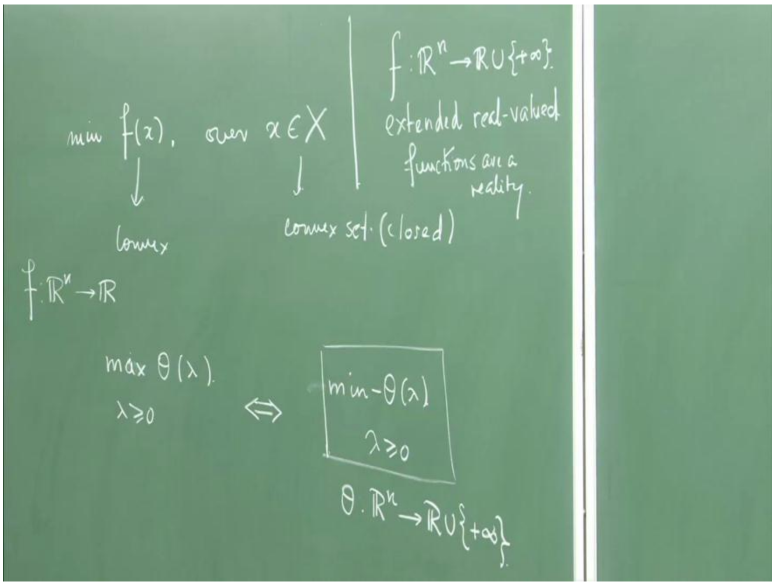
**Foundation of Optimization**
**Prof. Dr. Joydeep Dutta**
**Department of Mathematics and Statistics**
**Indian Institute of Technology, Kanpur**

# Lecture - 34
# Optimization

Do not worry about the dissents here and my looks possibly, because I am giving the lecture at the end of the day of one particular day, which of course, is a continuation of the lectures I was giving. We have just been able to establish at the dual is a possibly a nice idea in optimization. In the sense that it can provide at least in the convex case under certain constant qualification like the Slater constant qualification essentially, that the primal value and the dual value are equal, this is precisely very important linear programming.

Because what happens if the large number of constants in linear programming in the dual the number of variables would increase constants would decrease or vice versa. So whichever you are comfortable in computing, you just set the problem accordingly. So, there is hardly much ever difference in linear programming, whether you are solving the dual or you are solving the primal.

(Refer Slide Time: 01:30)



$\min f(x)$, over $x \in X$
convex
convex set (closed)
$f : \mathbb{R}^n \to \mathbb{R}$
$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$
extended real-valued functions are a reality.
$\max \theta(\lambda)$, $\lambda \geq 0$
$\Leftrightarrow$
$\min -\theta(\lambda)$, $\lambda \geq 0$
$\theta : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

The question of having the dual horse back to two important issues, one is that it sends back sends us back to investigating the question of minimizing effects over x element of some convex set x. So, minimizing a convex function may be it we can take a closed convex set, but here there is a word of (( )) standard to what we were doing of always assuming functions running from R n to R that is taking a function taking of a point in x and pushing it on point point on R.

Let me be very precise that if you look at the structure of the dual problem, which says that the minimum, this is my dual problem. And theta lambda sorry sorry max of which I can write as this problem can be equivalently posed as so, you can forget the minus sign and just consider on this problem. Maximizer of this if it is achieved the minimizer of this if it is achieved they, will be achieved at same points.

So, now this is a convex function minimizing over a closed convex set, but there is a cracks here theta is not essentially from R n to R. But theta in the sense, because it can take this particular value, because theta lambda would be minus infinity, because you find, find it out to minimization. So, minus theta lambda can take plus infinity value. So, essentially a general problem here my f has to be but you know that just having plus infinity values, we are admitting the fact that the plus infinity values do come like the ones here and a you cannot necessarily throw this fact out.

So, one has to get use in optimization of the fact that extended real valued functions these are called extended real valued functions are a reality. So, these are called the extended real valued functions and these are a reality and you really need to appreciate these things, if you really want to work with optimization. Now, what is important to know about how to play with infinity that we may restrict ourselves for certain time the only thing that we need to know, if you have a convex function of this form that what is 0 into infinity.

