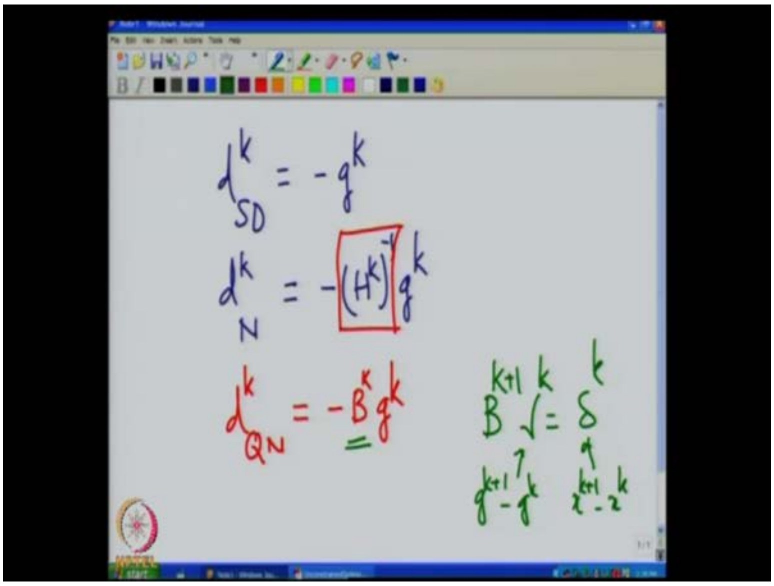
Numerical Optimization
Prof. Shirish K. Shevade
Department of Science and Automation
Indian Institute of Science, Bangalore

Lecture - 16
Quasi-Newton Methods - Rank One Correction, DFP Method

Hello, welcome back to this series of lectures on numerical optimization. In the last class we started looking at Quasi-Newton methods. So, the idea behind Quasi-Newton methods is very simple.

(Refer Slide Time: 00:37)



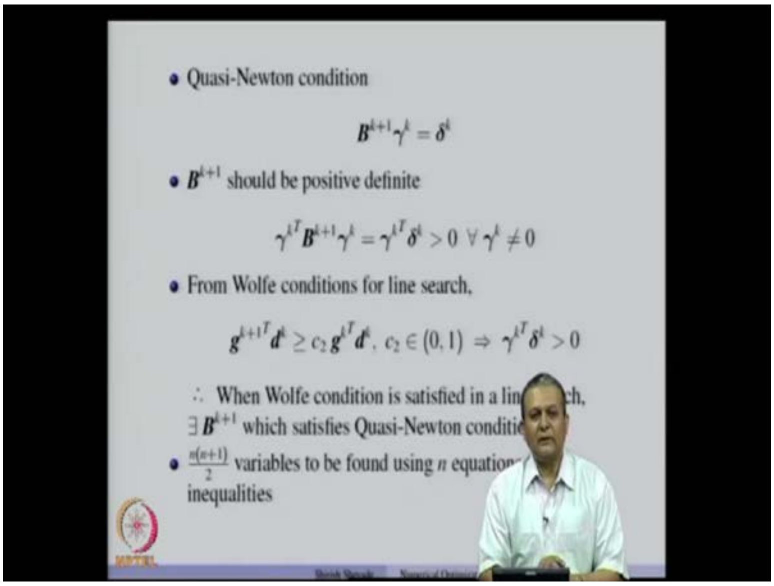
So, we know that in the visual steepest descent method the direction at the iteration k is given as minus g k, where g k denotes the gradient at the current iteration. Then we also saw that in the Newton iteration, the direction is minus H k inverse g k, where H k is the Hessian matrix at x k and thus the Newton method requires the first order as well as the second order information, and you look at this computation which needs to be done for Newton method. We will see that type requires inverse of a computation of inverse of a Hessian matrix at every iteration and that computation is computation is very expensive, it requires order n cube effort, if n is the size of the Hessian matrix or the Hessian matrix is of the size n by n Hessian matrix is always a square matrix.

So, to overcome this problem it was propose to use method which we are going to call Quasi-Newton method, and the idea of Quasi-Newton method is that to approximate the

Hessian inverse by some matrix which is positive definite matrix. So, if we can get a good approximation of the Hessian inverse at a given iteration and represent it by a matrix B k, and if B k is symmetric and positive-definite, then we definitely will get a descent direction. And we also saw that this B k should satisfy certain condition and that is called a Quasi-Newton condition and that says that B k plus 1, gamma k is equal to delta k, where this gamma k is g k plus 1 minus g k and delta k is x k plus 1 minus x k.

So, if the matrix B satisfies this condition which is called Quasi-Newton condition and if it is symmetric and positive-definite then we can use the Quasi-Newton method. Now let us see how to get this B k plus 1 from B k. So, B k plus 1 needs to be obtained from B k, x k, x k plus 1 g k and g k plus 1 or in other words gamma k and delta k and today we are going to see some methods which let you define a symmetric positive-definite B k plus 1 using B k and the other first order information at x k.

(Refer Slide Time: 03:56)



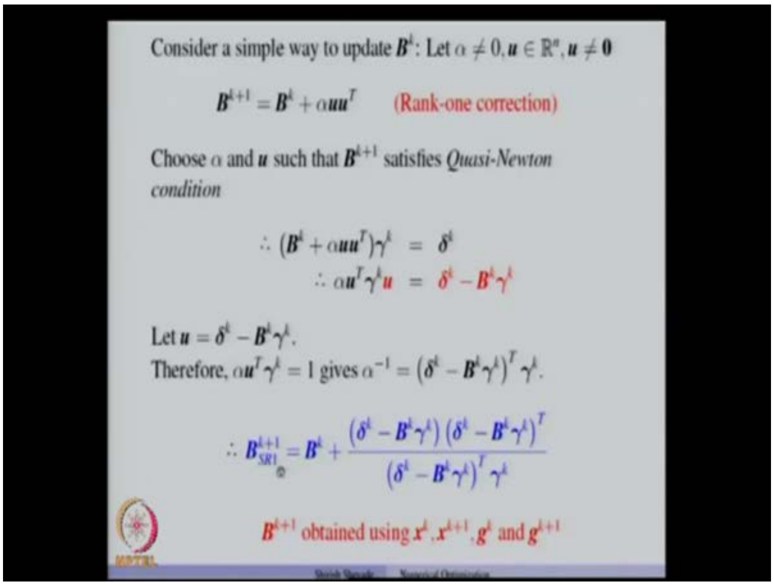
So, we saw that this is the Quasi-Newton condition that we need to satisfy and we also want B k plus 1 to be positive-definite. Now, if we take a scale which is represented as gamma k transpose B k plus 1 gamma k where gamma k is non-zero and if Quasi-Newton condition is satisfied then we can write this as gamma k transpose delta k and we have to ensure that this condition is satisfied. So, that this matrix becomes positive-definite.

Now how do we ensure that, this condition is satisfied we have already seen that if Wolfe condition is satisfied in line search. Then we have this expression g k plus 1 transpose d k greater than or equal to c 2, g k transpose g k where c 2 is positive fraction in fact to this picking Armijo condition is also satisfied, where and if c 1 is a constant corresponding to Armijo condition then c 2 is in the open interval c 1 to 1.

So, here I have just indicated it to be in the open interval 0 to 1 because that is what we need for this purpose. Now if we use this condition then we can write that gamma k transpose delta k is greater than 0, because we just have to subtract g k transpose d k from both the quantities or both the sides and that will result in m r k transpose delta k greater than 0. So, if Wolfe conditions are satisfied in a line search then this is satisfied and therefore, we have B k plus 1 to be positive-definite.

But, are there what the ways to get B k are plus 1 and this is the question we started answering in the last class. So, note that, this B k plus 1 is symmetric matrix. So, it has n into n plus 1 by 2 variable and these are the n equalities that 1 has to satisfy and more over B k plus 1 should be positive-definite and therefore, all principle minors should be positive that results in n inequalities. So, you have n equations and n inequalities corresponding to positive place of n principle minus and we have n into n plus 1 by 2 variables. So, we will see that there are lot many variables than the numbers of equations and inequalities. So, many solutions exist.

