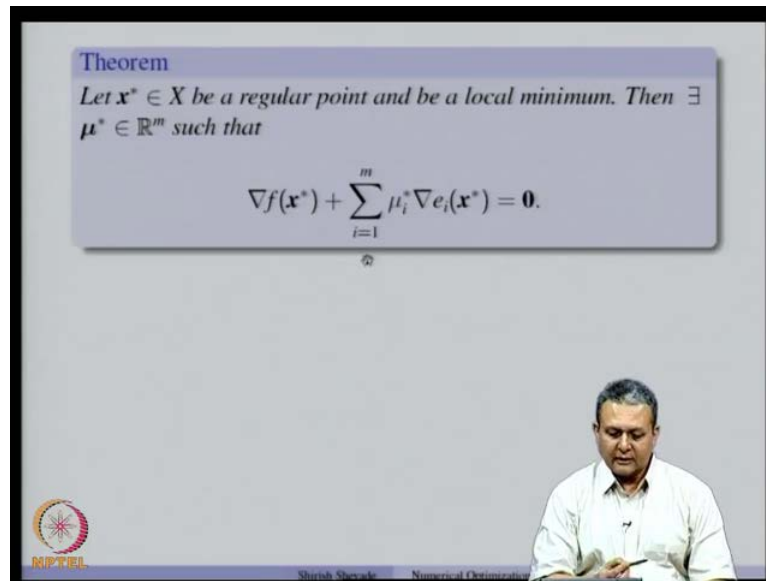


Numerical Optimization
Prof. Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Lecture - 24
Convex Programming Problem

(Refer Slide Time: 00:36)



Theorem

Let $x^* \in X$ be a regular point and be a local minimum. Then $\exists \mu^* \in \mathbb{R}^m$ such that

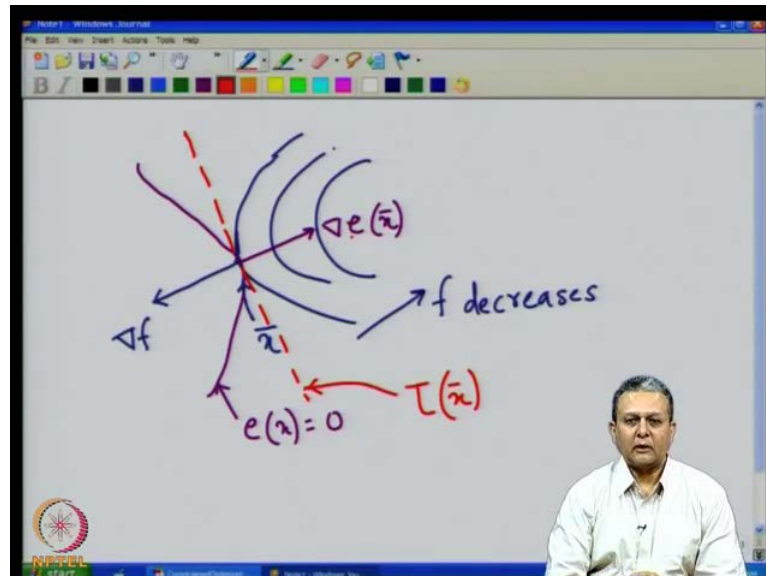
$$\nabla f(x^*) + \sum_{i=1}^m \mu_i^* \nabla e_i(x^*) = \mathbf{0}.$$

Hello welcome back in the last lecture, we started discussing about equality constrained optimization problems; and we looked at these theorem we says that. Let x^* belong to X be a regular point and also be a local minimum, then the necessary condition for that local minimum is that there exists. Some μ^* in m -dimensional space, note that m is the number of equality constraints that we have and the equality constraints are of the type $e_i(x) = 0$.

So, there exist μ^* in m dimensional space such that the gradient of f of x^* is a linear combination of gradient of $e_i(x^*)$ or all i is going from 1 to m . Now, remember that at x^* all the equality constraints are active. So, that is why unlike the inequality constrained problems, we have consider all the equality constraints. Because, they are activate in a x^* belong to X and by regular point what, we mean is that the gradient $e_i(x^*)$ for all i is going from 1 to m are linearly independent. So, the only way that there will be a trivial combination, which will make this component 0 and that is all $\mu_i^* = 0$. So, we wanted to avoid that so, we use the regular regularity assertion, now what

this result means is that. So, let us consider set of for objective objective function contours.

(Refer Slide Time: 02:16)



So, objective functions or contours are like this, so let us consider a point and let us call this as \bar{x} , now the objective function value. So, f decreases in this direction, so the gradient at this point assuming that the function is differentiable. So, the gradient of f will be in this direction. Now, let us look at a contour of an equality constrained problem. So, that contour is suppose, like this. So, this is the surface $e(x) = 0$, now if you look at the the gradient $\nabla e(\bar{x})$ is pointing in this direction. So, we will see that the if, we draw the tangent plane to this equality constraint at at the point \bar{x} , it will be like this. So, this is $T(\bar{x})$ and we will see that at this point the gradient of f can be written as a some multiple of gradient $\nabla e(\bar{x})$.


(Refer Slide Time: 04:10)

Theorem
 Let $\mathbf{x}^* \in X$ be a regular point and be a local minimum. Then $\exists \boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}.$$

Proof.
 Let $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), \dots, e_m(\mathbf{x}))$. $\mathbf{x}^* \in X$ is a local minimum.
 $\therefore \{ \mathbf{d} : \nabla f(\mathbf{x}^*)^T \mathbf{d} < 0, \nabla e(\mathbf{x}^*)^T \mathbf{d} = 0 \} = \phi$.
 Let $C_1 = \{ (y_1, \mathbf{y}_2) : y_1 = \nabla f(\mathbf{x}^*)^T \mathbf{d}, \mathbf{y}_2 = \nabla e(\mathbf{x}^*)^T \mathbf{d} \}$ and
 $C_2 = \{ (y_1, \mathbf{y}_2) : y_1 < 0, \mathbf{y}_2 = \mathbf{0} \}$
 Note that C_1 and C_2 are convex and $C_1 \cap C_2 = \phi$.

If C_1 and C_2 are nonempty convex sets in \mathbb{R}^n and $C_1 \cap C_2 = \phi$,
 $\boldsymbol{\mu} \in \mathbb{R}^n (\boldsymbol{\mu} \neq \mathbf{0})$ such that $\boldsymbol{\mu}^T \mathbf{x}_1 \geq \boldsymbol{\mu}^T \mathbf{x}_2 \forall \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2$.



Shreshth Shrivastava, Numerical Optimization

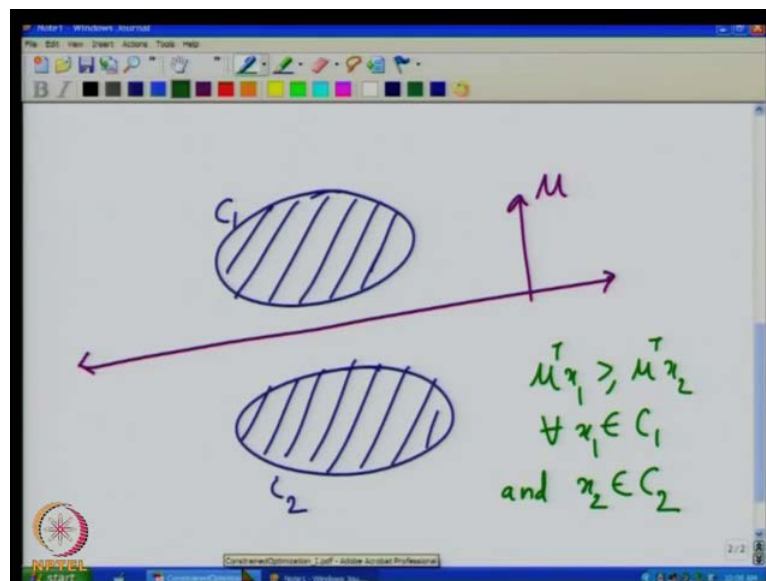
So, in general when, we have more number of equality constraints than at a local minimum the gradient of the objective function is written as a linear combination of the gradients of all the equality constraints. So, let us look at the proof of this so, let us arrange all the equality constraints in $\mathbf{e}(\mathbf{x})$ and then since \mathbf{x}^* , which is a feasible point is also a local minimum. What, we have is that the set of all directions \mathbf{d} , which make an obtuse angle with gradient $\nabla f(\mathbf{x}^*)$ and which are orthogonal to gradient $\nabla e(\mathbf{x}^*)$.

