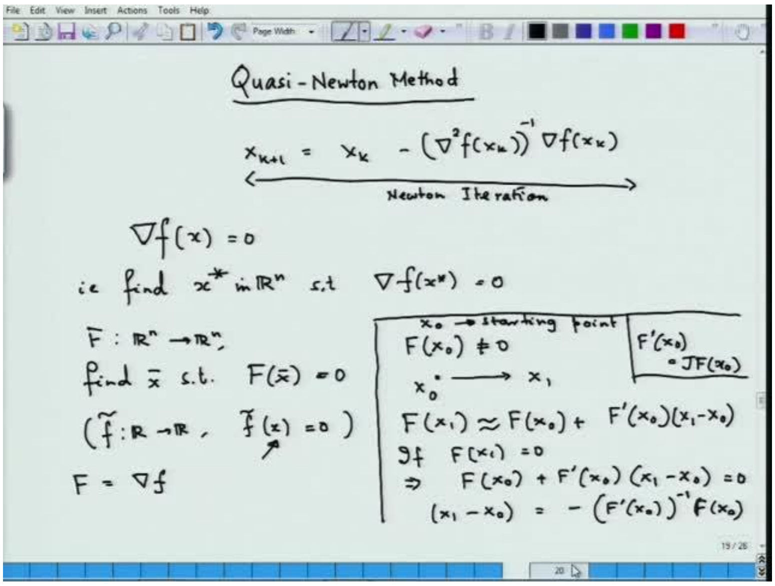
Foundation of Optimization

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

Lecture - 23

So, we now start as promised the Quasi Newton method's, we will see the home works later on. Now, if you go back and remember your Newton's method that we have discussed, that in Newton's method what we had essentially wanted to say is that instead of taking the instead of looking at the problem the function f itself at the point x k, current iterate, I look at a quadratic approximation of that f. And then I try to study the problem by trying to minimize that quadratic approximation, and that gives rise to what is called the Quasi, sorry the Newton equation Newton iteration, which I write as…

(Refer Slide Time: 00:59)



So, this is as you know we recall that this is what is called the Newton Iteration. Now, I can look at it from a very different point of view, what essentially I want? Why I am trying to solve an unconstraint optimization problem is to solve this equation, that is find some x star in R n such that the gradient of f x star is equal to 0. A more general problem can be that take a function from R n to R n, find x star or x bar such that F of x bar maps to the 0 vector.

Now, this how do I do, because we have we already know that equation solving, you you learn about Newton's iteration, essentially when you try to learn in calculus how to solve functions f, if you have functions from R to R say f tilde, then you want to have f tilde x equal to 0. So, you want to find the x here, so there you had learnt about Newton or Newton Raphson method whatever you want to call.

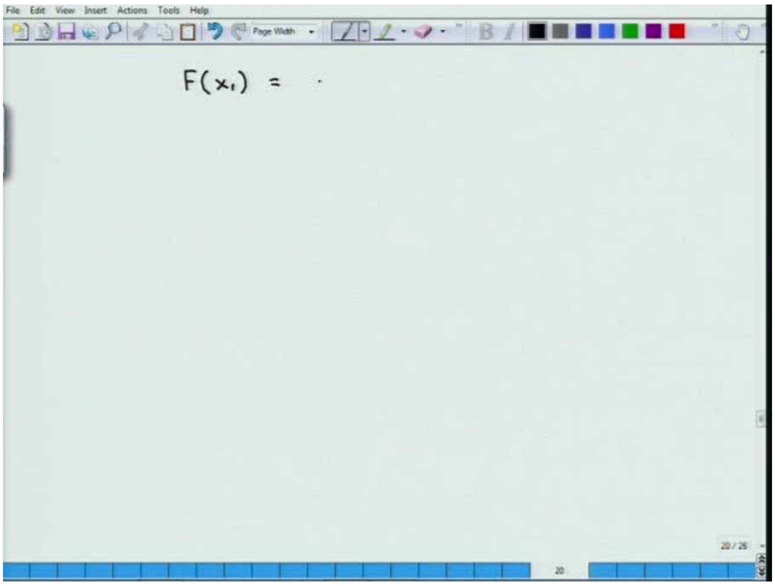
Now, how do the question is, how do if you have when n is strictly bigger than 1? The question of course, is how to find solution of this equation when essentially I can, now apply the same logic here, but only the important thing in optimization is to know that every iteration the function objective function value decreases, that we will fix up by fixing a descent direction. Now, if I have, supposed I start with a point x naught. So, x naught is my starting point, start point starting point.

