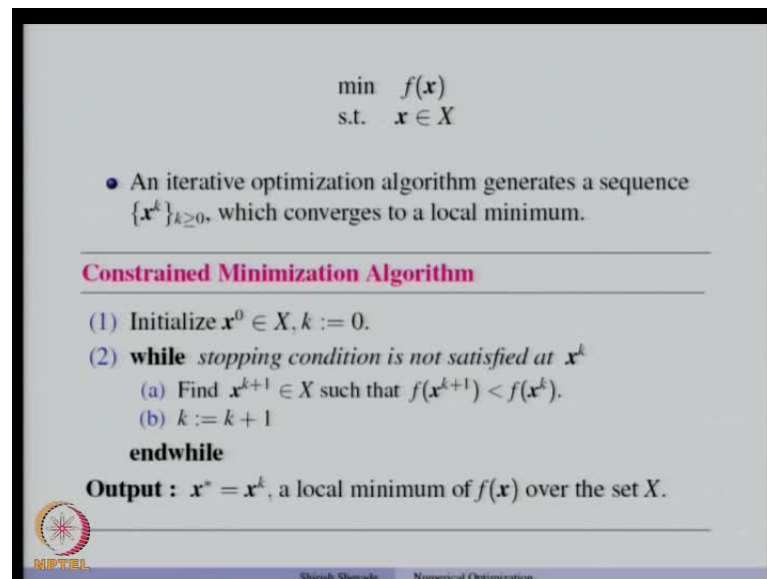


Numerical Optimization
Prof. Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Lecture - 21
Feasible and Descent Directions

(Refer Slide Time: 00:33)




$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in X \end{aligned}$$

- An iterative optimization algorithm generates a sequence $\{\mathbf{x}^k\}_{k \geq 0}$, which converges to a local minimum.

Constrained Minimization Algorithm

- (1) Initialize $\mathbf{x}^0 \in X, k := 0$.
- (2) **while** *stopping condition is not satisfied at \mathbf{x}^k*
 - (a) Find $\mathbf{x}^{k+1} \in X$ such that $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$.
 - (b) $k := k + 1$**endwhile**

Output : \mathbf{x}^* = \mathbf{x}^k , a local minimum of $f(\mathbf{x})$ over the set X .

 Shirish Shevade Numerical Optimization

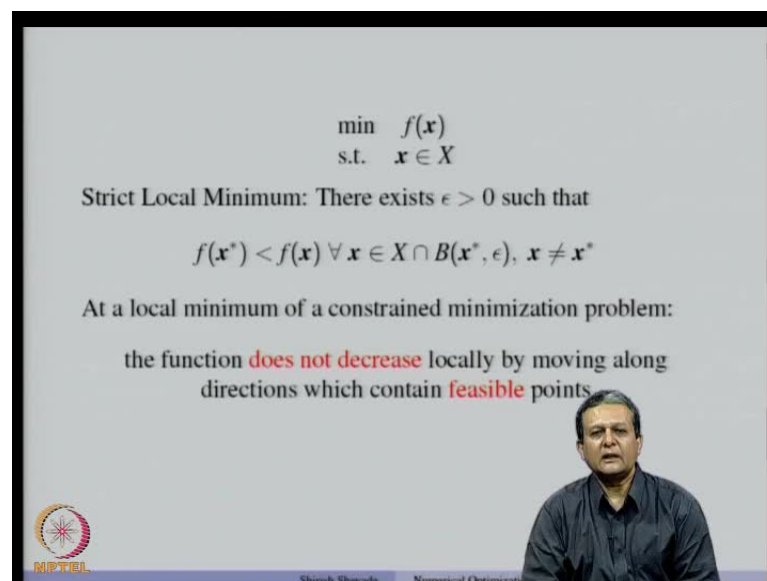
Hello, welcome back to this series of lectures on numerical optimization. In the last class, we started discussing about constrained optimization problems. And in particular way, we are looking at problem of this type, where we want to minimize an objective for f of x subject to the constrained that the every variable belongs to the set X . So, this set X is the feasible set or a constrained set, this is the objective function. And we are interested in giving some algorithm which is iterative algorithm that generates a sequence x^k , which converges to a local minimum of this problem.

So, a conceptual constrained minimization algorithm would look like this that given x^0 , which is feasible and setting a iteration counter to 0. We continue the algorithm till the stopping condition it satisfied at x^k . And at every step of the algorithm, what we do is that find x^{k+1} in the feasible set X , such that the value of the objective function at that point x^{k+1} is less than the value of the objective function at the current point. And then, we increase the iteration counter and the process is repeated. And finally, at

the end when the algorithm terminates we get a local minimum of $f(x)$ over the set X which is the feasible set so, so this is one of the solutions of this algorithm this problem.

Now, what we are interested in is that, what is this stopping condition that next to be satisfied at x_k ? If you recall for, for an unconstrained problem, we choose your stopping condition to be the norm of the gradient to be less than some epsilon. But that condition cannot be directly used for a constrained problem, because the gradient of the objective function may not vanish at a local minimum of a constrained problem. So, we need to come up with a different condition which can be use as a stopping criteria for this algorithm. And later on we will find a see different ways of finding x_{k+1} belong to belong to the feasible set such that the value of the objective function decreases. So, let us start looking at this stopping condition that needs to be satisfied at local minimum.

(Refer Slide Time: 02:48)



$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Strict Local Minimum: There exists $\epsilon > 0$ such that

$$f(x^*) < f(x) \quad \forall x \in X \cap B(x^*, \epsilon), \quad x \neq x^*$$

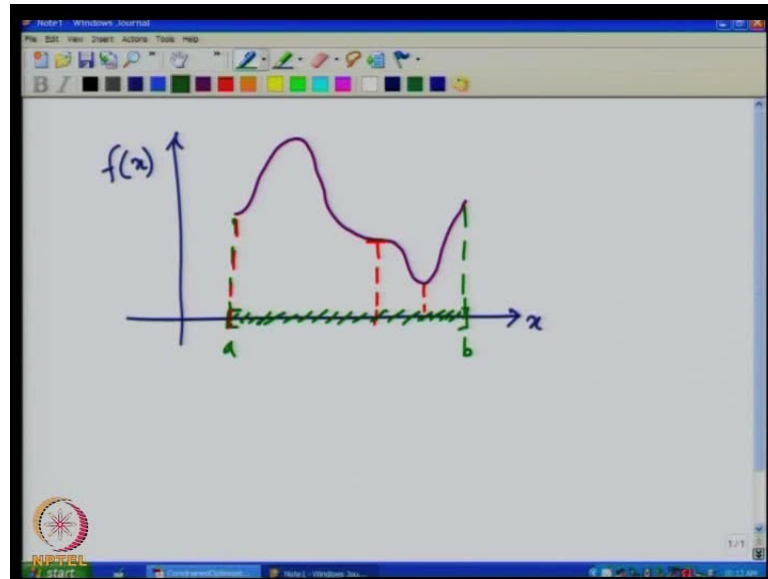
At a local minimum of a constrained minimization problem:
the function **does not decrease** locally by moving along
directions which contain **feasible** points

NIPTEP

Shreshth Shrivastava Numerical Optimization

So, consider this problem and recall the definition of strict local minimum that we discuss in the last class. So, if x^* is a strict local minimum if there exist some epsilon which is the positive quantity such that in the neighborhood of x^* in the set X , the value of the objective function f of x^* is strictly less than the value of $f(x)$ for every x in the neighborhood where x is not equal to x^* . So, so at a local minimum of constrained minimization problem you can see that the function does not decrease locally by moving along the directions which contain feasible point.

