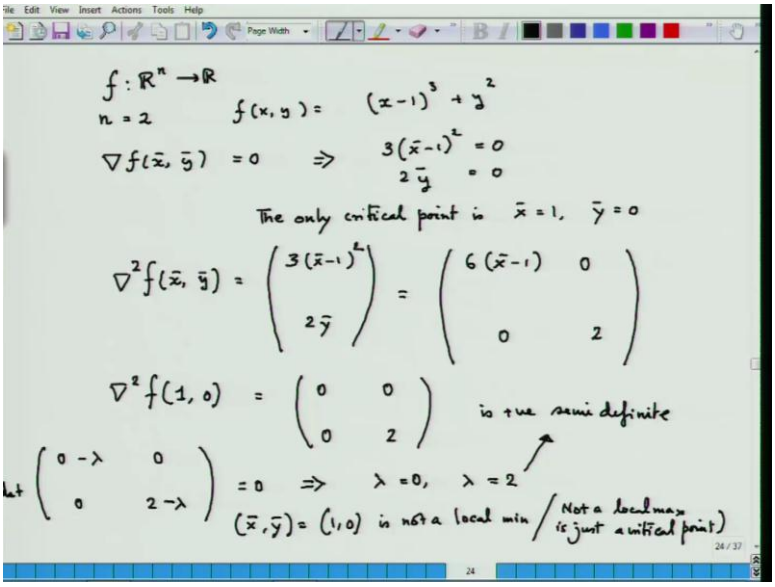
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 6

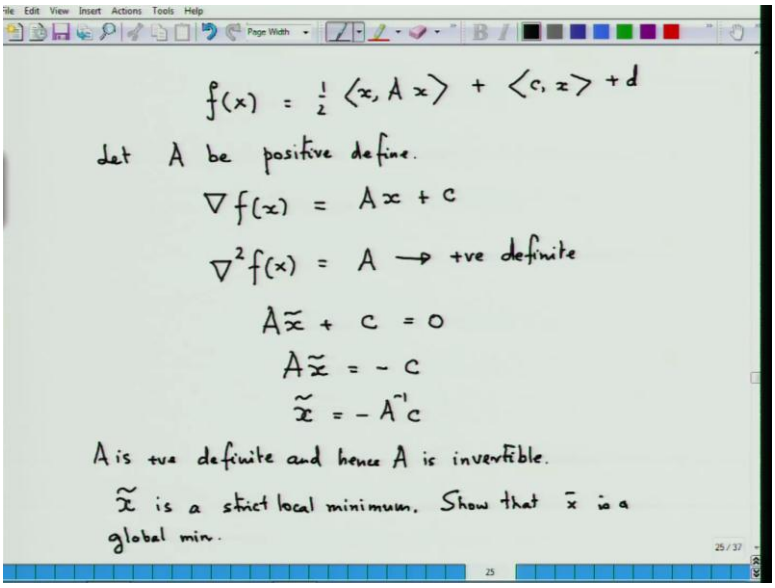
Now, we have spoken about the fact that if the Hessian matrix at a given point, a point which is a critical point is positive definite, then we can say that this point is a strict local minimum.

(Refer Slide Time: 00:37)



Now, we have demonstrated in the last class through an example that if we do not have positive definiteness may be just a positive semi definiteness, then it is not at all clear whether this point is our local minima or not; actually this point of this particular example that we have done in the last class, this point is not a local minimum. So, let us see how the theorem allows us to decide for some other case whether a point is minimum or maximum, and you see how sometimes second order conditions can become quite useful.

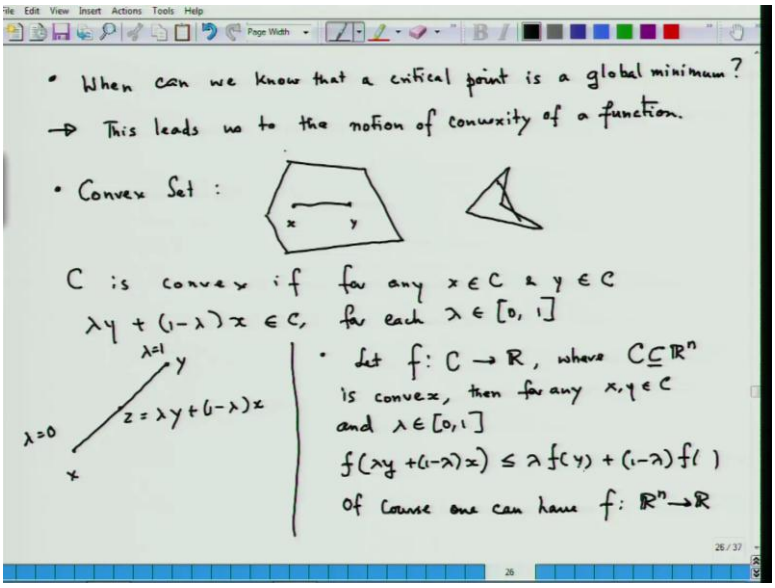
(Refer Slide Time: 01:07)



