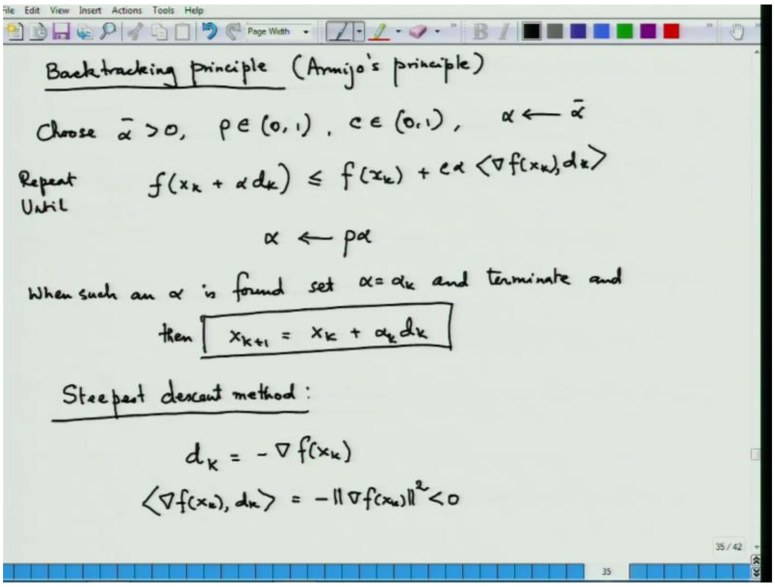
Foundation of Optimization

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

Lecture - 5

(Refer Slide Time: 00:20)



Backtracking principle (Armijo's principle)

Choose $\bar{\alpha} > 0$, $\rho \in (0,1)$, $c \in (0,1)$, $\alpha \leftarrow \bar{\alpha}$

Repeat Until $f(x_k + \alpha d_k) \leq f(x_k) + c\alpha \langle \nabla f(x_k), d_k \rangle$

$$\alpha \leftarrow \rho\alpha$$

When such an α is found set $\alpha = \alpha_k$ and terminate and then $\boxed{x_{k+1} = x_k + \alpha_k d_k}$

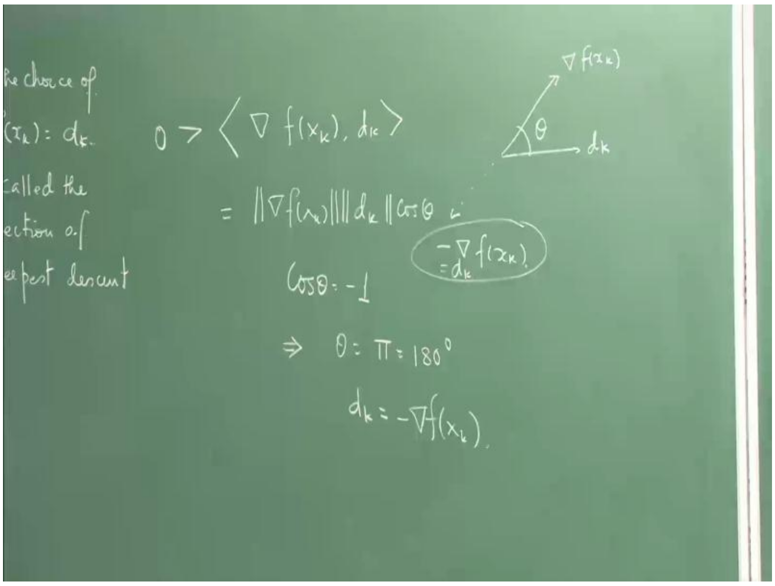
Steepest descent method:

$$d_k = -\nabla f(x_k)$$

$$\langle \nabla f(x_k), d_k \rangle = -\|\nabla f(x_k)\|^2 < 0$$

Today we start with this special choice of the direction of descent d k is negative of the gradient of f of x k. It is clear that if grad of f x k is not equal to 0, then norm of grad f x k is not equal to 0 and grad f x k into d k in this particular case comes down to. So, this of course is strictly less than 0. Now, why such a method would be called the method of steepest descent.

(Refer Slide Time: 01:02)



he choice of
$(x_k) = d_k$.
called the
ection of
eepest descent

$$0 > \langle \nabla f(x_k), d_k \rangle$$

$$= \|\nabla f(x_k)\| \|d_k\| \cos\theta$$

$$\boxed{-\nabla f(x_k) = d_k}$$

$$\cos\theta = -1$$

$$\Rightarrow \theta = \pi = 180^\circ$$

$$d_k = -\nabla f(x_k).$$

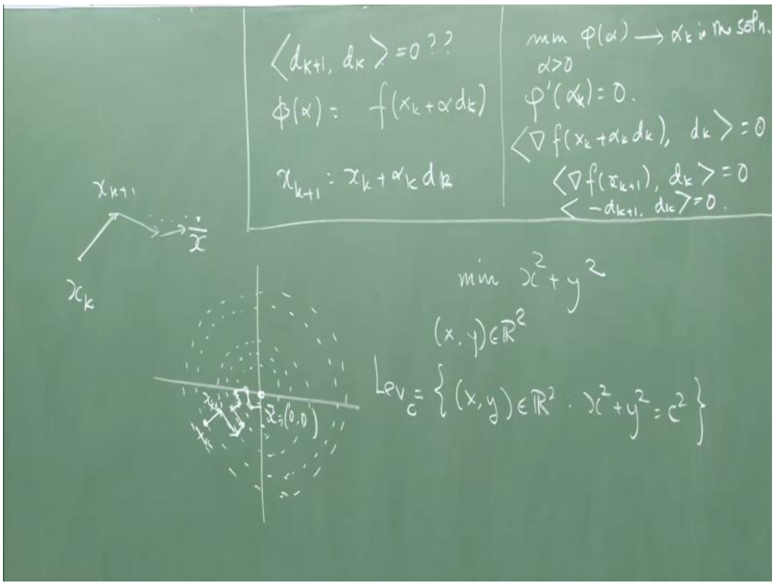
$\nabla f(x_k)$
θ
d_k

If you look at this thing, so what do you have here is you want to multiply this quantity. So, this is... So, why this choice my question is this; why the choice of minus grad f x k equal to d k is called that direction of steepest descent. So, we are now going to study various types of algorithms depending on various types of descent directions we choose; that is various types of line search methods; we are going to study Newton's method, we are going to study quasi-Newton method and so on with of course several examples. Now, in the last one I have given you that this example of Rosenbrock functions which we will see. So, we can apply that on the steepest descent technique. We can use the steepest descent technique on that and see what it illustrates?