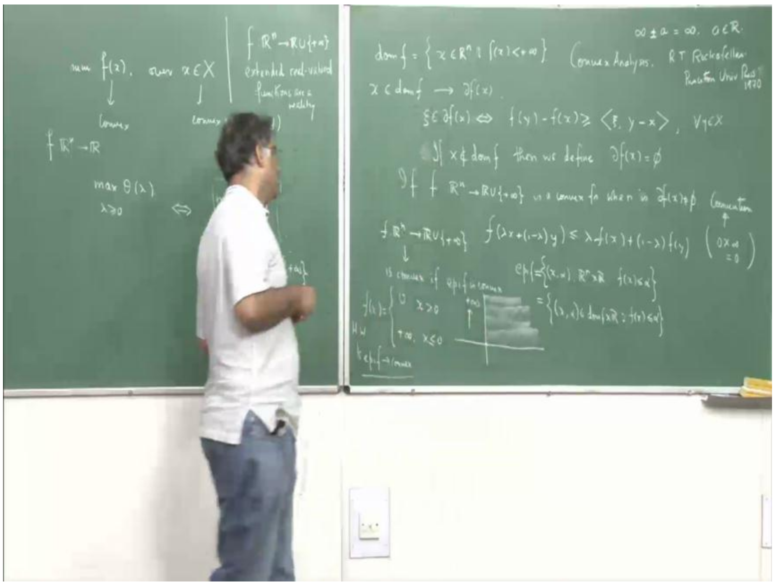
So, 0 into infinity is usually assumed to be 0 as per the a, as per the structure taken by Docofolar and Widch which is possibly on the finest epiphyses on which modern optimization is been built. See, we have to remember and I would also like to tell the engineers who are engineering students who would be watching this course. Then apart from learning the fundamentals of algorithms, one has to appreciate that optimization that you are taught largely in Indian universities largely not everywhere is of the 1960’s being teached.

But. one has to understand like any other part of mathematics optimization also has a life of it is own and it has progressed a lot it is answers questions which are pretty important;

and we would see that actually not ready in 1960’s been treasured but, actually the stuff. So, 19 late 40’s early 50’s vintage and in India those things came in 1960’s and the similar sort of things are been taught we will hardly find extended a route functions we really talked about in graduate courses in optimization if there is any in India.

So, as a result of this that leads me to actually emphasize that we have to appreciate that optimization as a subject has advanced. An advanced quite a bit not quite a bit heavily rather and here we do not really have a chance to talk about the advancement that optimization has done but, we we are really trying to tell you the glimpses and bits of the most interesting things in optimization. Now, the question is if this is my scenario, how do I, how do I actually think of discussing the solution of this problem; obviously, the points at which x is equal to infinity, you really do not need to bother about that.

(Refer Slide Time: 07:35)



min f(x). over x ∈ X
convex
convex
f: ℝⁿ → ℝ
max θ(λ)
λ ≥ 0
⇔
f: ℝⁿ → ℝ ∪ {+∞}
extended real-valued functions are a reality
∞ ± a = ∞, a ∈ ℝ
dom f = { x ∈ ℝⁿ : f(x) < +∞ }
Convex Analysis. R T Rockafellar
Princeton Univ Press 1970
x ∈ dom f → ∂f(x)
ξ ∈ ∂f(x) ⇔ f(y) − f(x) ≥ ⟨ξ, y − x⟩, ∀y ∈ X
If x ∉ dom f then we define ∂f(x) = ϕ
If f: ℝⁿ → ℝ ∪ {+∞} is a convex fn when is ∂f(x) ≠ ϕ
Convention
f: ℝⁿ → ℝ ∪ {+∞}   f(λx + (1−λ)y) ≤ λf(x) + (1−λ)f(y)
(0 × ∞ = 0)
is convex if epi f is convex
epi f = {(x, α): ℝⁿ × ℝ : f(x) ≤ α}
= {(x, α) ∈ dom f × ℝ : f(x) ≤ α}
f(x) = { 0, x > 0 ; +∞, x ≤ 0
+∞
HW
epi f → convex

So, the point of important point for such problems is to know that we can consider our self or put ourselves in a set x in a set of all x where the function value is finite. So, for every such x in dom f I can define the sub differential f of x as what is the sub differential. So, psi belongs to del f x, if and only if f of y minus f of x, now you see f of x is always finite if y is a point where f of y is plus infinity, then you really do not bother because because infinity minus infinity infinity.

Infinity minus of finite number sorry not infinity minus infinity, infinity minus infinity is still something we do not want to talk about at least in this course. So, infinity minus some finite number, so infinity minus a in the axioms of the arithmetic that we do with infinity, infinity minus a is infinity where a is a real number. It does not matter whether a is a positive number or a negative number basically infinity plus minus a is plus infinity. So, if this side is infinite and; obviously, this side is finite and the inequality is valid, if x is not in dom f then we define now at every x point of x element of dom f.