(Refer Slide Time: 03:50)



So, suppose we want to minimize a function, so suppose that this is an interval that we are looking at and the function is one dimensional, if suppose the function is something like this. So, we saw in the last class that this point a is a local minimum then this point is also local minimum. So, so if we consider this point; this point is also a local minimum and this point is also local minimum in addition to the point a. So, we saw that these are local minimum and this point and this point is a strict local minimum, while this point is not a strict local minimum, because if you see that in the neighborhood the function is a constrained.

So, if, if you consider this point and this is going to be our feasible region, so our feasible region is closed interval a b so which denoted here by this portion, so this is a our feasible region. So, if you move along the feasible region from the point a, the function strictly increases same is to here. So, if you move either from this point if you move either in the increasing direction of x or in the decreasing direction of x the function value strictly increases. So, and if you look at this point then in the neighborhood of this point, if you move the function at least, does not decrease. So, what we are interested in is finding out those points in whose neighborhood if we make a movement in the feasible set the function at least does not decrease.

(Refer Slide Time: 06:44)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in X \end{aligned}$$

Strict Local Minimum: There exists $\epsilon > 0$ such that

$$f(\mathbf{x}^*) < f(\mathbf{x}) \quad \forall \mathbf{x} \in X \cap B(\mathbf{x}^*, \epsilon), \mathbf{x} \neq \mathbf{x}^*$$

At a local minimum of a constrained minimization problem:
the function **does not decrease** locally by moving along directions which contain **feasible** points

- How to convert this statement to an algebraic condition?

NPTTEL Shreshth Shrivastava Numerical Optimization

So, in other words we are at a local minimum of constrained minimization problem, the function does not decrease locally by moving along directions which contain feasible points. Now, this is a statement which describes a local minimum. Now, how do we convert this statement into algebraic condition? That is what we want to see.

(Refer Slide Time: 07:08)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in X \end{aligned}$$

Definition

A vector $\mathbf{d} \in \mathbb{R}^n, \mathbf{d} \neq \mathbf{0}$ is said to be a *feasible* direction at $\mathbf{x} \in X$ if there exists $\delta_1 > 0$ such that $\mathbf{x} + \alpha \mathbf{d} \in X$ for all $\alpha \in (0, \delta_1)$.

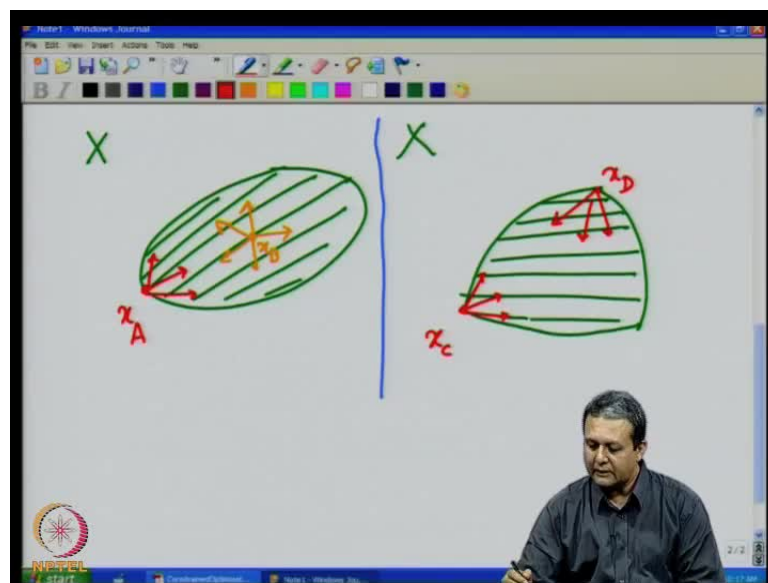
NPTTEL Shreshth Shrivastava Numerical Optimization

So, consider the same problem and it is define a set of feasible directions, so a vector a non zero vector \mathbf{d} in \mathbb{R}^n is said to be a feasible direction at point \mathbf{x} belong to set X . If there exist some δ_1 which is a positive quantity such that $\mathbf{x} + \alpha \mathbf{d}$ belongs to

the set X for all α in the range 0 to Δ . Note that we are not considering d equal to 0, because we are assuring that x is always belongs to the set X . So, we are considering all those points which are feasible.

And therefore, if we take d equal to 0 that is a trivial that this condition is trivially satisfied. So, we will not consider d equal to 0 in this case. So, we are interested in finding those direction is d such that x plus αd belongs to the set X for some α in the range 0 to Δ where Δ is a positive quantity. Now, let us see this definition.

(Refer Slide Time: 08:37)

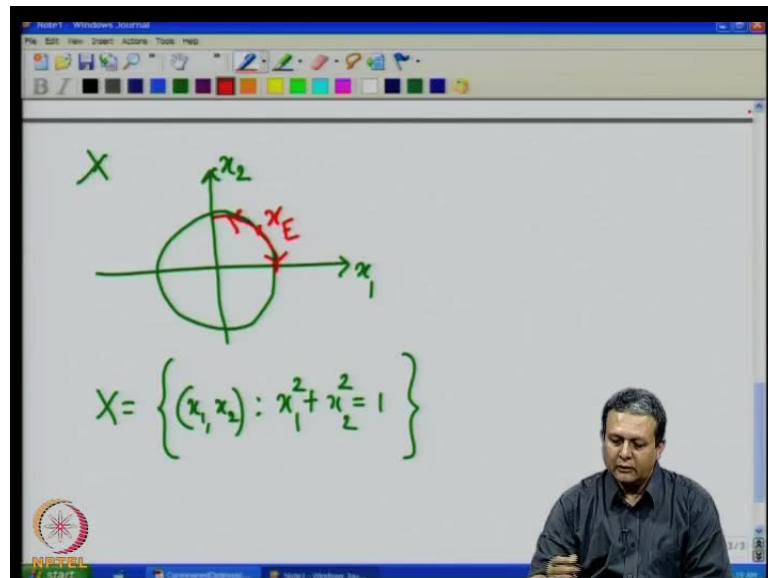


So, if we have a constrained set X to be set like this and let us consider a point. So, let us say that this point is x_A , so at x_A we are interested in finding out what are the feasible directions. So, you would see that some feasible directions are like this. So, if we make a small movement along these directions, we still retain feasibility of the set. On the other hand suppose if we take this point and then we can move in any direction in the neighborhood. So, these are all some of the feasible direction that we are draw here, so from this point if you consider the point x_A , we cannot move in the direction in the other directions, because then we will not return feasibility while if you consider the point x_B , locally we can move along any direction in the input space and we still remain feasible.