Now, theta is of course the angle between grad f x k. So, this is the descent direction. So, this is the direction of descent. So, this is strictly negative. Now even if it is this is strictly negative, what is the most negative value of this? The most negative value of this is nothing but when cos theta takes the most negative value, because here these are fixed, cos theta takes the most negative value and cos theta is equals to minus 1. So, if cos theta is minus one which means what when cos theta is minus 1 cos theta equal to minus 1 gives the most negative value of this. Then cos theta equal to minus 1 implies at this and this angle theta it implies that theta is pie hundred eighty degrees which implies that d k is nothing but in this direction; d k becomes minus grad f x k, this is my d k now.

So, d k of course the angle has to be 180. So, it will come in this direction. So, that is why it is called the direction of steepest descent because the value of this becomes most negative when I choose d k equal to minus grad f x k, because this corresponds to the minimum angle for which cos theta value is minimum. Now this is not the only issue with steepest descent method. The question is, is the steepest descent method very good whether it is very fast; we need to think about all this. How does the steepest descent method move?

(Refer Slide Time: 04:56)



$$\langle d_{k+1}, d_k \rangle = 0 \;??$$

$$\phi(\alpha) = f(x_k + \alpha d_k)$$

$$x_{k+1} = x_k + \alpha_k d_k$$

$$\min_{\alpha > 0} \varphi(\alpha) \rightarrow \alpha_k \text{ is the soln.}$$

$$\varphi'(\alpha_k) = 0.$$

$$\langle \nabla f(x_k + \alpha_k d_k), d_k \rangle = 0$$

$$\langle \nabla f(x_{k+1}), d_k \rangle = 0$$

$$\langle -d_{k+1}, d_k \rangle = 0.$$

$$\min_{(x,y) \in \mathbb{R}^2} x^2 + y^2$$

$$\mathrm{Lev}_c = \{ (x,y) \in \mathbb{R}^2 : x^2 + y^2 = c^2 \}$$

That is suppose this is the actual minimum unconstraint minimizer of the function and this is my current chosen point x k. Now I have to find a direction which will take me somewhere here, then possibly it should take me along this direction, some in this direction in this way. This is the way line search method would work, this is x k plus 1 and so on and it will move towards x bar; that is the basic idea. Then this is my d k and this is my d k plus 1 the direction of descents, but in the case of steepest descent method very important thing to note is that the direction of descent d k and the direction of descent d k plus 1 is perpendicular to each other; that is if you for example consider minimizing the function x square plus y square over x y in R 2.

If you note this can you see here is 0 x bar is equal to 0 is 0 0 is the actual minimum minimize then what you draw around 0 are level curves. So, level curves are for this particular case is a set of all x y in R 2 such that x square plus y square is equal to some c square or some constant. So, this is called level c.

Now this level curves. So, this is of radius c. So, you can have level curves like this and level curves like this smaller and smaller. Now the interesting fact is that if you are here in any x k, suppose you are on any of the level curves at any time you will be on one of the level curves which is obvious because if you know x, you know you will see what is x square plus y square if you know x y so you know x square plus y square. So, you know at what level curve you are in.

So, the level curve you are on a level curve x k. Now from here you move to one point x k plus 1 this is x k plus 1, you will see what happens. Now you cannot move once it comes here in the case of steepest descent the next movement would be in direction perpendicular to d k. The next movement would be in another direction perpendicular to d k, the next could be like this and it could be like this, then it would be like this, then it would be like this, then like this. See the direction of descent cannot be in this direction because then it will increase the function value. So, it is decreasing function value, but you see this zigzagging procedure this sort of procedure where you are trying to maintain the perpendicularity that every point it is perpendicular.

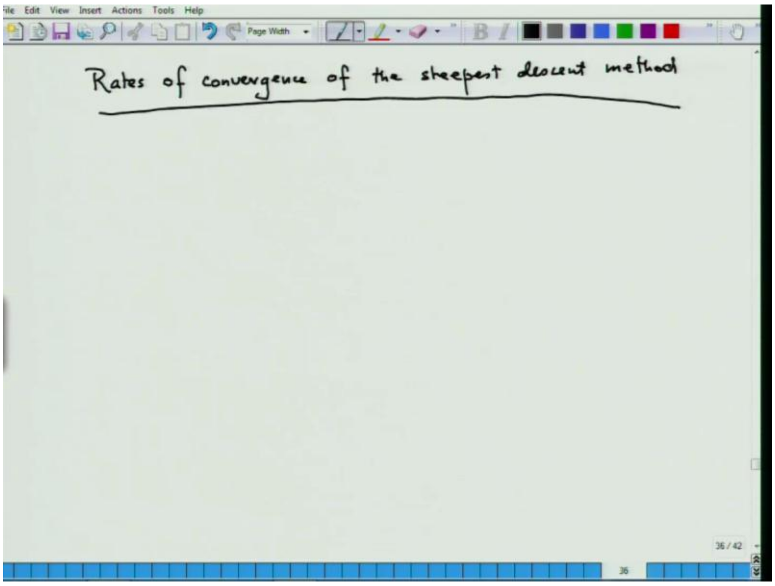
So, this d k plus 1 is perpendicular with d k d k plus one and d k plus 2 are perpendicular and so on. So, this zigzagging procedure slows down this algorithm very much; though there is an inherent simplicity in this algorithm we do not tend to get its benefit because it slows down because of maintaining this zigzag business. So, what I have to show is what I have claimed that this is zero. If you look at this thing I can tell you that but if you look at this thing just straight it is not easy to prove this; one has to come through a very different route.

Now what we have been doing here is we have been considering this function from R to R at the given k. Now to find this alpha this required step length, our main job was to minimize phi alpha over alpha strictly bigger than 0. Now you see you are actually faced with a constrained optimization problem which you have not studied till now in this course. So, how do I find an alpha which will minimize pie alpha if there is one such? Your first step is to take the derivative equal to 0; you might say okay, this is not a unconstraint problem, this is a constrain problem alpha is strictly bigger than 0, but alpha is strictly bigger than 0 means we are just considering the positive real line and that is an open set.