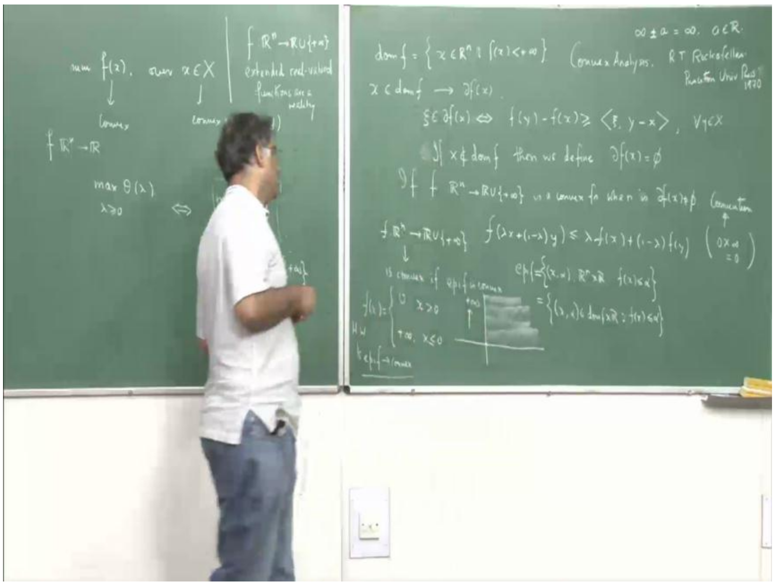
If you have a extended valued functions your dom f need not be empty may be non empty. So, if f from R n is a convex function when is, so here we are coming in to very, very, very deep questions actually which possibly I have not discussed in such detail in my convex optimization course. I had but not there is not much scope, because there also you need to cover a lot of, lot of thing linear programming a lot of lot of space had been given to that.

Now, a very, very important thing to realize at this point is a following now the question is how do you define a convex function itself. And you function if you say that the function is extended valued of course, you have to go in for you can write down the same definition and if you have to go down with the same arithmetic rule that 0 into infinity is 0. So, this this is the definition of convex function.

Suppose f y is infinity and these are non zero from there everything is plus infinity. So, this is valid now. So, f x is x is a point where f x is plus infinity and lambda is 0. So, when lambda is 0, I just have y. So, lambda into 0 this part will, so this will be f y and this lambda into an m zero into infinity is infinity is sorry again I apologize 0 into infinity is zero. So, it will just become f y square again the inequality will be valid. So, this is valid under the convention here convention is also used in measured theory a lot of thing in mathematics is usually about convention.

If you read this book mathematics very short introduction by Timothy Gowers that is what he wants to say lot of things about convention. Now, what I want to emphasize is that there is another way of going about defining a convex function from R n to just say f, this f is convex if epi f is convex epigraph. Now, epigraph is only defined about the finite part that is we can define it about the epigraph infinite part but, it do not make much sense, so epi f of a convex function here.

(Refer Slide Time: 014:03)



$\min f(x)$, over $x \in X$

$\downarrow$ convex $\quad \downarrow$ convex

$f: \mathbb{R}^n \to \mathbb{R}$

$\max_{\lambda \geq 0} \theta(\lambda) \quad \Leftrightarrow$

$f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

extended real-valued functions are a reality

$\infty \pm a = \infty, \ a \in \mathbb{R}$

$\text{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$ Convex Analysis. R T Rockafellar Princeton Univ Press 1970

$x \in \text{dom} f \longrightarrow \partial f(x)$

$\xi \in \partial f(x) \Leftrightarrow f(y) - f(x) \geq \langle \xi, y - x \rangle, \ \forall y \in X$

If $x \notin \text{dom} f$ then we define $\partial f(x) = \phi$

If $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex fn then in $\partial f(x) \neq \phi$ Convention

$f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \quad f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \left( 0 \times \infty = 0 \right)$

$\downarrow$ is convex if epi f is convex

$\text{epi} f = \{(x, \alpha) : \mathbb{R}^n \times \mathbb{R} \ f(x) \leq \alpha\} = \{(x, \alpha) \in \text{dom} f \times \mathbb{R} : f(x) \leq \alpha\}$

$f(x) = \begin{cases} 0 & x > 0 \\ +\infty, & x \leq 0 \end{cases}$

HW

Is epi f → convex

You are taking basically x alpha where x is in R n and alpha is in r such that f x is less than equal to alpha. So, this value has to be finite, so it essentially it means epigraph sorry it means. So, if the epigraphical set is convex, then you know that the function is convex for example, if you define a convex function as 0, when x is greater than or equal to 0 is equal to plus infinity plus the epigraph this is epigraph as beyond this, this is plus infinity. So, if the epigraph is a convex set then the function is a convex function, so this function is a convex function.