(Refer Slide Time: 06:39)



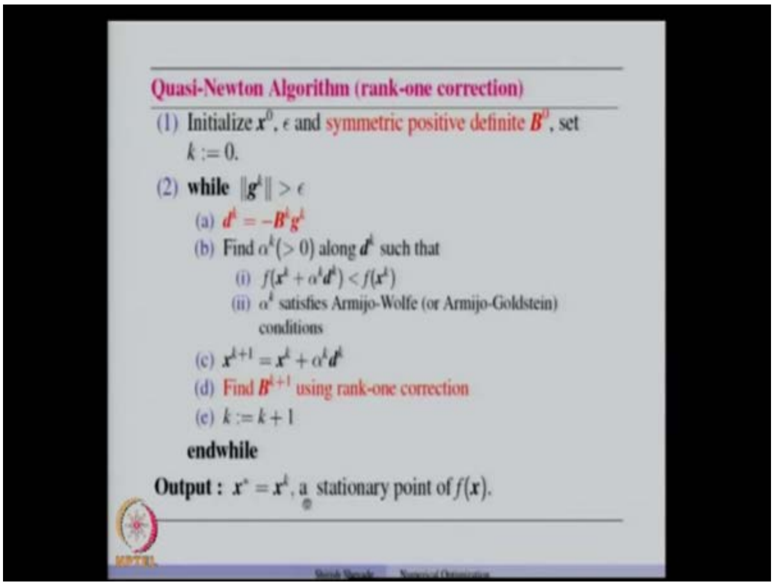
So, let us look at the simple way of updating B k. So, let us choose some alpha which is non-zero let u be the n-dimensional vector which is non-zero and suppose we simply add in the matrix alpha u u transpose to the matrix B k to get the B k plus 1. Now note that, u transpose is a rank-one matrix, and therefore this is called the rank-one correction of B k to get B k plus 1, u u transpose in addition to be a rank-one is also symmetric matrix therefore, if B k is symmetric then we have a symmetricity in B k plus 1 more over if alpha is suppose greater than 0 and B k is positive-definite then certainly this matrix also will be positive-definite.

So, what we need is alpha to be greater than 0 and u to be non-zero, that will and if B k is symmetric and positive-definite then that will guarantee B k plus 1 will be symmetric and positive-definite now. So, this is about the correction but we also want B k plus 1 to satisfy the Quasi-Newton condition and that condition if we use B k plus 1 into gamma k is equal to delta k. So, substitute this quantity here in that equation. So, B k plus 1 gamma k delta k will be written in this form and the variables here now are alpha and u because B k is known gamma k is known and delta k is known. So, how do we get alpha and u. So, suppose we rearrange the terms in this equation and we write this as alpha u transpose gamma k into u is equal to delta k minus B k gamma k.

Now u is a vector delta k minus B k gamma k is a vector suppose we assume that u is nothing but delta k minus B k gamma k. So, that makes this quantity the multiple the scalar multiple of this scalar multiple of u to be 1. So, which means that alpha u transpose gamma k becomes 1 if we choose u to be delta k minus B k gamma k. So, if we do that then what we get is 1 over alpha is this quantity delta k minus B k, gamma k transpose gamma k and if you plug in this value of alpha in this equation and this value of u in this equation then what we get is basically called the symmetric rank-one update for B k. So, the subscript SR1 here indicates it is a symmetric rank-one update. So, we have B k then alpha is replace by 1 over delta k minus B k gamma k transpose gamma k such basically the denominator here and we have u u transpose which is nothing but delta k minus B k gamma k into delta k minus B k gamma k transpose. So, a very simple way of updating the B k to B k plus 1 and if alpha is greater than 0 and u is non-zero.

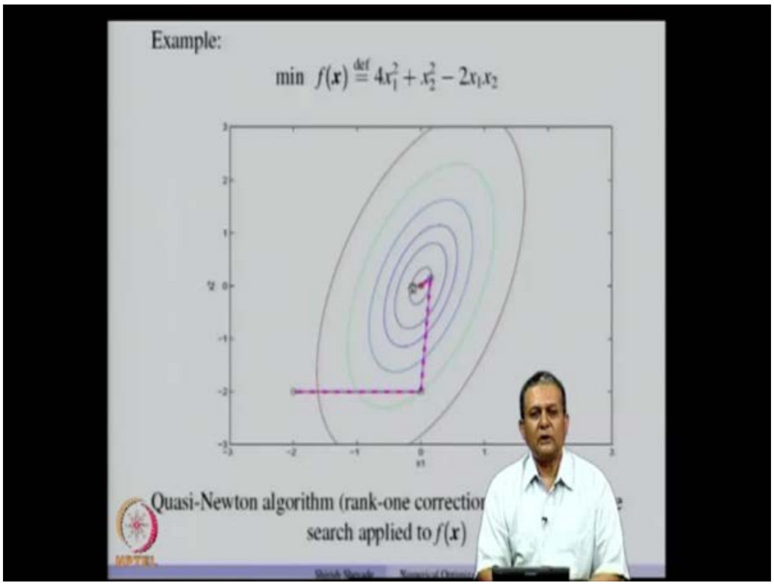
Then we will see that the new matrix that we get B k plus 1 is also symmetric and positive-definite provided B k is symmetric and positive-definite. So, we could start with a symmetric positive-definite matrix b 0 and use this formula to get B 1, B 2, B 3 and so on. So, one good thing about this update is that B k plus 1 is obtained using B k delta k and gamma k or in other words B k, x k, x k plus 1, g k and g k plus 1. So, the information at the current iteration and the previous iteration is used to update B k to B k plus 1. So, this is a very important point to that one needs to remember.

(Refer Slide Time: 11:05)



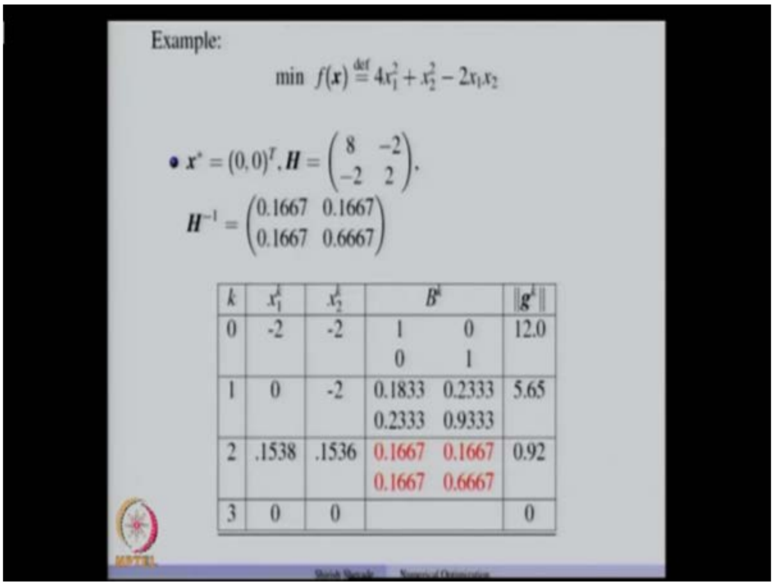
Now, let us look at the Algorithm. Now, compared to other algorithms that we have studied so far this Quasi-Newton algorithm needs a symmetric positive-definite matrix B 0 to start with. Now, one can also use identity matrix as B 0. Now the stopping condition remains the same. So, the direction d k that we get is minus B k into g k then the we do the line search such that the Armijo-Wolfe or Armijo-Goldstein condition are satisfied and then we move to the new point x k plus 1 using this method and at x k plus 1. We calculate g k plus 1 now we have information of x k, g k then x k plus 1 and g k plus 1 with then use all this information along with the knowledge of B k to update B k to B k plus 1. So, we use rank-one correction to find B k plus 1 the formula for this we have already seen and then the iteration counter is increased and the whole procedure is repeated till some stopping condition is satisfied and finally, we get a stationary point of f x.

(Refer Slide Time: 12:38)



Now, let us see some examples to see how this method works. So, we will consider a simple quadratic function which is define here as f x equal to 4 x square plus x 2 square minus 2 x 1, x 2. Now, we first look at the contours of this function we have already seen this function when we used Newton method to minimise this function or steepest descent method to minimize this function. So, we know that the minimum of this function exists at this point. Now on the figure you will see that the Quasi-Newton method apply to this function tresses this path which is shown by the merge inter line. So, suppose we start with the point minus 2 minus 2 where works the x and y coordinates such or a both x 1 and x 2 are minus 2 then the first step of Quasi-Newton algorithm with in exact line search. I am not using exact line search here. So, with in exact line search we come to this point and then from this point we go to this point and then in the third iteration we reach the solution. So, this is the initial point, this is the x 1 this is x 2 and x 3. So, at the end of the third iteration we have reached the solution in this case.