So, if you minimize over the open set and you minimize over the whole space then the optimality conditions necessary optimality conditions are the same; that is what we really have to do in order to find a step length. So, if alpha star is my solution or alpha k is my solution, then phi dash alpha k is equal to 0 because the required step length alpha k is the solution. So, if alpha k is the solution of this problem then alpha k is my required step length. So, that is I want the next point to be x k plus alpha k d k. So, phi dash alpha k is equal to 0 because that is the solution of this problem. So, which means that by definition you want to take the gradient that gradient of f of x k plus alpha k d k into d k is equal to 0, but what is this alpha k x plus alpha k d k, because the solution of this problem would be considered step length value.

So, the alpha which solves this problem is the step length value at the k th stage. So, this point obviously you know that we write x k plus 1 as x k plus alpha k d k because alpha k is the solution of this problem. It minimizes the function value on this ray; that is a ray in the direction of d k. So, this means this we have nothing but grad of f x k plus 1 into d k is equal to 0, but note that what is grad of f x k plus 1; by steepest descent method this is nothing but minus d k plus 1 and hence this is 0. So, d k plus 1 into d k is also 0 proving the fact that we are always in the perpendicular mode that the directions are perpendicular to each other. So, well we are now going to get into slightly more technical issues; we are going to talk about rates of convergence.

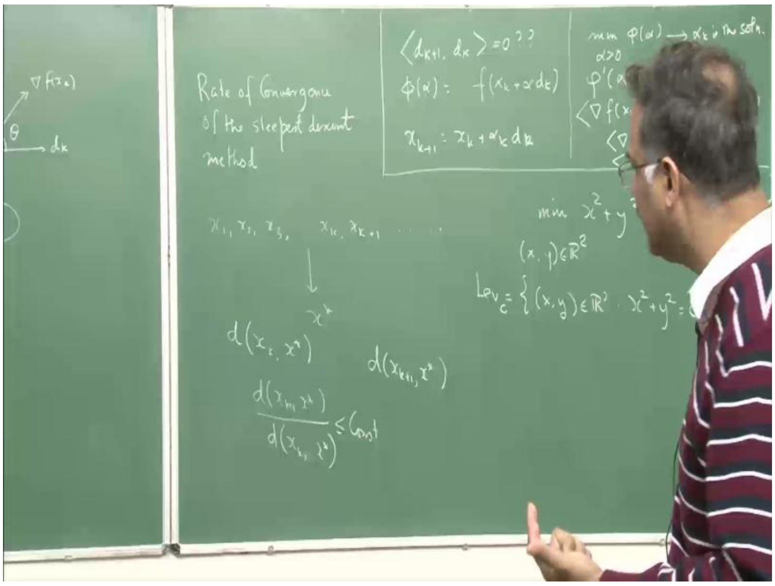
(Refer Slide Time: 13:01)



Rates of convergence of the steepest descent method

So, rates of convergence of the steepest descent method. So, what do you mean by the rate of convergence? Rate of convergence is some sort of ratio which gives you an understanding of how fast you are progressing towards the solution. So, let me just rub the board and let me erase.

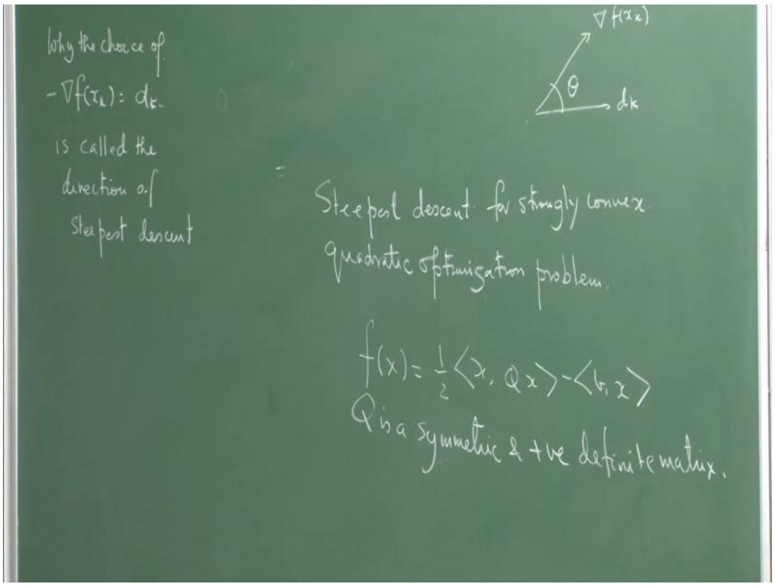
(Refer Slide Time: 14:05)


$\nabla f(x_k)$
θ
d_k
Rate of Convergence of the steepest descent method
$\langle d_{k+1}, d_k \rangle = 0$??
$\phi(\alpha) = f(x_k + \alpha d_k)$
$x_{k+1} = x_k + \alpha_k d_k$
$\min_{\alpha>0} \phi(\alpha) \to \alpha_k$ is the soln.
$x_1, x_2, x_3, \ldots x_k, x_{k+1} \ldots$
x^*
$d(x_k, x^*)$
$d(x_{k+1}, x^*)$
$\frac{d(x_{k+1}, x^*)}{d(x_k, x^*)} \leq$ Const
$\min_{(x,y)\in\mathbb{R}^2} x^2 + y^2$
$Lev_c = \{(x,y)\in\mathbb{R}^2 : x^2+y^2 =$

So, when you have this. So, what do I mean by this term rate of convergence of the steepest descent method; you will find this very much in most numerical optimization books. See I am actually running an algorithm which is generating these iterates and which we want to converge. We want this to converge this sequence of iterates converge to the true solution x star. So, when k is very large we can choose one of these elements from the sequence and we can say we are sufficiently happy with such a solution. So, what one has to do is to find the distance between x k and x star and the distance between x k plu1 and x star and one is to find the relation between them; that is really looking at the ratio which is basically the relative change if this is less than some constant. So, then the constant is less than one then we say that as we are, this is x k.