So, orthogonal to gradient $\nabla e(\mathbf{x}^*)$ means, they are orthogonal to each and every constraint in the constraint equality constraint set. So, they are e_1 to e_m \mathbf{x} , so this set is a null set because, \mathbf{x}^* is a local minimum, we have seen this earlier. So, let us define 2 sets C_1 and C_2 . So, C_1 is a set in $m+1$ dimensional space, where which consists of 2 components y_1 and \mathbf{y}_2 . y_1 is a scalar and \mathbf{y}_2 is a vector and y_1 is written as gradient $\nabla f(\mathbf{x}^*)^T \mathbf{d}$ and \mathbf{y}_2 is written as gradient $\nabla e(\mathbf{x}^*)^T \mathbf{d}$ and C_2 be that set of y_1, \mathbf{y}_2 , such that $y_1 < 0$ and $\mathbf{y}_2 = \mathbf{0}$.

Now, clearly since \mathbf{x}^* is a local minimum the intersection of C_1 and C_2 is a null set and more over C_1 and C_2 both are convex sets. Now at this point, we would like to recall some the results that, we studied when m we, discussed about convex sets and convex functions and one of the results is related to 2 non empty, non intersecting convex sets.

So, if we have 2 non-empty, non-intersecting convex sets then there exists a separating hyperplane, we separates the 2 sets. So, let us recall that result that, if we have C_1 and C_2 is are nonempty convex sets and they have empty intersection. So, if the 2 sets have empty intersection then, there exist μ which, is non zero, such that $\mu^T x_1$ is greater than or equal to $\mu^T x_2$ for all x_1 in C_1 and x_2 in C_2 or in other words.

(Refer Slide Time: 07:23)



So, if we have 2 sets which, are convex and which, are not intersecting which, have empty intersection. So, this is our set C_1 and this is our set C_2 then there exists a separating hyperplane, that μ be the normal to this hyperplane. So, $\mu^T x_1$ where, x_1 belongs to C_1 will be always greater than or equal to $\mu^T x_2$. So, what we have is $\mu^T x_1$ is always greater than or equal to $\mu^T x_2$ for all x_1 belongs to C_1 and x_2 belong to C_2 . So, this is the result that, we saw last time when, we talked about the convex sets.


(Refer Slide Time: 08:40)

Theorem
 Let $\mathbf{x}^* \in X$ be a regular point and be a local minimum. Then $\exists \boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}.$$

Proof.
 Let $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), \dots, e_m(\mathbf{x}))$. $\mathbf{x}^* \in X$ is a local minimum.
 $\therefore \{ \mathbf{d} : \nabla f(\mathbf{x}^*)^T \mathbf{d} < 0, \nabla e(\mathbf{x}^*)^T \mathbf{d} = 0 \} = \emptyset$.
 Let $C_1 = \{ (y_1, y_2) : y_1 = \nabla f(\mathbf{x}^*)^T \mathbf{d}, y_2 = \nabla e(\mathbf{x}^*)^T \mathbf{d} \}$ and
 $C_2 = \{ (y_1, y_2) : y_1 < 0, y_2 = \mathbf{0} \}$
 Note that C_1 and C_2 are convex and $C_1 \cap C_2 = \emptyset$.

If C_1 and C_2 are nonempty convex sets in \mathbb{R}^n and $C_1 \cap C_2 = \emptyset$,
 $\boldsymbol{\mu} \in \mathbb{R}^n (\boldsymbol{\mu} \neq \mathbf{0})$ such that $\boldsymbol{\mu}^T \mathbf{x}_1 \geq \boldsymbol{\mu}^T \mathbf{x}_2 \forall \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2$.



Shreshth Shrivastava Numerical Optimization

So, let us use this theorem in this case. So, now we have this 2 sets C_1 and C_2 and their intersection is empty, they are convex.

(Refer Slide Time: 08:53)

Proof. (continued)
 Therefore, $\exists (\mu_0, \boldsymbol{\mu}) \in \mathbb{R}^{m+1}$ such that

$$\mu_0 \nabla f(\mathbf{x}^*)^T \mathbf{d} + \boldsymbol{\mu}^T (\nabla e(\mathbf{x}^*)^T \mathbf{d}) \geq \mu_0 y_1 + \boldsymbol{\mu}^T y_2 \forall \mathbf{d} \in \mathbb{R}^n, (y_1, y_2) \in C_2$$


Letting $y_2 = \mathbf{0}$, we get $\mu_0 \geq 0$.
 Letting $(y_1, y_2) = (0, \mathbf{0})$, we get

$$\mu_0 \nabla f(\mathbf{x}^*)^T \mathbf{d} + \boldsymbol{\mu}^T (\nabla e(\mathbf{x}^*)^T \mathbf{d}) \geq 0 \forall \mathbf{d} \in \mathbb{R}^n$$

If we take $\mathbf{d} = -(\mu_0 \nabla f(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla e(\mathbf{x}^*))$, we get
 $-\|(\mu_0 \nabla f(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla e(\mathbf{x}^*))\|^2 \geq 0$. Therefore,

$$\mu_0 \nabla f(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla e(\mathbf{x}^*) = \mathbf{0} \text{ where } (\mu_0, \boldsymbol{\mu}) \neq (0, \mathbf{0})$$

Note that, $\mu_0 > 0$ since \mathbf{x}^* is a regular point. Hence,

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\mu}^{*T} \nabla e(\mathbf{x}^*) = \mathbf{0}$$


Shreshth Shrivastava Numerical Optimization

So, therefore, there exist vector in $m+1$ dimensional space which, has 2 components μ_0 and $\boldsymbol{\mu}$ note that, this μ_0 will be associated with gradient $f(\mathbf{x}^*)$ and $\boldsymbol{\mu}$ will be associated with gradient $e(\mathbf{x}^*)$. So, there exists non zero vector, I repeat that this is a non zero vector not all components are 0.

Such that $\mu_0 \text{ into gradient } f(x^*)^T d + \mu^T \text{ into gradient, } e(x^*)^T d$ is greater than or equal to $\mu_0 y_1 + \mu^T y_2$ for all d , so for all d in \mathbb{R}^n and y_1, y_2 in \mathbb{C}^2 . So, for all d s in \mathbb{R}^n , we are able to get this result, because of the separability of the two convex sets.

Now, if we let y_2 equal to 0 here then, what happens is that y_1 can be made arbitrarily large negative quantity and because of which, this number if μ_0 is less than 0. This number will be very large and we are saying that the quantity on the left is greater than this large number. And μ_0 could be infinity minus infinity and y_1 is very large negative number. So, this quantity will be close to infinity and not the that means, the left quantity is greater than or equal to infinity, which is not possible. So, this by letting y_2 equal to 0 and noting that y_1 can be made arbitrarily large negative number, we have to have μ_0 greater than or equal to 0.

So, we note that this μ_0 has to be greater than or equal to 0, now if, you let y_1, y_2 both close towards 0 then what, we get is that the left hand side quantity should be greater than or equal to 0, for all d belong to \mathbb{R}^n . And since this quantity is greater than or equal to 0, if we take d to be so, this holds for all d . So, we are free to choose any d in \mathbb{R}^n . So, if we chose d to be $-\mu_0 \text{ gradient } f(x^*) + \mu^T \text{ gradient } e(x^*)$ then, what we get is that a minus norm of $\mu_0 \text{ gradient } f(x^*) + \mu^T \text{ gradient } e(x^*)$ whole square equal is greater than or equal to 0.