Suppose, f x naught is not equal to 0, then I must move from x naught to some x 1, how do I do that? So, I can write the Taylor's expansion for functions from R n to R n vector valued functions. So, F of x1 the next iteration F of x naught plus, the Jacobian mapping which I am writing at as F dash x naught which actually if you if you want to be true truly and everything is nice, you should write this as Jocobian of F at x naught. So, I am just writing this as F dash x naught.

Then I have x 1 minus x naught, so I have made a Taylor's expansion, but it is not exactly a Taylor's expansion, because I have ignored the error term there is some error term here which I am ignoring suppose, I can do this that is if x 1 is somewhere near to x naught possibly, I can write here in this form here this is the matrix now. Technically this should be a number between there should be a vector between x 1 and x naught, but we are just or we should just briefly write, if you do not want to write equality you just can think of that I can write F x 1 approximately I can write it as this.

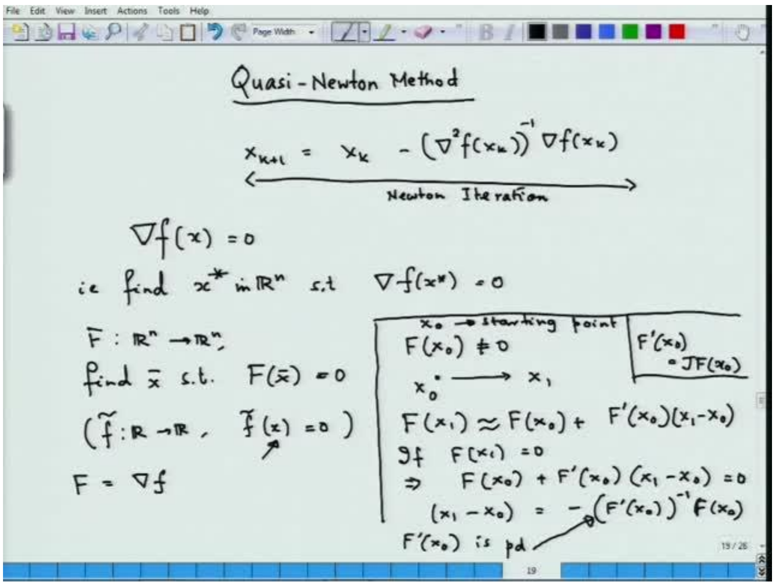
Now, once you can do that then you observe that now I want that x 1 to be my solution that. So, if x 1 F of x 1 is equal to 0, it would imply that we can approximately. Hence we can just forgetting the approximation for a certain time, you can have this expression which will lead to provided you can make a inverse of this matrix. Now, just put F capital F as your gradient of F. Then you get a Newton iteration then you get a Newton iteration. Now, what I do that, if I can do this sort of thing, when I want to move to the case where I need not have this 2 to be positive definite at every x k. Then I try to do something else that else is as follows.

(Refer Slide Time: 06:56)