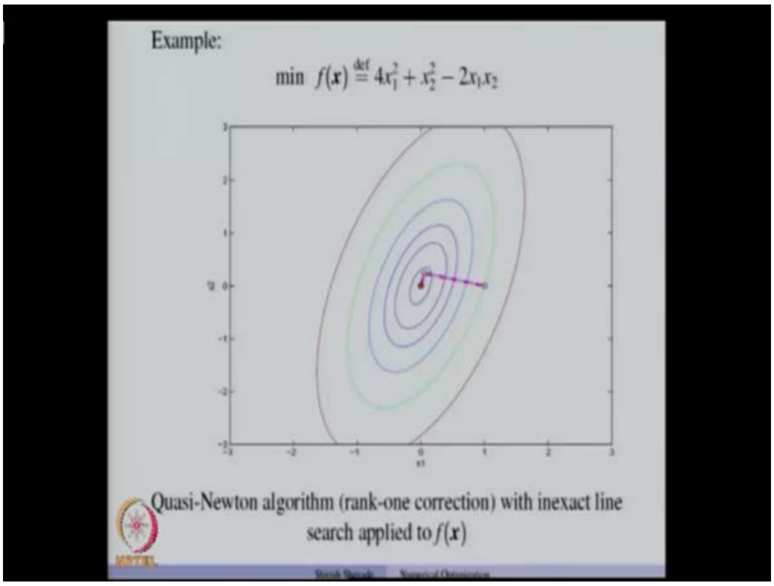
(Refer Slide Time: 14:12)



Now, let us see some more details about this experiment. So, we know that the minimum of this function occurs at the origin the function is quadratic. So, the Hessian matrix is constant irrespective of the value of x and you can also see that the Hessian matrix is positive-definite. Now, I have also given here the inverse of the Hessian matrix which is shown here. Now, let us look at the iterates obtained using the Quasi-Newton algorithm with symmetric rank-one correction.

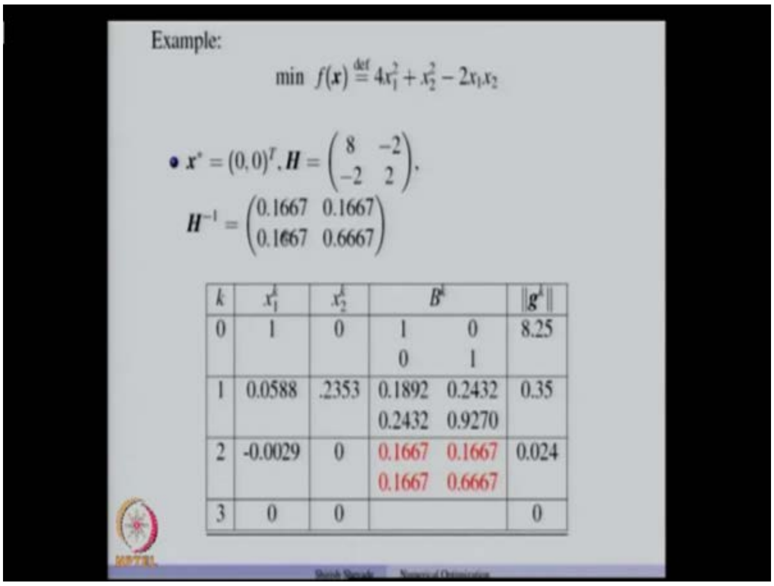
So, suppose we start with minus 0.2, minus 2 and we initialise B k 2 identity matrix and the norm of the gradient is 12.0 at this x k now the first iteration of Quasi-Newton algorithm with symmetric back one update would take us to the 0.0 minus 2 if the corresponding B k matrix updated like this and the norm of the gradient is 5.65. So, at the end of the second iteration the algorithm moves to the 0.15, 0.15 and we get the B k matrix as same as the Hessian inverse and corresponding norm of the gradient is 0.92 and in the third iteration at the end of the third iteration we have reach the point to 0 where the norm of the gradient is 0. So, you will see that. We started with a identity matrix B 0, and in the in the course of the algorithm we reached an iteration, where the matrix B k turned out to be same as the matrix achievers.

(Refer Slide Time: 16:33)



Now, let us take one more example. So, let us start with a different point. So, this point is 1 0 and this is the path trace by the Quasi-Newton algorithm with rank-one correction and using in exact line search. Now, here also the algorithm needs three iterations the third iteration is difficult to see here in this figure because the second and the third iteration points are very close to each other.

(Refer Slide Time: 17:07)

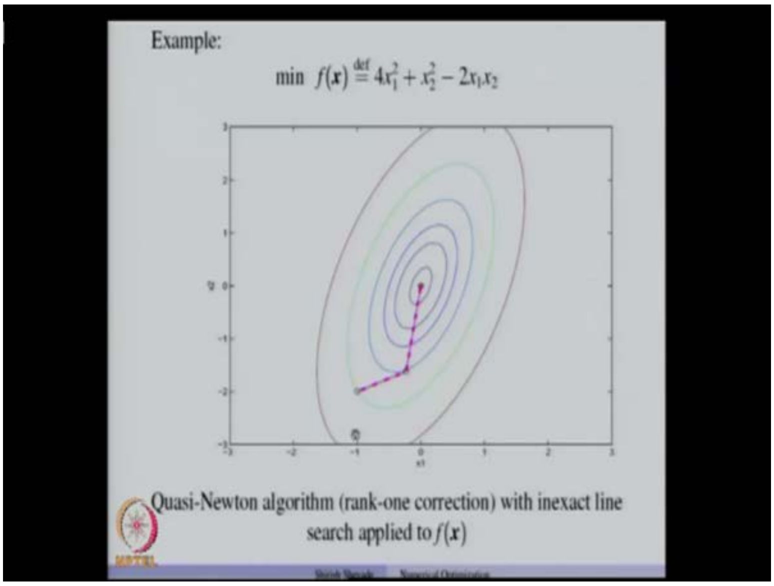


k	x_1^k	x_2^k	B^k		$\lVert g^k \rVert$
0	1	0	1 0	0 1	8.25
1	0.0588	.2353	0.1892 0.2432	0.2432 0.9270	0.35
2	-0.0029	0	0.1667 0.1667	0.1667 0.6667	0.024
3	0	0			0

So, let us again analyze the path trace by the algorithm to minimize this using symmetric rank-one update. So, we start with a 0.10 and the initial matrix B 0 as a identity matrix the norm of the gradient is 8.25. We then move to 0.05, 0.23 and this is going to be the matrix B k and then we move to the point minus 0.0029 and 0 and the corresponding B k at the iteration 2 is same as the H inverse corresponding norm of g k is 0.0 to 4.

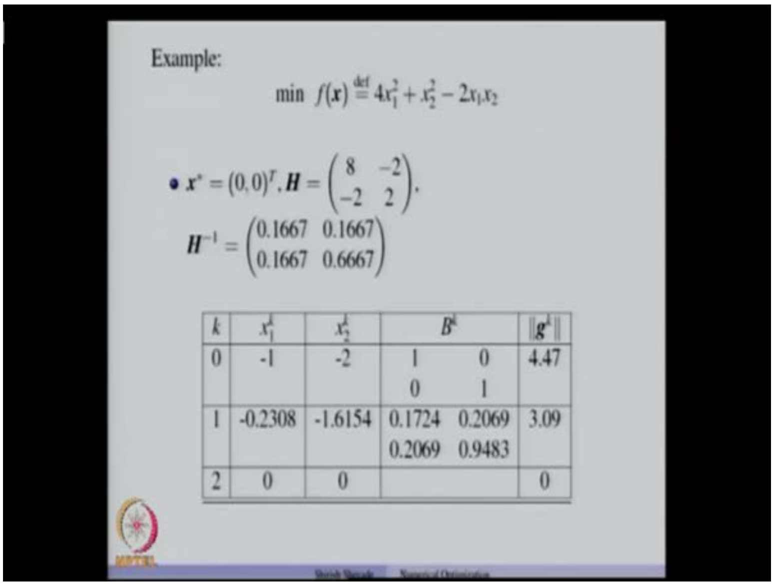
Now, I have used 0.001 as the stopping criteria epsilon in the algorithm. So, that means that norm g k should be less than 0.001. So, this is still higher. So, the algorithm will further make a progress and you will see that in the at the end of the third iteration we have reached the optimal point which is x star and the corresponding norm of the gradient is 0 here again you will see that we have got the matrix B k at 1 stage which is same as the Hessian inverse.