Now, we have already seen that the norm any vector is a non negative quantity. So, this quantity is non negative and the negative of that has to be greater than or equal to 0, now the only way this inequality satisfied is when this quantity is 0. So, that means, we have $\mu_0 \text{ gradient } f(x^*) + \mu^T \text{ gradient } e(x^*) = 0$, where μ_0 and μ are not they do not constitute 0 vector.

Because, we are constitute 0 vector then this relationship is trivially satisfied and why we, are not interested in this relationship in such trivial relationship. Now, remember that we, have assume that x^* is a regular point and earlier, we showed that if x^* is a regular point this μ_0 has to be greater than 0. Because otherwise again we, will end up in a contradiction.

So, μ_0 has to be greater than 0, since x^* is a regular point and therefore, we we can divide the equation by μ_0 and what, we get is $\text{gradient of } x^* + \mu^*$

transpose gradient $e^T x^*$ equal to 0 and this is what we wanted to prove. So, by ensuring that because, of the regularity of x^* μ^0 is greater than 0, we are able to say that at if x^* is a local min then gradient $f(x^*) + \mu^* \text{transpose gradient } e^T x^*$ has to be 0 of gradient $f(x^*)$ is written as a linear combination of gradient $e_i^T x^*$, i going from 1 to m . Now unlike the inequality constraints note that there are no sign restrictions on μ . μ is a vector in m dimensional space and therefore, for equality constraint problems, we can have any linear combination of equality constraints and that is equal to a gradient $f(x^*)$ in this case.

(Refer Slide Time: 14:07)

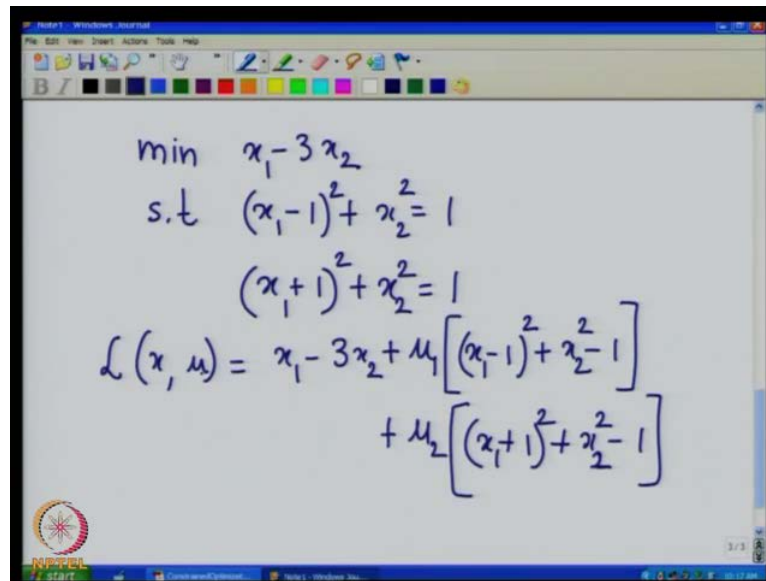
Examples:

$$\begin{aligned} \min \quad & x_1 - 3x_2 \\ \text{s.t.} \quad & (x_1 - 1)^2 + x_2^2 = 1 \\ & (x_1 + 1)^2 + x_2^2 = 1 \end{aligned}$$

The slide also features the NPTEL logo in the bottom left and the text 'Shrinath Shrivastava - Nonlinear Optimization' in the bottom center. A lecturer is visible in the bottom right corner of the frame.

Now, let us take some examples. So, here we have 1 example where, we want to minimize $x_1 - 3x_2$ subject to $(x_1 - 1)^2 + x_2^2 = 1$ and $(x_1 + 1)^2 + x_2^2 = 1$.

(Refer Slide Time: 14:29)

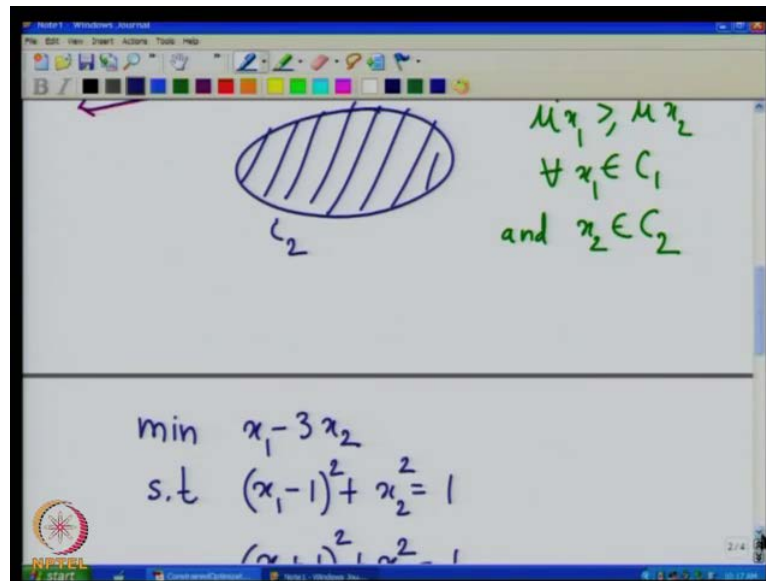


The image shows a whiteboard with handwritten mathematical expressions. At the top, it states the objective function and constraints: $\min x_1 - 3x_2$ and $s.t. (x_1 - 1)^2 + x_2^2 = 1$. Below this, a second constraint is written: $(x_1 + 1)^2 + x_2^2 = 1$. The Lagrangian function is then defined as $\mathcal{L}(x, \mu) = x_1 - 3x_2 + \mu_1 [(x_1 - 1)^2 + x_2^2 - 1] + \mu_2 [(x_1 + 1)^2 + x_2^2 - 1]$. The whiteboard also features a toolbar at the top and a logo in the bottom left corner.

So, minimize $x_1 - 3x_2$ subject to $x_1 - 1$ square plus x_2 square equal to 1 and x_2 x $x_1 + 1$ square plus x_2 square is equal to 1. So, let us look at this problem, now we have 2 equality constraints. So, both are active at a given point and so, let us first write the lagrangian, so lagrangian of x, μ to be $x_1 - 3x_2 + \mu_1 [(x_1 - 1)^2 + x_2^2 - 1] + \mu_2 [(x_1 + 1)^2 + x_2^2 - 1]$.

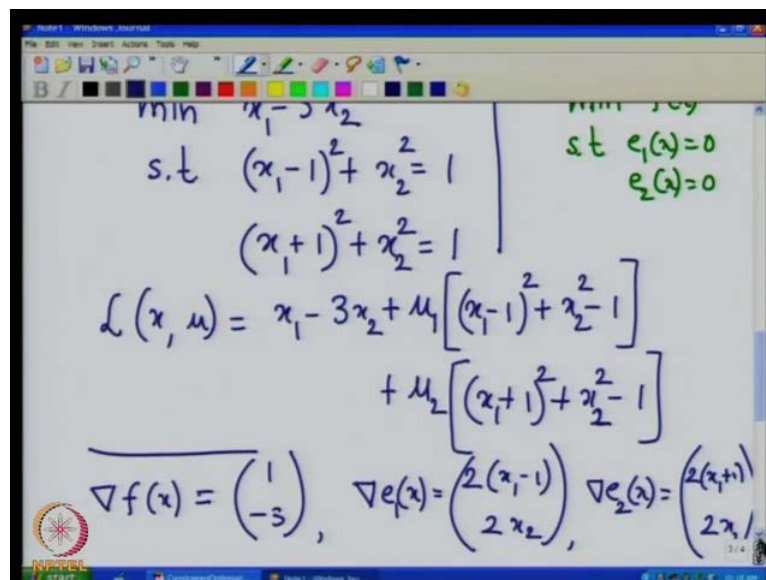
So, remember that this constraint the first constraint is written in the form $e_1 x$ equal to 0 and that is why, that $e_1 x$ is $(x_1 - 1)^2 + x_2^2 - 1$ that is multiplied by the lagrangian multiplier for that constraint plus μ_2 into $(x_1 + 1)^2 + x_2^2 - 1$. So, this is our lagrangian function, which is a function of both x and μ remember that, this x is a vector and μ is a vector x contains x_1 and x_2 as it is components μ contains μ_1 and μ_2 as it is components.