So, if the constant is less than one than we say that this is moving at a linear rate. Distance is usually given by the norm, then we say that it is moving at a linear rate and if the square of the norm of this distance square is less than some constant into distance square, then we say we have a quadratic convergence. In the sense what we are trying to show is that we are trying to assert in that what is the distance between x k plus 1 and x star in relative to the distance between x k and x star is x k plus one nearer to x star than x k. So, these are the question that one asks. So, this particular thing determines the speed at which the algorithm moves; for the steepest descent method this is very, very slow but we will consider this steepest descent method now for a very special class of problems called the convex strongly quadratic programming problem. So, we will consider the study of steepest descent for strongly convex quadratic problem.

(Refer Slide Time: 17:43)



Why the choice of $-\nabla f(x_k) = d_k$ is called the direction of steepest descent

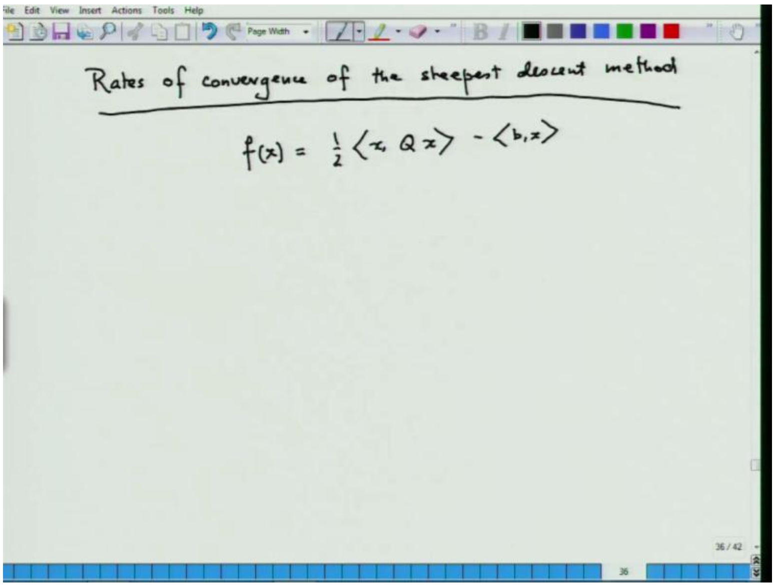
Steepest descent for strongly convex quadratic optimisation problem.

$$f(x) = \frac{1}{2}\langle x, Qx\rangle - \langle b, x\rangle$$

Q is a symmetric & +ve definite matrix.

So, here we will consider where Q is a symmetric positive definite matrix; we will explain why we are calling all this. So, we will diverge we will just take a little detour into convexity right now matrix. Of course, I expect everybody knows the definition of positive definiteness. So, we can look in to the thing little bit more in detail.

(Refer Slide Time: 19:15)

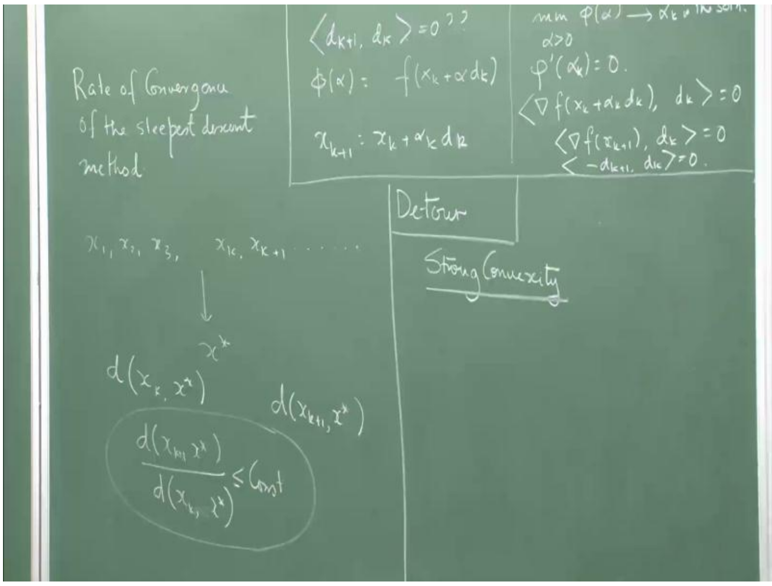


Rates of convergence of the steepest descent method

$$f(x) = \frac{1}{2}\langle x, Qx\rangle - \langle b, x\rangle$$

Now this is the function we have written on the board and we have claimed that this is the strongly convex quadratic optimization problem. So, we have spoken a bit about convexity in the very beginning and we have said that how does a convex function looks like when it is differentiable and that for a convex function every local minimum is global, and hence we need to show why this is strongly convex function. In fact we have to make a little definition of what strong convexity is and so that is our first task now before we try to analyze the steepest descent method for this particularly simple looking quadratic optimization problem.

(Refer Slide Time: 20:12)



Rate of Convergence of the steepest descent method

$x_1, x_2, x_3, \quad x_k, x_{k+1} \cdots\cdots$

$\downarrow$

x^*

$d(x_k, x^*)$

$d(x_{k+1}, x^*)$

$\frac{d(x_{k+1}, x^*)}{d(x_k, x^*)} \leq Const$

$\langle d_{k+1}, d_k \rangle = 0$??

$\phi(\alpha) = f(x_k + \alpha d_k)$

$x_{k+1} = x_k + \alpha_k d_k$

$\min_{\alpha > 0} \varphi(\alpha) \rightarrow \alpha_k$ is soln.