(Refer Slide Time: 18:37)



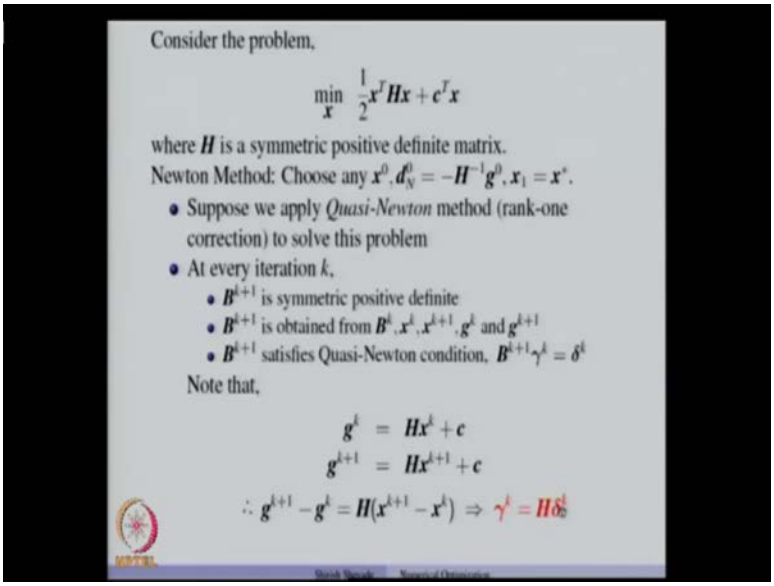
Now, let us take one more example. Now, let us start with some different initial point which is whose x 1 coordinate is minus 1 and x 2 coordinate is minus 2 and you will see that this algorithm the Quasi-Newton algorithm in this case converges to x star in two steps this is unlike the previous two examples, where the Quasi-Newton algorithm needed three iterations or three steps to reach the solution. In any case we have seen that the algorithm required not more than three steps in these two-dimensional cases of a quadratic function.

(Refer Slide Time: 19:30)



So, if we look at the last example. So, we started with the - 0.1, -0.2 with the same initial matrix B k to be identity matrix and at the end of first iteration the B k turned out to be like this and at the end of the second iteration we go to the solution, where the norm of the g k is 0. Now, let us try to analyze about what is going in this case, when we applied Quasi-Newton algorithm with symmetric rank-one update to quadratic function.

(Refer Slide Time: 20:14)

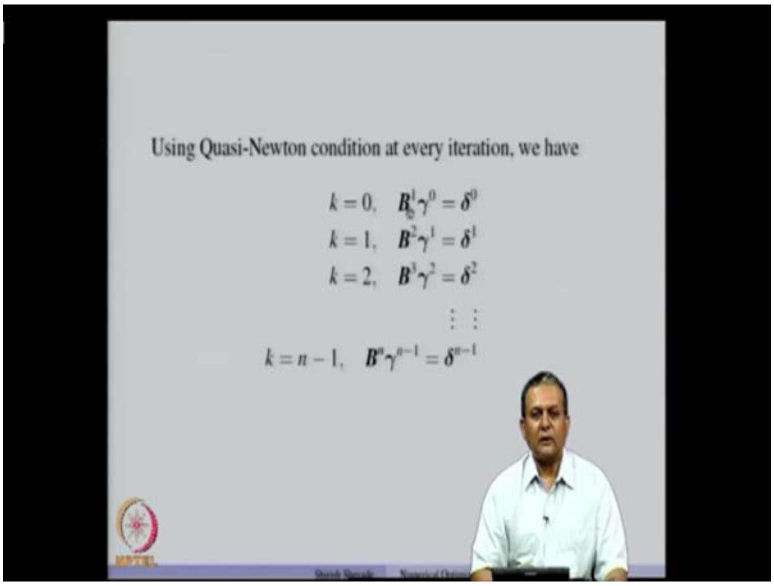


So, let us consider problem where we want to minimize half x transpose H x plus e transpose x at general quadratic function, where H is a symmetric positive-definite matrix and suppose we want to use Quasi-Newton method but before we do that let us look a Newton method. Now as we saw in the case of Newton method that if we choose any x naught then the Newton direction is minus H inverse g naught. Where g naught is the gradient of the function at given point x naught and since this is a quadratic function the Hessian is always the fixed Hessian and does not depend on k.

So, does not depend on iteration number here. So, the Newton direction is minus H inverse g naught and if we use exact line search. Then what happens is that in the next iteration we will get x inverse. So, starting from any point for a convex quadratic function the Newton method gives the solution in one step we have seen this result early. So, we get x 1 equal to x star now let us see what happens when we apply Quasi-Newton method with a symmetric rank-one collection a correction to solve this problem. Now, at every iteration let us assume that we have some mechanism to get B k plus 1 from B k and B k plus 1 is symmetric and positive-definite at every iteration. So, suppose that is ensured and B k plus 1 is obtained from B k, x k, x k plus 1 g k and g k plus 1. Further we also assume that B k plus 1 satisfies the Quasi-Newton condition which is B k plus 1 gamma k is equals delta k. So, all these conditions are satisfied at every iteration k.

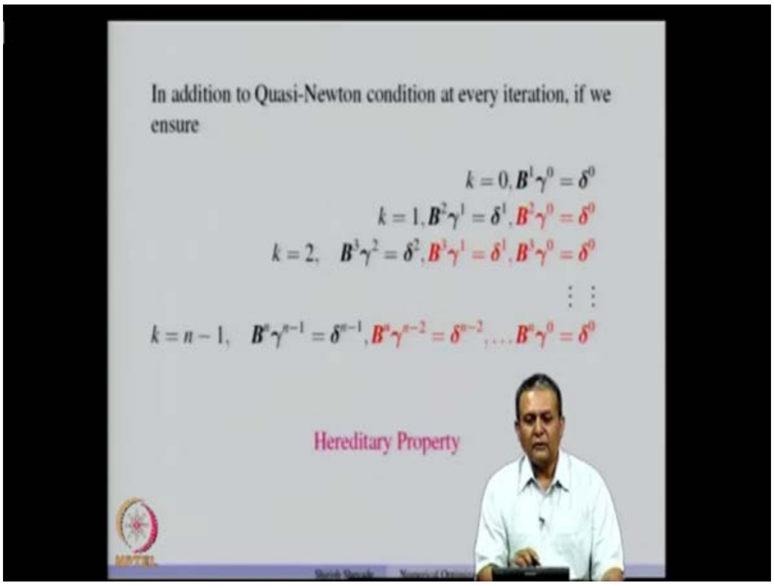
Now if we take the gradient of this function at the iteration k that will be H x, k plus c and we are going to denote it by g k similarly, at the iteration x k plus 1 we have g k plus 1 to be H x, k plus 1 plus c. Now if we subtract the first equation from the second equation what we get is g k plus 1 minus g k to be H into x k plus 1 minus x k and by our definition g k plus 1 minus g k is nothing but gamma k and x k plus 1 minus x k is nothing but delta k. So, what we have is gamma k to be equal to H into delta k. So, there is a interesting relationship that we have got. So, Quasi-Newton condition says that B k plus 1 gamma k is equal to delta k and gamma k is equal to H delta k.

(Refer Slide Time: 23:22)



Now, if we use Quasi-Newton condition at every iterations then at 0-th iteration we need B 1 that satisfies B 1 gamma 0 is equal to delta 0 then at iteration 1, B 2 should satisfy B 2 gamma 1 equal to delta 1 and so on and finally at the n minus 1 iteration. We have B n which satisfies B n gamma n minus 1 is equal to delta n minus 1. So, at each of this iteration the matrix B, the new matrix B that we get should satisfy this Quasi-Newton conditions.