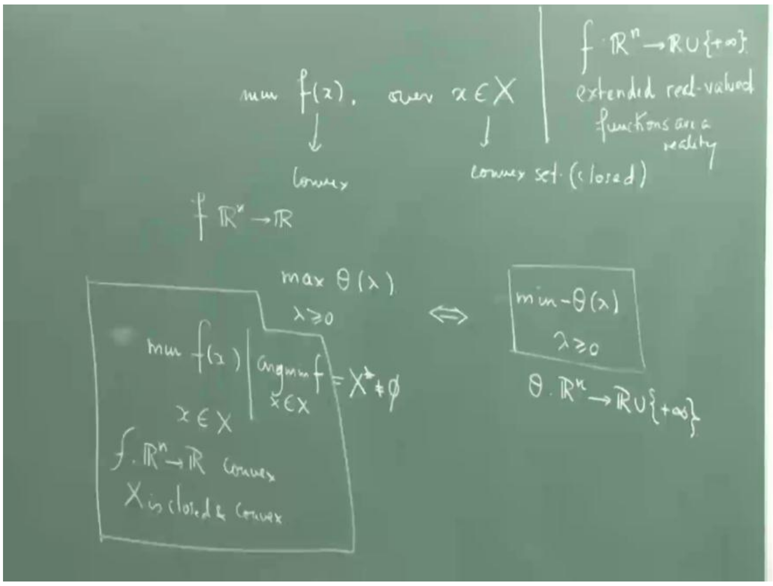
If you even, suppose I do a little bit this strike my question as a homework to you is is epi f convex. Now, what is very, very important to admit that if I really want to go into a depths then I have to talk about the lower semi continuity. And all other issues which might be beyond the scope of understanding on many people who are watching the this course specially from the engineering stream.

But, in the engineering stream you are also faced with not only on not only on need not always bother about the Lagrangian duality you are also faced with the problem of minimizing this over this, where f is convex and this is a convex set. But, f is a finite valued function this is a very common problem but, it might be that your f now is not differentiable. So, I am making the problem not problem slightly simpler.

So, we are really not going to address and try to solve or write an algorithm of how to solve this problem which is of a function of this form, with the same extent valid form which has all complications. You will see the in declasses that are coming in and this actually needs a training in mathematics to appreciate what is, what is going on here I would I would not get into the things.

But, I would rather refer to you the book of con those who are interested to know more about it I would rather refer to you a book of convex analysis by R F Rockefeller it is Princeton university press 1970. And is a landmark book it was reissued in 1974 I guess that is Princeton landmarks in mathematics 1970. So, this is a book; obviously, this book cannot be read from cover to cover what wherever you need some information of a convex analysis. You just need to go and look at this book every researcher what is solved in optimization needs this book.

(Refer Slide Time: 17:42)



$\min f(x)$, over $x \in X$

convex — convex set (closed)

$f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

extended real-valued functions are a reality

$f: \mathbb{R}^n \to \mathbb{R}$

$\max \theta(\lambda)$

$\lambda \geq 0$

$\Leftrightarrow$

$\min -\theta(\lambda)$

$\lambda \geq 0$

$\theta: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

$\min f(x)$

$x \in X$

$\underset{x \in X}{\operatorname{argmin}} f = X^* \neq \phi$

$f: \mathbb{R}^n \to \mathbb{R}$ convex

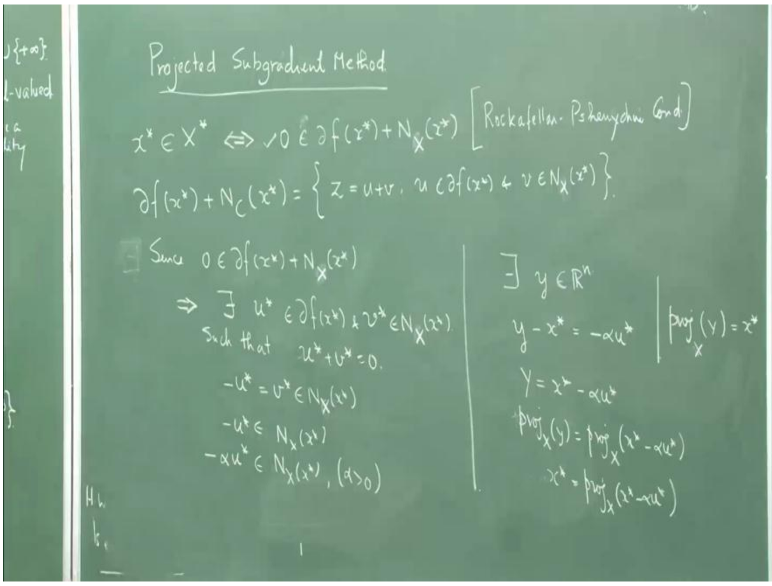
$X$ is closed & convex

So in rather than scaring you we will now, like to talk about how I can develop an algorithm for solving this problem at this is f is R n to r and convex x is closed and convex. And we will assume that a aargmin means the solution set it is usually denoted like this, this set which let us denote this by x star this is not equal to phi let it as a solution at least from the solution, so the minimizer exist.

Now, the question would be how to find a minimizer how would you go about doing this now you see here we have not mentioned. What is the differentiability aspect of this function we have not mentioned anything about differentiability the function may not be differentiable. And usually non differentiability would be exactly at the usually at the solution points now this require a generic property of convex function.

Now, once that is done then it seems that we can do something which even engineering students can appreciate and actually implement them in their work. Now, this would again bring forth the usefulness of the projection operated the projection mapping. So, which we will now write down which will call the projected sub gradient method and we will try to talk about the projected sub gradient method first write down what should be the projected sub gradient method and then I will start analyzing it.