$\varphi'(\alpha_k) = 0.$

$\langle \nabla f(x_k + \alpha_k d_k), d_k \rangle = 0$

$\langle \nabla f(x_{k+1}), d_k \rangle = 0$

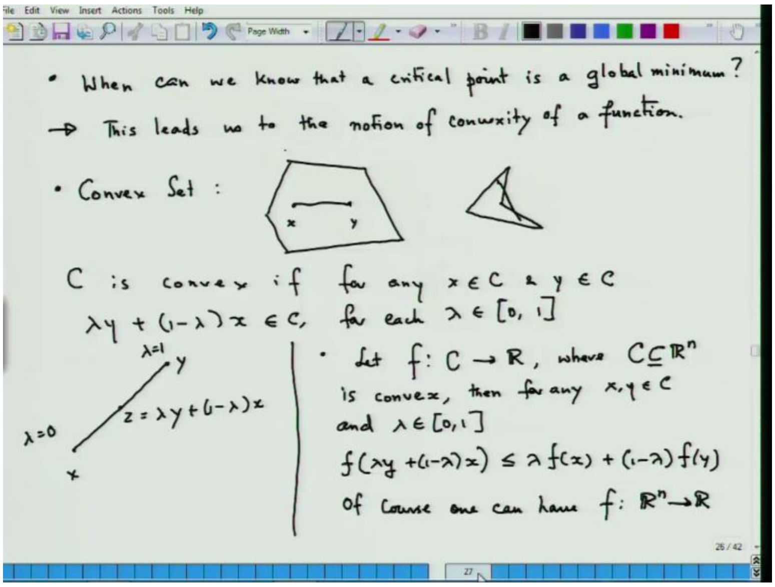
$\langle -d_{k+1}, d_k \rangle = 0.$

Detour

Strong Convexity

So we just take a detour, maybe I should write here detour. You should take a detour and let us talk about strong convexity. So, what is strong convexity means? So, you have I think if I just go back and try to show you the definition of the convex function.

(Refer Slide Time: 20:54)



- When can we know that a critical point is a global minimum?

→ This leads us to the notion of convexity of a function.

- Convex Set :

C is convex if for any $x \in C$ & $y \in C$
$\lambda y + (1-\lambda)x \in C$, for each $\lambda \in [0, 1]$

$z = \lambda y + (1-\lambda)x$

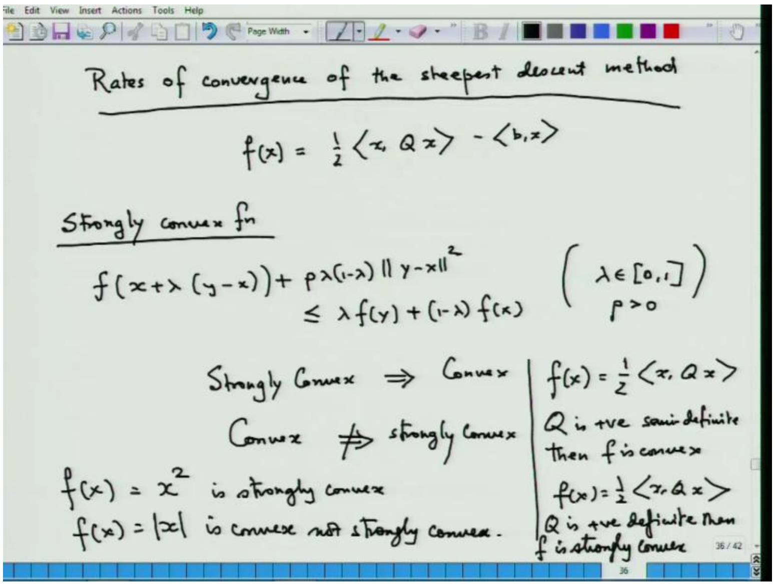
- Let $f : C \rightarrow \mathbb{R}$, where $C \subseteq \mathbb{R}^n$ is convex, then for any $x, y \in C$ and $\lambda \in [0,1]$

$f(\lambda y + (1-\lambda)x) \leq \lambda f(x) + (1-\lambda) f(y)$

Of Course one can have $f : \mathbb{R}^n \rightarrow \mathbb{R}$

So, here is the definition of convex function. So, please keep on looking at the definition of the convex function as we write down the definition of a convex function here.

(Refer Slide Time: 21:07)



Rates of convergence of the steepest descent method

$$f(x) = \frac{1}{2}\langle x, Qx\rangle - \langle b, x\rangle$$

Strongly convex fn

$$f(x+\lambda(y-x)) + \rho\lambda(1-\lambda)\|y-x\|^2 \leq \lambda f(y) + (1-\lambda)f(x) \quad \left(\begin{array}{c}\lambda \in [0,1] \\ \rho > 0\end{array}\right)$$

Strongly Convex $\Rightarrow$ Convex

Convex $\not\Rightarrow$ strongly Convex

$f(x) = x^2$ is strongly convex

$f(x) = |x|$ is convexe not strongly convex.

$f(x) = \frac{1}{2}\langle x, Qx\rangle$

Q is +ve semi definite then f is convex

$f(x) = \frac{1}{2}\langle x, Qx\rangle$

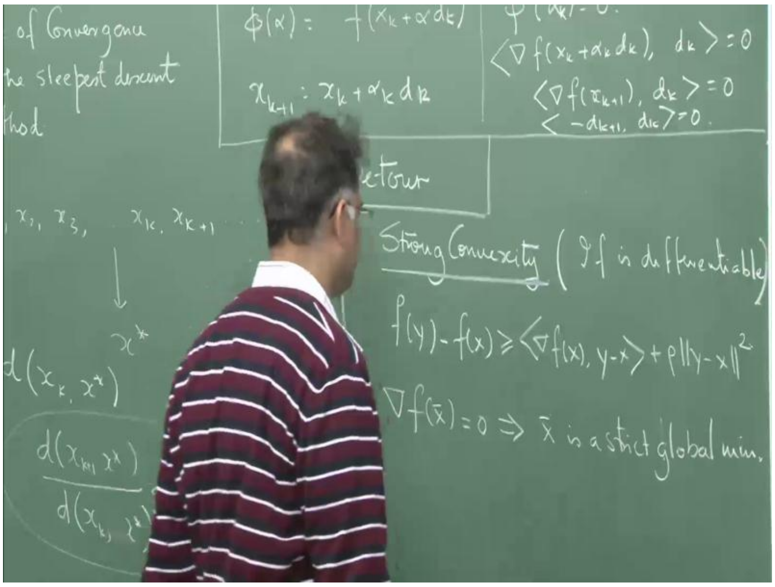
Q is +ve definite then f is strongly convex

So strongly convex function, I am not going to write too much of details, but this is what I am going to write. So, this quantity which is this one plus, now this is nonnegative quantity. So, what you have that this thing is obviously bigger than this little part this part, it is a nonnegative quantity. So, of course here lambda is between 0 and 1 and x y is in any r, strongly convex you write implies convex, but convex need not imply strongly convex.