(Refer Slide Time: 24:10)

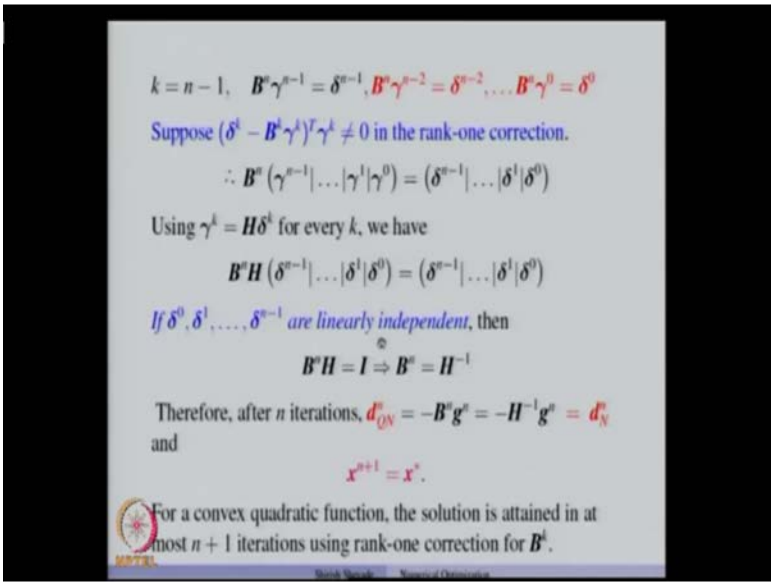


Now, let us look at these conditions in more detail. Suppose, in addition to this Quasi-Newton conditions that we have seen. So, far in addition to those conditions at every iteration, if we also ensure that. So, at k is equal to 0 we just have to satisfy the Quasi-Newton condition but at k equal to 1 in addition to this Quasi-Newton condition B 2 gamma 1 is equal to delta 1 if we also make sure that B 2 gamma 0 is equal to delta 0. So, not only that the matrix B 2 satisfies the Quasi-Newton condition but it also satisfies the extra condition which is B 2 gamma 0 is equal to delta 0. So, for k equal to 2 the Quasi-Newton condition is B 3 gamma 2 is equal to delta 2 and if the addition to that, if we also satisfy B 3 gamma 1 is equal to delta 1 and b 3 gamma 0 is equal to delta 0.

So, in other words at every iteration the B matrix, if it is satisfies B k gamma j is equal to delta j where j goes from 0.2 k minus 1. So, if we ensure that it happens then at n minus one-th iteration at the end of the n minus 1 iterations we have b n gamma n minus 1 is equal to delta n minus 1 and all the other conditions are satisfied up to b n gamma 0 is equal to delta 0.

So, the first quantities which are shown here are the Quasi-Newton conditions and there are suppose some extra things that are also true which are shown here that B n gamma n minus 2 is equal to delta n minus 2 and. So, on up to B n gamma 0 is equal to delta 0. Now if this happens then we say that B satisfies hereditary property. So, not only with respect to the current iteration that B k plus 1 gamma k equal to delta k but we can say that B k plus 1 gamma j is equal to delta j for all j going from 0 to k. So, this property is called hereditary property.

(Refer Slide Time: 26:33)



Now, if this hereditary property holds then and in addition to that if suppose in the symmetric rank-one correction if we make sure that the denominator in the second term. So, this quantity is nothing but 1 over alpha. So, this quantity is not 0 in the rank-one correction if you ensure that. So, that means that every update is define properly without any numerical issues, then what we have we can write this expression compactly as B n into a matrix whose n columns are gamma n minus 1 up to gamma 0 and that is equal to another matrix whose n columns are delta n minus 1 up to delta 0.

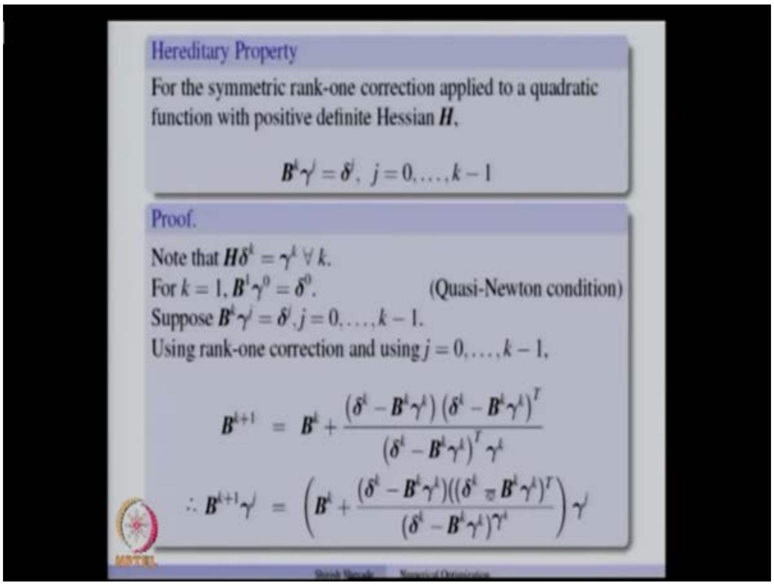
So, this entire system of equations has been written compactly in this form. Now, we know that we have already seen that in this problem where the function is quadratic we can write gamma k to be H delta k for every k and therefore, let us replace each gamma k by H delta k. And therefore, what we have is Y n into H, we can take H common from each of this and what remains in the inside the matrix are the n columns delta n minus 1 to delta 0 and the right hand side remains the same. Now, you will see that this matrix is same as this matrix.

Now, if we further make some assumption that delta 0 to delta n minus 1 are linearly independent. Then this matrix is a full rank matrix and therefore, it is invertible matrix and if you post-multiply throughout by inverse of this matrix. Then what we get is identity matrix on the right side and on the left side. We will get B n into H because this matrix with its inverse will give us identity matrix.

So, therefore, we have B n into H equal to identity matrix and this happens if delta 0 to delta n minus 1. This n vectors are linearly independent and then we have B n into H is equal to identity matrix or in other words B n is nothing but H inverse. So, at the end of n iterations we have the matrix B n to be H inverse, which is nothing but the inverse of the Hessian. So, remember that we did not make any assumption about B 0. So, we start form any B 0 and if we ensure that the hereditary property holds and delta 0 to delta n minus 1 are linearly independent. Then at the end of n iterations we have B n to be H inverse and if at that point if you look at our Quasi-Newton direction. So, Quasi-Newton direction for the n plus 1th iteration is minus B n into g n and this minus B n into g n is nothing but minus H inverse into g n because you have seen that B n to be H inverse.

Now, if you recall these directions is same as the Newton direction. So, after n iterations the directions that. We get using Quasi-Newton method in this case turned out to be same as a Newton direction and then it is as good as a applying Newton method to the given problem and therefore, what we get is that in the next iteration we get x star to be our minimum. So, x n plus 1 will be equal to the actual x star because at the end of n iterations we have got the Newton directions and we know that for a convex quadratic problem from starting from any point the Newton method takes us to the solution exactly in one iterations for a quadratic function. So, we get x 1 plus 1 to be x star. Therefore for convex quadratic function the solution is attained in at most n plus 1 iterations using rank-one correction for B k, now what this requires is that the hereditary property holds and these are linearly independent. So, let us postponed the discussion on the linear independence of the vectors generated let us look at the hereditary property.

(Refer Slide Time: 31:40)

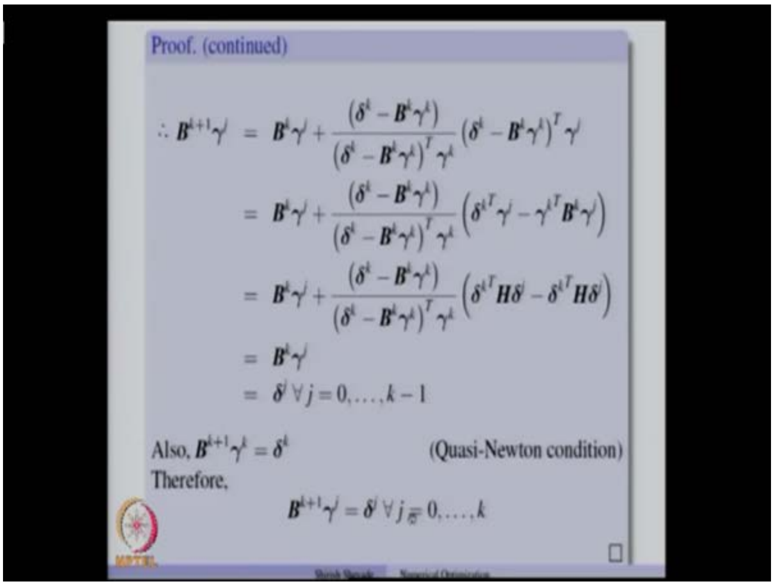


Now, the hereditary property says that for the symmetric rank-one correction, apply to a quadratic function with positive-definite Hessian matrix H B k gamma j is equal to delta j. So, remember that B k, gamma k minus 1 is equal to delta k minus 1 that is true, because of the Quasi-Newton condition. But, the hereditary property says that in addition to the Quasi-Newton condition B k gamma j equal to delta j for all j going from 0 to even all go all j is going from 0 to k minus 2, because for k minus 1 automatically holds because of the Quasi-Newton condition. So, not only for the case of k minus 1 but all j from 0 to k minus 2, this condition holds and this is the hereditary property and let us see how to show that this property holds for a given problem now we have already seen that, H delta k is gamma k for this quadratic problem with the Hessian matrix H.