(Refer Slide Time: 19:45)



Projected Subgradient Method

$x^* \in X^* \Leftrightarrow 0 \in \partial f(x^*) + N_X(x^*)$ [Rockafellar-Pshenychni Cond]

$\partial f(x^*) + N_C(x^*) = \{ z = u+v,\ u \in \partial f(x^*)\ \&\ v \in N_X(x^*) \}$.

Since $0 \in \partial f(x^*) + N_X(x^*)$

$\Rightarrow \exists\ u^* \in \partial f(x^*)\ \&\ v^* \in N_X(x^*)$

Such that $u^* + v^* = 0$.

$-u^* = v^* \in N_X(x^*)$

$-u^* \in N_X(x^*)$

$-\alpha u^* \in N_X(x^*)$, $(\alpha > 0)$

$\exists\ y \in \mathbb{R}^n$

$y - x^* = -\alpha u^*$ | $\text{proj}_X(y) = x^*$

$y = x^* - \alpha u^*$

$\text{proj}_X(y) = \text{proj}_X(x^* - \alpha u^*)$

$x^* = \text{proj}_X(x^* - \alpha u^*)$

So, we do not have gradient we have already talked about gradient projected gradient method now we are talking about projected sub gradient method and you have you already know, what is a sub gradient? See only knowing the very basic definition you could understand the quiet involved things, now if the problem is not differentiable then x star. So, I will call this problem c p again, so x star is an element of x star that is a solution set if and only if.

So, this is a Rockefeller (( )) condition we have already discussed about the normal cone it is called the Rockefeller (( )) condition. So, this is true we had also discussed about this and this fact is true. So, we had discussed in a quiet a wage, way the problem is that the analysis of this is quiet involved and we have to use separation theorem and all those details have not been a part of this course in, so much of detail it. You need to have a graduate course in optimization to go through all this, because this is largely a course where you mix up some undergrad stuff largely undergrad stuff with some grad stuff.

So, basically you are bringing in undergrad students to the grad level and you could possibly now try to do some analysis out of this condition. So, what does this mean, this means the two sets and zero must belong to their Minkowski addition. So, these are these any element of this particular set the sum of these two sets is that you take an element of vector from here and a vector from here and you add those two vectors, say make all such possible combinations to form the set.

So, basically if you if you are still not comfortable with this writing but, I am still may writing this for you also those who are just opt into the program late is that this is this consists of all z given in the form of u plus v is at u. So, if 0 belongs to the set, so which means there exist if, so this is true if this is true then since zero is a element of Minkowski addition on this two sets. See, today we are doing some advanced task as I told you we have made maximum grad stuff graduate stuff is now undergraduate stuff a last part of the course was essentially undergrad but, this part is grad.

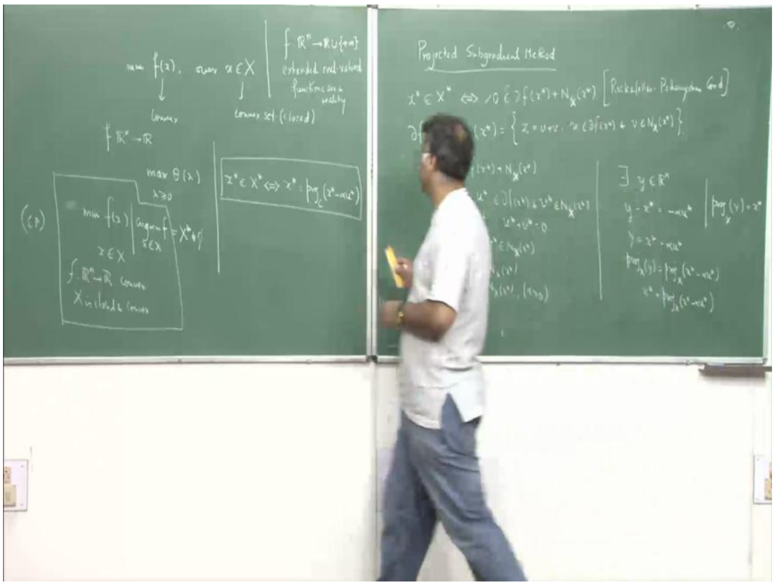
Once this projected sub gradient method ends we will basically do pot (()) of the or have the last two lectures as miscellaneous. So, we would talk about a lot of things in optimization giving a very brief idea on which you actually go on and expand. So, now you see this is what you have, so which means is implies such from these definition there exist for ze your 0 is now the z there exist u star element of del f x star and v star element of n c x star such that u star plus v star equals to 0.

