to 0. Now, how do we get this direction d 1.

(Refer Slide Time: 23:20)

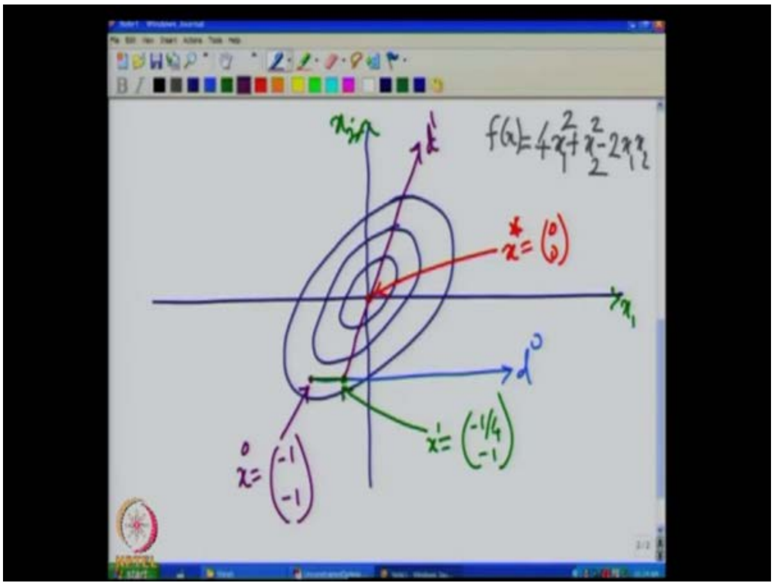


So, let us assume that the direction d 1 is a b, the 2 components of that direction and what we want is that d 1 transpose H d 0 to be 0. Therefore, what we want is that a b into hessian matrix into the direction d 0 which is 1 comma 0 that should be 0. So, if we now, rewrite this in the equation form what we get is that, the a and b of this direction d 1 should satisfy this property at a minus 2 b equal to 0. So, you can see that we have 1 equation and 2 variables and therefore there exists infinitely many solutions to this equation. Now, we can choose 1 of those its solutions so for example, suppose we choose a to be 1 and b to be 4. So, if you recall if we choose a to be 1 and b to be 4 that satisfies this equation.

(Refer Slide Time: 24:27)

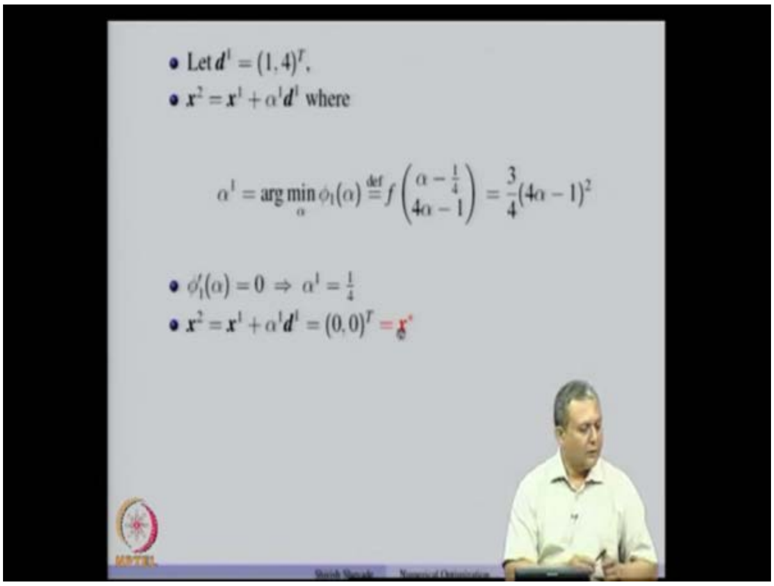
So, suppose we choose d 1 to be 1 comma 4.

(Refer Slide Time: 24:50)



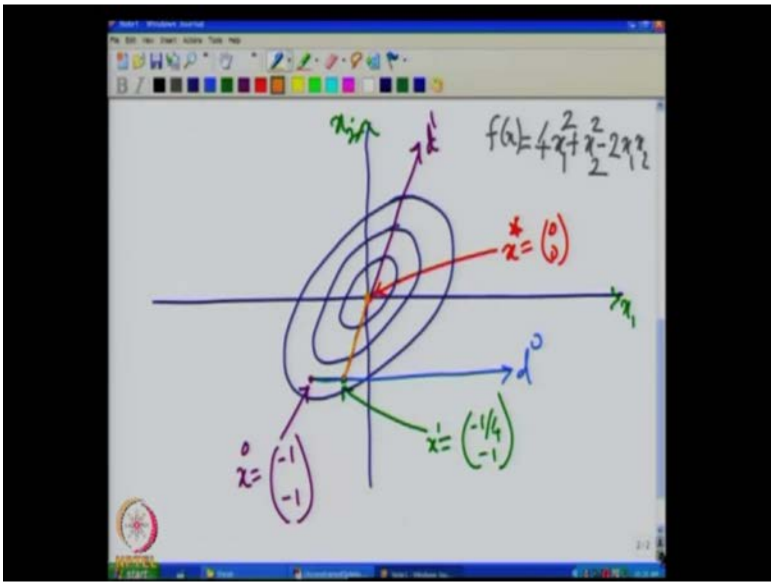
Now, if you use that direction. So, that 1 comma 4 direction will be, this is going to be our d 1 this is the direction 1 comma 4 from this x 0.