Now if we take k equal to 1 in this case we have B 1 gamma 0 is equal to delta 0 and that is nothing but Quasi-Newton condition. So, we are going to show this hereditary property by principle of induction and for that purpose we have first indicated that for k equal to 1 this property holds. Now, suppose this property holds for some k which is greater than 1 then we show that it holds for k plus 1 also. So, suppose B k gamma j is equal to delta j for all j going from 0 to k minus 1. So, that means this property holds for sum k and we now show that it holds for k plus 1. So, that means B k plus 1 gamma j is equal to delta j for all j going from 0 to k. So, let us see how to do that.

Now, if you use rank-one correction and if we use j to be 0 to k minus 1, then what we have is B k plus 1 to be B k plus. This is the u u transpose divided by alpha into alpha. So, one over alpha is the quantity in the denominator or alpha is nothing but one over this quantity. So, we have already seen this. Now, we have to show that B k plus 1, gamma j is equal to delta j for all j going from 0 to k minus 1 because B k plus 1, gamma k is equal to delta k is true from the Quasi-Newton condition. So, we have to show that, this is right hand side quantity is nothing but delta j for all j going from 0 to k minus 1. So, if you expand this right side then. Let us see what we get. So, we have to rearrange the terms in the right hand side. So, we take B k gamma j out and then this quantity by the denominator is out and then we have the other quantity.

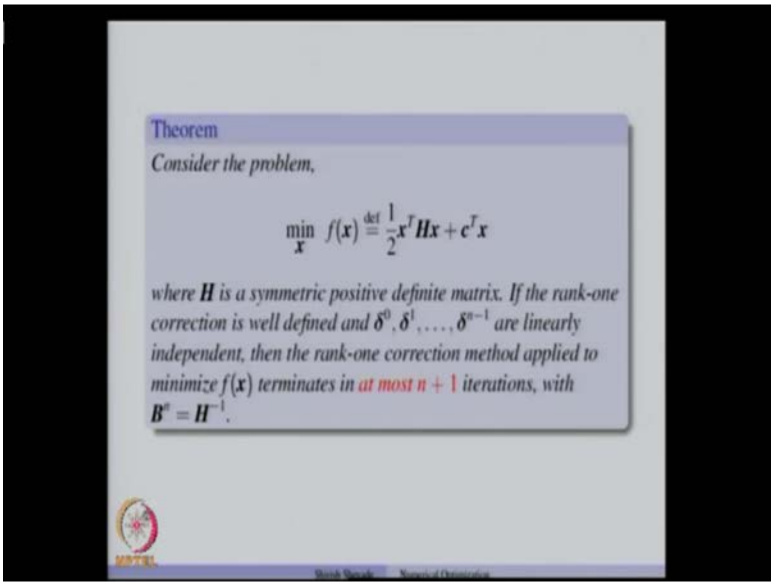
(Refer Slide Time: 35:01)



Now, we have B k plus 1 gamma j to be this quantity. Now, let us look at this quantity. So, we keep this quantity same now expand this the last term. So, this is delta k transpose gamma j minus gamma k transpose B k gamma j at remember that B k is a symmetric matrix. So, B k transpose is same as B k now we use the property that we saw earlier and the property is that H delta k is nothing but gamma k for all k. So, let us use this property here. So, gamma j will be replaced by H delta j. So, therefore, what we get is delta k transpose H delta j and similarly, we have B k gamma j is equal to delta j and gamma k is nothing but H delta k. So, this becomes delta k transpose H delta j.

These two quantities are same and therefore, this second term is equal to 0 and therefore, what we get is B k plus 1, gamma j to be equal to B k gamma j for all j going from 0 to k minus 1 and B k gamma j equal to delta j for all j going from 0 to k minus 1, because we have assume that the hereditary property holds for k. So, this quantity is nothing but delta j therefore, we have shown that B k plus 1 gamma j is equal delta j for all j going from 0 to k minus 1 now the only thing that remains is that what happens when j equal to k but we already know that B k plus 1 satisfies Quasi-Newton condition. So, that means that B k plus 1 gamma k is equal to delta k this is because of the Quasi-Newton condition. So, if we combine these two, then what we get is that B k plus 1, gamma j is equal to delta j for all j going from 0 to k. And that proves that the hereditary property holds for the update symmetric rank-one update.

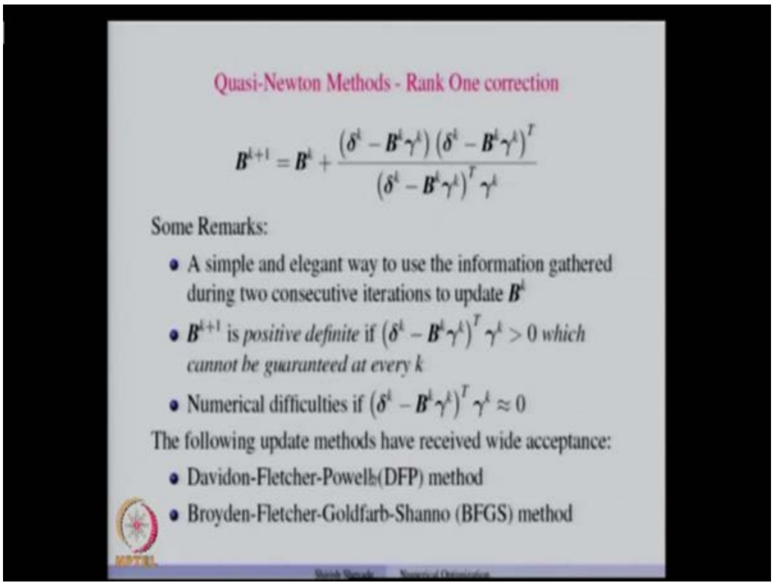
(Refer Slide Time: 37:36)



So, if we consider the problem to minimize f x where H is symmetric positive-definite matrix and if the rank-one correction is well defined and the delta 0 delta 1 up to delta n minus 1 are linearly independent then the rank-one correction method applied to minimize f x terminates in at most n plus 1 iterations with B n to be H inverse and we saw this in those couple of examples, that we considered that the algorithm in those cases in those two-dimensional cases required at the most three iterations and at the end of the second iteration we saw that b n was actually H inverse.

So, this is a good thing about the symmetric rank-one correction that hereditary property holds and if these are ensure to be linear independent then the algorithm requires at the most nth plus 1 iteration and not only that at the end of the n-th iteration if the algorithm does require n plus 1 iterations then at the end of the n-th iteration we have b n to be H inverse. So, this is the very simple way of updating matrix B k to get B k plus 1.

(Refer Slide Time: 38:49)



And this is what we saw earlier. So, let us see some remarks on this. So, this is very simple way to use the information gathered during two consecutive iterations consecutive iterations to update B k. Now, remember that B k plus 1 is positive-definite if B k is positive-definite. So, all throughout the discussion we have assume that the B k positive-definite.

So, let us not worry about that part now how do you make sure that B k plus 1 is positive-definite. So, that will be true only when this quantity is greater than 0 and if you look at our analysis now where we ensured that this quantity is greater than 0. So, this is not always guaranteed at every iteration. So, if that is not guaranteed at every iteration that means that B k plus 1 may not be a positive-definite matrix at every iteration and that can.

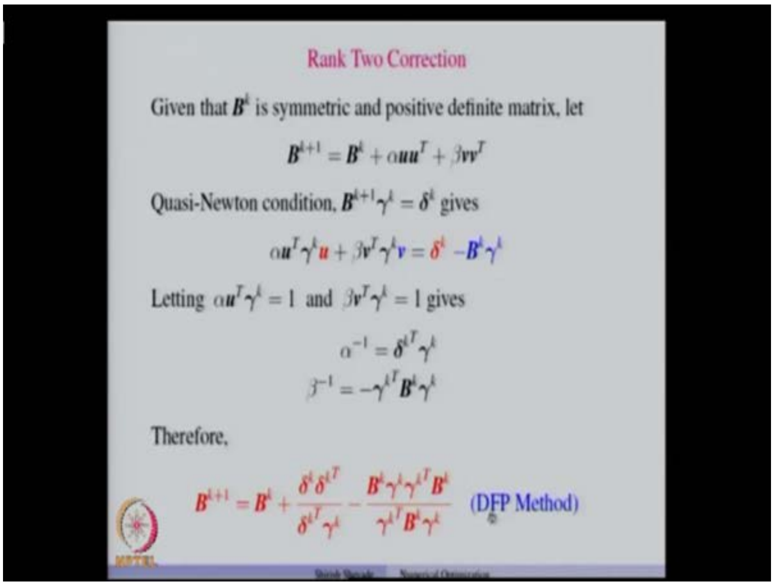
That is not a desired thing that can result into a problem and moreover if the quantity in the denominator becomes close to 0 then the some numerical issues will come up and this matrix can become very large matrix because this quantity is close to zero.