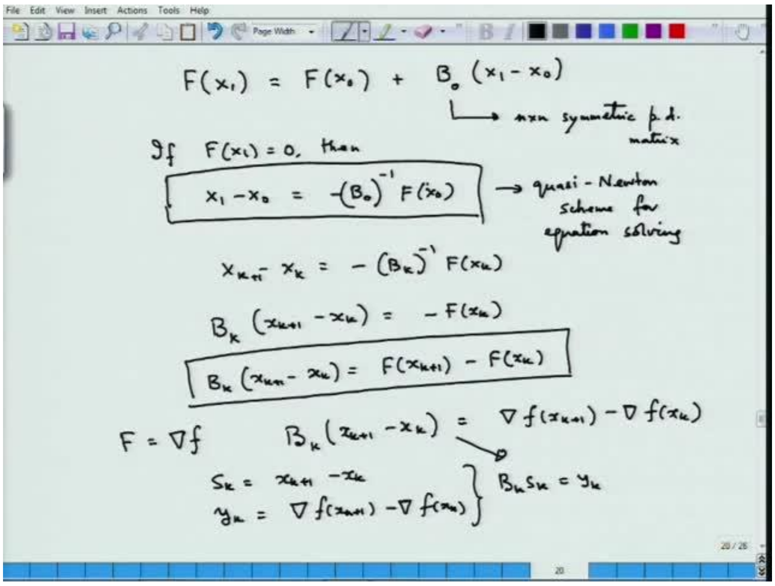
Now, instead of writing F x 1 in terms of the this derivative, I bring some B k, which is a positive definite matrix. I write because here, I would have really needed a here in this case, I would have really needed a positive definite matrix, so I can replace this by a positive definite matrix.

(Refer Slide Time: 07:50)



So, if I had a positive definite matrix, then if F F dash x naught was positive definite is p d, then it is invertible, you have got you get this, right? The, it is very important that, now what I do? I will make this sort of approximation.

(Refer Slide Time: 08:04)



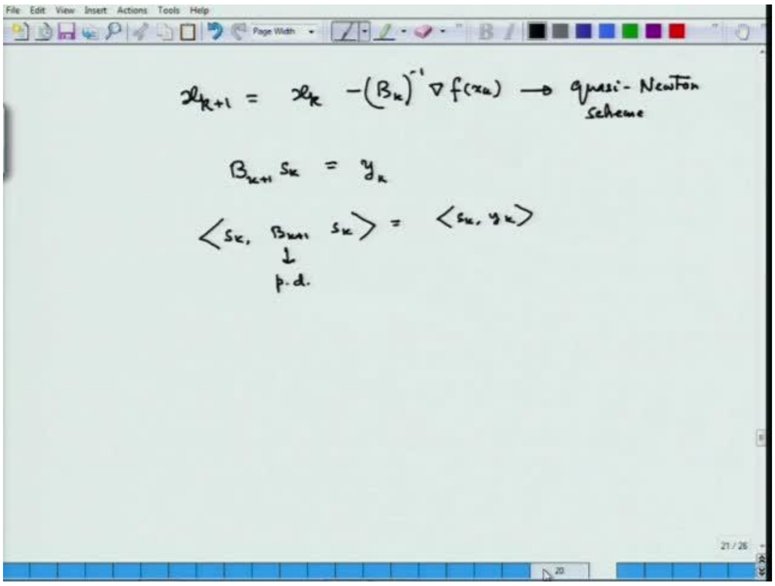
So, what I have is some p d matrix into with an n cross n symmetric p d matrix. It is an n cross n symmetric p d matrix positive definite matrix. So, once I do that it is very, very important to note the following. Now, I can write suppose my x 1 is my, if F x 1 is equal to 0, then I can write x 1 minus x naught minus B naught inverse F here. So, this is a very good scheme a Quasi Newton scheme or Quasi Newton scheme for equation solving.

So, I have replaced the because here, I need an inevitability which I cannot guarantee of the Jacobian every time. So, I so at every moment now, where I go from B naught at every iteration where I go from x 1 to x 2 I have to have a new B 1, which will be used to make the jump from x 1 to x 2. So, it is a Quasi Newton scheme for equation solving. Now, if you look at that what happens is that I can write more general thing as x k minus x k plus 1 minus x k is minus B k inverse f of x k, but I am making too much of an assumption, that I will take a p d matrix.

Suppose, I do not take a p d matrix, then I can write this thing in general as this is one way of writing the thing. So, essentially I can now make a more a nicer thing because I have already assumed that x K plus 1 is 0 at F x k plus 1 if it is 0, then I get this equation. So, it does not harm, if I write this as this is 0. So, from x k to x k plus 1, I come with a presumption that possibly x k plus 1 is a solution. So, if not then it it would not have a 0 value. So, this would be the iteration that I will use.