For example, you take a quadratic function f x. Now, let A be positive definite. So, this is a quadratic problem. So, this of course differentiable not only differentiability it is twice differentiable. So, now observe that grad of f x of this function is A x plus c while the Hessian matrix at any x whatever be your x is A and thus the Hessian is positive definite. So, any critical point of this problem any point x which satisfies this equal to 0 will be a solution of strict local minima of this problem; at least this information we would have. So, here A x plus c is equal to 0 would imply that A x is equal to minus c or x is equal to minus A inverse c.

A is invertible because A is also positive definite; any positive definite matrix is invertible. A is positive definite and hence A is invertible. So, when this is done; so this is one example an application. So, now the question is in this particular case, suppose this is my x bar which is the solution x hat let us take this as x tilde. So, what we can conclude that x tilde is a strict local minimum but actually in this particular case. So, x bar is a strict local minimum that information we already have from our theorem, but in this case may be as a homework I can ask you that show that x bar is a global minimum; this particular thing brings us to this very important question.

(Refer Slide Time: 04:18)



When can we know that a critical point is a global minimum? This will lead us to the notions of convexity which we will avoid at this movement. So, this leads us to the notion of convexity of a function. Now once this is known that we have to discuss something extra; we will take small d 2 and discuss what is the convex function and what is a convex set and what happens when a convex function is differentiable; for more details on the proofs of what we are going to establish, I would request, the viewer actually, the viewer to have a look at the course on convex optimization which I had given earlier to get a better understanding of all this.

So, here because our main notion is to put in plain simple words and plain simple approach the various basic tools in optimization we will not get into too much of mathematical issues which we got in the last lectures, and here for example I have not given a proof as to why the Hessian mean positive definite at a critical point leads to that critical point to be in a strict local minimum.

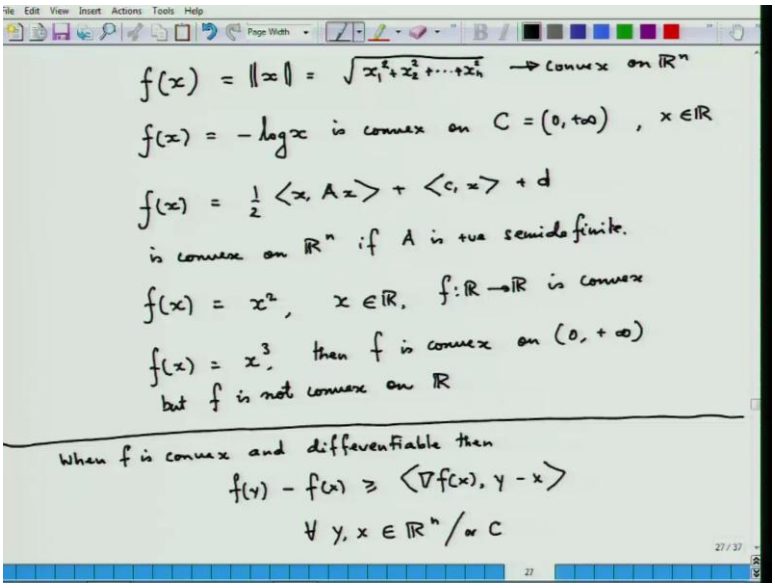
So, we have stated the result, showed one example. Now let me go and define first what is the convex set? Convex set means a set of this form means if you take any x and y points in the convex set and join them by a line segment and this line segment lies completely in the set. It is quite nice to look at. I give you two points, when I join them; the part of the line is outside the set.