(Refer Slide Time: 25:21)



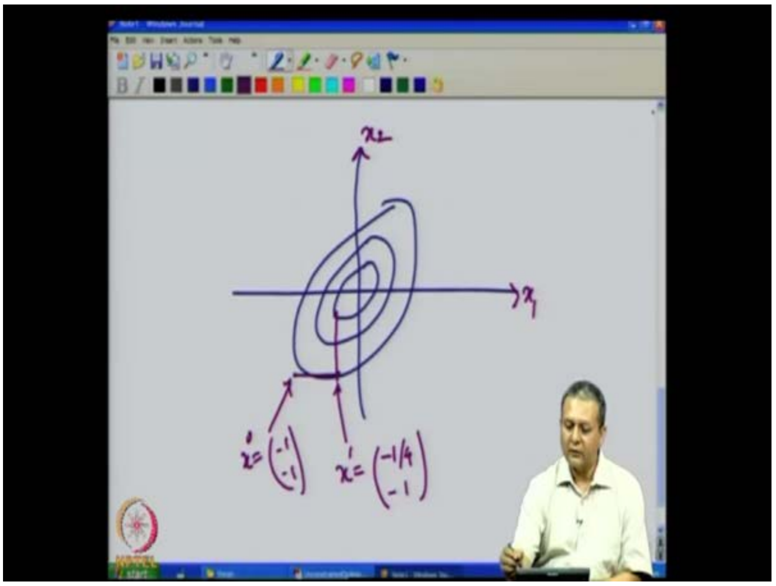
Now, x 2 is x 1 plus alpha 1 d 1, where alpha 1 is the minimum of phi 1 alpha with respect to alpha again I repeat that this alpha belongs to the set of real numbers not the set of positive real numbers only. And we get f of alpha minus 1 by 4 and 4 alpha minus 1, we plug in that in the original equation and what we get is 3 by 4 into 4 alpha minus 1 square. So, we take the derivative of phi 1 alpha with respect to alpha equated to 0, we get alpha 1 equal to 1 by 4 and when we use x 2 equal to x 1 plus alpha 1 d 1 and plug in the value of d 1 alpha 1 and x 1 which is minus 1 by 4 comma minus 1. What we get is 0, 0 which is nothing but our x star.

(Refer Slide Time: 26:23)



So, if we do the exact line search along this direction what we get is and we end up at x star. Now, compare this with our earlier approach where for the same function.

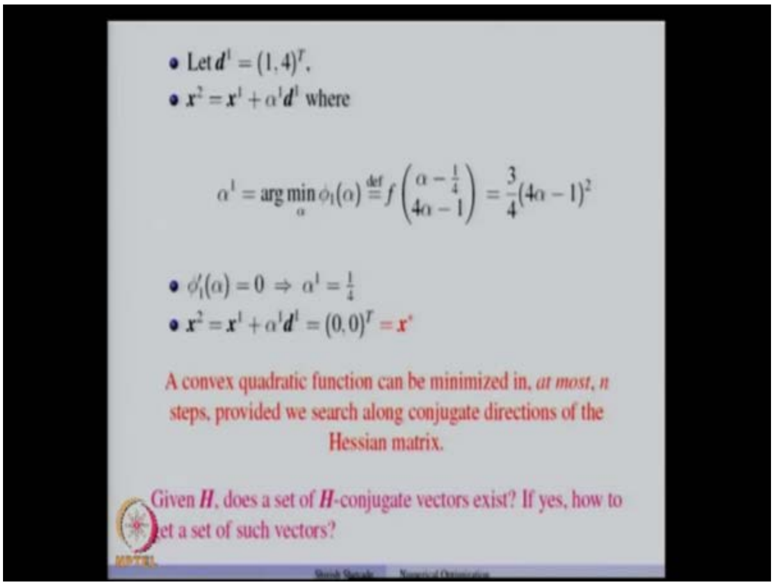
(Refer Slide Time: 26:41)



So, if we start from 1 minus 1. So, initially we went to… So, this was x 0 to be minus 1 minus 1 and reach the point x 1, x 1 was minus 1 by 4 minus 1 and then if we use the other coordinate. So, these are the coordinates. So, if we use the other coordinate we would reach some point which is here. Now, compare this so this was the coordinate descent method in the x 1 x 2 space.