Now, let us consider another example, so let us consider another X which is like this. Now, let us consider a point some point, let us call it as x_C , so you will see that one can

move along this lines if you consider a point x D, one could move along. So, what, what is draw here is only some set of feasible directions and likewise we can draw feasible directions for a given set. Now, let us consider another example.

(Refer Slide Time: 11:30)



So, let us assume that our set X , this circle, so let me call this is x_1 and this is x_2 , so x is equal to set of all x_1, x_2 , such that $x_1^2 + x_2^2 = 1$. So, this is circle of radius 1 which is drawn here and so let us take a point. So, let us call this point as x_E and now suppose we want to find out what are the feasible directions at x_E is belongs to the set X ? Then you will see that we cannot find a straight line direction which is feasible from the point x_E . But what we can do that we can form, we can get what are call the curvilinear directions. So, if you move along this curves we can return feasibility, but these are not the straight line directions. So, what we are interested in is getting the straight line directions d , such that $x + \alpha d$ belongs to the set X for sufficiently small, α , positive α .

So, we will not be interested in the directions like this. Now, remember that in this example, what we have consider is only a set of directions or what are indicated here are only subset of the directions which are possible at a given feasible point. Now, how do we get all possible feasible directions and then show them algebraically, because this was this is one way to show the directions geometrically. But then how do we represent these directions algebraically? That is what we want to see.

(Refer Slide Time: 14:02)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in X \end{aligned}$$

Definition
A vector $\mathbf{d} \in \mathbb{R}^n, \mathbf{d} \neq \mathbf{0}$ is said to be a *feasible* direction at $\mathbf{x} \in X$ if there exists $\delta_1 > 0$ such that $\mathbf{x} + \alpha \mathbf{d} \in X$ for all $\alpha \in (0, \delta_1)$.

- Let $\mathcal{F}(\mathbf{x}) =$ Set of *feasible* directions at $\mathbf{x} \in X$ (w.r.t. X)

Definition
A vector $\mathbf{d} \in \mathbb{R}^n, \mathbf{d} \neq \mathbf{0}$ is said to be a *descent* direction at $\mathbf{x} \in X$ if there exists $\delta_2 > 0$ such that $f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x})$ for all $\alpha \in (0, \delta_2)$.

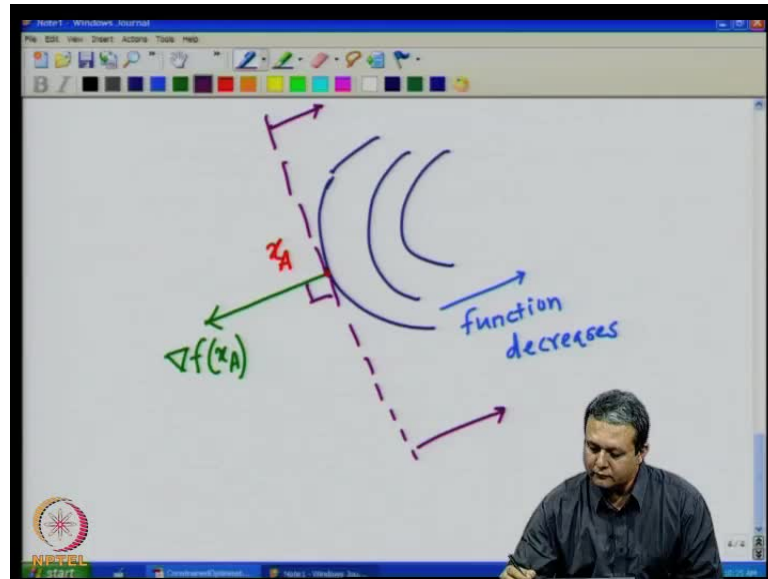
- Let $\mathcal{D}(\mathbf{x}) =$ Set of *descent* directions at $\mathbf{x} \in X$ (w.r.t. f)

NIPTEL Shreshth Sharma Numerical Optimization

So, let \mathcal{F} of \mathbf{x} the script \mathcal{F} of \mathbf{x} denote the set of feasible directions at \mathbf{x} belong to the feasible set X . So, remember that the feasible directions are always associated with the feasible set X so for given feasible set X and a given point \mathbf{x} which belong to the feasible set X , we can denote the set of feasible directions by \mathcal{F} of \mathbf{x} . Now, if you recall the definition of a local minimum in addition to the feasible direction, we need those directions which along which if we make a movement we can decrease the objective function. So, those, those directions are called descent directions. And we have seen this descent directions when we studied unconstrained optimization problems. So, nonzero vector \mathbf{d} in \mathbb{R}^n is said to be a descent direction at \mathbf{x} . If there exists δ_2 which is the positive quantity such that f of \mathbf{x} plus $\alpha \mathbf{d}$ is less than f of \mathbf{x} for all α in the range 0 to δ_2 .

So, note that the descent directions is descent direction is always associated with the objective function, while in the definition of feasible directions you will not see any mention of the objective function, because of feasible directions are always associated with a constrained set and the descent directions are always associated with objective function which is to be optimized. In this case we want to minimize this objective function f of \mathbf{x} , so the descent direction is associated with f . So, so let us denote by the script \mathcal{D} in the set of feasible directions \mathbf{x} belongs to \mathbf{x} and as I said that the descent directions are always with respect to some objective function f .

(Refer Slide Time: 16:28)



Now, now we have seen that the contours of a typical objective function would look like this, these are the contours of the objective function. Now, let us assume that the objective function is differentiable. So, at a given point, so let us take a point. And, let us call this is x_A , so let us call this is x_A . Now, if the function is differentiable, we can find out the gradient of the function. And, let us assume that the gradient of the function is pointing in this direction. Now, we are seen earlier in when we studied unconstrained optimization that the gradient points in the direction where the function increases. So, so along this direction the function decreases. Now, we also saw earlier when we studied unconstrained optimization problems in that we take tangent plane. So, this is orthonormal to at direction then so if we look at this set, set of all directions which make an off choose angle with a gradient off at x_A . So, if we make a movement along those directions then the function value decreases, so those are the descent directions and so we have seen this earlier.

