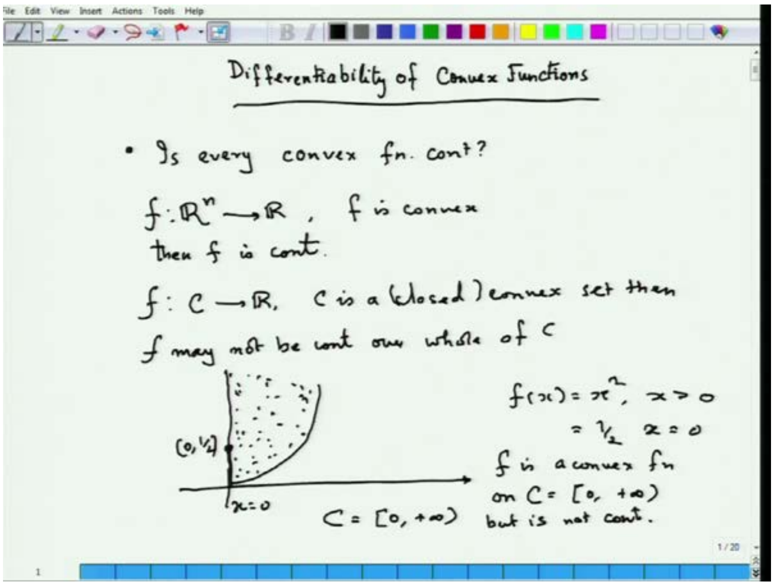
Convex Optimization

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

Lecture No. # 07

So, welcome once again to this course on convex optimization, in today we are going to speak about some very important aspect of convexity - the differentiability convex functions. Now, first question is whether every convex function is differentiable or whether it is continuous at all or is there something else.

(Refer Slide Time: 00:40)

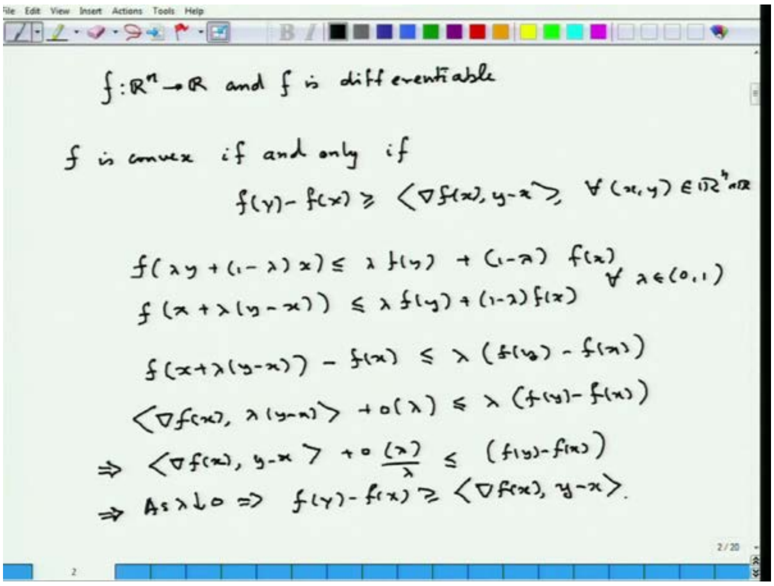


So, first question is, is every convex function continuous?; is a every differential function is continuous, so we will at least know that if a function is not continuous is not differentiable occurs. So, is every convex function continuous? The answer is surprising me nice, because here if you take a function of R n to R, and if f is convex then f is continuous. So, this is something quite interesting, but if you take a function f from C to R where C is a closed convex set. Convex set you need not bother word closed and this bracketing it. Then, f may not be continuous over whole of C.

(No audio from 01:57 to 02:07)

Let me take this very simple thing. Let us look at a function like this. Let us take this C as 0 plus infinity in real line and I define a function like this. So, ok. (No audio from 02:30 to 02:38) Say, this is just fx is equal to x square for x greater than 0 and is equal to half for x equal to 0. So, this is half, this is point 0 half, But if you look at the epigraph this function, this function is discontinues at a x equal to 0. But if you look at the epigraph of this function, this epigraph of this function is obviously convex. The epigraph is obviously a convex set. So, it is very clear that f(x) - f is a convex function on C equal to 0 to infinity, but is not continuous.