(Refer Slide Time: 16:57)



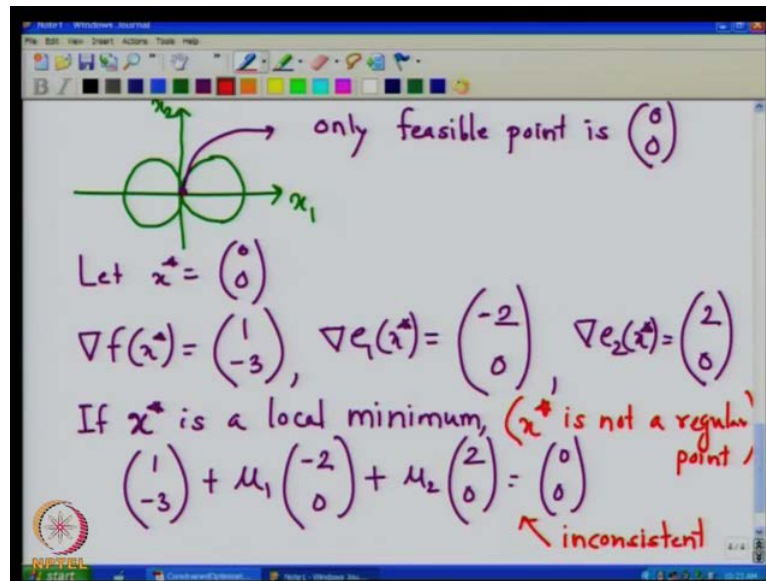
Now so, let us write down the gradient of the objective function and the gradients of the constraints. So, we write this function in this problem in the form.

(Refer Slide Time: 17:14)



Minimize $f(x)$ subject to $e_1(x) = 0$ and $e_2(x) = 0$. So, let us write down the gradient of $f(x)$. The gradient of $f(x)$ is equal to 1 minus 3. Then the gradient of $e_1(x)$ is equal to 2 into x_1 minus 1 and 2 x_2 and the gradient of $e_2(x)$ is equal to 2 x_1 plus 1 and 2 x_2 .

(Refer Slide Time: 18:20)



So, now, let us draw the feasible region. So, we have x_1 and x_2 . So, $x_1 - 1 + x_2^2 = 1$ is a circle with centre $(1, 0)$ and radius 1 . So, it is a circle with this centre and then the second constraint uses another circle, now if you intersect them. So, we get only feasible point is $(0, 0)$. So, let us consider this x^* , so let x^* be $(0, 0)$.

So, gradient of x^* since it is a linear function, this gradient is not going to change then gradient $e_1(x^*)$ is equal to $(-2, 0)$ and gradient $e_2(x^*)$ will be $(2, 0)$, now if x^* is a local minimum then, what we want is that. So, if x^* is a local minimum what, we want is that gradient $f(x^*)$. So, $(1, -3) + \mu_1(-2, 0) + \mu_2(2, 0) = (0, 0)$ now.

So, you will see that it is not possible to construct or to find μ_1 and μ_2 which, satisfy this. Because, the first equation gives $1 - 2\mu_1 + 2\mu_2 = 0$, while the second equation gives $-3 + 0 + 0 = 0$ and which, is not possible. Because, of the left side, we have -3 and the right side of the equation, we have 0 they cannot be same. So, the problem here, that we cannot use the KKT condition is that, x^* is not a regular point because, this is the only feasible point is not a regular point and therefore. So, this system is inconsistent. So, therefore, the regularity is very important when, we want to solve any optimization problem to get a

KKT point and then check whether that point is a local minimum or not. So, although x^* is a solution of this problem, we know that x^* is not a KKT point in this case.

(Refer Slide Time: 23:12)

Examples:

- $$\begin{aligned} \min \quad & x_1 - 3x_2 \\ \text{s.t.} \quad & (x_1 - 1)^2 + x_2^2 = 1 \\ & (x_1 + 1)^2 + x_2^2 = 1 \end{aligned}$$

$(0, 0)^T$ is the only feasible point; $(0, 0)^T$ is not a regular point.
- $$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

Now So, the only feasible point is the origin and is not a regular point and therefore, although it is a solution, it is not a KKT point. Now, let us look at another example minimize $x_1 + x_2$ subject to $x_1^2 + x_2^2 = 1$.

(Refer Slide Time: 23:35)

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

$$L(x, \mu) = x_1 + x_2 + \mu(x_1^2 + x_2^2 - 1)$$

$$\frac{\partial L}{\partial x_1} = 1 + 2\mu x_1$$

$$\frac{\partial L}{\partial x_2} = 1 + 2\mu x_2$$

So, minimize $x_1 + x_2$ subject to the constraint $x_1^2 + x_2^2 = 1$, now if we plot the constraint. So, this is a constraint, that we have and we look at the x_1

plus x_2 . So, x_1 plus x_2 equal to constant is are the. So, these are the lines x_1 plus x_2 equal to some constant. So, we will see that as, we move in this direction. So, f increases such, we move in this direction f increases.

Now So, let us write the lagrangian of this problem. So, the lagrangian x μ will be the objective function plus μ time the constraints, note let us take the partial derivative of l with respect to x_1 and that will be $1 + 2\mu x_1$ partial of l with respect to x_2 will be $1 + 2\mu x_2$. Now, if we look at the KKT conditions, what the KKT conditions demand is that.

(Refer Slide Time: 26:30)

The image shows a whiteboard with handwritten mathematical equations for KKT conditions. The text is as follows:

$$\begin{aligned} \text{KKT conditions:} \\ \left. \begin{aligned} \frac{\partial L}{\partial x_1} = 0 &\Rightarrow 1 + 2\mu x_1 = 0 \\ \frac{\partial L}{\partial x_2} = 0 &\Rightarrow 1 + 2\mu x_2 = 0 \end{aligned} \right\} \mu = -\frac{1}{2x_1} = -\frac{1}{2x_2} \\ \text{feasibility of } x \Rightarrow x_1^2 + x_2^2 = 1 \end{aligned}$$

So, we have partial of l with respect to x_1 equal to 0 and that implies $1 + 2\mu x_1$ is equal to 0, then partial of l with respect to x_2 is equal to 0, which means $1 + 2\mu x_2$ is equal to 0 and then we also want to satisfy the feasibility of the points x_1 and x_2 . So, the feasibility constraints x_1 square plus x_2 square equal to 1, so we have 3 equations and the unknowns are x_1 , x_2 and μ .

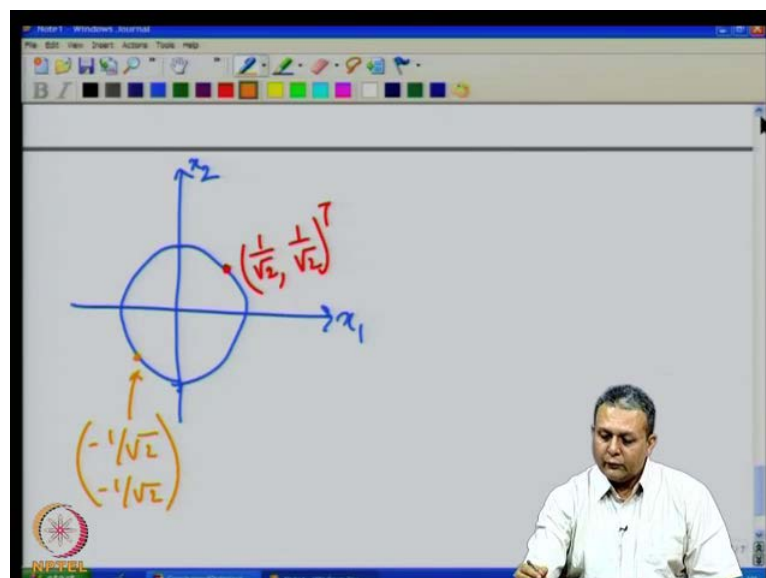
Now, these 2 equations together, they give 1 and equal to minus $2\mu x_1$. So, μ is equal to minus 1 by $2x_1$. So, μ is equal to minus 1 by $2x_2$. So, this means that x_1 and x_2 is equal to x_1 is equal to x_2 and if we substitute x_1 equal to x_2 here, what we get is $2x_1$ square equal to 1 and therefore, what we get is.