So, this is not a convex set; human body for example is not a convex set in all. So, this is x written as this is given as follows that c is convex if for any x element of c, any y element of c, lambda y plus 1 minus lambda x is element of c for each lambda belonging to 0 and 1. You see this for any lambda between 0 and 1 this presents a point on the line segment joining x y including x and y; that is here is x and here is y. Any point z on this line segment is lambda y plus 1 minus lambda x and you see when I put lambda equal to 0 I get x; when I put lambda equal to 1 I get y. So, as I make lambda move from 0 to 1 I am actually moving along the line segment from x to y.

So, that is what it says that if you move along the line segment from x to y, you still continue to remain on the set c and this is what is called a convex set. So, then we come to the notion of a convex function. This is all done as you know as I am trying to answer such a question which is quite natural because that question came from the last example of that quadratic convex function which has Hessian which is positive definite. Now let us take a convex let us take a function f from C to R where C this is in R n where C is subset of R n is convex.

Then for any x y in c and lambda in 0 1 f of lambda y plus 1 minus lambda x is less then lambda of f y plus 1 minus lambda of f x. Now you could of course, one can also have see this function is well defined because for example, this point in c because x and y is in c because of the convexity of the set c itself. So, of course one can have f from. Now, let us give some examples of this convex function and why this function is important on studying such a question.

(Refer Slide Time: 10:09)

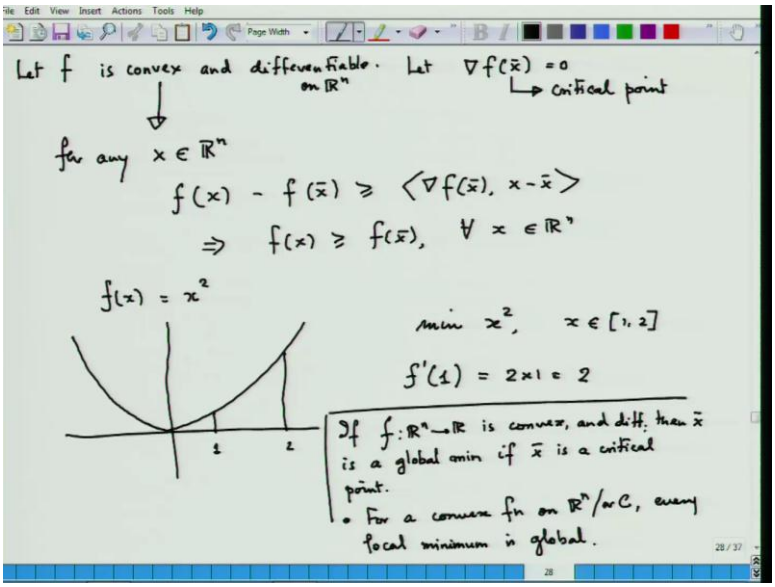


Now if you take a function f x norm of x, say, the Euclidean norm. So, these are examples of convex functions. So, f x is equal to minus log x is convex on c. So, you would see here that this set and check out is also a convex function. So, this convex function convex is on R n. So, another convex function f x, let me just go back the same form of A x c, x plus d is convex on R n if A is positive semi definite. There are many many such examples actually I do not say. So, if you take f x equal to x square of course, here x is in R, x is element of R then f from R to R is convex. Now you take f x equal to x cube then f is convex on 0 to plus infinity. Sorry, there is a mistake here; I just like to point this; it should be 0 to plus infinity, but f is not convex on r. So, function on a particular domain when they are seated at a particular domain could be convex but need not be convex on whole of R n.

Now it is important that you can look into some little property of these convex functions. So, when f is differentiable f is convex and differentiable, then now when I say make differentiable and if I define the convex function over the set c, then I meaning that the function f itself is differentiable may not be convex outside c, but it is differentiable on a neighborhood which contains the set c on an open set which contains the set c. So, because for differentiability openness of the set is important because would allow us to take limits in always; otherwise at the boundaries there could be difficulties. So, when f is convex and differentiable then would be on c could be on R n to whole of R n. Then what we have is this formula for all y and x in R n or c whatever

you want. So, whenever I am taking the set c instead of R n then I am actually telling that if the set c is closed, then I am assuming the differentiability of the function over a domain and over an open set which contains the set c.