Now, once you know that you immediately know that minus of u star is equal to v star where v star is itself an element of the normal cone to s s x at x star. So, I am inputting c oh sorry I think, I made a mistake it should be all x I am sure you can correct me on this, because we have taken x element of x. Possibly I am, so habitually writing c I will just put an x does not matter if you put. Now, once this is done what, what does this mean which means minus u star is element of N x x star now take any alpha which is positive and by very definition of the cone, the alpha gets absorbed in the cone. So, minus alpha, alpha minus u star is also an element of this by the very definition where alpha is bigger than 0. Now, what does it say, it says by the idea of projection it says that there exists some y in R n, such that y minus x star is equal to minus alpha u star, where the projection of y on x is nothing but x star and projection.

So, normal cone this is the nor nor this is the meaning of the normal cone. So, from x star and there is a point outside you have drawn the normal. So, projection of y on x is the point x star go back to your you can go back to the lectures that I have spoken about projection. So, y is nothing but, x star minus alpha u star, so again it means projection of x of y is equal to projection of x of x star minus alpha u star but, what is projection of y, projection of y is nothing but, x star.

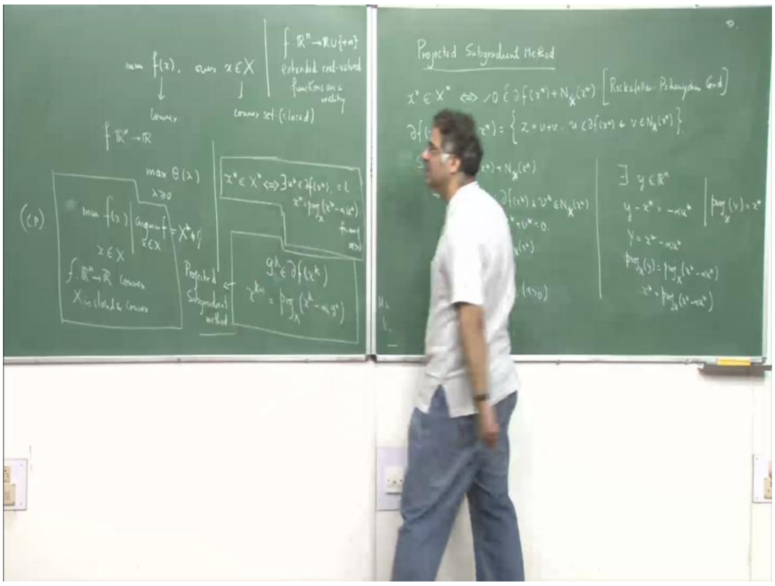
So, the Schenychle Rockefeller condition can be now put into the form, put in terms of the projection mapping and that is why we use algorithm mentally. You might ask x can have various forms and how do I compute the projection map, there is a very simple ways of computing the projection map, because computing a projection as you know is again computing a small optimization problem involving the distant function or square of the distant. But we would show that for very certain standard class of sets you can have here computation of the projection right of the shelf you can just it is, it is it is known. So, we will just tell you that tomorrow or later at not today and some of it can be homework. So, now what we have done, what we have, what we have proved is the following.

(Refer Slide Time: 29:00)



So, now we I have rewritten Rockefeller pshenychne conditioning as follows x star is element of x star, I am only proving one direction you can prove the other one. If and only if that it look right no. you cannot just write this you have to add something more and what is that something more. You have to understand that this story is not true for every u star it is only true for this u star for which u star plus v star is equal to 0.

(Refer Slide Time: 29:56)


Projected Subgradient Method
$x^* \in X^* \iff 0 \in \partial f(x^*) + N_X(x^*)$ [Rockafellar-Pshenychne Cond]
$\exists\, y \in \mathbb{R}^n$
$y - x^* = -\alpha u^*$
$\mathrm{proj}_X(y) = x^*$
$y = x^* - \alpha u^*$
$\mathrm{proj}_X(y) = \mathrm{proj}_X(x^* - \alpha u^*)$
$x^* = \mathrm{proj}_X(x^* - \alpha u^*)$
$g^k \in \partial f(x^k)$
$x^{k+1} = \mathrm{proj}_X(x^k - \alpha_k g^k)$
Projected Subgradient method
$f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$
extended real-valued functions
$f: \mathbb{R}^n \to \mathbb{R}$
convex set (closed)