While if you look at this, that the first iteration was along the direction d 0. But, second iteration was along the direction d 1 and that d 1 where chosen such that it is h conjugate to d 0. And if we do the exact line search along with d 0 we would end up at x star. So, this method required the most 2 iteration to reach the solution. So, this is a very important point.

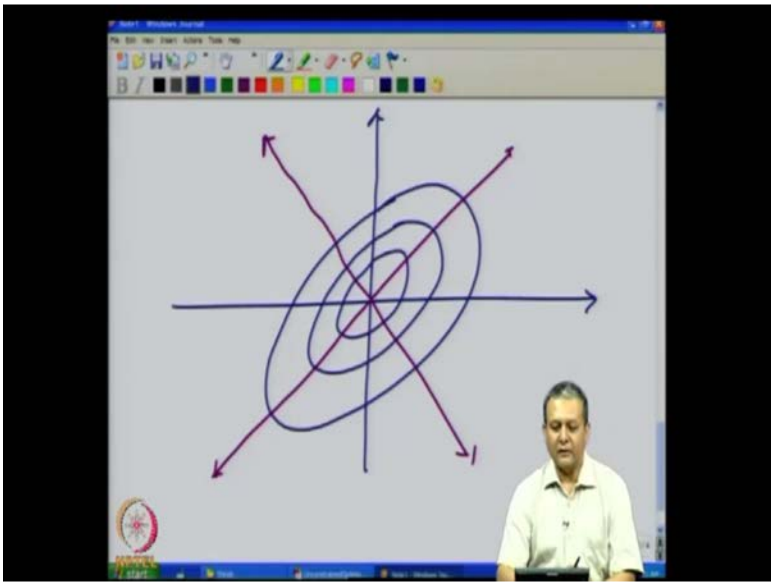
(Refer Slide Time: 28:28)



So, in 2 steps we got the solution x star. So, in general we can say that a convex quadratic function can be minimized in at most n steps provided we search along conjugate directions of the hessian matrix. So, if we have this information of conjugate directions for the hessian matrix then if we do the search along those conjugate directions at the most n steps, we will reach the solution of a convex quadratic function.

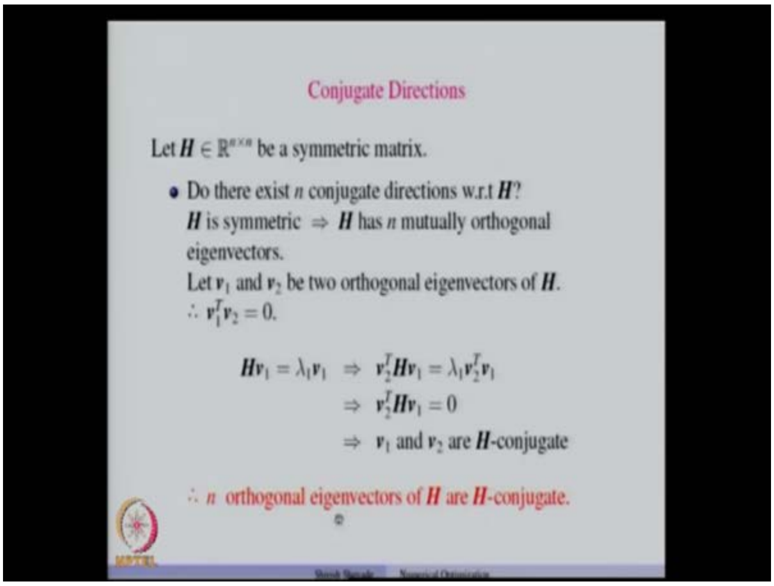
Now, there are some important questions that need to be answered and 1 of the questions is that. Given h does a set of h conjugate vectors exist? So, what is the guarantee that there exist H conjugate vectors? And suppose there is a guarantee that H conjugate vectors exist, then how do you get 1 such set of h conjugate vectors? So, these are the important questions that need to be answered and we will now, answer these questions.

(Refer Slide Time: 30:02)



Now, if you consider a quadratic function this convex, now if we look at the eigenvectors of this function. So, suppose if we transform the original system into our original coordinates into the new coordinates and then if we move, if we use these as our coordinates. Then we can use coordinates descent method in this new space of coordinate vectors. And if we use the coordinate descent method in this space you can see that we can reach the solution in at the most n-th steps. So, this the eigenvectors of this matrix the hessian matrix H is the 1 possible set of H conjugate vectors. Now, we will see how to show that.