(Refer Slide Time: 28:44)

$$\frac{\partial L}{\partial x_2} = 0 \Rightarrow 1 + 2\mu x_2 = 0$$
$$\text{feasibility of } x \Rightarrow x_1^2 + x_2^2 = 1$$
$$x^* = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \quad \text{or} \quad x^* = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$
$$\mu^* = -\frac{1}{\sqrt{2}} \qquad \qquad \mu^* = \frac{1}{\sqrt{2}}$$

X star equal to 1 by root 2 1 by root 2 or x star equal to minus 1 by root 2, minus 1 by root 2, so these are the possible solutions and the correspondingly mu star. So, if x star is 1 by root 2 then in this case mu star will be minus 1 by root 2 and in this case mu star will be plus 1 by root 2. So, we get a KKT point, we in fact, we get 2 KKT points 1 is x star is 1 by root 2 1 by root 2 and mu star is minus 1 by root 2 and other 1 is x star is minus 1 by root 2 minus by root 2 and mu star is 1 by root 2.

(Refer Slide Time: 30:10)



Now, let us again re plot the constraint set x_1 and x_2 , now here is a point, which is $1/\sqrt{2}$ and $1/\sqrt{2}$ and here is a point, which is $-1/\sqrt{2}$ and $-1/\sqrt{2}$.

(Refer Slide Time: 30:59)

$$\frac{\partial L}{\partial x_1} = 0 \Rightarrow 1 + 2\mu x_1 = 0$$

$$\frac{\partial L}{\partial x_2} = 0 \Rightarrow 1 + 2\mu x_2 = 0 \quad \left\{ \begin{array}{l} \mu = -\frac{1}{2x_1} = -\frac{1}{2x_2} \end{array} \right.$$

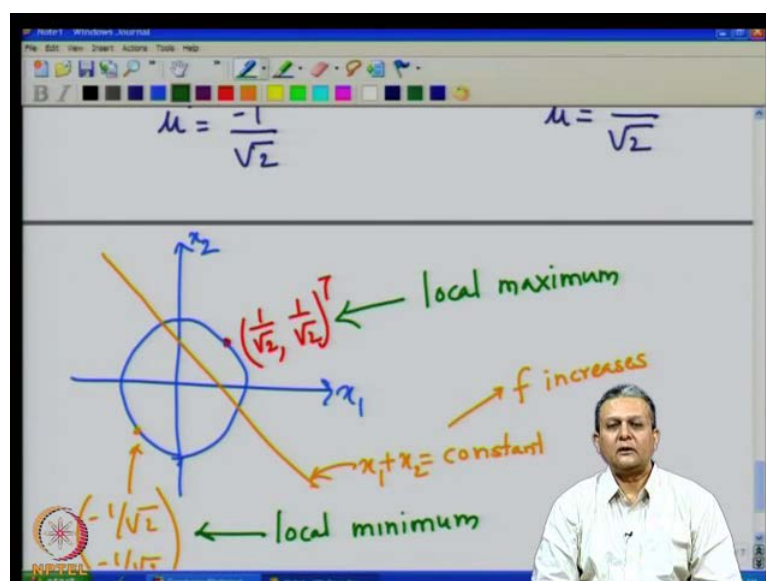
$$\text{feasibility of } x \Rightarrow x_1^2 + x_2^2 = 1$$

$$x^* = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \quad \text{or} \quad x^* = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)$$

$$\mu^* = -\frac{1}{\sqrt{2}}$$

So, if you look at our KKT points $1/\sqrt{2}$ and $1/\sqrt{2}$ and the corresponding μ^* was $-1/\sqrt{2}$. So, this point is shown here and the other point is shown here.

(Refer Slide Time: 31:16)



And our function is $x_1 + x_2$ is equal to constant is of this form and it increases in this direction f increases in this direction. So, you will see that, we got 2 KKT points and as a function increases in this direction, this point turns out to be a local maximum and this point turns out to be as the function decreases in this direction. This point turns out to be local minimum because, if we cannot go beyond this point because then, we will violate the constraint. So, we got 2 KKT points and one of them turned out to be a local maximum and the other 1 turned out to be a local minimum and the main reason for this is that the first order KKT conditions are just necessary conditions.

If you recall our discussion on unconstrained optimization, we saw that if x^* is a local minimum then gradient of x^* equal to 0, for differentiable function f . But, gradient f x^* equal to 0 does mean that, x^* is a local minimum unless, we are talking about some special functions like convex functions. So, the gradient of the lagrangian equal to 0 gives you a KKT point, which can be either a local minimum or local maximum. And so, we need to look for some second order conditions, which make use of the curvature of the lagrangian function and then only, we can conclude whether a given point is indeed a local minimum or not.

(Refer Slide Time: 33:29)

Examples:

1

$$\begin{aligned} \min \quad & x_1 - 3x_2 \\ \text{s.t.} \quad & (x_1 - 1)^2 + x_2^2 = 1 \\ & (x_1 + 1)^2 + x_2^2 = 1 \end{aligned}$$

$(0, 0)^T$ is the only feasible point; $(0, 0)^T$ is not a regular point.

2

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

local maximum : $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$
 local minimum : $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$

NPTEL
Srinath Sridhar

So, we have a local maximum, which is at $1/\sqrt{2}, 1/\sqrt{2}$ and a local minimum at $-1/\sqrt{2}, -1/\sqrt{2}$. Incidentally, both these points along with the respective μ^* where, the KKT points of this problem.

(Refer Slide Time: 33:45)

General Nonlinear Programming Problems

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & e_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- $f, h_j (j = 1, \dots, l), e_i (i = 1, \dots, m)$ are sufficiently smooth
- $X = \{\mathbf{x} : h_j(\mathbf{x}) \leq 0, e_i(\mathbf{x}) = 0, j = 1, \dots, l; i = 1, \dots, m\}$
- $\mathbf{x}^* \in X$
- Active set of X at \mathbf{x}^* :
 - $\mathcal{I} = \{j : h_j(\mathbf{x}^*) = 0\}$
 - All the equality constraints, $\mathcal{E} = \{1, \dots, m\}$

$$A(\mathbf{x}^*) = \mathcal{I} \cup \mathcal{E}$$

- Assumption: \mathbf{x}^* is a *regular point*. That is, $\{\nabla h_j(\mathbf{x}^*) : j \in \mathcal{I}\} \cup \{\nabla e_i(\mathbf{x}^*) : i \in \mathcal{E}\}$ is a set of independent vectors

NPTEL

Now, we have considered so far the problems of the type minimize affects subject to the inequality constraints and problems of the type minimize affects subject to the equality constraints and for both the cases. We derived the KKT conditions KKT conditions necessary for the local minimum.

Now, let us combine the equality as well as the inequality constraints and write a general non linear programming problem. And that problem is of the type minimize $f(\mathbf{x})$ subject to the l inequality, constraints of the type $h_j(\mathbf{x}) \leq 0$ and m equality constraints of the type $e_i(\mathbf{x}) = 0$. So, this is our general non linear programming problem.

Now, again we assume that f in all h_j 's and e_i 's are sufficiently smooth. So, for first order conditions, we just need this f, h and e to be continuously differentiable and for a second order conditions, we need all this function f, h and e to be twice continuously differentiable. So, depending upon the condition, we will require the sufficient smoothness of this functions. So, let us denote the feasible set by capital X . So, set of all points, which satisfy all $h_j(\mathbf{x}) \leq 0$ and all $e_i(\mathbf{x}) = 0$.