Now, if I put capital F as grad F, what I have here is that if I take some B some other matrix, which I do not bother, then it could be a positive definite matrix. Then you can put a b k that is it when you could get this iteration scheme, which is the Quasi Newton scheme. So, there is something I can write I can write this as grad F x k plus 1 minus grad F x k, this is sometimes called the secant equation or the Quasi Newton equation. So, I would put S k as x k plus 1 minus x k and y k as grad F x k plus 1 minus grad F x k. Then I can write this thing can be now written as, B K S k equals to y k.

(Refer Slide Time: 12:13)

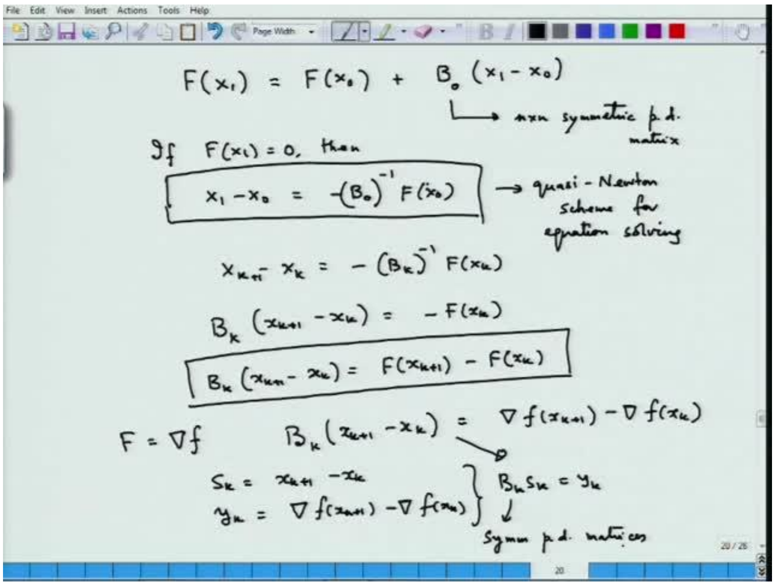


Now, actually the Quasi Newton iteration for the optimization problem can be written as because I am expecting this. I will update from B 0 to B 1 in such a way that all the B k's would be positive definite. So, you will have an inverse sometimes we need not bother about even competing the inverse, we can take an approximation of the inverse itself. So, this 1 is a Quasi Newton scheme, so also observed that if I have kept B k plus 1 S k is equal to k, then because B k plus 1 is positive definite, we will have S k B k plus 1 S k is equal to S k Y k, that now this is strictly bigger than 0, because this is p d.

So, in general we will always assume B naught to be p d because that will help us, but even if you have not assumed the p d. You could have still replaced the gradient by some

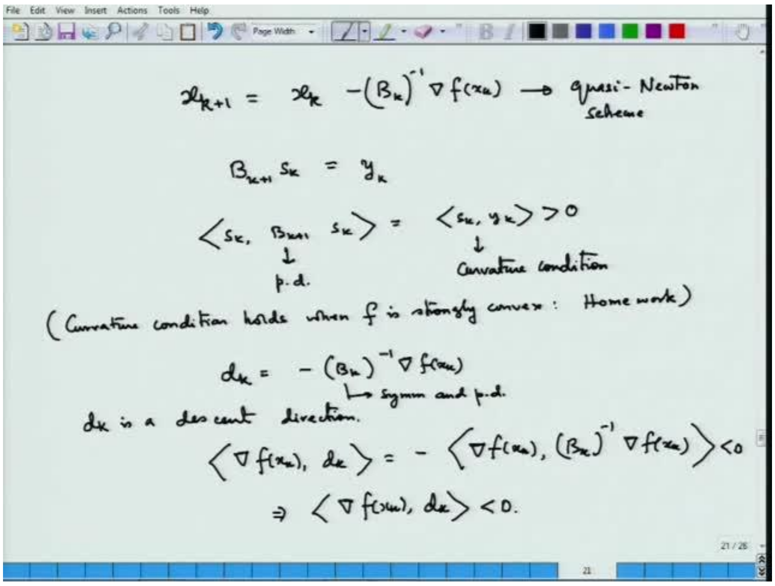
Hessian by some matrix and could have got this equation, but we will just continue that we are only taking symmetric p d matrices. So, all these B k's are symmetric p d matrices.