(Refer Slide Time: 31:43)

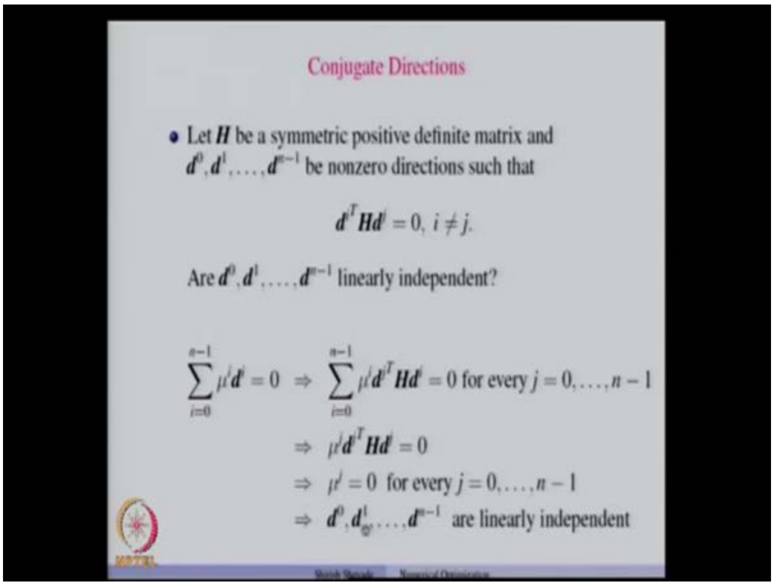


So, let us consider a matrix H which is the symmetric matrix and the question is that do there exists n conjugate directions with respect to H. Now, we know that H is symmetric, means that H has n mutually orthogonal eigenvectors. So, let us take a couple of eigenvectors which are usually orthogonal. So, let v 1 and v 2 b 2 orthogonal eigenvectors of H. And since H is symmetric we can choose v 1 and v 2 such that v 1 transposes v 2 equal to 0 or v 1 and v 2 are orthogonal. Now, since v 1 is eigenvectors and v 2 also Eigen vector.

So, of h... So, we can write H v 1 equal to lambda 1 where lambda is in Eigen value of H remember that this is the symmetric matrix. So, Eigen values are real so lambda 1 is a real number. So, h v 1 is equal to lambda 1 v 1. Now, if we multiply throughout by v 2 transpose H v 2 transpose. So, what we get is v 2 transpose H v 1 is equal to lambda 1 into v 2 transpose v 1. But, we know that v 1 transpose v 2 which is nothing but v 2 transpose v 1 with 0 so that, means that v 2 transpose h v 1 is equal to 0.

And v 1 and v 2 they are linearly independent and v 2 transpose H v 1 equal to 0. So, they form what are call the H conjugate vectors of H. So, v 1 and v 2 are H conjugate and therefore, we can say that the n orthogonal eigenvectors of H are H conjugate. So, we can extend this result for all the n orthogonal eigenvectors of H and these orthogonal eigenvectors of H. From this result you can see that there H conjugate. So, given a symmetric matrix there always exists H conjugate vectors and one set of H conjugate vectors is the orthogonal eigenvectors of H. Remember that this set of H conjugate vectors for a given symmetric matrix H did not be unique. This is the set of n orthogonal eigenvectors is just a 1 set.

(Refer Slide Time: 34:28)

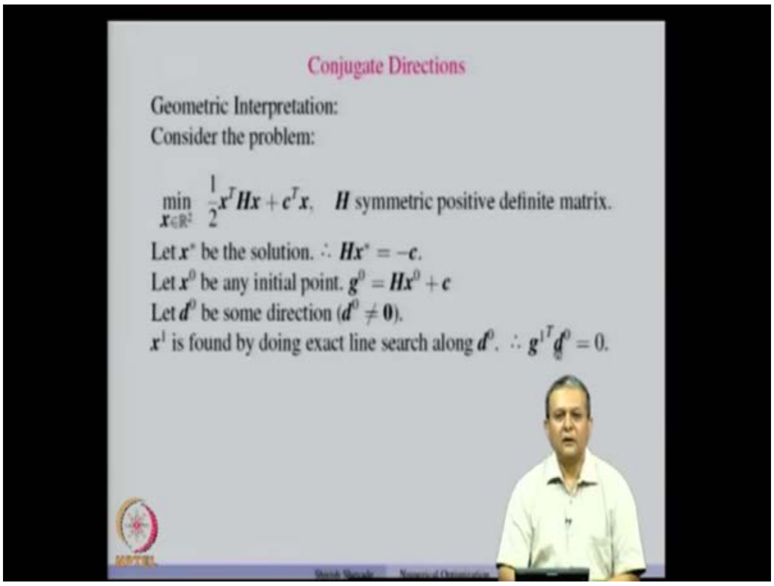


Now, let us look at some other properties of conjugate directions. So, suppose we have H to be a symmetric and positive definite matrix and suppose d 0 to d n minus 1 are non-zero directions such that, d i transpose H d j is equal to 0 for all i naught equal to j. Then the question is that are this directions linearly independent? Now, to show that the they are linearly independent what we have do is that, if you take a linear combination of this vectors and equate to 0. Then all the coefficients in that linier combination should also be 0 or in other words if you say sigma mu i d i equals to 0 for i going from 0 to n minus 1, then we have then if they are linearly independent then mu i has to be 0. So, let us consider this sigma mu i d i to be equals to 0. Now, that implies that if you multiply throughout by d j transpose H what we get is sigma mu i d j transpose H d i equal to 0, for every j going from 0 to n minus 1. Now, we know that d i transpose d j for all i naught equal to j.