(Refer Slide Time: 19:10)

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

- $X = \{\mathbf{x} \in \mathbb{R}^n : h_j(\mathbf{x}) \leq 0, e_i(\mathbf{x}) = 0, j = 1, \dots, l, i = 1, \dots, m\}$
- At a local minimum $\mathbf{x}^* \in X$, the function does not decrease by moving along feasible directions

NPTEL Shreshth Shrivastava Numerical Optimization

So, now let us try to would this conditions in algebraic form, now to start with let us consider a problem where we want to minimize the function f of x subject to the any quality constrained of the type $h_j(x)$ less than or equal to 0. And, the equality constrained of the type $e_i x$ equal to 0 where x is any vector in in dimensional space. And let us consider the feasible set X defined using the constrained and x , we mentioned earlier that at a local minimum of x star of x , the function does not decrease by moving along feasible direction.

(Refer Slide Time: 19:52)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in X \end{aligned}$$

Theorem
Let X be a nonempty set in \mathbb{R}^n and $\mathbf{x}^* \in X$ be a local minimum of f over X . Then, $\mathcal{F}(\mathbf{x}^*) \cap \mathcal{D}(\mathbf{x}^*) = \phi$.

Proof.
Let $\mathbf{x}^* \in X$ be a local minimum.
By contradiction, assume that \exists a nonzero $\mathbf{d} \in \mathcal{F}(\mathbf{x}^*) \cap \mathcal{D}(\mathbf{x}^*)$.
 $\therefore \exists \delta_1 > 0 \ni \mathbf{x}^* + \alpha \mathbf{d} \in X \forall \alpha \in (0, \delta_1)$ and
 $\exists \delta_2 > 0 \ni f(\mathbf{x}^* + \alpha \mathbf{d}) < f(\mathbf{x}^*) \forall \alpha \in (0, \delta_2)$.
Hence, $\exists \mathbf{x} \in B(\mathbf{x}^*, \alpha) \cap X \ni f(\mathbf{x}) < f(\mathbf{x}^*)$, for $\alpha \in (0, \min(\delta_1, \delta_2))$. This contradicts the assumption that \mathbf{x}^* is a local minimum. \square

NPTEL Shreshth Shrivastava Numerical Optimization

So, let us consider this problem and so we have a theorem which characterizes local minimum of this problem, so x is a nonempty set in \mathbb{R}^n and x^* belong to x the feasible region is a local minimum off over x then the set of feasible directions and a set of descent directions they have empty intersection. So this is the very important result and this gives a necessary condition for the local minimum of given constrained optimization problem. So, let us look at the proof of this theorem, so let us assume that a x^* is feasible and is a local minimum. Now, the proof will be given by the method of contradiction and for that we assume that there exists a nonzero d , such that that lies in the intersection of F_{x^*} and D_{x^*} . So, let us assume that the, such as D exist and that therefore, this set $F_{x^*} \cap D_{x^*}$ is nonempty.

Now, since d belongs to F_{x^*} the set of feasible directions. So, by the definition of the set of feasible directions, we can say that there exists δ_1 greater than 0 such that $x^* + \alpha d$ belongs to the set X for all α in the range 0 to δ_1 . And similarly, one can use the definition of D_{x^*} to say that there exists δ_2 greater than 0 such that $f(x^* + \alpha d)$ is less than $f(x^*)$ for all, all α in the range 0 to δ_2 . So, there exists δ_1 such that direction d is feasible and there exists δ_2 such that the direction d is also the descent direction. Now, if you take a minimum of δ_1 and δ_2 . Then you can see that along that along a direction d as long as α is in the range 0 to $\min(\delta_1, \delta_2)$ $x^* + \alpha d$, it was belongs to X and $f(x^* + \alpha d)$ is less than $f(x^*)$.

So, that means that there exists x in the neighborhood of x^* and which is we intersection that α neighborhood with x such that $f(x)$ is less than $f(x^*)$ for every α in the range 0 to $\min(\delta_1, \delta_2)$. And so that means that in the local neighborhood of x^* we are able to find some x which is not equal to x^* such that the value of the function at x is strictly less than the value of the function at $f(x^*)$, and that contradicts the fact that x^* is a local minimum. And therefore, because of this contradiction, we cannot have the direction d which is in the intersection of the feasible set and the descent set. If x^* is a local minimum so this contradicts our assumption that x^* is a local minimum.

(Refer Slide Time: 23:49)

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in X \end{aligned}$$

- $\mathbf{x}^* \in X$ is a local minimum $\Rightarrow \mathcal{F}(\mathbf{x}^*) \cap \mathcal{D}(\mathbf{x}^*) = \phi$
- Consider any $\mathbf{x} \in X$ and assume $f \in \mathcal{C}^2$
- $\lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha} = \nabla f(\mathbf{x})^T \mathbf{d}$
- $\nabla f(\mathbf{x})^T \mathbf{d} < 0 \Rightarrow f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x}) \Rightarrow \mathbf{d}$ is a descent direction $\Rightarrow \mathbf{d} \in \mathcal{D}(\mathbf{x})$
- Let $\tilde{\mathcal{D}}(\mathbf{x}) = \{\mathbf{d} : \nabla f(\mathbf{x})^T \mathbf{d} < 0\} \subseteq \mathcal{D}(\mathbf{x})$
- $\mathbf{x}^* \in X$ is a local minimum $\Rightarrow \mathcal{F}(\mathbf{x}^*) \cap \tilde{\mathcal{D}}(\mathbf{x}^*) = \phi$