(Refer Slide Time: 14:49)



Now let me observe a very simply thing. If f is convex and differentiable; no, let say let f is convex and differentiable, let us have a convex and differentiable function. Now let grade of f x bar equal to 0; that is x bar is my critical point. So, here I am trying to answer when can we know that a critical point is a global minima? A critical point here I want to recall again that is a point which satisfies the equation that the gradient of f at that point is equal to 0. So, it is a critical point. Now because f is convex and differentiable, then we know that this formula holds for every x y. Now if I can fix the x I can have this formula valid for every y. So, it shows that for any x in. So, let us in this case take the domain c to be R n then for any x in R n f of x minus f of x bar must be greater than grad of f x bar x minus x bar. But grad of f x bar is 0 which would imply immediately that f of x must be bigger than f of x bar for all x element of R n

Now you can ask can I do this for the set c if f was restricted to a set c and I have a point for which grad f x bar is equal to 0; then of course, it will be true. So, instead of x element of R n I will write x element of c, but the point is that the gradient values need not be 0 at the boundary points specifically or at the minimization points when you have a restriction to a set c. A very familiar example which I always like to demonstrate is

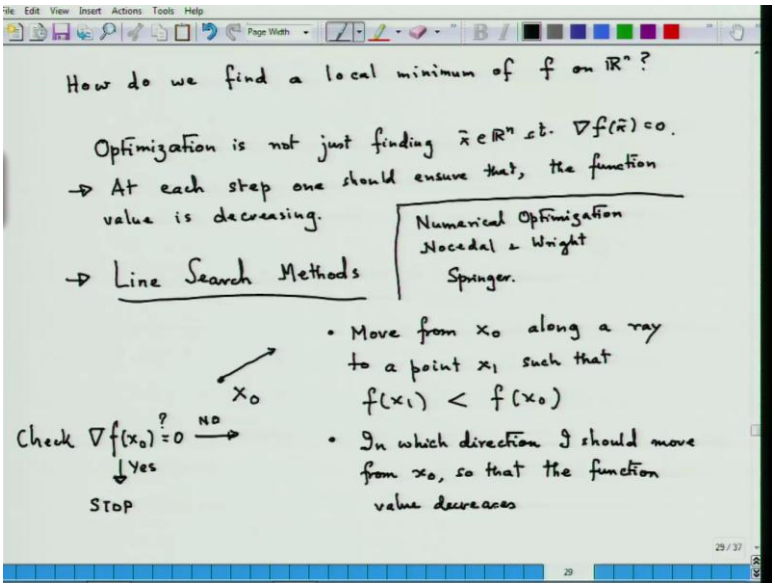
looking at the function f x x square. Now, let me take the graph of f x is equal to x square which is nothing but the parabola. Now, I restrict this parabola to 1, 2 and I ask this question minima is x square where x belongs to 1, 2. Now it is clear that the minimum point here it is achieved at the point 1; yet if it is so what is f dash of 1, it is 2 because this is two x the derivative.

So, f dash of 1 is 2; it is not equal to 0. So, it is not. So, if the problem has a constant is a constant problem and has a constant minimum; that constant minimum need not be a critical point. If it is a critical point, fine, but it not need be a critical point for very simple looking convex function whose graph is f x equal to x square. So, this fact is essentially a necessary and sufficient condition when we are considering an un constant convex problem.

So if R n to R, if f from R n to R is convex, so we have a neat nice condition if f from R n to R is convex; now we have a clear and little idea that if f is a convex function which is differentiable then x bar is a global minima if x bar is a critical point. Now if x bar is a global minima anyway it will become a critical point and if x bar is a critical point it is always a global minimum. Now the interesting part why convex functions had played such an important role and that is why I had a whole course on convex optimization is that for every convex function, local minimum is a global minimum; for a convex function R n or c every local minimum is global.

So, these are information which I am giving in a nutshell not in detail; just to enthuse you about this concept of basic fundamental concept of convexity and its usefulness in studying optimization theory, so optimization theory and algorithm. So, with this set up in place this little idea about that, yeah, there are functions for which it is very just enough for me to find a critical point where in that critical point can be declared as a global minimum. So, even it is very simple to hear, simple to listen. This is a very, very powerful result and that is what we should know to appreciate. Now we come to a mode of practical question. So, essentially given a ordinary function f which is differentiable and we want to find its unconstraint minimize, we essentially try to find a local minimum. Do we really find a local minimum? The question is how do we find a local minimum over function f on R n is the question.