So, all the terms in this expression vanish except the j-th term. So, what remains here is that mu j in to d j transpose H d j equals to a 0 because of the H orthogonality of the said d 0 to d i minus 1. Now, H is a positive definite matrix, d j is non-zero. So, d j transpose H d j is greater than 0 and therefore, mu j has to be 0, therefore mu j becomes 0 for a every j. So, which means that sigma mu i d i equal to 0 means that mu i equals to 0 for all i and that means that this set of vectors are linearly independent. So, if we have positive definite matrix and d 0 to d n minus 1 are non-zero directions remember that this non-zero is important because in this case, we want to show that d i transpose H d j is strictly greater than 0.

And that is possible because H is a positive definite matrix and d i is non-zero. So, d i transpose H d j strictly greater than 0 and therefore, mu j has to be 0. So, if these are non-zero directions which are h orthogonal to each other then d 0 to d n minus 1. The n are linearly independent vector in the space r n. So, they form a basis for r. Now, this is what we show that d 0 to d n minus 1 are linearly independent.

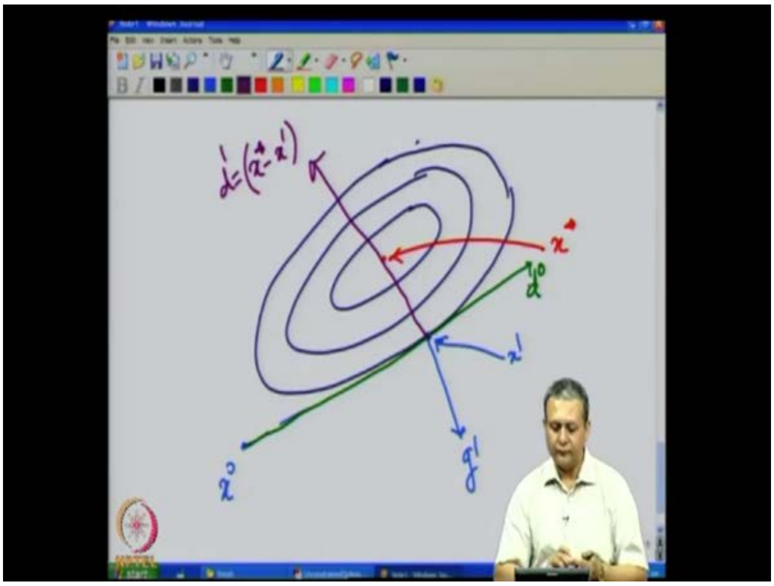
(Refer Slide Time: 37:53)



Now, let us look at the geometric interpretation of conjugate directions. Now, again let us consider the general problem in 2-dimensional space remember that, we are considering a 2-dimensional problem. So, the matrix H which is the symmetric positive definite matrix is a two-second matrix and we want to minimize half x transpose x plus c transpose x. Now, let us assume that x star be the solution of this problem therefore, we can say that H x star plus c is equal to 0, because we have to take gradient of this quantity and equality to 0. So, gradient is H x plus c and when we equate it to 0 that is satisfied at the solution so H x star equal to minus c.

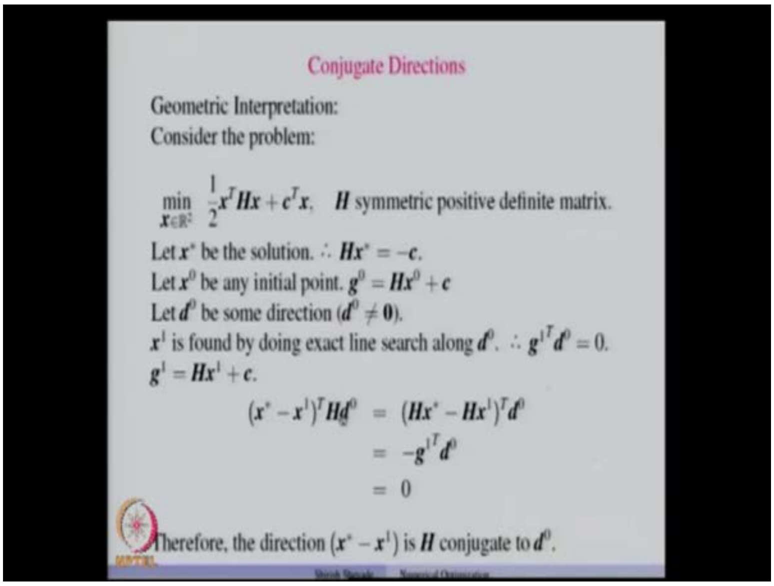
Now, let x 0 be any initial point and the gradient at that point is nothing but h x naught plus c and let d 0 be some direction d 0. So, the idea is that we start from a point x 0, use the direction d 0 and do the exact line search to go to the point x 1 and then at x 1. We choose a direction d 1 which is H conjugate to d 0 that is the idea and we want to see whether that direction is indeed a H conjugate direction or not. So, x 1 is found by doing exact line search along the d 0 and we know that in such a case g 1 transpose d 0 is 0. The gradient at the point of x 1 is orthogonal to the direction d 0.