Now, in the context of inequality constraint problem, we discuss about the regular point and we also discuss about the linear independence constrained qualification. So, if we take a feasible point and if we take all the inequality constrained, which are active then

we said that the linear independence constraint qualification holds. If the gradients of the active inequality, constraints are linearly independent.

Now, similarly we assume the linear independence of the gradient of e_i at x^* . Now we, have to combine those ideas to define the regular point for this problem. So, let us consider a feasible x^* and let us see, which sets are active at x^* . So, this let us denote by this script \mathcal{I} the set of all inequality constraints, which are active. So, that is $h_j(x^*) = 0$ or this constraint is satisfied with equality.

Now, as far as the equality constraints are concerned. If x^* is feasible then all the equality constraints need to be satisfied so that means, all equality constraints are active, so \mathcal{I} active at x^* . So, let us combine all the equality constraints in dashes and put them in the set script \mathcal{E} . Now this \mathcal{I} and script \mathcal{E} these 2 sets together form the active set of the constraint set at x^* and therefore, $\mathcal{A}(x^*)$ is nothing but $\mathcal{I} \cup \mathcal{E}$.


So, let us assume that x^* is a regular point, so that means, that gradient $\nabla h_j(x^*)$ belong to \mathcal{I} and gradient $\nabla e_i(x^*)$ belong to \mathcal{E} . The this put together are linearly independent vectors, so gradient $\nabla h_j(x^*)$ where, j is coming from \mathcal{I} and gradient $\nabla e_i(x^*)$ from the set \mathcal{E} , they form a linearly independent set of vectors. So, this assumption here is very important that, x^* is a regular point means that they form a linearly independent set of vectors.

(Refer Slide Time: 38:34)

General Nonlinear Programming Problems

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & e_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- $f, h_j(j = 1, \dots, l), e_i(i = 1, \dots, m)$ are sufficiently smooth
- $X = \{\mathbf{x} : h_j(\mathbf{x}) \leq 0, e_i(\mathbf{x}) = 0, \quad j = 1, \dots, l; \quad i = 1, \dots, m\}$
- $\mathbf{x}^* \in X$
- Active set of X at \mathbf{x}^* :
 - $\mathcal{I} = \{j : h_j(\mathbf{x}^*) = 0\}$
 - All the equality constraints, $\mathcal{E} = \{1, \dots, m\}$
- $\mathcal{A}(\mathbf{x}^*) = \mathcal{I} \cup \mathcal{E}$
- Assumption: \mathbf{x}^* is a *regular point*. That is, $\{\nabla h_j(\mathbf{x}^*) : j \in \mathcal{I}\} \cup \{\nabla e_i(\mathbf{x}^*) : i \in \mathcal{E}\}$ is a set of *linearly independent* vectors



Shrinath Shrivastava Numerical Optimization

So, we assume that, they are a linearly independent set.

(Refer Slide Time: 38:40)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & e_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- $X = \{\mathbf{x} : h_j(\mathbf{x}) \leq 0, e_i(\mathbf{x}) = 0, \quad j = 1, \dots, l; \quad i = 1, \dots, m\}$

KKT necessary conditions (First Order): If $\mathbf{x}^* \in X$ is a local minimum and a *regular* point, then there exist unique vectors $\boldsymbol{\lambda}^* \in \mathbb{R}_+^l$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^l \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^m \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

$$\lambda_j^* h_j(\mathbf{x}^*) = 0 \quad \forall j = 1, \dots, l$$

$$\lambda_j^* \geq 0 \quad \forall j = 1, \dots, l$$

- KKT Point: $(\mathbf{x}^* \in X, \boldsymbol{\lambda}^* \in \mathbb{R}_+^l, \boldsymbol{\mu}^* \in \mathbb{R}^m)$ satisfying above conditions
- First order KKT conditions also satisfied at a local max

NPTEL
 Shreshth Shrivastava Numerical Optimization

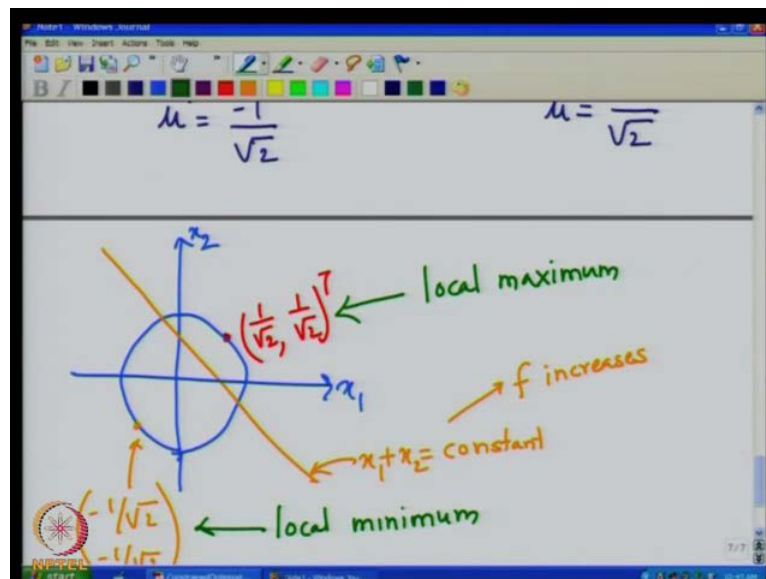
Now, let us look at this problem and we assume that X is the feasible set, now we can write down the KKT necessary conditions. So, if \mathbf{x}^* is a feasible point remember that always start with a feasible point. So, whenever I specify the KKT necessary conditions, I do not say explicitly that $h_j(\mathbf{x}^*) \leq 0$ and $e_i(\mathbf{x}^*) = 0$. Because, it is assumed here that \mathbf{x}^* belongs to X , \mathbf{x}^* is a feasible point. Now, if \mathbf{x}^* is a feasible point and also a regular point, now if \mathbf{x}^* is a local minimum then there exist unique vectors $\boldsymbol{\lambda}^* \in \mathbb{R}_+^l$. So, this plus indicates that, these $\boldsymbol{\lambda}^*$ are non-negative and $\boldsymbol{\mu}^* \in \mathbb{R}^m$.

So, remember that there are l inequality constraints. So, there is 1 Lagrangian multiplier λ corresponding to every e , inequality constraints and there is 1 Lagrangian multiplier μ corresponding to every equality constraint. So, the difference is that the $\boldsymbol{\lambda}^*$ come from \mathbb{R}_+^l or the λ each of the components of $\boldsymbol{\lambda}$ is non-negative well, there is no such restriction on $\boldsymbol{\mu}^*$'s.

So, the KKT first order conditions say that, if \mathbf{x}^* is a local minimum and is a regular point and \mathbf{x}^* is feasible then the gradient of $f(\mathbf{x}^*)$ plus $\sum \lambda_j^* \nabla h_j(\mathbf{x}^*)$ plus $\sum \mu_i^* \nabla e_i(\mathbf{x}^*)$ is 0. And $\lambda_j^* h_j(\mathbf{x}^*) = 0$ for all j 's. This is the complementary slackness condition that we saw earlier and all the $\boldsymbol{\lambda}^*$'s are non-negative.

This is stated here, but again, I have repeated here just for the completeness. So, these are called the first order KKT necessary conditions and any point x^* where, x^* is feasible λ^* is non-negative μ^* is belongs to \mathbb{R}^m and we satisfies this conditions. We call it as a KKT point. So, to find a solution of this problem, just amongst to finding the KKT points and then checking, which of them give rise to a local minimum or a local maximum. And as we saw earlier that, first order KKT conditions are also satisfied at a local max.

(Refer Slide Time: 41:50)



We saw that example where we, so this example we saw that both this points are KKT points, because they satisfy the first order KKT conditions. But one of them is a local maximum and the other one is a local minimum.