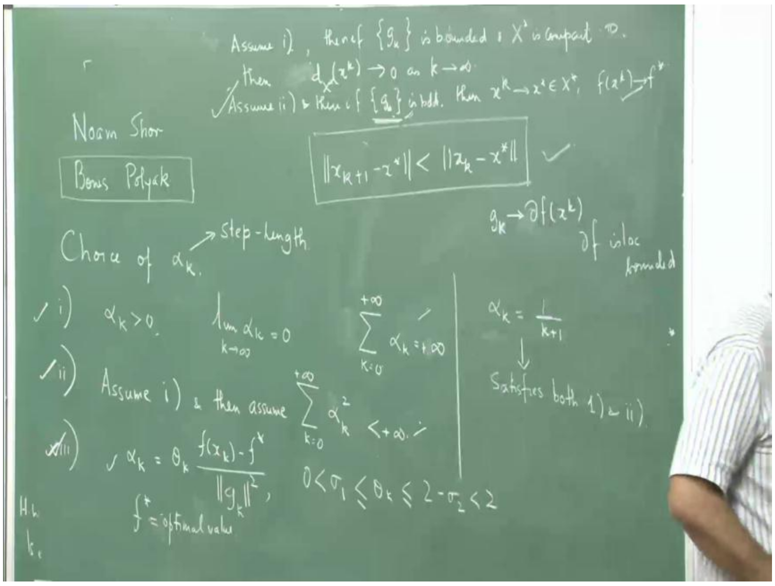
So, what you can say x star is an element of x star if, if and only if there exist u star element of del f x star, such that x star is equal to projection of x, x star minus alpha u star for any alpha greater than 0. So, now the Rockefeller pshenychne condition can be written in the, in this particular format, so this is the way Rockefeller pshenychne condition can be posed. Now, does it ring a bell if you go back and look at the differentiable situation where we have spoken about the projected gradient method does it give you a clue that how could I write down the projected sub gradient method. You see there is no way you can use the negative of a sub gradient to say that that is a direction of descent it is not possible, there are several other approaches one of them is a bundle method. We will not speak about them at all but, may we give you a hint while we are doing that miscellanea optimization in next few classes in few lectures.

So, just look at it the projected sub gradient method now has to include this because there is no gradient but, if there is a gradient if the function is differentiable then your u star is nothing but, grad f x star there is no other u star. So, the only possible u star is this. So, that will be the projected gradient, let me write down the scheme. So, projected sub gradient g k is one of the traditional ways where in optimization liter the algorithm literature the sub gradient had been always referred to.

So, g k is a vector x k where x k is k either it and k plus one either it is a projection on x of x k minus alpha k g k. Remember when alpha is when the when we are writing sequences of real numbers we are writing the subscript below, when we are writing sequences of vectors, we are writing the sequence, so sequence notation above as a superscript.

So, this is what is called the projected sub gradient method. So, this is this is what is called the projected sub gradient iteration method, projected sub gradient method. But, here there is a whole clue the idea is who knows what sort of, so sub gradient would actually satisfies those things I do not know. So, I just take a sub gradient and just put in something put in an alpha pumping. But, here my choice of alpha would become fundamental that is what sort of alpha I have to choose, so that the ideate that is generated from this would actually leave me to the solution. So, now there are several ways of choosing that alpha, now this was done by Russian school of optimization in the mid 60's and early 70's.

(Refer Slide Time: 34:21)



Assume i), then if $\{g_k\}$ is bounded & $X^*$ is compact $\neq \emptyset$.

✓ then $d_{X^*}(x^k) \to 0$ as $k \to \infty$

✓ Assume ii) & then if $\{g_k\}$ is bdd. then $x^k \to x^* \in X^*$; $f(x^k) \to f^*$

Noam Shor

$\boxed{\text{Boris Polyak}}$

$\boxed{\|x_{k+1} - x^*\| < \|x_k - x^*\|}$ ✓

$g_k \to \partial f(x^k)$

$\partial f$ is loc bounded

Choice of $\alpha_k$ → step-length

✓ i) $\alpha_k > 0$. $\lim_{k \to \infty} \alpha_k = 0$ $\sum_{k=0}^{+\infty} \alpha_k = +\infty$ ✓

✓ ii) Assume i) & then assume $\sum_{k=0}^{+\infty} \alpha_k^2 < +\infty$. ✓

iii) ✓ $\alpha_k = \theta_k \frac{f(x_k) - f^*}{\|g_k\|^2}$, $0 < \sigma_1 \leq \theta_k \leq 2 - \sigma_2 < 2$

$f^*$ = optimal value

$\alpha_k = \frac{1}{k+1}$

$\downarrow$

Satisfies both i) & ii)

And they had a huge influence on the development of sub gradient methods for convex optimization with professor Noam Shore who has passed away is who was a leading light in this and we had Boris Polyak one of the two major players in this area. So, we will also name some few more. So, these two people developed this projected sub gradient algorithm apart from other sub gradient algorithm based on the simplest decision type technique, which we will not discuss in this course.

So, here we would essentially concentrate on the approach finally, which Boris Polyak had taken which is a very, very interesting approach. So, how would we choose alpha k. So, our algorithm would really depend or behavior of the algorithm will really depend on the choice of alpha k. So, we would now tell you the choice of alpha k alpha k is of course, positive number that you know may be I should be more eddied and precise rather putting alpha k.