(Refer Slide Time: 40:05)



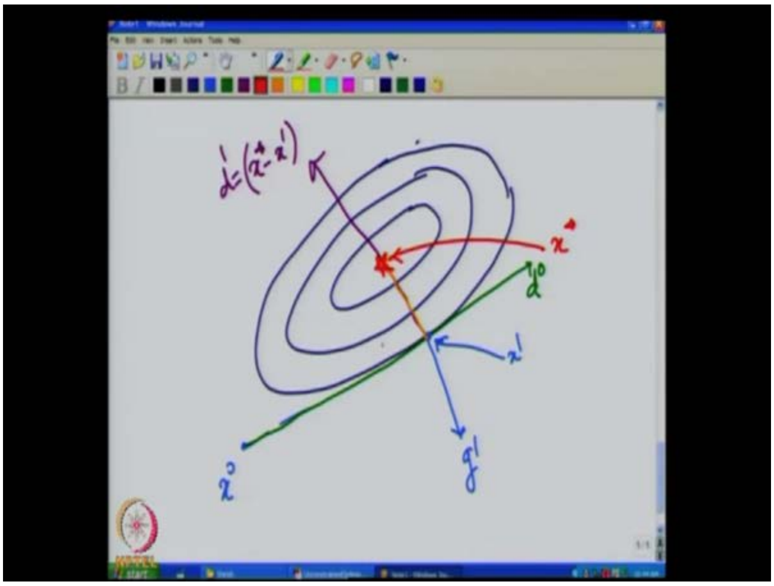
So, suppose these are the contours of quadratic function and this point is our x star. So, suppose that this is x naught and this direction is d 1. So, since we are doing exact line search. So, what we get is the point x 1 be this part and this g 1 the gradient at x 1 is it should be d 0. So, you will see that g 1 is orthogonal to d 0. Now the… Now, let us look at this direction, the direction which will get from x 1 to x star. So, let us look at this direction. So, this is the direction x star minus x 1 and this is going to be our d 1. Now, we want to see whether x star minus x 1 is orthogonal to d 0 or d 1 is orthogonal to d 0.

(Refer Slide Time: 42:32)



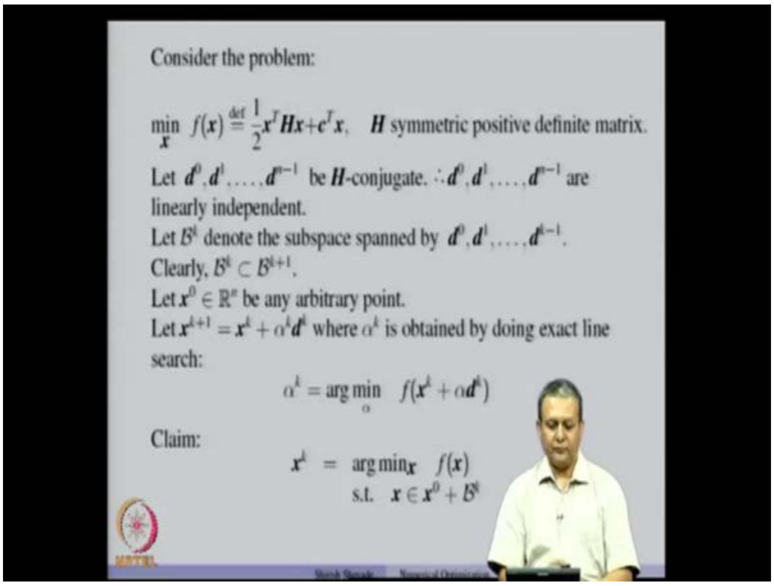
Now, remember that g 1 is nothing but the gradient of this at x 1 and that is nothing but h x 1 plus c. So, x star minus x 1 transpose H d 0 that is what we are interested in finding out. What is this direction? We know this direction x star minus x 1, we had earlier chosen d 0. So, what is this expression x star minus x 1 transpose d 0. Now, if you look at this direction so x star is nothing but minus H inverse c or we can bring in H inside and write this as H x star minus H x 1 transpose d 0. And this H x star is nothing but minus c. So, what we have here in this expression in the parenthesis is minus c minus H x 1 and minus c minus H x 1 is nothing but minus g 1. So, what we have is minus g 1 transpose d 0 and we know that x 1 was found by doing exact line search along d 0 therefore, g 1 transpose d 0 is equal to 0 and therefore, this quantity is 0 and therefore, x star minus x 1 is H conjugate to d 0. So, in other words…