(Refer Slide Time: 03:57)



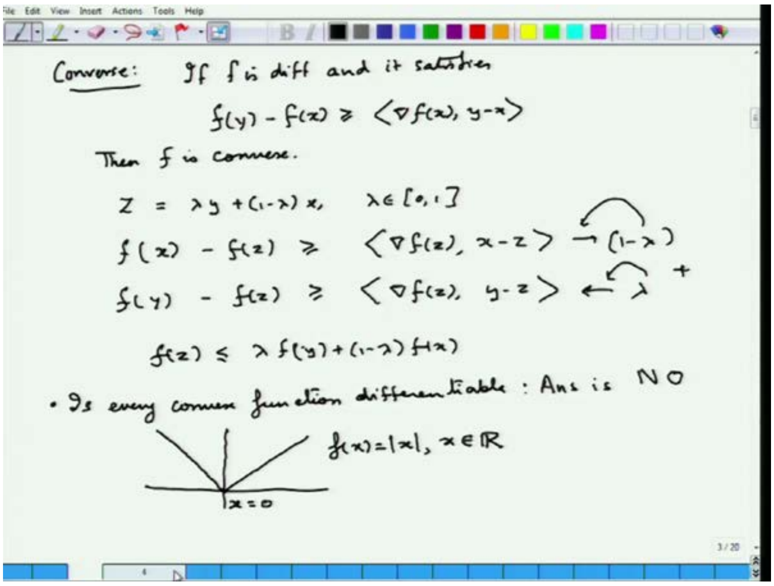
So, with this basic idea let us come down and consider the case when f is from R n to R and f is differential (No audio from 04:08 to 04:17) Then you can characterize convex functions through this you know through the basic notion of a gradient. That is f is convex if and only if f(y) minus f(x) is greater than the gradient times f(x) into y minus x for all (x,y) in R n cross R n. There is for all (x,y) in R n. Now, I guess I have already prove this fact once move once earlier, but let me just recollect and do the prove. The prove of this fact is absolutely simple. It relies on the notion of convexity. So, if function is convex you write lambda y plus 1 minus lambda x is lambda times f(y) plus 1 minus lambda times f(x) for all lambda time between 0 and 1. Of course, it is including 0 and 1, but ok. So, of course, 2 for every lambda between 0 and 1.

So, now you get rearrange this you know and write this as a f of x plus lambda times y minus x from that less than lambda f(y). Now, if you look at this one, then transfer one effects to the other sites. So, if it f of x plus lambda y minus x minus f(x) is less than lambda times f(y) minus f(x). Now, differentiability would have again allow us to expand this thing in the form of a Taylor's theorem and that would need to grad of f(x) lambda y minus x plus small o of lambda is less than lambda times f(y) minus f(x). So, this would imply that grad of f(x) lambda times y minus x plus… So… So, it will divide by lambda. Thus lambda is between 0 and 1, we can divide by lambda and we can have this.

(No audio from 07:12 to 07:25)

Now, this will immediately show us something. It will immediately show us lambda goes to 0 as lambda is positive and goes to 0, this will goes to 0, it will imply that f(y) minus f(x) is greater than equal to gradient of f(x) times y minus x right.

(Refer Slide Time: 08:04)



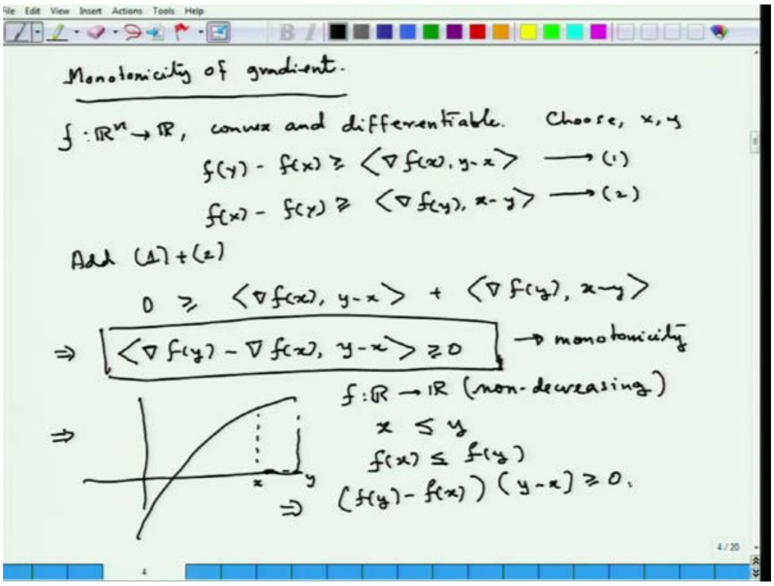
Now, the question just look at the converse. If f is differentiable and it is satisfies following. (No audio from 08:19 to 08:34) Which satisfies the following, then f is convex. So, I would prove it, but I am just going to give a hint of the prove, hint is as follows. Considered z or maybe I will just do the prove; considered z as lambda y plus 1 minus lambda x that lambda is something number between 0 and 1. Then if this is what it is true and f(x) minus f(z) is greater than gradient of f(z) into x minus z. Similarly, you can have f(y) minus f(z) is greater than equal to gradient of f(z) into y minus z. See the job would be do add this, multiply this with 1 minus lambda, multiply this with 1 plus lambda. And once you do this when you add up you will simply get that f(z) is less than equal to lambda times f(y) plus 1 minus lambda times f(x).