So, all though the symmetric rank-one corrections simple method to get B k plus 1 from a symmetric positive-definite B k but then every time B k plus 1 is not positive-definite or that positive-definiteness of B k plus 1 is not guaranteed and secondly the quantity in the denominator can become close to 0 and that can result in numerical difficulties. So, these are some of the drawbacks of a rank-one correction and we have to look for some alternative method which does not have these drawbacks. So, there were a couple of update methods suggested in the literature and they have received quite acceptance and those methods are Davidson–Fletcher-Powell method in short it is call D F P method and the second method is called Broyden–Fletcher- Goldfarb-Shannon method this called B F Gs method.

So, this D F P method and the B F Gs method they have become quite popular Quasi-Newton methods compare to the symmetric rank-one update all though symmetric rank-one update is a very simple and elegant way and among these two the B F Gs method has become quite popular because it was found that this method works better than the D F P method.

So, Davidson-Fletcher and Powell these are the inventors of this D F P method and Broyden–Fletcher–Goldfarb-Shannon are the inventor inventors of B F Gs method. So, let us first look at the D F P method and then see the connection between the D F P and B F Gs method and the term will see some ways to combine D F P and B F Gs methods.

(Refer Slide Time: 42:22)



So, we have already seen the rank-one correction and we saw that, these are some problems associate with the rank-one correction and therefore, now let us look at the rank-two correction. So, given that B k is symmetric and positive-definite matrix. Let us write B k plus 1 as B k plus alpha u u transpose plus beta v v transpose, where alpha and beta are non-zero scalars and u and v are non-zero vectors.

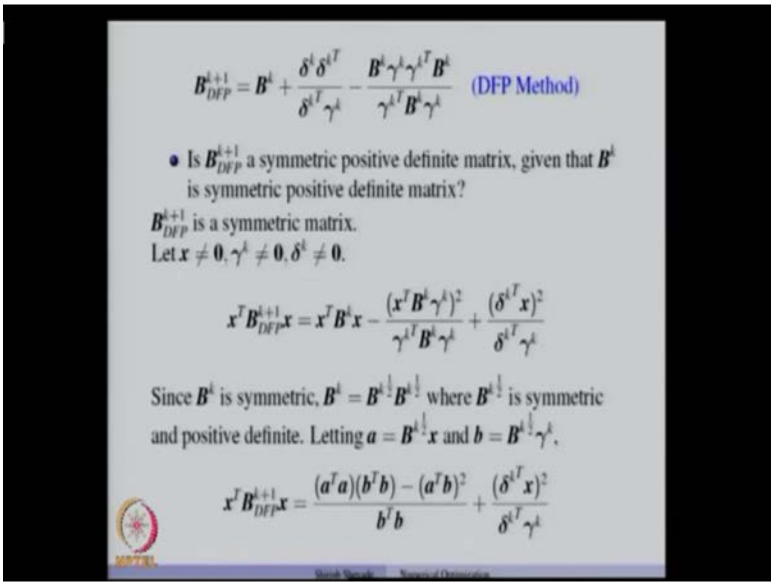
So, we are adding two symmetric rank-one matrices to B k to get B k plus 1. So, if B k plus 1, B k is symmetric this matrix alpha u u transpose is a rank-one symmetric matrix beta v v transpose is a rank-one symmetric matrix i. If alpha and beta are 1 not 0 and u and v are not zero. So, naturally the matrix B k plus 1 is going to be symmetric matrix. Now, the question is that, how do you choose alpha u, beta and v such that, if B k is positive-definite, B k plus 1 is also positive-definite and not only that the matrix B k plus 1 also should satisfy the Quasi-Newton condition.

So, if we apply Quasi-Newton condition to the matrix B k plus 1. So, B k plus 1 gamma k is equal to delta k. So, substitute the right hand side in this equation and what we get is alpha u transpose gamma k into u plus beta v transpose gamma k into v that is equal to delta minus B k gamma k. So, now we have two vectors delta k and minus B k gamma k on the right side. So, let us equate u to delta k and v to minus B k gamma k. So, then naturally multiplies of u and v have to be one.

Because the multipliers of delta k and minus B k gamma k are 1 here on the right side therefore, this quantities become 1 and this gives us the one of the values of alpha beta u and v. So, if we use u to be delta k v to be minus B k gamma k let alpha u transpose gamma k to be 1 and beta v transpose gamma k to be 1 then we get 1 over alpha to be delta k transpose gamma k and beta to be minus 1 over gamma k transpose B k gamma k.

And therefore, what we have is the new update rule corresponding to the rank-two correction. So, we have B k and then alpha is nothing but one over delta k transpose gamma k and u is nothing but same as delta k. So, that we have delta k delta k transpose here in the numerator and then we have beta to be minus 1 over gamma k transpose B k gamma k. So, that quantity is here and v v transpose. So, v is minus B k gamma k. So, that becomes B k gamma k gamma k transpose b k. So, this is the symmetric matrix symmetric matrix and if B k is symmetric then certainly this is going to be the symmetric matrix. So, this method of rank-two correction is calling the D F P method. So, this is a named after the inventors of this method Davidson-Fletcher and Powell. So, this method will be called D F P method.

(Refer Slide Time: 46:15)



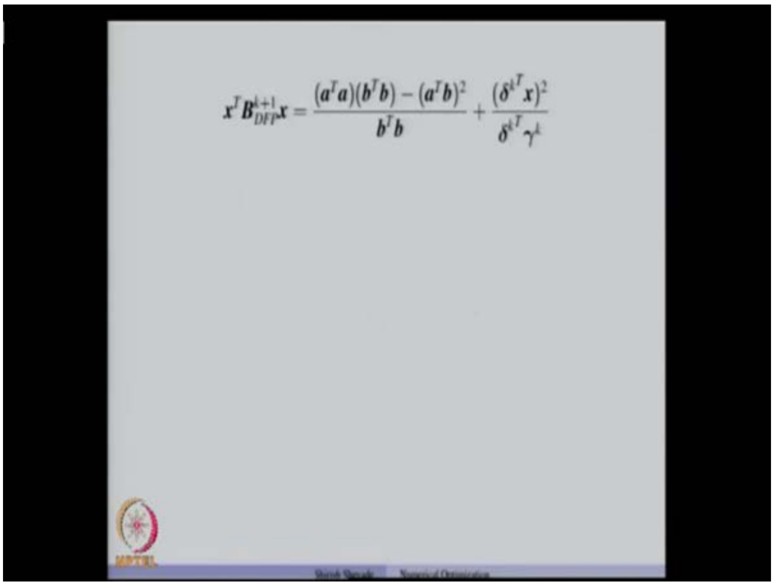
Now, let us see more about this method. So, we have this formula and the question that we would like to answer is that is the matrix on the left side symmetric positive-definite given that this matrix B k is positive-definite and the matrix now left side is obtained using this D F P update rule or D F P method. Now, clearly that if B k is symmetric, then this matrix is also and symmetric and this is also symmetric. So, if B k is symmetric in the addition of symmetric matrices will give also symmetric matrix. So, that is a clear answer for symmetricity now what about positive-definiteness is this matrix positive-definite provide given that B k is positive-definite and that is the question that we would like to answer now.

Now, to show that matrix is positive-definite what we have to do is that we have take a non-zero vector x and show that x transpose B k plus 1 into x is greater than 0 for all x non-zero now let us choose some x which is non-zero then gamma k is also non-zero. So, which means that g k and g k plus are not equal and delta k is also non-zero that means that x k and x k plus 1 are not equal now.