(Refer Slide Time: 42:07)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & e_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- $X = \{\mathbf{x} : h_j(\mathbf{x}) \leq 0, e_i(\mathbf{x}) = 0, \quad j = 1, \dots, l; \quad i = 1, \dots, m\}$

KKT necessary conditions (First Order): If $\mathbf{x}^* \in X$ is a local minimum and a *regular* point, then there exist unique vectors $\boldsymbol{\lambda}^* \in \mathbb{R}_+^l$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^l \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^m \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

$$\lambda_j^* h_j(\mathbf{x}^*) = 0 \quad \forall j = 1, \dots, l$$

$$\lambda_j^* \geq 0 \quad \forall j = 1, \dots, l$$

- KKT Point: $(\mathbf{x}^* \in X, \boldsymbol{\lambda}^* \in \mathbb{R}_+^l, \boldsymbol{\mu}^* \in \mathbb{R}^m)$ satisfying above conditions
- First order KKT conditions also satisfied at a local max

So, it is important to note that, these are just the necessary conditions, they do not guarantee that \mathbf{x}^* , if this that a KKT point is a local minimum.

(Refer Slide Time: 42:20)

Consider the problem (CP):

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & e_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- Assumption: $f, h_j, j = 1, \dots, l$ are smooth convex functions
- $e_i(\mathbf{x}) = \mathbf{a}^T \mathbf{x}_i - b_i, \quad i = 1, \dots, m$
- CP is a **convex programming problem**
- $X = \{\mathbf{x} : h_j(\mathbf{x}) \leq 0, e_i(\mathbf{x}) = 0, \quad j = 1, \dots, l; \quad i = 1, \dots, m\}$
- Assumption: **Slater's Constraint Qualification holds for X.**

There exists $\mathbf{y} \in X$ such that $h_j(\mathbf{y}) < 0, \quad j = 1, \dots, l$

- If X satisfies Slater's Constraint Qualification, then the first order KKT conditions are necessary and sufficient for a global minimum of a convex programming problem CP

Now, there are some sets of problems for, which the first order KKT conditions are necessary and sufficient. So, consider the problem again a general non linear programming problem that we saw earlier. But, now there are some restrictions on this function. So, let us let us assume that the f and h_j 's are smooth convex functions. So, here smooth convex functions mean, that f and h_j 's are continuously differentiated.

Now, let us assume that $e_i^T x$ is of the type $a_i^T x - b_i$, for all i is going from 1 to m . So, that means, that all the constraints, equality constraints are affine constraints, now we know that, if $h_j(x)$ is a convex function the set $h_j(x) \leq 0$ is a convex set. So, the intersection of all convex sets is a convex set. So, the set of all inequality constraints under the assumptions that h_j is the convex function will form a convex set.

Now, if you take a intersection of those sets with the affine sets which, are of the type $e_i^T x = a_i^T x - b_i$ then, that intersection is also a convex set. Now here is a problem where, we want to minimize a convex function subject to a convex set. So, minimization of a convex function subject to a convex set is called a convex programming problem. And therefore, this program is also called a convex program and that is why, we are going to denote it by CP, CP stands for convex problem, programming problem or convex program.

So, CP is a convex programming problem, because we want to minimize a convex function with subject to a convex set. Now if you denote the feasible set by x then, this set x is a convex set, now let us assume that Slater's constraint qualification holds for x . So, we have seen this Slater's constraint qualification earlier. So, what the Slater's constraint qualification says is that the constraint set x does not have empty interior, so that means, there exists at least 1 y in the constraint set, such that $h_j(y) < 0$.

So, for all j is going from 1 to l , so this is a very important thing that for all j is going from 1 to l $h_j(y)$ should be less than 0. Now, here we have assumed that y belongs to x . So, which automatically means, that $e_i^T y = 0$ for all i going from 1 to m . So, that assumption is always true when, we talk about y belongs to x . So, it is just that, we have to ensure that, there exist at least 1 point which, is in the interior of the set.

So, if x satisfies Slater's constraint qualification then, the first order KKT conditions for this problem are necessary and sufficient for a global minimum of a convex programming problem CP. So, we saw a similar result for the case when, f was a convex function and $h_j(x)$ was $h_j(x) \leq 0$, where the constraint, where the inequality constraints, where $h_j(x)$ were convex functions. Now, we have added equality functions or equality constraints of the type $a_i^T x - b_i$. So, how do we extend the result that, we proved earlier.

(Refer Slide Time: 46:51)

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } h_j(x) \leq 0, \quad j=1 \rightarrow l \\ & \quad e_i(x) = 0, \quad i=1 \rightarrow m \end{aligned}$$

Let $x^* \in X$, $x \in X$. (x^*, λ^*, μ^*) is a KKT point

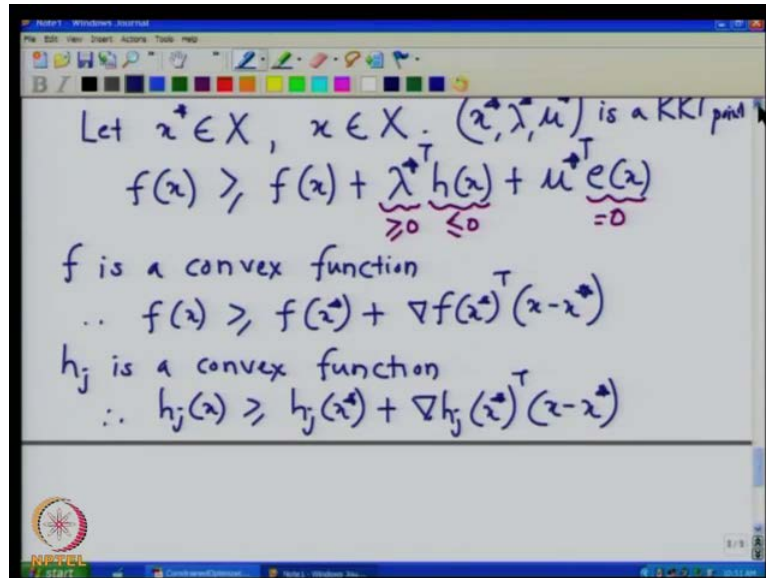
$$f(x) \geq f(x) + \underbrace{\lambda^{*T} h(x)}_{\geq 0} + \underbrace{\mu^{*T} e(x)}_{=0}$$

f is a convex function

$$\therefore f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*)$$

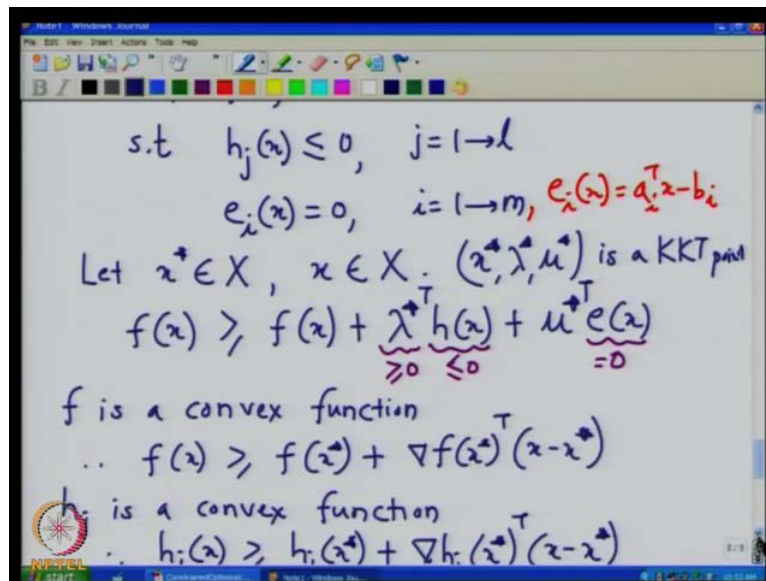
So, this is the problem that, we want to solve minimize $f(x)$ subject to $h_j(x) \leq 0$, j going from 1 to l and $e_i(x) = 0$, i going from 1 to m . Now, So, earlier case, we saw that $f(x) \geq f(x^*) + \lambda^{*T} h(x) + \mu^{*T} e(x)$. Now, here $h(x)$ contains all the inequality constraint $h_1(x) \leq 0, \dots, h_l(x) \leq 0$ and $\mu^{*T} e(x)$, now the reason for this is that let us assume that (x^*, λ^*, μ^*) is a KKT point. Now, if you look at this quantity, since x is feasible $h(x)$ is less than or equal to 0 and λ^* is greater than or equal to 0 and if, we look at $e(x) = 0$. So, this quantity is 0, this quantity is non positive quantity and therefore, we have $f(x) \geq f(x^*)$. Now, here we use our result from convex functions and we say that if f is a convex function. So, so f is convex and therefore, we can say that $f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*)$.