Now… So, this operation was 1 minus lambda into this equation and lambda into this equation; 1 minus… So, multiply with 1 minus lambda and here multiply with lambda and then add up this, to get this result, and z you know is what that is convexity exactly. Now, the question is, is every convex function differentiable? (No audio from 10:38 to 10:52) So, how do I answer this question? Answer is answer is no, because just look at the most well-known convex function which does not have a derivative. That is f(x) is equal to the absolute value of x when x is in R and at x equal to 0 is the point where the function does not have a derivative and also it is the point where function attends a minimum.

And in fact most function which are differentiable attends a minimum only at the most convex function which is not differentiable precisely attends the minimum, in the place where the derivative does not exist. So, this property of convex functions which are not differentiable for property of minimum been attend at non differentiable points is a generic property. And it is work for all convex function one of is almost all. That is the meaning of general city in a way used of, but...

So, the question is, what happens if I do not have differentiability? That question we will answer slightly letter, but let us show that the convex functions can also be characterized through the notion of gradients through a interesting property called monotonicity. And currently there is a huge study on the relation between monotonicity and convex functions. It is a huge resist topic and it is worth wise exploring.

(Refer Slide Time: 12:51)

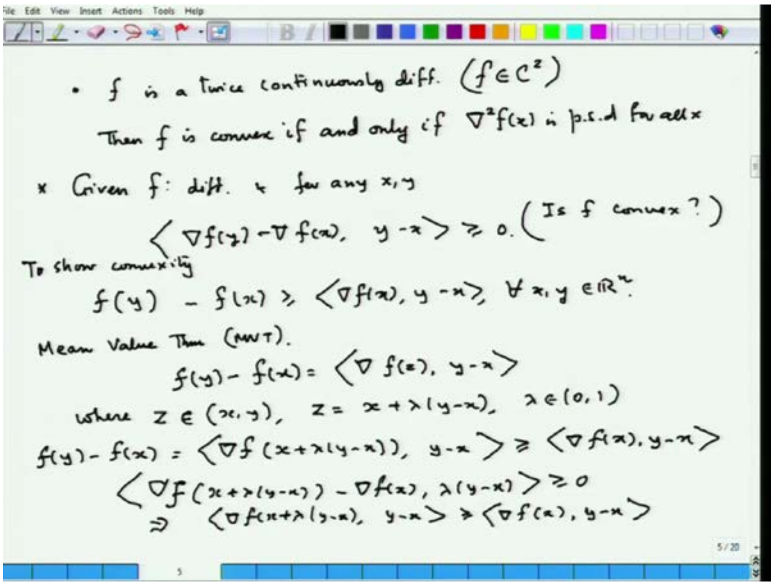


So, let us look at the notion of monotonicity of gradients. (No audio from 13:53 to 13:03) So, look at the convex function, take a convex function f from R n to R is convex and differentiable. (No audio from 13:15 to 13:25) Now, then I can take it to an into (x,y), let me choose an into (x,y). What (( )) you want? Then I can write f(y) minus f(x) is gradient of f(x) into y minus x and then I can also write just by swapping the positions are x and y; this inequalities also true, because in state inequalities true for any (( )) (x,y).

So, if I call this equation 1, when I call this equation inequality 2. So, if I add up 1 plus 2 what I get is 0 is bigger than grad of f(x) into y minus x plus grad of a f(y) into x minus y. So, this would immediately show that a grad of f(y) minus grad of f(x) into y minus x is greater than equal to 0, this is this is the monotonicity property of the gradient. This property is called monotonicity. Now, we must also show that why it is called monotone property, because we dealing with that functions which are vector functions. Note that you take any function which is increasing. Say is just in R to R, take a function from R to R, need not be convex. Here I would not a convex function, is increasing. See increasing or non-decreasing basically. Take a f, f is non-decreasing; More fashionable it now; Increasing is what would one call strictly increasing.

This means that you have a x lesser than y, f(x) must also we less or than f(y). If you look at this, this would imply that f(y) minus f(x) is non-negative. So, is y minus x. So, the product is also non-negative. So, this is actually generalize to this set up, because here we have vector function. So, instead of multiplication we have inner product as we know inner product is a generalization of the notion of multiplication in vector spaces. So, this… That is y it is called monotonicity property, this is coming from the increasing idea.