(Refer Slide Time: 44:02)



So, this direction x star minus x 1 is indeed H conjugate to the direction d 0. So, this direction is… So, this d 1 is chosen which is H orthogonal to d 0, and if you do know the exact line search along this direction d 1. So, what we get is that we move from this point to this point and end up at the minimum which is x star. So, at the most 2 steps where needed here to reach the solution. So, this is the geometrical interpretation of conjugate directions.

(Refer Slide Time: 44:57)



Now, let us again consider the same problem, quadratic programs problem where the hessian matrix is symmetric and positive definite matrix. And let us assume that d 0 to d n minus 1 are H conjugate. So, suppose we have already got this H conjugate directions. Now, how does this method work when we use this H conjugate directions. Remember that d 0 to d n minus 1 are H conjugate and therefore, linearly independent and we have shown that. So, this span the entire n-dimensional space.

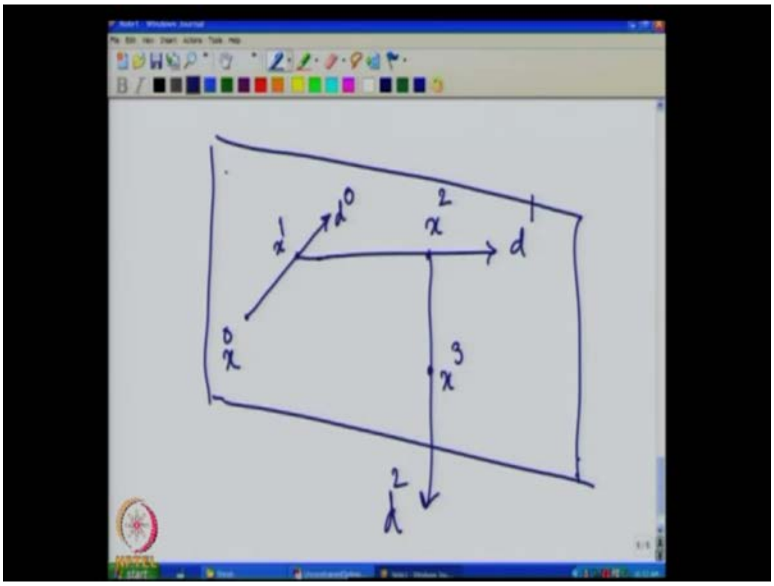
So, is it possible to do the minimization with respect to each of the directions 1 at a time? So, we start from x 0 and move along the direction d 0 go to x 1, and then move along d 1 go to x 2 and so on. So, is it possible to get a minimum at the end of searching through d n minus 1 and that is possible. And this is what we are going to see now. So, we have already shown that these directions which are H conjugate means that they are linearly independent.

So, let us denote by script B k, the sub space of r n spanned by d 0 to d k minus 1. So, these are k vectors and they span k-dimensional space of real numbers and which is the sub space of r n. Now, clearly the B k space is the sub set of B k plus 1, because B k plus 1 will have a extra vector d k which is independent of this. So, the dimension of this sub space is 1 more than the dimension of this sub space and this is the sub set of the said B k plus 1.

Now, let us choose any point x 0 in n-dimensional space, any arbitrary point. And as usual we use this formula, where x k plus 1 is said to x k plus alpha k d k. Where alpha k is obtain by doing exact line search and I repeat that this alpha is over all possible set of real numbers. So, one has to search over all set of real numbers alpha which minimizes f of x k plus alpha d k. And that minimum value of alpha or that value of alpha which gives minimum of f let us generated by alpha k and if you use that alpha k here and find x k plus alpha d k we get the x k plus 1. Now, the claim is that, the x k that we obtain by minimizing f x over the space which contains x 0 and spanned by d 0 to d k minus 1. So, x k is the minimum over the subspace.

So, if we want to extend that result what it means is that, x n will be the minimum of f x where x belongs to x 0 plus b n and b n is nothing but the space spanned by d 0 to d n minus 1 and that is the entire space because d 0 to d n minus 1 are linearly independent set of x so that, span the entire r n. So, what this result means is that if you search along the directions d 0 to d n minus 1 which are H conjugate. And every time if you do the exact line search then after doing at the most exact end line searches, assuming that we have reach the solution in that n line searches at the end of n the n-th line search we will have reach the solution. So, let us see the interpretation of this result.