NIPTEL Shrish Shrivastava Numerical Optimization

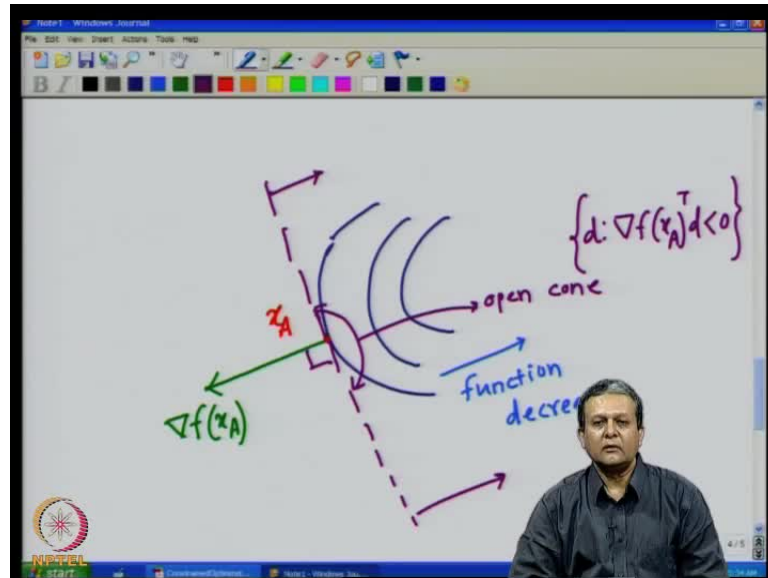
Now, first let us look at the way of characterizing on that this set of descent directions. And as I mentioned earlier that if we assume that the function is sufficiently smooth then the set of the descent directions can be characterize using the gradient of the objective function. So, let us consider any \mathbf{x} in the feasible set X and assume that f is in \mathcal{C}^2 , in fact for current analysis we do not need f to be in the class of \mathcal{C}^2 function, we just need class of f , f belongs to \mathcal{C}^1 that is the class of continuously differentiable function. But later on when we move on to the second order condition, we will need this condition so I have assume that f is in the class of twice continuously differentiable functions.

Now, we already know by the definition of directional derivative that gradient effect transpose \mathbf{d} is nothing but limit is α tends to 0 on the positive side f of \mathbf{x} plus $\alpha \mathbf{d}$ minus f \mathbf{x} by α . So, if this quantity is less than 0 then we can say that in the neighborhood of \mathbf{x} where α is greater than 0 in a sufficiently small neighborhood of \mathbf{x} . This numerator will be less than 0, so that is gradient effect transpose the less than 0, it implies f of \mathbf{x} plus $\alpha \mathbf{d}$ less than f \mathbf{x} . And this means that, so this this will be true for sufficiently small positive α . And therefore, we can say that \mathbf{d} is a descent direction and therefore, we can say that \mathbf{d} belongs to descent $\mathcal{D} \mathbf{x}$.

So, if we are able to find a direction \mathbf{d} such that the direction, mention off choose angle with gradient effects then we can say that \mathbf{d} is a descent direction. So, let us define set \mathcal{D} tilde to be the set of all directions of \mathbf{d} such that those directions make and acute angle

with gradient of f of x so clearly $D \tilde{x}$ is a subset of $D x$. And therefore, we can say that x^* is a local minimum if F of x^* intersection $D \tilde{x}^*$ is a null set or they do they have empty intersection. And as, as I mentioned earlier that.

(Refer Slide Time: 27:09)



So, if we look at this figure then this open cone, cone of directions these are the directions the set of d such that gradient f of x_A transpose d is less than 0. So, this open cone uses the set of descent directions at x_A for the function f so if you make a movement along any direction in this cone that will be a descent direction.

(Refer Slide Time: 27:59)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

- $x^* \in X$ is a local minimum $\Rightarrow \mathcal{F}(x^*) \cap \mathcal{D}(x^*) = \phi$
- Consider any $x \in X$ and assume $f \in C^2$
- $\lim_{\alpha \rightarrow 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha} = \nabla f(x)^T d$
- $\nabla f(x)^T d < 0 \Rightarrow f(x + \alpha d) < f(x) \Rightarrow d$ is a descent direction $\Rightarrow d \in \mathcal{D}(x)$
- Let $\tilde{\mathcal{D}}(x) = \{d : \nabla f(x)^T d < 0\} \subseteq \mathcal{D}(x)$
- $x^* \in X$ is a local minimum $\Rightarrow \mathcal{F}(x^*) \cap \tilde{\mathcal{D}}(x^*) = \phi$
- If $\mathcal{F}(x^*) = \mathbb{R}^n$ (every direction in \mathbb{R}^n is locally feasible), $x^* \in X$ is a local minimum $\Rightarrow \{d : \nabla f(x^*)^T d < 0\} = \phi \Rightarrow \nabla f(x^*) = \mathbf{0}$
- Can we characterize $\mathcal{F}(x^*)$ algebraically for a constrained optimization problem?

Shreshth Shekhar | Numerical Optimization

So, D_x is a way to characterize D_x in terms of the gradient of the function, so we were able to convert the D_x in the original condition by D_x as D_x is a subset of D_x and moreover D_x can be written in terms of the gradient of f at a given point x . Now, so if you consider f of x^* to be \mathbb{R}^n that means that every direction \mathbb{R}^n is locally feasible then x^* belongs to X is a local minimum that implies that this set has to vanish, because if this set is \mathbb{R}^n . Then the D_x x^* should x^* belong to X is a local minimum like that D_x x^* should be a null set and that is the set of all descent set gradient effect star transposes less than 0 is a null set.

And this will be possible when gradient effect star is equal to is a 0 vector. And this condition confirms with our unconstrained optimization, condition optimization problems where we discuss about the local minimum of unconstrained optimization problem. So, there we saw that x^* is a local minimum implies that gradient effect star is equal to 0. So, so, we have seen one way to characterize the said x^* . Now, how do we characterize f of x^* algebraically for a given constrained set X ? So that is what we will be seen now.

(Refer Slide Time: 30:05)

Consider the problem:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

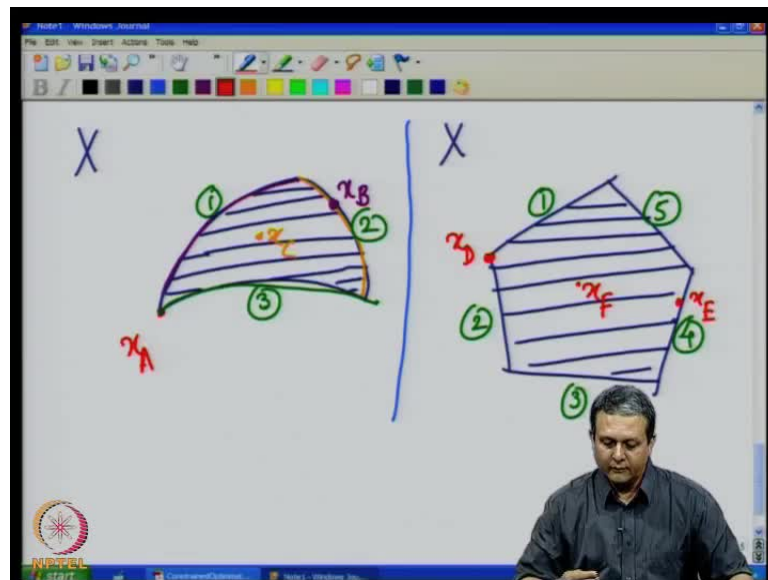
- Assume $f, h_j \in C^2, j = 1, \dots, l$
- $X = \{\mathbf{x} \in \mathbb{R}^n : h_j(\mathbf{x}) \leq 0, j = 1, \dots, l\}$
- **Active constraints:** $\mathcal{A}(\mathbf{x}) = \{j : h_j(\mathbf{x}) = 0\}$

NPTEL
Shreshth Shrivastava Numerical Optimization

So for the time being let us assume that we are interested in solving this problem where we want to minimize f of x subject to only the inequality constrained of the type h_j x less than or equal to 0 and there are one such inequalities and x is a point in n dimensional space of real numbers. Now, again we assume that f and h_j is belongs to C

2 although for the present analysis it is enough that f and x_j belongs to C^1 . But later on as I mention earlier that we will need the twice continuous differentiability, so will assume that the functions are twice continuously differentiable. Now, so let us denote X the set of all x_j less than or equal to 0 j going from 1 to l .