(Refer Slide Time: 16:44)



Now, of course, we have already prove this fact which we are not going to prove that if f is a twice continuously differentiable convex function. So, twice continuously differentiable that is symbolically f can be written to be in c 2 - twice continuous differential. Then f is convex. So, this is one I am repeating, I have already proved it. If and only if (( )) matrix is positive semi definite for all x. And you might be q s, I would ask this question. The gradient satisfies this property, what about a differentiable functions whose gradient satisfies this property. Is it a convex function? That is the question. There is figure it out.

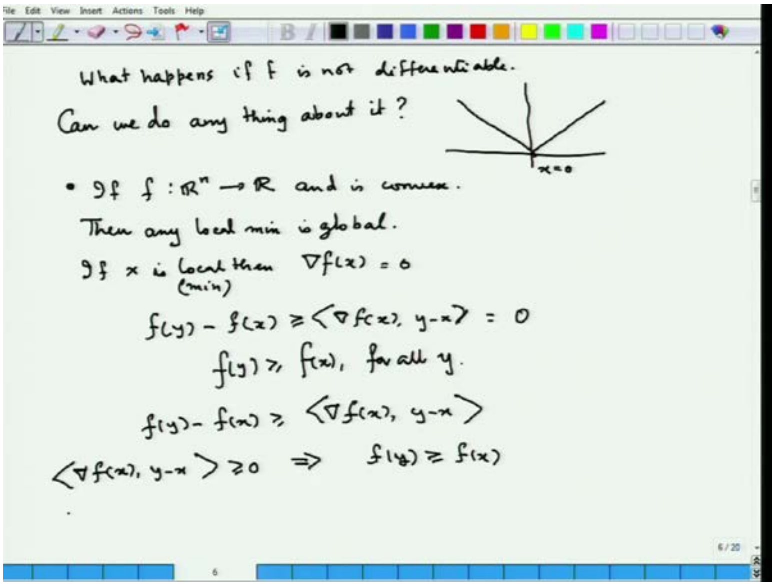
So now, I am given f differentiable and for any (x,y). (No audio from 18:03 to 18:15) This is given to me. Now, let me look into let me look into this. See what I have to do to prove that a differentiable function is convex, is to show that there is my job now to show that f is convex. So, my question is is, is f convex? That is my question. So, to convexity I must show something like this. Of course, because the function is differentiable; is enough to show that this is true. Because this is for any pair (x,y). So, this is of this result would be for any pair (x,y). This is what I have to show.

Now, my question of course is can I show that. To do this, I will use the mean value theorem. Now, what does MVT does? So, if you remember for function are R n to R the MVT's where z and is an element of the open line segment between the x and y that is z is written as x plus lambda pi minus x while lambda is strictly between 0 and 1. So, there is it is some lambda. So, there is a lambda between 0 and 1 such that the z can be written like this.

If I write it like this, I would have immediately f(y) minus f(x) is equal to the gradient of z is x plus lambda y minus x into y minus x. So, this is nothing but greater than. So, if I subtract out x. So, now let let me look into this. Now, I use this monotonicity property, fairly well, what I will do is, I will prove the following I mean I will see that. See look at this f of the gradient of f(x) plus lambda y minus x minus a gradient of f(x) into the difference lambda y minus x; this by the monotonicity property has to be greater than equal to 0 again. I can obviously take the lambda out and divide by lambda, because lambda is between 0 and 1. So, I would immediately have the fact that grad of f(x) plus lambda y minus x into y minus x.

The inner product is bigger then grad of f(x) into y minus x. So now, again just sleep it up and show that this, is nothing but greater than grad of f(x) into y minus x. So, what does it show, f(y) minus f(x) is bigger then grad of f(x) into y minus x. This is exactly this fact, but since x and y an arbitrary. So, this is to for any (x,y) thus showing that the function is convex.

(Refer Slide Time: 22:17)