In the first case I would choose alpha k strictly bigger than zero I would want the limit of this sequence of alpha k as k goes to infinity to be 0 I would want, at the same time I would want the sum of alpha k is sorry is plus infinity then it diverges. So, it is plus infinity the second rule says assume. So, this is and this alpha k is again called the step plain, alpha k in the same way what we have done in the line search method for un constant case. See, here we are talking about constant case and the projection is taking care of the constants, because the projections on x will always give you a point in x. So, that is taking care of the constants and that is very, very important to note.

So, assume one that is assume all this and then assume one to infinity or zero to infinity does not matter this is finite. So, this infinite series does not converge this infinite series converges but, here limit alpha k is equal to 0, you can immediately understand what sequence alpha k is a model of this, because if I take alpha k to be 1 by k plus 1. See if it is started with k equal to 1, then you could have just taken k then it is clear that this is holding and it is clear that this is, this some sort of harmonic services part of a a. So, this is and also summation 1 by k plus 1 whole square will go like that k k plus one whole square every time you take the square now k then k plus 2 and so and so; k is here 1 my k plus next 1 would be k plus 1 k is 1 then it is k plus 2.

So, instead of having the sequence one by k I am having the sequence 1 by k plus 1. So, if you look at this sequence this one satisfies both 1 and 2, the third one. So, this is this are very im these are you see the type of strange assumptions you you might feel it absolutely strange to assume such a thing keeping one by n as 1 by k as a model at this. So, surprising of sort of step length choice actually leads you to the solution and that that is what we are going to understand assumptions but, the most important choice was given by this.

Where you have chosen some sigma 1 and sigma 2 and so, you choose theta can like this that is fine you can choose sigma 1 and sigma 2. But, though theoretically this is very, very powerful you see what happen the see, the problem is that once you make a jump from smooth to non smooth there is a huge paradise shift. Here you are assuming f star because you are assuming that the problem has a solution but, in reality if you because the problem has a solution theoretically I know f star but, actually I do not know f star I am trying to find that f star actually I will never find.

Then how would I actually calculate alpha k what would be my f star that that is is a major question. So, this is a very, very important thing that this will show this humid as a whole we do not know the f star what I am trying to what sort of alpha k you are producing. See, instead of f star you can replace it by some lower bound you can make some guess work of the loser bound by nature of the problem etcetera. And then put f star but, what it shows theoretically is that as we will show tomorrow we will start discussing convergence analysis.

We will show tomorrow that the sequence of iterates move towards the global minima not based on the fact that we have to reduce the function value just like in other algorithms it would decrease the function value. But rather than it is based on the fact that we have to decrease the distance of the iterates from the objective value that if this statement is followed. Then the distance of x k from x k plus one from the actual solution from a solution x star is strictly less than it is distance, distance of x star from x k.

So, it is something like that this is this is something that will happen. So, if you follow the Boris as Boris Polyak step length this is what is going to happen now what happens if you assume this is what we will prove. Now, what happens if we assume this what happens if we assume this and what happens if we assume this, if we assume this then we have to assume.

Suppose we assume one assume the step length choice one then if this sequence of g k in that that is bounded and x star is compact then the distance function you take x k. And try to find it is distance from x star this would go to 0 as k tends to infinity. Now, what would happen if I choose step length two where is my assume two, if I assume 2 I think which you will prove. And then if g k is bounded then x k goes to x star element of x star and f of x k goes to f x f star f star is nothing but, the solution f star is the solution of at the value of the problem this optimal value of the problem. And if you choose the Boris Polyak step length then you do not have to take this, if you assume 3 then this will happen. But, remember the Boris Polyak step length is essentially a theoretical guide but, these are implementable step lengths.

So, once you have implementable step lengths the second is a better one which has choice like this you have to make a now how do you know it this is bounded how do you know that g k is bounded. So, g k is belonging to del f x k and there is a notion of local boundedness. So, this is actually required when you are talking about these sort of functions but, when you are talking about extended this functions the finite valued function then these thing that g k is bounded comes out of the fact that the sub differential map is locally bounded.

So, I have not told you locally bounded. So, you need not bother about it let us assume that this this is true and tomorrow we will start by proving this and then writing down this Boris Polyak's result then get into details first we will prove this and then we will prove this. So, tomorrow's class would be the proof of these two and which will be one of the most advanced lectures given in this course and as a result of which you know you will see the issues that come up the art of optimization algorithm. Say if optimization algorithm is a spring optimization algorithm is as a art see you cannot just write down something.

And so just compute and this is the solution this will give you the solution no you cannot do that you have to say what I have what you have whatever you have written. Whatever tampering you might have done this would ultimately lead me to the solution or somewhere some, some at least the values would lead me to the objective optimal value some are it should take me which is meaningful where makes it meaningful.

So, that is the art the convergence analysis is the art of optimization algorithms. So, it is very important in this class that we have done something with the Newton's case, Newton's where if un constant case but, for the constant case in the convex situation, it is also important that we do thorough analysis at least one of them, so I will just do this.

Thank you very much.