Now if we write x transpose B k plus 1 x. So, that will be x transpose B k x. So, I have just brought in the third term as a second term here and second term becomes your third term in this case, that is just for convenience. So, this quantity on the right side is basically x transpose B k plus 1 x. Now, we have B k to be a symmetric matrix. So, we can write B k as b to the power half into B k to the power half where B k to the power half is a symmetric and positive-definite. So, we can write is as a product of two symmetric positive-definite matrices note that B k also positive-definite. So, naturally B k plus B k to the power half will be positive-definite and B k is symmetric. So, we can always write B k as a. So, these are product of these two matrices which are factors of the matrix b k.

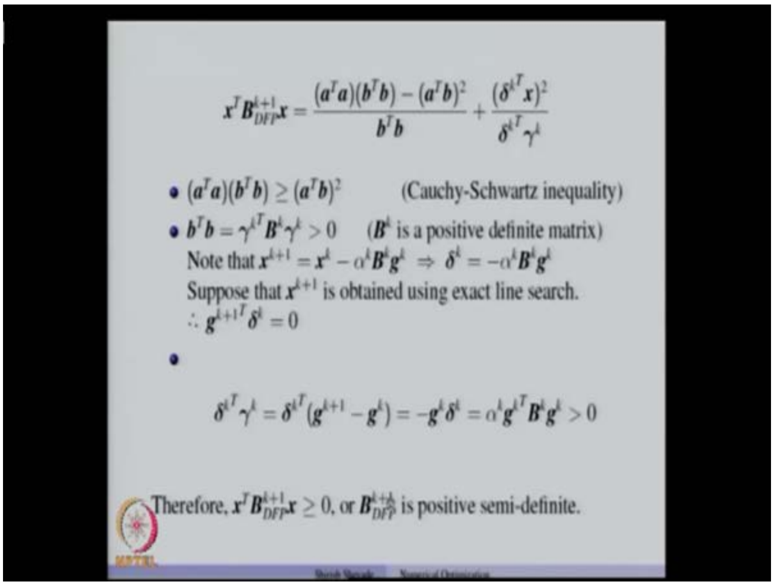
Now, let us define two new vectors called a and b, a is B k to the power of x and b to be the B k to the power half gamma k. So, x transpose B k plus 1 x it can be written as. So, this quantity here becomes a transpose a. Then, this quantity becomes a transpose b square and if we take the l c m is b transpose b and therefore, that a transpose a here gets multiplied by the b transpose b and these two quantities become a transpose a into b transpose b minus a transpose b square by b transpose b plus the other quantity remains as it is. Now what we have to show is that this quantity is greater than 0. Now first we show that, this quantity cannot be negative. So, that means that first we show that the matrix B k plus 1 is positive semi-definite and later on we show that it is positive-definite.

(Refer Slide Time: 50:22)



Now, let us look at the first term. Now, if you look at the numerator we will see that the numerator is always greater than or equal to 0 because of the Cauchy-Schwarz inequality. This numerator is also greater than or equal to 0. Now, if you look at the definition of b. So, b is like this. So, b transpose b is nothing but gamma k transpose B k gamma k and if B k is positive-definite then b transpose b is which is nothing but gamma k transpose B k gamma k that will be greater than 0 even gamma k is not equal to 0 and we have already assume that gamma k is not equal to 0. So, b transpose b is greater than 0.

(Refer Slide Time: 51:10)

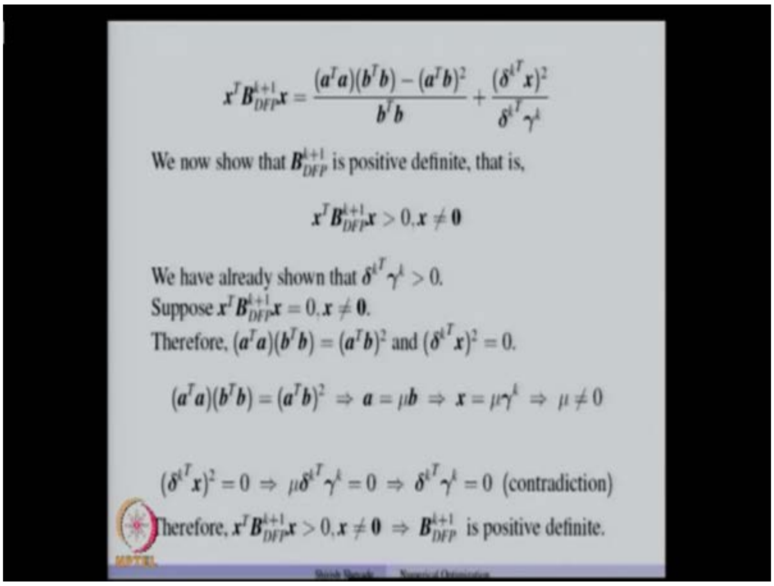


And what we have to do is that, we have to show that delta k transpose gamma k is also greater than 0. So, let us see how to do that. So, as I said earlier that the quantity in the numerator is always non-negative because of the Cauchy-Schwarz inequality. Then b transpose b is greater than 0 because B k is positive-definite matrix then delta k transpose x is square greater than 0 is greater than or equal to 0. Now, we have to look at the delta k transpose gamma k for that purpose we use the fact that x k plus 1 equal to x k minus alpha k, B k, g k remember that this is same as our formula that x k plus 1 is nothing but x k plus alpha k, d k and d k is nothing but minus B k, g k.

So, we can write this as delta k which is nothing but x k plus 1 minus x k and that is nothing but minus alpha k, B k, g k. And we suppose that x k plus 1 is obtained using exact line search. So, if we use exact line search then g k plus 1 transpose delta k is 0. So, now let us look at this quantity delta k transpose gamma k is nothing but delta k transpose g k plus 1 minus g k now if you have use exact line search in g k plus 1 transpose delta k is 0 and this quantity becomes minus g k transpose delta k and. So, this should be g k transpose delta k.

That is nothing but alpha g k transpose B k, g k now alpha k is greater than 0. Because of the step size is greater than 0 g k. We have assume that is not 0 and B k is a positive-definite matrix. So, g k transpose B k, g k is greater than 0 and therefore, what we have is that, this quantity on the right hand side cannot be negative because each of these quantities in the numerator is non-negative this is also non-negative this is positive and this is positive. So, the right hand side is non-negative that means that matrix B k plus 1 is positive semi-definite but we have to show that B k plus 1 is positive-definite. So, let us see how to do that.

(Refer Slide Time: 53:47)



So, let us again look at this expression and you have to show that B k plus 1 is positive-definite. Now for a moment we assume that B k plus 1 is positive semi-definite or not only that we assume that x transpose B k plus 1 x is 0 given that x equal to x not equal to 0. So, this is what we want to show but let us assume that x transpose B k plus 1 into x is 0, where x is not equal to 0 and note that we have already shown that the quantity in the denominator is greater than 0 the delta k transpose gamma k. Now if you assume that, this quantity is 0 and we have already shown that this quantity cannot be negative. Each of the terms cannot be negative. So, the only way that this quantity will be 0 is that each of the terms is 0. So, the first term is 0 and second term is 0 or in otherwise the numerators in both the terms are 0.

Now if suppose numerator in the first term is 0. So, that means a transpose a into b transpose, b a transpose b square and this in the numerator and the denominator is 0 means that delta k transpose x square is equals to 0. Then what we have is that if this holds then a has to be a scalar multiple of b and a and b are not 0 and therefore, mu is also not 0 and if we recall the definition x a and b what we can write is that x is equals to mu gamma k and since none of these are 0.

Mu is not equal to 0. So, x is not equal to 0 because we have assumed it here that we have already assumed that gamma k is non-zero. So, which means that mu is not equal to 0. So, the first the numerator in the first quantity is 0 implies that is the scalar multiple of b and that scalar is non-zero now let us look at the second term now delta k transpose x square equal to 0. Since x is nothing but mu gamma k. So, we can write this as mu delta k transpose, gamma k is equal to 0. Now, we already know that mu is not equal to 0 and that means delta k transpose, gamma k equal to 0 but then that contradicts the fact that we have already shown that is delta k transpose gamma k is greater than 0.

So, we get a contradiction and therefore, we cannot have x transpose B k plus 1 x equals to 0 and x not is equal to 0 and we have already shown that this B k plus 1 is a positive semi-definite matrix. So, this contradiction also ensures that x transpose B k plus 1, x has to be greater than 0 when x is not equal to 0 and which means that B k plus 1 is a positive-definite matrix. So, what we have shown here is that the D F P updated method results in a positive-definite B k plus 1. Now, let us look at some examples related to this D F P method and application of D F P method to some problems in the next class.

Thank you.