(Refer Slide Time: 31:00)



Now, let us consider a constrained set that we saw earlier the set X suppose a set X is so this is a set X . Now, let us consider a point now this constrained set is made up of says 3 inequalities which are shown here. So, one inequality is introduced in this the other is denoted using this and then the third one is denoted is in the green color. So, we have 3 inequalities which where use to form this constrained set and if we look at a point x_A will see that the only two constrained are used to represent the point x_A . So, this third constrained even if you perturb a little bit that is not going to change the representation of the point x_A . So, in this case these two constraints are set to be active constraints as where as the point x_A is concern, so if let us number this constraints so let us call this is 1; this is 2 and this is 3.

So, at x_A , the constraints 1 and 3 are active. Now, if you take a point, so let us take a another point x_B which is also feasible point, so at the point x_B the constraints 2 is active, so even if we move the constraints 1 and 3 a little bit point x_B is not affected while a slide perturbation of the constraints 1 and 3. And if we consider a point x_C so let us consider a point x_C which is in the interior of the set.

Now, we will see that this point at this point $x \in C$ no constraints are active that means that even if you perturb any of these constraints a little bit, the point x still remains feasible and does not get rejected. So, at a given point feasible point when we want to characterize the set X , the feasible set algebraically or when we want to characterize the set of feasible directions at $x \in A$ for the feasible set X algebraically, what do we have to do that? We just have to consider what are called active constraints, sometimes they are also called binding constraints.

So, if we consider another example, so suppose we, if we have a feasible set X . And let us number the constraints, so we have 5 inequality constraints and the set is found using the intersection of those 5 inequality constraints and if we consider a point. So, let us call this point as $x \in D$, we will see that the only the constraints 1 and 2 are active while if you consider a point $x \in E$ only the constraint 4 is active as far else the point x is concerned and in the interior, if we consider a point $x \in F$ no, no constraints are active.

Or in other words we can define active constraints as those in constraints which are satisfied at a given point with equality. So, if you look at $x \in A$ the constrained set 1 and 3 are satisfied with equality, at $x \in D$ the constraints 1 and 2 are satisfied with equality. And all the other constraints are inactive the constraints 3, 4 and 5, because they are not satisfied with equality sign. And similarly, at $x \in F$, none of the constraints is active as the point does not satisfy these constraints with strict inequality.

(Refer Slide Time: 36:45)


Consider the problem:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

- Assume $f, h_j \in C^2, j = 1, \dots, l$
- $X = \{\mathbf{x} \in \mathbb{R}^n : h_j(\mathbf{x}) \leq 0, j = 1, \dots, l\}$
- **Active constraints:**

$$A(\mathbf{x}) = \{j : h_j(\mathbf{x}) = 0\}$$

Lemma
For any $\mathbf{x} \in X$,

$$\tilde{\mathcal{F}}(\mathbf{x}) \triangleq \{\mathbf{d} : \nabla h_j(\mathbf{x})^T \mathbf{d} < 0, j \in A(\mathbf{x})\} \subseteq \mathcal{F}(\mathbf{x})$$


Shreshth Shrivastava Numerical Optimization

So, let us define the set of active constraints $A(x)$ to be the set of all j 's such that $h_j(x)$ is equal to 0, so though set of all constraints which are satisfied with equality. Now, here is an interesting result that for any feasible x , if you defined $\tilde{F}(x)$ will be the set of all directions d such that $\nabla h_j(x)^T d < 0$ that means that the directions which make an obtuse angle with a gradient $\nabla h_j(x)$ for all j in the set $A(x)$ the set of active constraints.

So, if you collect all the active constraints and find out the set $\tilde{F}(x)$, the set of directions which makes an obtuse angle with all the gradients of the active constraints. So, let us denote that set by $\tilde{F}(x)$ and the result says that this $\tilde{F}(x)$ is a subset of $F(x)$ or is a set of all feasible directions at x . So, the important thing about this result is that now we are able to get a set in terms of the gradient of the active constraints and that set is shown to be a subset of $F(x)$. And this set $\tilde{F}(x)$ will be useful for writing optimality conditions.

(Refer Slide Time: 38:32)

Lemma
For any $x \in X$,

$$\tilde{F}(x) \triangleq \{d : \nabla h_j(x)^T d < 0 \quad j \in \mathcal{A}(x)\} \subseteq \mathcal{F}(x)$$

Proof.
Suppose $\tilde{F}(x)$ is nonempty and let $d \in \tilde{F}(x)$. Since $\nabla h_j(x)^T d < 0 \quad \forall j \in \mathcal{A}(x)$, d is a descent direction for h_j , $j \in \mathcal{A}(x)$ at x . That is,

$$\exists \delta_1 > 0 \quad \exists h_j(x + \alpha d) < h_j(x) = 0 \quad \forall j \in \mathcal{A}(x).$$

Further, $h_j(x) < 0 \quad \forall j \notin \mathcal{A}(x)$. Therefore,

$$\exists \delta_3 > 0 \quad \exists h_j(x + \alpha d) < 0 \quad \forall \alpha \in (0, \delta_3), \quad \forall j \notin \mathcal{A}(x)$$

Thus, $x + \alpha d \in X \quad \forall \alpha \in (0, \min(\delta_1, \delta_3))$, and $\therefore d \in \mathcal{F}(x)$. \square

NPTEL Shresh Shekhar Numerical Optimization

So, let us look at the proof of this lemma, so the claim is that $\tilde{F}(x)$ is a subset of $F(x)$. Now, what we have to show is that if we take a new vector d which is in $\tilde{F}(x)$ that vector d is also in the set $F(x)$. So, let $\tilde{F}(x)$ be nonempty, so that there is this some direction non zero direction d such that d is in $\tilde{F}(x)$. Now, since that its since $\nabla h_j(x)^T d < 0$ so that is a d makes an obtuse angle with the other gradients of the active constraints, we know that d is a descent direction for the

function h_j where j belongs to x we have seen this result to when we talked about the objective function. So similar results holds here that d is a descent direction for the for the function h_j or in other words there exists some δ_1 greater than 0 such that h_j of x plus αd is less than h_j of x for all j in the set of active constraints set x . And we know that since j belongs to A h_j of x equal to 0 that means that h_j of x plus αd is less than 0 which means that d is a, or these also a feasible direction.