(Refer Slide Time: 20:51)



So, we start with this question how do we find a local minimum of f on R n? It is an unconstraint minimization; how do we go about finding a local minimum? So, let us go and do some natural steps through which this procedure will pass. So, if I take a point on R n. So, my first step is that how do I find a local minimum. You can say try to find grad effects bar equal to 0, but when we do algorithms it is not always possible to compute out grad effects bar equal to 0 because to find out grad effect bar equal to 0, you have to run another algorithm to find an x bar which will solve that equation grad effect bar equal to 0. Now remember that there are a whole lot of things to say about how to solve nonlinear equations.

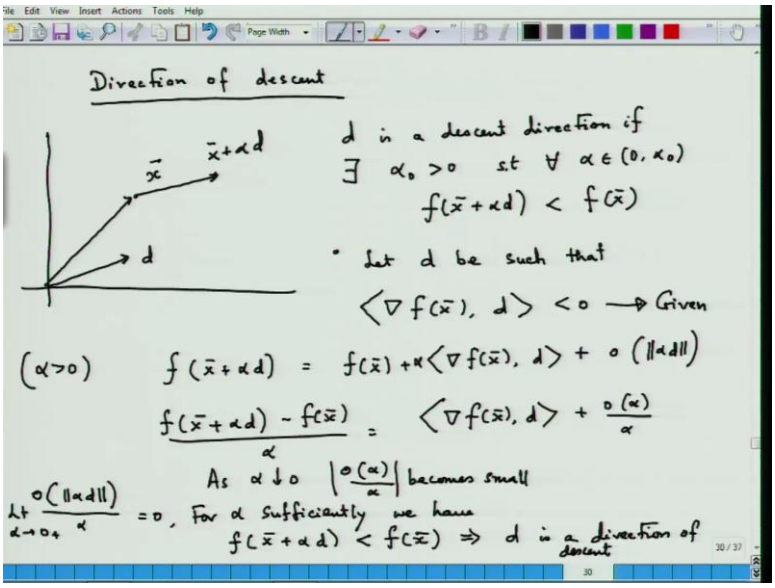
So, which we are not going to entertain ourselves with such things here, but we have to realize that optimization is not just finding x bar element of R n such that. You know when we run an ideality of algorithm in optimization, we have to remember that every step I have to improve the objective function value and if I am minimizing at every new point that I get when I start with the point I test whether it is an optimum or not. If it satisfies this I can stop the process; if not I need to find a point where the function value would decrease and this process of decreasing the function value should remain on as till you terminate the algorithm because that is what you intend to do because you want to minimize. So, next step is that at each step one should ensure that function value is decreasing.

Now how do we ensure that these two things the solving of this equation and this are done simultaneously? So, the game starts like this that how do we proceed to do that. A fundamental way of doing it is as follows and that technique is called the line search methods. One of the most important references and which we will use in this lecture is a book by Nocedal and Wright called Numerical Optimization published by Springer in the series in operations research and I would like to show you this book so that you can actually write down its name. So, please have a look at this book. So, I will just write down the name of the author whose books we will consult. There is another book called practical optimization by Fletcher, but it is largely for research rather than student who is at the advanced graduate and graduate level.

So, the name of the book is Numerical Optimization and the beautiful thing is that Indian addition is available. Now what does line search method mean? So, I get a point say x naught, I choose a point in R n; whatever happens in R n is fine, the story has to be told by drawing in R 2. Now my first part is to check because if grad f x not equal to 0 we know that we have at least found a critical point and once the critical point is found we can start doing this sort of trying to check positive definiteness and all those sort of things. Now if yes, suppose yes, I am asking this question whether it is 0; if yes then stop. Now if it is no, the question is what we are supposed to do if it is no; once this is no, then we must know that which direction I should move. So, I should move; shall I move in this direction, this direction, this direction, there are infinite ways to move, but I want to move along the line along a ray emanating at x 0. I have to move in some direction so that my function value sufficiently decreases.

So, the policy or the strategy is to move from x naught along a line to a point x 1, not really along the line or along a ray if that makes you comfortable, along a ray to a point x 1 such that f of x 1 actually sufficiently less than f of x naught; that is the strategy of the line search. Now here the first thing to know is in which direction I should move. I am telling okay, you move in some direction, move along a ray in certain direction, move along a ray. So, along a ray you move in a certain direction. So, in which direction I have to move; my next question is in which direction I should move from x naught so that; in which direction I should move from x naught so that my function value decreases, the function value decreases. So, I will take a descent along that direction function value descent. This brings us to the notion of a direction of descent or a descent direction.