(Refer Slide Time: 50:13)



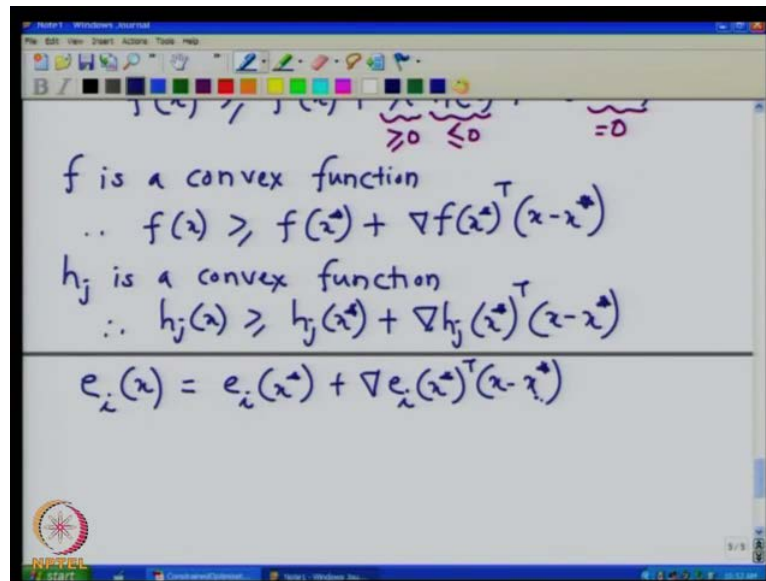
So, similarly we have h_j is a convex function and therefore, $h_j(x)$ is greater than or equal to $h_j(x^*) + \text{gradient } h_j(x^*)^\top (x - x^*)$, now $e_i(x)$ are of the type.

(Refer Slide Time: 51:02)



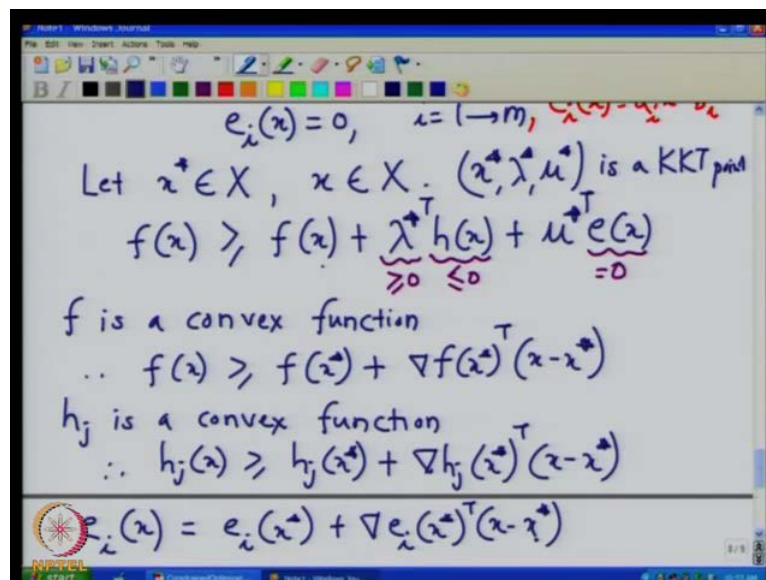
The $e_i(x)$ are of the type $a_i^\top x - b_i$.

(Refer Slide Time: 51:19)



And therefore since, they are of the type $e_i^T x - b_i$, we have $e_i^T x$ is equal to $e_i^T x^* + \text{gradient } e_i^T x^T (x - x^*)$.

(Refer Slide Time: 51:51)



So, now, if we combine all this, if we combine all this, then this function can be written as $f(x) \geq f(x^*) + \sum \lambda_j^* h_j(x^*) + \sum \mu_i^* e_i(x^*) + \text{the gradient of the lagrangian } \nabla L^T (x - x^*)$. So, if we combine this what we get is.

(Refer Slide Time: 52:21)

$$\dots f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*)$$

$$h_j \text{ is a convex function}$$

$$\therefore h_j(x) \geq h_j(x^*) + \nabla h_j(x^*)^T (x - x^*)$$

$$e_i(x) = e_i(x^*) + \nabla e_i(x^*)^T (x - x^*)$$

$$\therefore f(x) \geq f(x^*) + \sum_j \lambda_j^* h_j(x^*) + \sum_i \mu_i^* e_i(x^*)$$

$$+ \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*)^T (x - x^*)$$

So, therefore, f of x is greater than or equal to f of x^* plus sigma j lambda j^* $h_j(x^*)$ plus sigma i mu i^* $e_i(x^*)$ plus gradient of the lagrangian eval with respect to x , evaluated at x^* lambda * mu * transpose $(x - x^*)$. Now what, we do is that, we use the KKT conditions. So, since x^* lambda * mu * is a KKT point.

(Refer Slide Time: 53:24)

$$\min f(x)$$

$$\text{s.t. } h_j(x) \leq 0, \quad j=1 \rightarrow l$$

$$e_i(x) = 0, \quad i=1 \rightarrow m, \quad e_i(x) = a_i^T x - b_i$$

$$\text{Let } x^* \in X, \quad x \in X. \quad (x^*, \lambda^*, \mu^*) \text{ is a KKT point}$$

$$f(x) \geq f(x^*) + \underbrace{\lambda^{*T}}_{\geq 0} \underbrace{h(x^*)}_{\leq 0} + \underbrace{\mu^{*T}}_{=0} \underbrace{e(x^*)}_{=0}$$

$$f \text{ is a convex function}$$

We made this assumption earlier that, x^* lambda * mu * is a KKT point. So, we use this assumption along with the KKT conditions to see what happens to this expression.

(Refer Slide Time: 53:43)

$$\begin{aligned} \therefore f(x) &\geq f(x^*) + \nabla f(x^*)^T(x-x^*) \\ h_j &\text{ is a convex function} \\ \therefore h_j(x) &\geq h_j(x^*) + \nabla h_j(x^*)^T(x-x^*) \\ e_i(x) &= e_i(x^*) + \nabla e_i(x^*)^T(x-x^*) \\ \therefore f(x) &\geq f(x^*) + \sum_j \lambda_j^* \underbrace{h_j(x^*)}_{=0} + \sum_i \mu_i^* \underbrace{e_i(x^*)}_{=0} \\ &\quad + \underbrace{\nabla_x L(x^*, \lambda^*, \mu^*)^T}_{=0}(x-x^*) \\ \therefore f(x) &\geq f(x^*) \quad \forall x \in X \end{aligned}$$

So, since it is KKT point, this quantity is 0 then since x^* is a KKT point, because of the complimentary slackness condition. This quantity is also equal to 0. So, we have this quantity equal to 0, this quantity equal to 0 and x^* is the feasible point and therefore, this quantity is also equal to 0 and therefore, what we get is. So, therefore, $f(x)$ is greater than or equal to $f(x^*)$ for all x belongs to X and therefore, we get that x^* is a global minimum of CP the convex program. So, under the Slater's constraint qualification x^* , the first order KKT conditions are necessary and sufficient for a convex programming problem. Now, we will see some results related to second order conditions in the next class.

Thank you.