(Refer Slide Time: 13:56)



Now, the interesting thing that we really need to look in here is that this would mean.

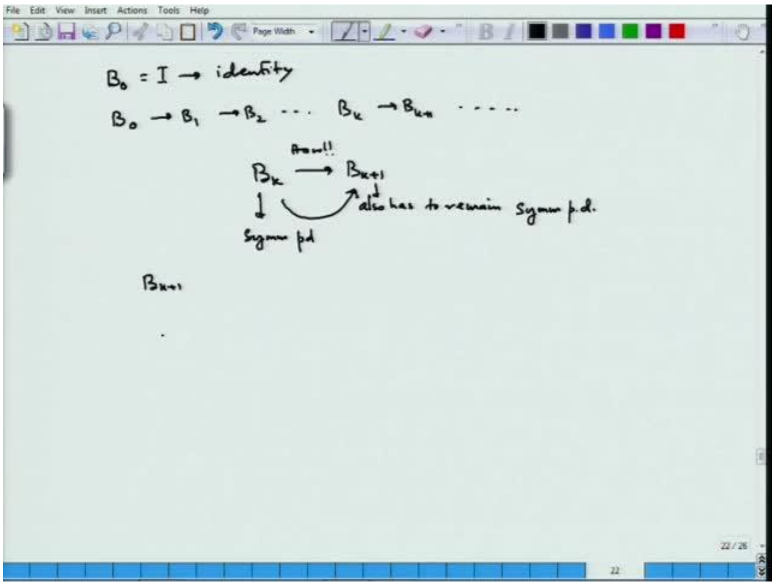
(Refer Slide Time: 14:13)



That this is strictly bigger than 0, so this is sometimes called the curvature condition, which is guaranteed when b is positive definite which is not guaranteed ,when it is not. So, when F is strongly convex such a thing would actually occur, but the curvature condition holds. So, this goes as your homework for this part. So, essentially now observe that if I take d k is equal to minus b k inverse grad f x k where B k is symmetric p d symmetric positive definite.

Then d k is a descent direction then d k is a descent direction in the sense, that if you write grad F x k into d k, this is nothing but minus because B k's p d's inverse is also p d positive definite. So, this part without the minus would be strictly greater than 0, so the whole thing this one is strictly less than 0 implying that… So, the important question here is not how important question here is, how do I update the matrices B naught?

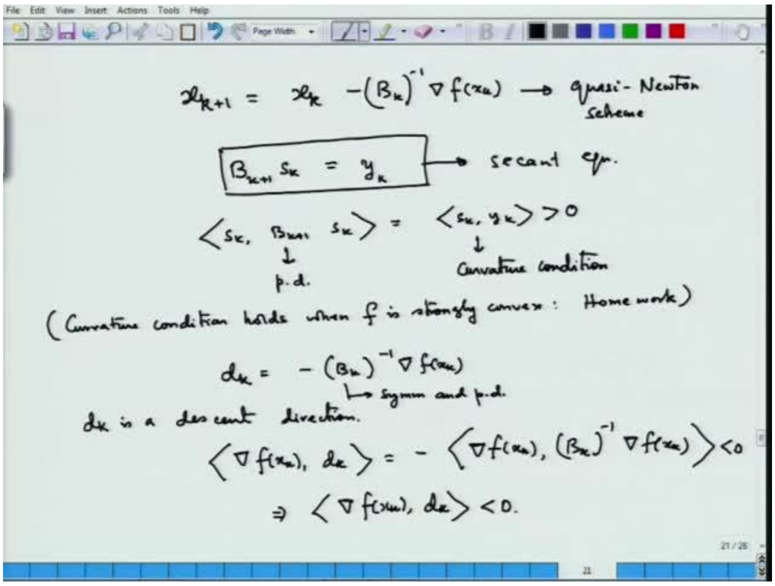
(Refer Slide Time: 16:25)