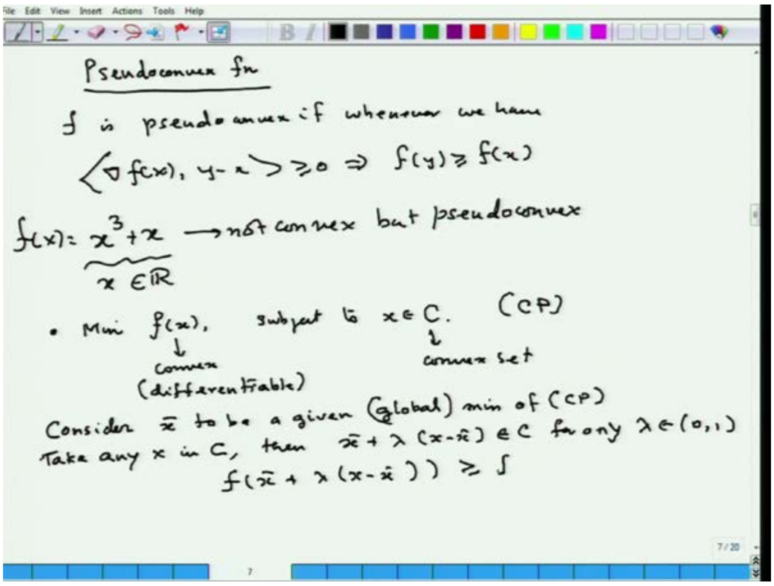


Now, in the next question is, what happens if f is not differential? Can we do anything about it. (No audio from 23:40 to 23:53) Just take the diagram of this and try to play with it for while, as I go on speaking of bit more. So, what how much I can play with the derivative from an optimization point of view. That is ok. Now, if f is… I will talk about this after some times. So, I give you already time to play a play around. Take this non differential function and see that, because here there is no derivative, you standard f dash x equal to 0 type technique want (( )) while trying to figure out the minimum. So, which means that there must be some way by which one can figure out the minimum. So, if that is so, then what should be that way?

So, can there be something which is replace the derivative in this scenario. I will give you some time to thing can play around. But then let me tell you something bit more about differentiability of convex function optimization. So, if f is some are in toward and is convex then any local min is global. For example… So, if x is local then we have proved then grad of f(x) is equal to 0 is a is a local min so. Then by using the convexity inequality for any y convex fix the any other y you can always write this. Now, once you plug in grad f(x) equal to 0 that will give you see there will show that f(y) for all x sorry for all y. So, showing that x is a real minimum.

Now, this also triggers this expression f(y) minus x is greater than equal to grad f (x) into y minus x. This also triggers. So, slide generalization of the notion of convexity. Because if you observe, if I take this. (No audio from 25:30 to 25:37) Then what do I see? I see that f(x). (No audio from 25:42 to 25:52) So, whenever this is greater than 0, f(y) is greater than equal to f(x).

(Refer Slide Time: 26:03)

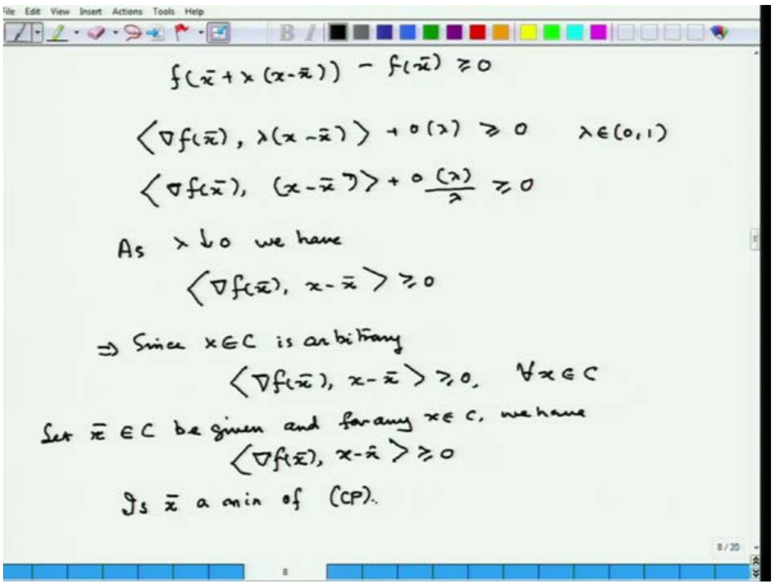


So, this would imply then definition of the so called pseudo convex function. Every convex function naturally would be pseudo convex, but the pseudo convex function is not convex. So, let me just give this. This is just for slide d 2, but we will just was not bother about this function is a new. f is pseudo convex, if whenever we have grad of f(x) into y minus x is greater than 0, we should imply f(y) bigger than x. For example, if you take the function f(x) equal to x cube plus x. This is not convex. Of course, x is here is in R. This is not convex, but pseudo convex from the mid seventies to immediate is also even late in the nineties and some in the current decade. They have been vast amount of work on pseudo convex functions and how they are apply to areas like mathematical, economics have been explode.

So, now once we know a bit about this going beyond convexity, but we less scale back or convexity, because it is not easy to detect functions in R n to R which of this form. So, we go back to our standard convex function again. In here, now we talk about this problem; minimize effects subject to x element of C, which means that here I considered f to be a convex function and C to a close convex set or just a convex set. And I will consider f is a convex and also differentiable. Now, the question is, can I write a necessary and sufficient optimality conditions? Here we have observed that every local minimize global. So, once you have local minimum the necessary condition is grad of f(x) equal to 0. This is necessary. But one the function is convex, it is also sufficient; because once I have this I know that this corresponding x is the optimum.