Of course you can say, yeah, it is already clear from the name. See what we are expecting in this case of strong convexity that f is not only this part is not only bigger than this; this is bigger than some quantity which is bigger than this part is much more stronger than ocean. So, of course here rho is strictly bigger than 0 called the modulus of strong convexity and the rho has to remain same for every x y. This is something you have to know. So, f x equal to x square is strongly convex; f x equal to mod x is convex not strongly convex. So, these two examples are for this assertion.

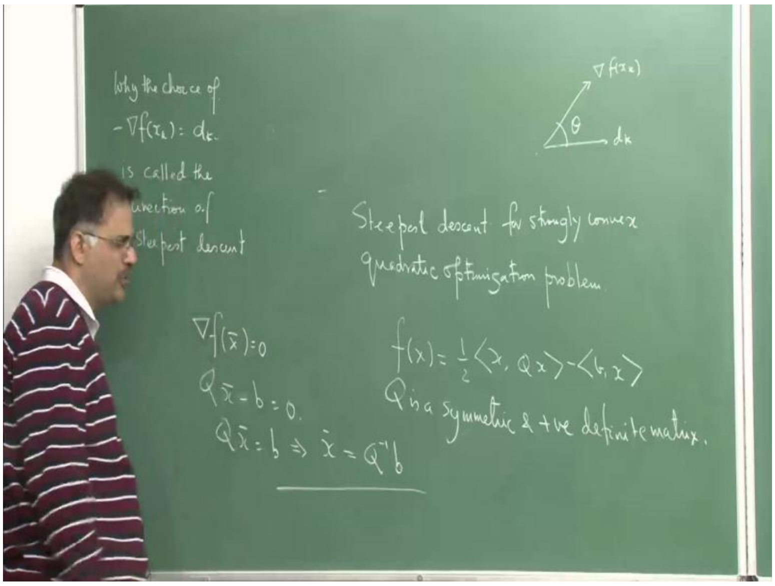
So, let us look at more or less at the quadratic function. Every linear function every linear part every linear function is both concave and convex; concave is of course the negative of convex, if f is convex minus f is concave. So, this part is always convex, but if you add two convex functions you will generate one more convex function. So, for a strongly convex function suppose you have taken Q is positive semi definite then f is convex, f x is half x Q x where Q is positive definite then f is strongly convex. Then f is strongly convex; that is for a strongly convex function the Hessian is always positive definite that is the idea the Hessian matrix. If there is twice continuously differentiable convex function whose Hessian matrix if it is strongly convex then its Hessian matrix is always positive definite, and if the Hessian matrix is positive definite we can say clearly that it is strongly; it is definitely it will be strictly convex in this case, and about also be strongly convex that it will satisfy this.

(Refer Slide Time: 25:22)


of Convergence
the steepest descent
thod
$\phi(\alpha) = f(x_k + \alpha d_k)$
$x_{k+1} = x_k + \alpha_k d_k$
$\langle \nabla f(x_k + \alpha_k d_k), d_k \rangle = 0$
$\langle \nabla f(x_{k+1}), d_k \rangle = 0$
$\langle -d_{k+1}, d_k \rangle = 0$
etour
$x_1, x_2, \ldots x_k, x_{k+1}$
x^*
$d(x_k, x^*)$
$\frac{d(x_{k+1}, x^*)}{d(x_k, x^*)}$
Strong Convexity (If f is differentiable)
$f(y) - f(x) \geq \langle \nabla f(x), y-x \rangle + \rho \|y-x\|^2$
$\nabla f(\bar{x}) = 0 \Rightarrow \bar{x}$ is a strict global min.

In fact if f is differentiable, then you will have a strong convex function this is to be true for any pair x y, one can take rho by 2 also it does not matter. So, this is what we get. So, grad f x equal to 0 would imply or grad x by 0 would imply x is a strict global minimum. So, if I want to find the minimum of this problem. So, how do I find the minimum of this problem?

(Refer Slide Time: 26:43)



Why the choice of $-\nabla f(x_k) = d_k$ is called the direction of steepest descent

Steepest descent for strongly convex quadratic optimisation problem

$f(x) = \frac{1}{2}\langle x, Qx\rangle - \langle b, x\rangle$

Q is a symmetric & +ve definite matrix.

$\nabla f(\bar{x}) = 0$

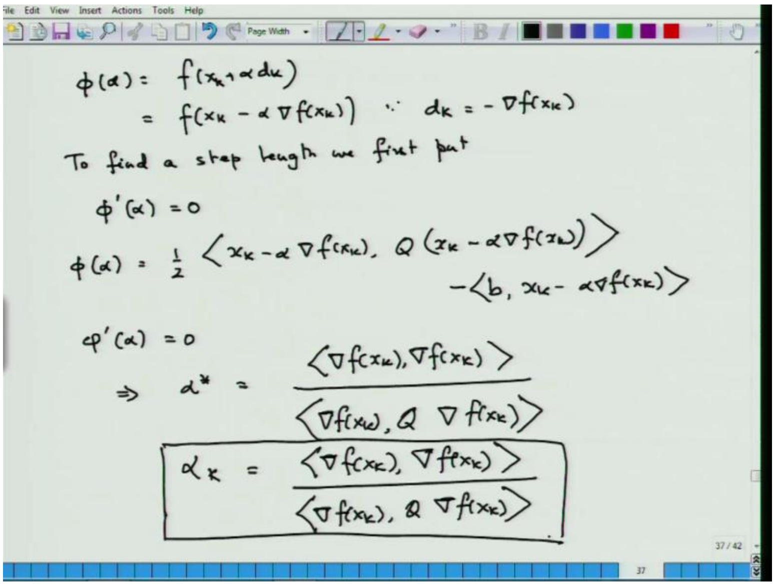
$Q\bar{x} - b = 0.$

$Q\bar{x} = b \Rightarrow \bar{x} = Q^{-1}b$

So, here because it is a strongly convex quadratic optimization problem minimum which would be unique in this case, and also for any strongly convex function it will be unique and it will be a strict global minimum. So, in this case if I take grad of f x if I want to find the solution, see here that grad f x bar equals to 0 would give me the solution since these are convex function. So, for a convex function every critical point is a global minimum, I think which we have told few classes back and so here you have this which means you have Q of x bar plus b equal to 0. So, minus b equal to 0 or Q x bar is equal to b, but since Q is positive definite it is also invertible and that would imply that x bar is equal to Q inverse b and that is the solution to this problem.