Now, if we consider all the, now although constraints which are not active at the current point x for them we can write that there exists some δ_3 greater than such that h_j of x plus αd is less than 0 for all α in the range 0 to δ_3 and j not in the set of active constraints. So, for the set of active constraints we were able to find a direction d of we were able to find δ_1 such that h_j of x plus αd is feasible and for those constraints which are not active, we were able to find δ_3 such that h_j of x plus αd is less than 0. Now, if you take a minimum of δ_1 and δ_3 then you we can see that h_j of x plus αd is less than 0 which means that it belongs to the feasible set.

So, x plus αd belongs to the feasible set X for all α in the range 0 to minimum of δ_1 and δ_3 and that means that d is a feasible direction, because by definition. We have found some δ_2 which is greater than 0 such that x plus αd is a feasible point for all α in the range 0 to δ_2 , δ_2 is nothing but we know δ_1 and δ_3 . So, this important lemma let us characterize a feasible set by using the gradients of active constraints at a given point at a given feasible point.

(Refer Slide Time: 42:27)


$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Let $X = \{\mathbf{x} \in \mathbb{R}^n : h_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, l\}$. For any $\mathbf{x} \in X$,
 $\tilde{\mathcal{F}}(\mathbf{x}) \triangleq \{\mathbf{d} : \nabla h_j(\mathbf{x})^T \mathbf{d} < 0 \quad j \in \mathcal{A}(\mathbf{x})\} \subseteq \mathcal{F}(\mathbf{x})$ and
 $\tilde{\mathcal{D}}(\mathbf{x}) \triangleq \{\mathbf{d} : \nabla f(\mathbf{x})^T \mathbf{d} < 0\} \subseteq \mathcal{D}(\mathbf{x})$.

$$\begin{aligned} \mathbf{x}^* \in X \text{ is a local minimum} &\Rightarrow \mathcal{F}(\mathbf{x}^*) \cap \mathcal{D}(\mathbf{x}^*) = \phi \\ &\Rightarrow \tilde{\mathcal{F}}(\mathbf{x}^*) \cap \tilde{\mathcal{D}}(\mathbf{x}^*) = \phi \end{aligned}$$

$\mathbf{x}^* \in X \text{ is a local minimum} \Rightarrow \tilde{\mathcal{F}}(\mathbf{x}^*) \cap \tilde{\mathcal{D}}(\mathbf{x}^*) = \phi$

- This is only a necessary condition for a local minimum
- Utility of this condition depends on the constraint representation
- Cannot be directly used for equality constrained problems



Shreshth Shrivastava Numerical Optimization

Now, we can now combine this results which we saw earlier, so let us denote the constraints set by the capital by the set capital X. And for any feasible set X in the feasible set, we have already defined F tilde x to be the set of all the directions d such that those directions d make an off choose angle with the gradients of the active constraints and that F tilde x is a subset of F x. And similarly, D tilde x we defined it to be set of all directions d which make an off choose angle with a gradient f x and that the subset of the set of descent directions.

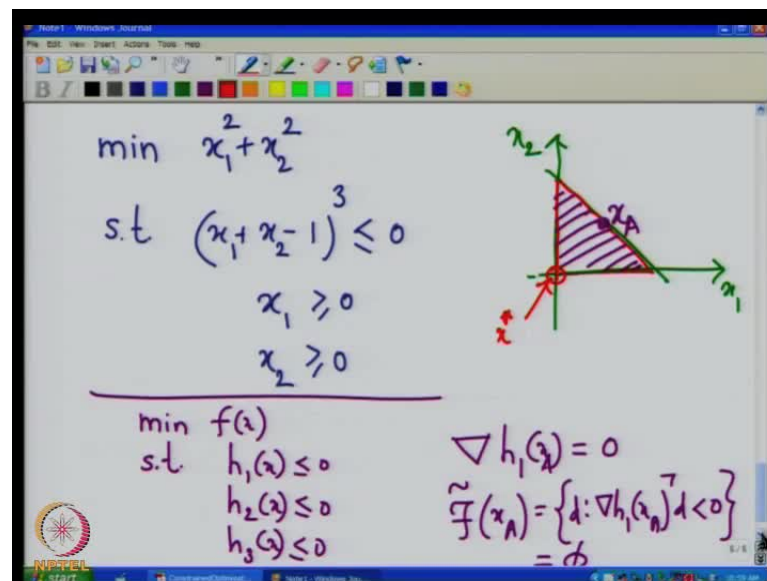
And discuss an important result, and we saw earlier that x star in a feasible set is a local minimum implies f f x star. The feasible set at x star and the descent, the set of descent direction set X star they have a empty intersection and since a F tilde x is a subset of F x D tilde x is a subset of D x. We can write this as the condition where the, the sets F tilde x star and the sets D tilde x star they have a empty intersection because F tilde is a subset of f and D tilde is a subset of d.

Now, the reason for doing this is that this condition is written in terms of the gradient of the objective function and the gradients of the active constraints and so it can be written in algebraic form. So, this is the result that we are interested in that x star which is feasible is the local minimum implies that F tilde x star intersection D tilde x star is a null set. Now, remember that this is only a necessary condition for local minimum, because will see later that there could be situations where F tilde x star is a null set or D

\tilde{x}^* is a null set and which will automatically mean that this set is a null set, but that may not mean that x^* is a local minimum, so this is just one necessary condition. And the utility of this condition depends on the constraint representation. So, this is a very important point and we will give some examples about that and this condition cannot be directly used for equality constraints problem say if you write the equality constraints as a set of inequality constraints problem. Now, so let us look at this point first that this is only a necessary condition.

Now, suppose at, at a given feasible point x the gradient of the function vanishes so which means that if the gradient of a function vanishes, so gradient effect is 0. Then the set \tilde{D} is a null set and if the set \tilde{D} is null set then this set $\tilde{F} \cap \tilde{D}$ is a null set. And that a does not guaranteed at x is a local minimum, so this is a very important point that needs to be remember. Now, the utility of the condition depends on the constraints representation, so let us look at the some example.