So, similarly I am coming to this question. So now, if I take global minimum of this problem, because of convex function over convex set, if it has to be minimized every local minimize global that you already know. So, is the any characterization in terms of the gradient and terms of the elements of C, which would be always giving me and necessary and sufficient condition. How to do this is fact? So now, begin by considering consider x bar to be a global minimum of this problem which is the standard convex programming problem CP; x bar to be a given global minimum CP. I am writing global in the bracket, because it is always global minimum. Now, take any x in c, then x bar plus lambda x minus x bar is element of c for any lambda which is between 0 and 1 strictly lying between 0 and 1. We have obviously, any lambda between 0 and 1. Now, because x in c and (x,y) is a global minima.

(Refer Slide Time: 30:54)

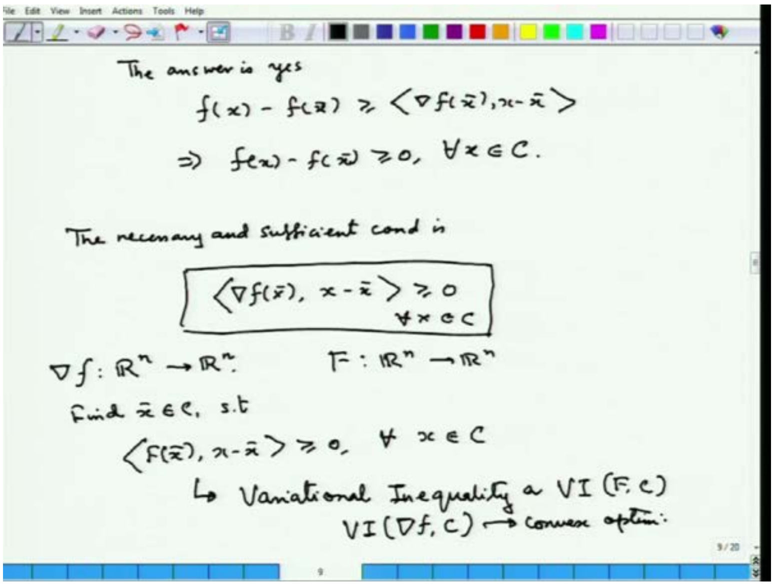


So, I would have f of x bar plus lambda x minus x bar to be bigger than f(x) bar right. Once I have this I can repeat that thing f of x bar thus lambda x minus x bar. So here, from here I will take it to this side to write this as. Now, once I know that Taylor's for the function is differentiable, the Taylor's you have would be invoke for the (( )) other definition of differentiability would be invoked to give me this expression. (No audio from 31:24 to 31:36) Now, I have taken lambda, this is true for whatever lambda you take between 0 and 1.

So, I can now divide by lambda on both side, because lambda is a positive quantity. So, I beyond dividing by lambda should lead to a equation on this form. So, as lambda goes to 0, we have… (No audio from 32:11 to 32:21) Now, you see this is true for any arbitrary x of c I have chosen, I have not chosen any particular x in c of particular structure just an x and c. So, this would imply since x is arbitrary, x element of c is arbitrary, this evoke condition. this condition though looks analytic is actually geometric condition, but we are not in a position to tell you what is that geometry.

So, the beauty of optimization of convex optimization lies in the interaction between analysis and geometry - analytic notions and geometric notions. And one hand and the beautiful interplay of matrices and optimization. This are the hallmarks of convex convex optimization and it is to and it is truly exiting. Anyone who wants to enter the field, this is one of the high times and because of the field is going at a very fast rate and lot of exiting thinks are coming out. And now the question is, so I have a necessary condition, I have to taken x bar to be a global minimum and then I figure out this is what x bar should satisfy.

(Refer Slide Time: 34:33)