Of course, then why I am going to use steepest descent method to solve such a simple looking problem, but though it is a simple looking from the mathematical point of view, but there are lot of computational issues which can make taking the inverse of a matrix very computationally intensive, and also computational expensive, because lot of data has to be stored. And so we cannot always use this direct techniques to get an answer or the solution to a problem of this sort and that is why we have to resort to iterative methods by which in several steps we can solve the problem, and thus we will start with studying the case where we are looking for the function phi alpha.

(Refer Slide Time: 28:29)



$$\phi(\alpha) = f(x_k + \alpha d_k)$$

$$= f(x_k - \alpha \nabla f(x_k)) \quad \because \; d_k = -\nabla f(x_k)$$

To find a step length we first put

$$\phi'(\alpha) = 0$$

$$\phi(\alpha) = \frac{1}{2}\langle x_k - \alpha \nabla f(x_k), Q(x_k - \alpha \nabla f(x_k))\rangle - \langle b, x_k - \alpha \nabla f(x_k)\rangle$$

$$\phi'(\alpha) = 0$$

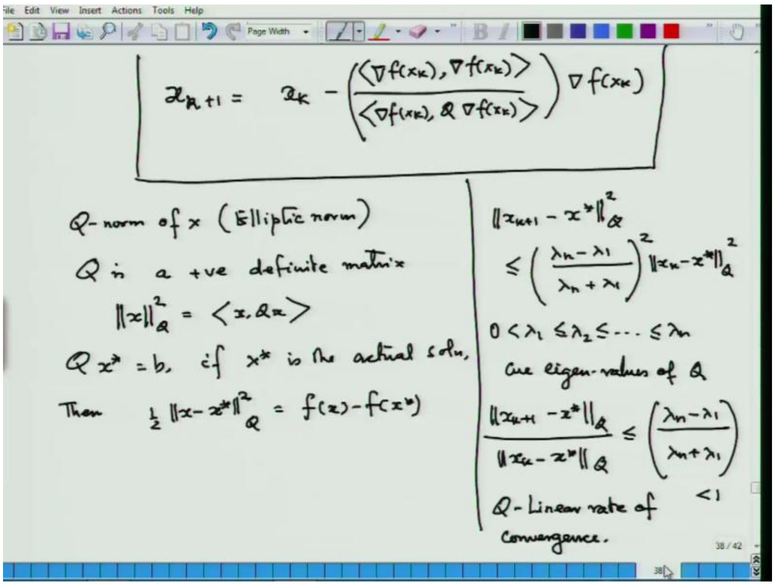
$$\Rightarrow \alpha^* = \frac{\langle \nabla f(x_k), \nabla f(x_k)\rangle}{\langle \nabla f(x_k), Q \nabla f(x_k)\rangle}$$

$$\boxed{\alpha_k = \frac{\langle \nabla f(x_k), \nabla f(x_k)\rangle}{\langle \nabla f(x_k), Q \nabla f(x_k)\rangle}}$$

So, we are now looking at the quadratic case. Now what happens here? Here this is nothing but because you are using the steepest descent. Now how do I find a step length? To find a step length we first put phi dash alpha is equal to 0; once I put phi dash alpha equal to 0 in this particular case, we may now write down the function value phi alpha particular case is written as half of.

Now, you have this minus, your job here is to compute phi dash alpha equal to 0 and this would implies some alpha star solution of this which I will leave as homework is given as. So, here because the function is convex quadratic any critical point is a global minimum because the function is strongly convex, the global minimum must be unique. So, unique the critical point that you get is your unique global minimum. So, this is my exact line search. Here actually I have to put alpha k this is alpha k is now computed in this form. So, alpha star is my alpha k. So, this is my required step length. Hence my iterative scheme now would look like this.

(Refer Slide Time: 31:32)



$$\boxed{x_{k+1} = x_k - \left(\frac{\langle \nabla f(x_k), \nabla f(x_k)\rangle}{\langle \nabla f(x_k), Q\, \nabla f(x_k)\rangle}\right) \nabla f(x_k)}$$

Q-norm of x (Elliptic norm)

Q is a +ve definite matrix

$$\|x\|_Q^2 = \langle x, Qx\rangle$$

$Qx^* = b$, if x^* is the actual soln,

Then $\frac{1}{2}\|x - x^*\|_Q^2 = f(x) - f(x^*)$

$$\|x_{k+1} - x^*\|_Q^2 \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \|x_k - x^*\|_Q^2$$

$0 < \lambda_1 \le \lambda_2 \le \dots \le \lambda_n$ are eigen-values of Q

$$\frac{\|x_{k+1} - x^*\|_Q}{\|x_k - x^*\|_Q} \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right) < 1$$

Q-Linear rate of convergence.

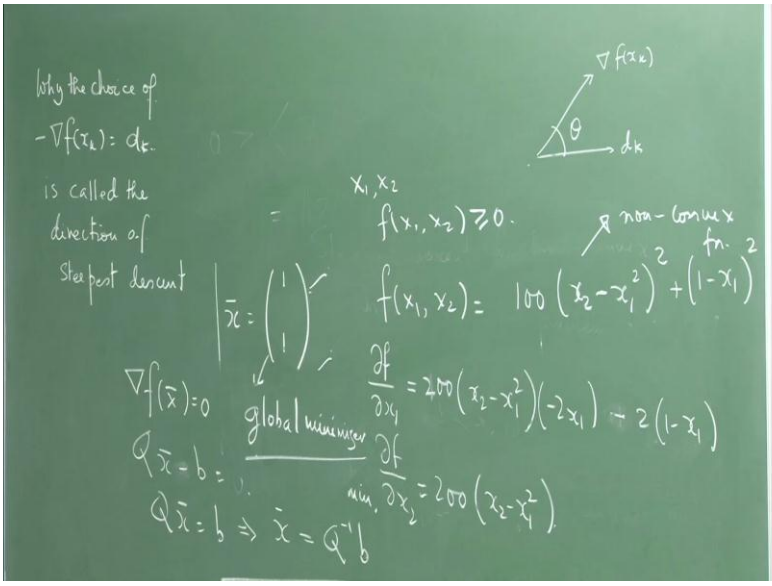
So, I am replacing the alpha. So, this is my iterative scheme for solving the strongly convex quadratic optimization problem. So, instead of see what I have done; I have now avoided the computation of the minimum. I have now avoided the computation of the minimum. Why using this, sorry, not the minimum the inverse. I have avoided the complete computation of the inverse of Q by taking iteration in this form. In fact, let us introduce what is called the Q norm of x or the elliptic norm. So, if Q is a PD matrix Q is a positive definite matrix, then we define the elliptic norm or the Q norm. This is what you have. Now, because you know that if x star is my actual solution, then Q x star is equal to b if x star is the actual solution then one can show half of. So, this is what you observe. So, this norm you see is measuring the difference between current value say x k and the x.