(Refer Slide Time: 29:03)



Now what is the direction of descent? So, if you now let me make the x axis and y axis also. So, with this is my x bar now if I am moving this is my x bar if I am moving in a given direction say d. So, this is I will move parallel to this vector basically from x bar the given direction d and I come here and stop; I have moved, say, x bar plus some alpha d. So, basically d is a descent direction. If d is a descent direction if there exits alpha naught, say, strictly greater than 0 such that for all alpha element of 0 to alpha naught f of x bar plus alpha d is strictly less than f of x bar. So, this is called a descent direction at x bar. Now how do I assure how can I find the d.

So, this is the definition. So, I have to keep on searching, keep on searching; is there a much simpler criteria to tell me which d I would like to consider; that criteria is used by in the following way. Let d be such that grad f of x bar d is strictly less than 0, then you know the simple fact about directional derivatives or from vector calculus or if you write down the expression for the derivative or the Taylor's theorem in dimension. So, by the definition of the derivative which I think, yeah, you see I have given you the definition; this is the definition of the derivative, the Frachet derivative I have basically. So, this is actually the definition of the derivative. Now this would mean that if I take x bar plus alpha d where alpha is sufficiently small I can write this as f of x bar plus grad f of x bar d plus order of norm of lambda of d.

When you have norm whatever norm of lambda of d this is same as order of lambda. So, I can write this as f of x bar sub quads give an alpha, not lambda d. So, it is mistake; it should be alpha d. So, it should be alpha d. Now if you take the small o of alpha d this is same as alpha norm d. This is same as because alpha will come out; it will be same as o norm of alpha. This little asymptotic because you need to understand this fact that if you have because alpha would come out whatever be the multiple of norm d the same norm alpha or same value in same powers alpha would have. So, if you divide by alpha and do the things that would be again going to 0 basically what happens is o of norm alpha d is same as o of alpha. Basically in this case what would happen if you divide this by alpha and take limit alpha tending to 0 you would actually this would become 0; if you do this, this would actually become 0. So, that is why I can replace this by the term o of alpha. So, I would have this and divide this by alpha and write.

Now here comes the interesting part of the reasoning. Now this grad f x bar into d, this grad of f x bar into d is strictly negative; this is given to me and then what happens is that you see as I make alpha as alpha. So, alpha is taken to be greater than 0 here. So, as alpha goes to 0 means this symbol means alpha is positive and going to 0, then this quantity actually goes to 0. So, this quantity becomes smaller and smaller and smaller and smaller whether it is negative positive it does not matter. If it is negative and going towards 0 then also this sum total would be negative for some alpha after certain period some alpha. Even if it is positive, this negative will start dominating because this will become very small which will go beyond this thing and the negative will dominate. So, alpha tends to 0 o alpha by alpha becomes small, this becomes small. So, whichever way whether this when alpha becomes goes towards 0 from the negative side or from the positive side, this quantity would be 0 for some alpha.

So, for alpha sufficiently small which is very shorthand for saving lot of writing for alpha sufficiently small alpha is greater than 0; of course, let me put it this way. So, for alpha sufficiently small, 0 plus I have written here so that you remember alpha is strictly greater than 0. So, for alpha sufficiently small we have this whole thing strictly less than 0; for alpha sufficiently small means what that I have found there is an alpha naught such that for every alpha naught which is below alpha naught and between 0 and alpha naught, I would actually have thing going on; I would actually have this whole thing strictly less than 0 which will give me strictly things strictly less than 0. So, this would

implies that by this definition, if this is satisfied, if there is a d which satisfies this, d is a direction of descent and this alpha that we have seen here is called the step length.

So, we will talk about this in detail tomorrow as more detail about the step length, more detail about finding the step length and we will talk about the Wolf conditions and Armizo conditions which are very, very fundamental and then we will talk about a very specific type of method that is called the steepest descent method and the quadratic programming problem that we had studied earlier we would try to study quadratic programming. In fact, we would try to study this notion of how to handle particular quadratic function with the positive definite Hessian in more detail. So, thank you for today and I would like to close for today's discussion. We would get on to this details of finding step length tomorrow, I repeat, and also trying to understand the steepest descent method that is when d is chosen to be the negative of the gradient and trying to understand the problem of quadratic optimization.

Thank you very much.