So, if I know B naught say usually the B naught the starting one is identity matrix because we know that it is anyway positive definite. Then I have to go go from B naught to B 1, B 1 to B 2 and so on. B k to B k plus 1 and so on let us stop. The question is how do I make this updates, so how to make an update from B k to B k 1 B k plus 1. So, question is how to do this and then if I want to do this, this B k plus 1 also has to remain this 1, has to remain p d positive definite positive definite, when I already know that B k is symmetric positive definite.

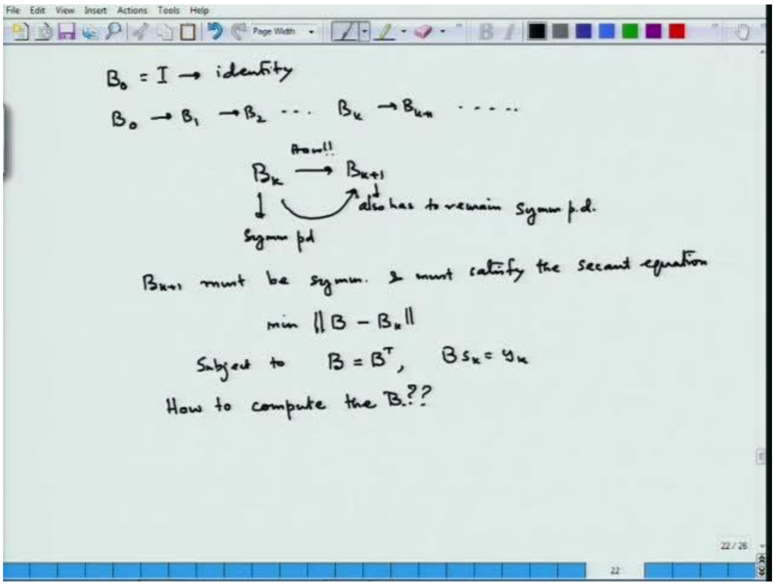
Of course, it has to remain symmetric p d which I am not writing every time. But like this, I will write in short that, now how do I find that right 4 minutes. Now, suppose I want to get a new B all right, I know the B k and I have to get a new B. So, what should that new B do the new B should at least be symmetric and satisfy and satisfy the same sort of equation that is basically I am expecting the new B k plus 1.

(Refer Slide Time: 18:27)



The same sort of thing that we had just figured out, so this is the equation B k plus 1 has in terms with S k and y k. So, k plus 1 and k is link through this, so called Secant equation.

(Refer Slide Time: 18:46)



Now, what I am trying to say is that, if I want to find a B k from B k, if I want to find a B k plus 1 B k plus one must be symmetric and must satisfy the Secant equation or the Quasi Newton equation. Now, what we want is that there should not be a huge change between B k and B k plus 1 because we want to also reduce our computational effort if there is a huge change between B k and B k plus 1. Then then that is not fair. So, what we really do is we want to find d the B k plus 1 through finding a problem of minimization in terms of matrices.

So, when I have a B k. So, I now want to look into this particular kind of problem, so I have to find the B this the B that will minimize this norm the difference between the norm, which I will say what sort of norm. We will use in the next class, so I have to have this the new the B k plus one should have been satisfying this. It has to has this, so basically I I want to now use this approach to chose my B. So, the thing is choosing different sort of norms, I would be able to choose a different sort of B which will be all p d.

So, tomorrow's class we will discuss how to compute the B and that would again take us back to the Karush, Kuhn Tucker conditions and the Fritz John conditions that we have learnt. So, and we will talk in detail about the type of norms etcetera that is used here so we will now try to solve may be we will look look into first the more simplified version of this, but then try to apply it here for a part left as homework. So, the question is now how to compute the B and that is exactly what we are going to discuss in the next lecture.

Thank you very much.