So, this difference this particular norm is measuring how far is the current objective value from the actual optimal objective value and that is measured by this Q norm of the current objective current iterate minus the solution which we really do not know because the solution is actually Q inverse b and we not know we do not want to find it by taking the inverse. As I told you it is very important to know the ratio of the distance between x k plus 1 and x star divided by x the distance of x k with x star. What one can prove this will be this proof of this will be given in your FAQ; at the end of the course the template will be attached to the course file and that you can see later on where the detail proof of this whole thing will be given because this depends on an inequality called conquer of inequality and our course means beyond the scope of our class to really go on proving such things though I would be happy to do.

This distance that is distance in terms of this elliptic norm the distance of x k plus 1 from x star is this is what you have. Now this lambda 1 is a positive definite matrix are Eigen values of Q. So, here we have a quadratic norm quadratic convergence with respect to the Q norm. Sorry not a quadratic convergence, I would say linear convergence with respect to the Q norm because what would we have is now x k plus 1 minus x star divided by x k minus x star is less than equal to lambda n minus lambda 1 by lambda n plus lambda 1 and this would be anyways strictly less than 1. So, here this by this has the constant which is strictly less than 1 then we at least in the Q norm. So, this called Q linear rate of convergence and this rate of convergence is very slow.

So, for the quadratic optimization problem what we have is called the Q linear rate of convergence. I had just told you that here I have told you that if this constant is strictly less than one then this is called this then x k this sequence is going towards x k x star at a linear rate. So, what you have shown for the quadratic case that we are going towards x star we are going towards the solution at a linear rate, but it is called Q linear convergence or Q we can call it Q linear we call or just call it the linear rate of convergence with respect to the ellipsoidal norm. So, with this idea we stop here and we would continue doing this studying these basic algorithms by doing the Newton’s method. But let us go back for a minute to the Rosenbrock function that I had given you from the book of Nocedal and right and let me see to what extent one can discuss about it.

(Refer Slide Time: 38:40)



Why the choice of

$-\nabla f(x_k) = d_k$

is called the direction of Steepest descent

$\nabla f(x_k)$

θ

d_k

x_1, x_2

$f(x_1, x_2) \geqslant 0.$

non-convex fn.

$f(x_1, x_2) = 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2$

$\bar{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

global minimizer

$\frac{\partial f}{\partial x_1} = 200(x_2 - x_1^2)(-2x_1) - 2(1 - x_1)$

min. $\frac{\partial f}{\partial x_2} = 200(x_2 - x_1^2)$

$\nabla f(\bar{x}) = 0$

$Q\bar{x} - b = 0$

$Q\bar{x} = b \Rightarrow \bar{x} = Q^{-1}b$

So, yesterday we gave an assignment where we are considering a function of this form. So, if you take the gradient of this one if you take the del f del x 1. So, it is 200 x 2 minus x 1 square into minus 2 x 1; this is one thing and del f del x 2 is plus 2 into 1 minus x 1. So, this is your del f del x 1 and del f del x 2 is nothing but here this is. So, if you put both equal to 0 what would happen? You see if I put both x 2 and x 1 equal to 1 both are equal to 0. So, x 2 equal to 1 1 is the only possible solution of this. So, I am asking you again to solve this at home which I will not solve. So, x bar equal to 1 1 is the only solution of this problem. Now it is the only critical minimum. So, here if you try to solve this, this will be equal to if you try to solve this thing you see 1 1 is a solution and this 1 1 is the only local minimizer.

You can calculate the Hessian at this point and the Hessian matrix will be positive definite at that point and it is not a global minimizer that you have to check out that this is a local minimizer and not a global minimizer. So, you calculate out and check if I put this one. So, I have x 2 equal to x 1 square. So, if I put x 2 equal to x 1 square this will become 0. So, from here I will get x 1 equal to 1. So, once I get x 1 equal to 1 I will also get x 2 equal to 1. So, 1 1 is the only solution of this thing and but still the critical point that we get is not a global minimum it is a local minimum. It is very important to know that this function is not a convex function; it is a non-convex function. It is differentiable, but this Rosenbrock function is a non-convex function. So, I would like you to try this at home.

So, for example if I put 1 1 here if I put 1 1 then x 2 is 1 and x 1 is 1 then I get 0. Of course, this is a non-negative function; this value is always non-negative for any x 1 and x 2. So, 0 is the minimum and that 0 value is at in 1 1. So, this is whole square this is whole square. So, 1 1 is where you are obtaining 0. Now the question is whether that is a global minimum. So, if you find. So, now if you put x 2 equal to 1 x 1 equal to 1, then now if you put x everything is equal to 0 you get 1. Sorry, it is not only a local minimize, I think I made a mistake. x 1 1 is not a local minimize; it is in fact because the Hessian is positive definite, you can definitely conclude that it is a strict local minimizer, but because you see that this function for any x 1 and x 2 f of x 1 x 2 is greater than equal to 0.

So, what if I would have x 1 equal to 1 and x 2 equal to 1 the function value is becoming 0. So, which means that this is a not just a local minimizer, it is a global minimizer and it is a strict global minimize; sorry, I made a mistake I think I was thinking this as minus 1. So, here is an interesting example of non-convex function for which you can attain a global minimizer at this point 1 1 and the steepest descent method on this would become very slow. There are lots of all the text books usually give examples where the steepest descent method on this class of functions works very, very slowly. So, with this I will end the talk today and tomorrow we will start discussing Newton’s method.

Thank you very much.