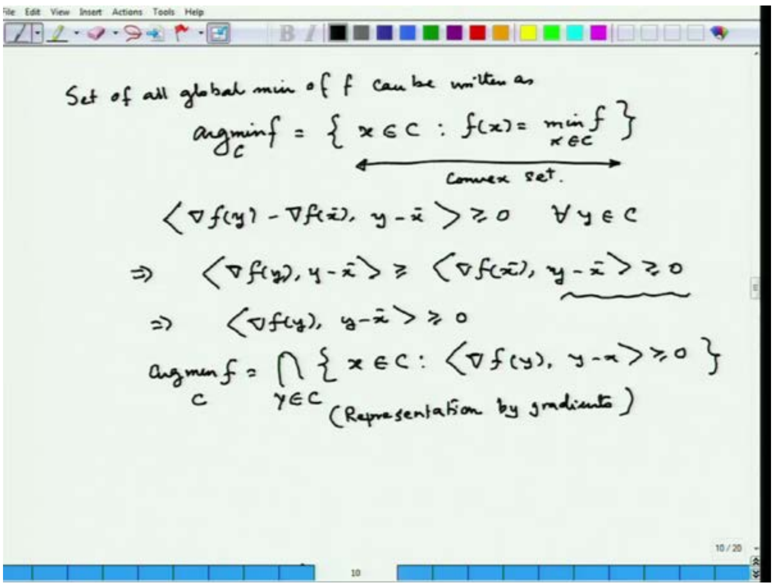


Now, what about sufficient sufficiency; let x bar element of c be given, let x bar element c be given and for any x in c we have. So, this is what, is the reverse we are asking for. Question is, is x bar minimum of CP? The answer is yes, because you can immediately see. Because then you know f(x) minus f of x bar by differentiability of the convex function. So, I have just said that this is given to be greater than equal to 0 for all x in c; so, which could immediately imply that f(x). So, the necessary and sufficient condition is the following.

(No audio from 35:13 to 35:37)

So, x bar is a global minima of convex programming problem CP if and only if, this holds for all experiment of c. So, this particular sort of representation of a necessary and sufficient optimality condition for a convex minimization problem - minimizing a convex problem over convex set c is usually told, usually termed as representation y of variation on inequality. Now, this probe in the… You see grad f, if you observe of vector function from R n to R m. Now, inside of this, if I choose some arbitrary vector function from R n to R n, and I post the question find x bar element of c such that, c is convex the same c, such that grad of f(x) bar x minus x bar is greater than equal to 0 for all x in c. So, this is what is called a variational inequality or VI problem, usually denoted as VI F and C. So, over convex programming optimization problem can be also written as a VI grad f c; so, this nothing but over convex optimization problem.

(Refer Slide Time: 37:35)



You know, if I have convex optimization problem, so the set of all global minima of f set of all global min of f can be written as or usually written as argmin f, if th argmin at which means x values for which attains the minimum. Argmin f - the set of all x in c such that f(x) is equal to min of f, it is what is call argmin f. Sometimes we will denote it by c to say that ok and. So, this is set of all global minimum. Now, of course, when f is convex this is the convex set. How will you prove this? A very… Of course, you can use convexity and then try to prove this; this is very simple, convexity (( )).

Another way to look at it is that if the function is differential, what would happen? Then suppose x is in c and then you know y differentiability, the we are invoking the monotonicity property, we x bar is the optima that is x bar is an argmin c and this is what I have for all y you see. So, this one immediately imply away the necessary and sufficient optimality condition I know that this grad f(x) bar into y minus x bar, this is greater than equal to 0. This is exactly what we have just; this is exactly what we have just studied this one.

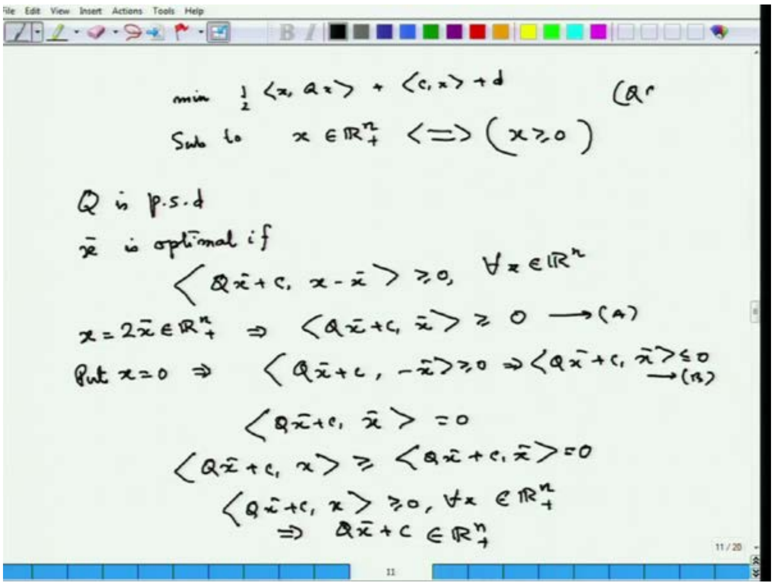
So, which would imply that grad of f(y) into y minus x bar is greater than equal to 0. So, you see this is an alternative way of looking at necessary and sufficient optimality condition. That is a convex function f has a global minima at f if an only if this is true. So, the argmin of f can also be written as it can also be written as grad of y belonging into c.