(Refer Slide Time: 46:48)



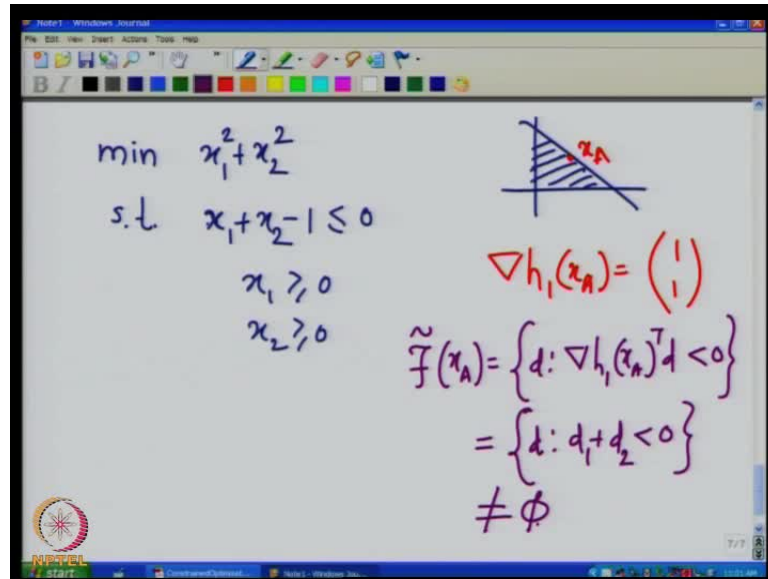
So, let us consider a problem minimize $x_1^2 + x_2^2$ subject to $x_1 + x_2 - 1$ cube is less than or equal to 0, x_1 greater than or equal to 0, x_2 greater than or equal to 0. Now, if we consider the feasible region, so we have x_1 and x_2 along the axis. And so this constraints does mean that $x_1 + x_2 - 1$ has to be less than or equal to 0, so that x cube is also less than or equal to 0. So, this constraints is represented using this and then a so this is the first constraints and let the second constraints is the x_1

greater than or equal to 0, so it is this. And let the third constraints, so together this forms the constraints set. And now if we write this function of the problem is minimize $f(x)$ subject to $h_1(x) \leq 0$, $h_2(x) \leq 0$ and $h_3(x) \leq 0$.

Now, if you take any point, so let us take a point on this, so let this point be x . Now, at this point if you take gradient at $h_1(x)$ so let me call this point x_A . So, gradient $h_1(x)$ is equal to 0, because this point satisfies $x_1 + x_2 - 1 = 0$ and the gradient of this quantity with respect to that will be a 0 vector. So, once we have a 0 vector if we start looking at the set of a feasible directions at x_A , and then represent them using the said that we saw earlier. So, we saw that at given point x_A the one of the gradients in fact there was only 1 active constraints so that for that active constraints the gradient is 0. And therefore, the $F_{\tilde{x}}$ will be a null set and that is clear from this figure.

So, if you write $F_{\tilde{x}_A}$ is equal to set of all d 's such that gradient $h_1(x_A)^T d$ is less than 0 and since gradient of this active. Remember that at x_A this is the only active constraints, the first constraints is only the active constraints and this is a null set. And therefore, $F_{\tilde{x}_A} \cap D_{\tilde{x}_A}$ will be a null set. But if you look at this problem formulation, so if you look at this problem formulation what we are interested in is minimizing $x_1^2 + x_2^2$ subject to this. So, you will see that this point the origin this is going to be our x^* and certainly x_A is not a local minimum, but suppose if we represent this constraints.

(Refer Slide Time: 51:59)



So, if we consider a same problem minimize x_1 square plus x_2 square subject to x_1 plus x_2 minus 1 less than or equal to 0 and x_1 greater than or equal to 0 and x_2 greater than or equal to 0. But if you look at this problem formulation, so if you look at this problem formulation what we are interested in is minimizing x_1 square plus x_2 square subject to this. So, you will see that this point the origin this is going to be our x^* and certainly x_A is not a local minimum, but suppose if we represent this constraints. Now, again if we take the point x_A , so this is going to be our feasible region. And if we take the point x_A , so gradient $h_1(x_A)$ will be 1 1 and which is not 0. So, the same feasible region if we represent it using a different constraints then we get a gradient vector which is a non-zero vector and this is different, this is unlike the previous case where we have the gradient vector to be 0.

(Refer Slide Time: 53:11)

$$\text{s.t. } (x_1 + x_2 - 1) \leq 0$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$

$$\min f(x)$$

$$\text{s.t. } h_1(x) \leq 0$$

$$h_2(x) \leq 0$$

$$h_3(x) \leq 0$$

$$\nabla h_1(x) = 0$$

$$\tilde{F}(x_A) = \{d \mid \nabla h_1(x_A)^T d < 0\}$$

$$= \emptyset$$

$$\min x_1^2 + x_2^2$$

So, remember that this is a 0 vector, so you will see that the same constraints said if it is represented in a different way. Then we could avoid the $\tilde{F}(x_A)$ from not becoming a null set, so if you write the $\tilde{F}(x_A)$ here.

So, $\tilde{F}(x_A)$ will be the set of all directions d such that $\nabla h_1(x_A)^T d < 0$. And this is nothing but the set of directions d such that $d_1 + d_2 < 0$. And this set is not a null set, so we were able to get a $\tilde{F}(x_A)$ which is not a null set and clearly we know that x_A is not a local minimum of this problem. Now, another important point that I have to notice that this condition cannot directly be used for equality constraints problems and we will see that now.

(Refer Slide Time: 54:39)

The image shows a whiteboard with the following handwritten content:

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

$$h_1(x) \leq 0 \rightarrow -(x_1 + x_2) \leq -1$$

$$h_2(x) \leq 0 \rightarrow x_1 + x_2 \leq 1$$

$$\tilde{F}(x) = \{d : \nabla h_1(x)^T d < 0, \nabla h_2(x)^T d < 0\} = \emptyset$$

So, if we minimize x_1 square plus x_2 square subject to x_1 plus x_2 equal to 1 this is our equality constraints problem. Now, this problem we can write this as minimize x_1 square plus x_2 square subject to x_1 plus x_2 greater than or equal to 1 and x_1 plus x_2 less than or equal to 1. Now, this constrain, we can write this as minus x_1 plus x_2 less than or equal to minus 1. So, so $h_1(x)$ less than or equal to 0 is a given is this constraints and $h_2(x)$ less than or equal to 0 is the constraints x_1 plus x_2 less than or equal to 1.

Now, you will see that for a given feasible point if we take the gradient $h_1(x)$ transpose d less than 0 and gradient $h_2(x)$ transpose d less than 0 that set will be null set. So if we write a equality constraints in this form then we will always get $\tilde{F}(x)$ to be a null set. And so the any feasible point $\tilde{F}(x)$ is becomes a null set and that does not guaranty that x is the local minimum. Now, will see more about this condition and how to write it in the actual algebraic form in terms of the gradients of f and then the gradient of the active constraints in the next class.

Thank you