(Refer Slide Time: 49:51)



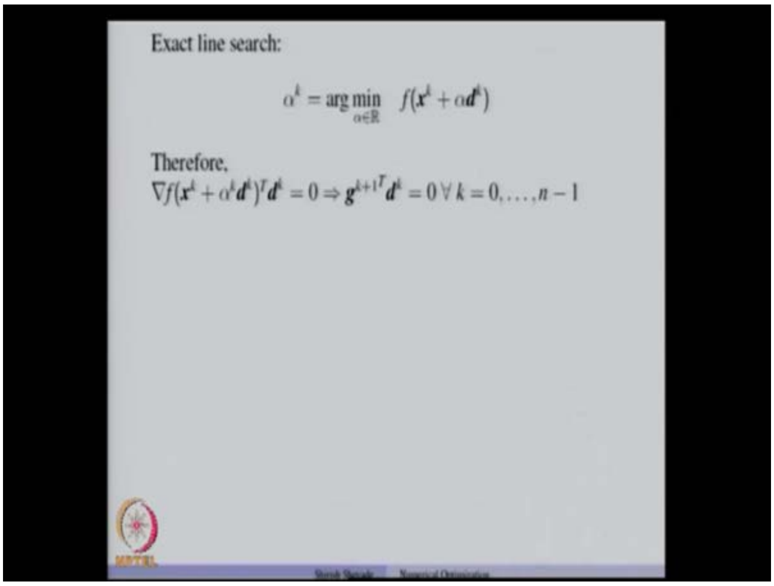
So, suppose this is the point x 0 and this is the direction d 0. Now, we do the exact line search along this direction to get the point x 1 and then we so what it essentially means that. So, this d 0 is non non-zero vector. So, it spans a 1 dimensional space. So, in this 1-dimensional space x 1 is the minimum. Now, we take another vector which is say d 1. Now, d 1 and d 0 are independent.

So, they form a 2-dimensional space. Now, what the claim is that when we find x 2 by doing exact line search along the direction d 1. We have minimized the function f in the 2-dimensional space spanned by d 0 and d 1 so that, space so let us denoted by this and then we go to the vector d 2 so that, d 2 could be a vector which is independent of d 1 and d 0. But, remember that d 2 will know be H conjugate to both d 1 and d 0. So, this will be a vector d 2 and now, that d 0 d 1 and d 2 being linearly independent this spanner 3-dimensional space. So, again the claim is that will when we move from x 2 to x 3. To get the new point x 3, we will have minimize the function f x over the space spanned by d 0 d 1 and d 2 except the 3-dimensional space.

So, every time we add the H conjugate vector to the said d 0, d 1, d 2 and so on. And we have to minimize the function in the respective k-dimensional space. And if you repeat these procedure n times and if at any time of the gradient does not up the function at end of vanish that means at the end of n-th iteration, when we reach x n we will have found the solution of the given problem. So, this property is called expanding sub space property.

And the important point that one has to remember is that, whenever we get x k we have solve the problem completely with respect to the k-dimensional space spanned by d 0 to d k minus 1 and that space which contains x 0. Note also that, we have chosen x 0 to any arbitrary point. So, the initial point is not going to matter when we do this and if the process is repeated n times then the claim is that, at the end of n-th iteration or n line search we will have got x star the minimum of this objective function.

(Refer Slide Time: 53:53)



Now, let us prove this result. So, suppose that alpha k is obtain as minimum of f of x plus alpha d k where alpha belongs to r. Now, since it is obtain using in exact line search. So, if we define phi alpha to be f of x k plus alpha d k, then phi alpha equal to 0 implies that gradient of f of x k plus alpha k d k transpose to d k equal to 0.

And this is nothing but x k plus alpha k d k is nothing but our new point x k plus 1. So, g k plus 1 transpose d k is equal to 0 for all k going from 0 to n minus 1. Now, the important point that you have to remember is that, so the gradient at the new point which is g k plus 1 is orthogonal to the previous direction d k and this is true for all k going from 0 to n minus 1. Now, in addition to this what happens is that, g k plus 1 transpose d j equal to 0 for all j going from 0 to k minus 1. So, not only that g k plus 1 is orthogonal to d k but g k plus 1 is orthogonal to all the previous directions d 0 to d k which are the k plus 1 directions.

So, the gradient at a new point is always orthogonal to the sub space spanned by d 0 to d k. So, this is what we are going to show. And then we show that when we minimize f of x over the sub space we will get x k that is obtained by minimizing f of x over this sub space. So, if we do this we indeed get x k which minimizes f x over x belongs to x 0 plus d k. So, we will see that proof in the next class.

Thank you.