(No audio from 40:35 to 40:44)

Now, each of this for a fixed y, each of this is a convex set. This is very simple to see this. And so, if you take intersection of arbitrary number convex sets you got convex sets are been f c is convex sets. This is another way to look it. This is an alternative optimality condition. So, this are all interesting areas which one can look into this leads to what is call the minty variation inequality, we will not get into those thing at all. So, this is one alternative. So, you see if a convex function is differentiable I can use the gradient to represent the optimality condition. So, it is representation of through up.

So, if my f - the convex function is differentiable, I can use the gradient itself - gradient of wave itself to represent the set of all global minimizes. Representation, here I am representing it through representation through optimality conditions. So, representation through gradient by gradients not optimality conditions sorry. So, optimality condition can be used in fact to represent the argmin; so, the 2-D. This argmin can has have this representation. So, the two ways of representing.

(Refer Slide Time: 42:17)

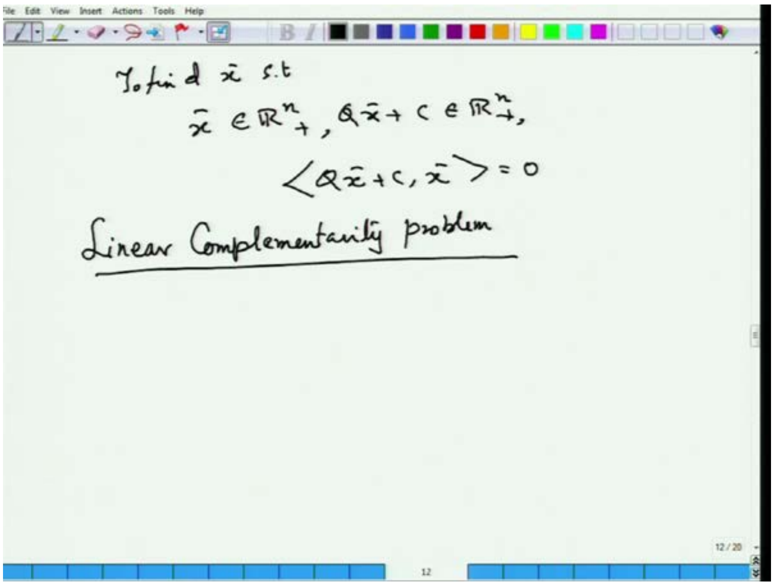


Now, I said something interesting. I will try to employ what we have just learned this optimality condition in one particular case. Considered and we will end the end today's lecture with this. Minimize half… So, I am talking about quadratic programming problem; (No audio from 42:26 to 42:35) x element of R n plus means x is in the cone R n plus that is x is greater than 0. This means that every component of x is greater than or equal to 0. Q is positive semi definite and then this is a convex function which we are minimizing over a convex set R n plus, we see is now R n plus.

Now, x bar is optimal, if the gradient of this at x bar which is Q x bar plus c times x minus x bar is greater than equal to 0 for all x in R n. Now, if you put x is equal to twice of x bar which is obliviously R n plus, there is x bar is optimal solution right; x bar is optimal means x bar is in r n plus, twice of x bar is also in r n plus. So, x is equal to twice of x convex set is particular x. So, this would imply Q of x bar plus c times x bar is greater than equal to 0. Now, put x equal to 0. So, from this equation it will imply that Q of x bar plus c putting I am plugging in first I plugged in 2 x bar here and then plugging in 0 here, I will have minus x bar greater than equal to 0 which implies Q of x bar plus c into x bar is less than equal to 0. So, if I call this as A and call this as B; so, A and B combine will give me Q of x bar plus c for x bar is equal to 0.

Now, what do I get from here immediately from this condition which will immediately show me Q x bar plus c x is greater than equal to Q x bar sorry plus c into x bar. Q x bar plus c into x bar is nothing but this is equal to 0 which is already know. So, what you get Q x bar plus c x is greater than equal 0 for all x in R n plus. So, this is the 2 for any x, because we just braking the inner product and pulling one to the other. So, this would immediately imply, because this is true with the every x R n plus, Q of x bar plus c is also in R n plus.

(Refer Slide Time: 45:35)



So, process of finding a minimum this quadratic programming problem or QCP - quadratic convex problem is to find x bar such that x bar should belong to R n plus Q x bar plus c should also belong to R n plus and Q x bar plus c in a product with x bar is equal to 0. This is called the complementarity condition. So, this problem is called the linear complementarity problem and has huge applications in many different areas and is very well studied and still people are looking into this.

So, you see that just from that simple optimality condition, we have developed; we have got some quite new and interesting information. So, with this we stop here and tomorrow we will talk a bit more about what happens when the convex function fails to be differentiable, a question which are asked you. So, you till the next lecture you just thing how will you handle the situation and then we can speak about it in the next class.