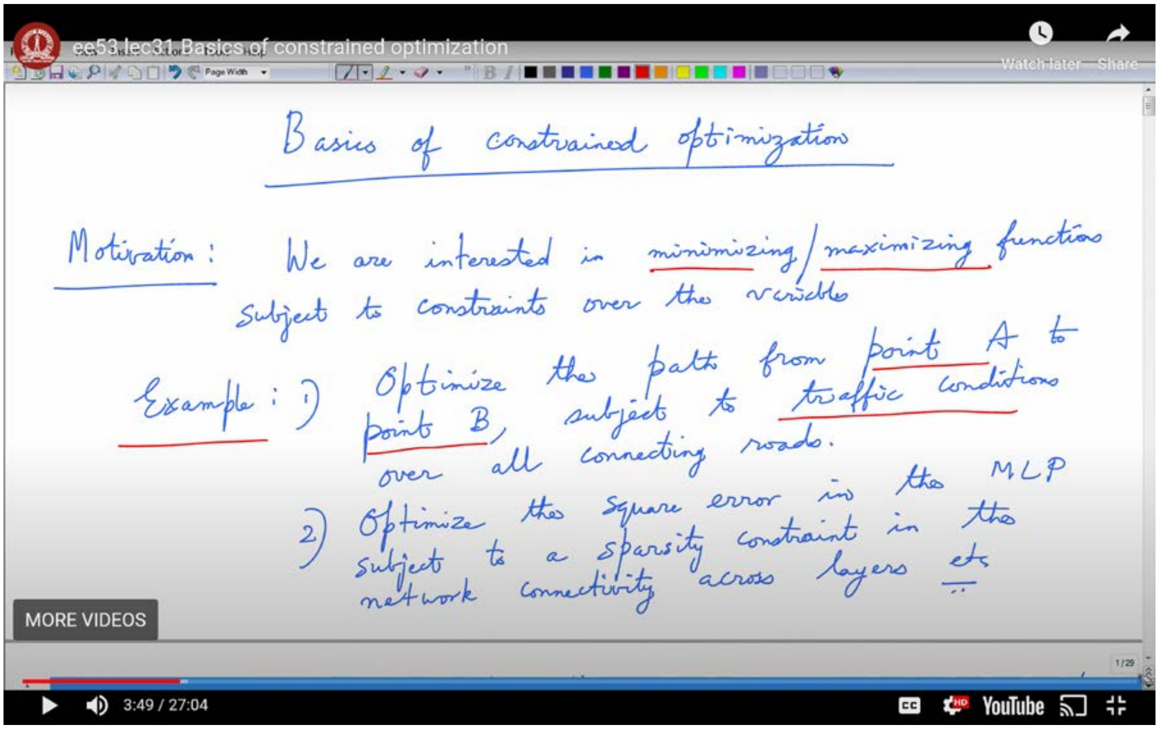
Neural Networks for Signal Processing-I

Prof. Shayan Srinivasa Garani
Department of Electronic System Engineering
Indian Institute of Science, Bengaluru

Lecture – 31
Basics of Constrained Optimization

Let's delve into the basics of constrained optimization. The central focus of our discussion will be on setting up the Lagrangian function for problems with equality and inequality constraints, specifically how to determine the appropriate sign in the constraint equations. This is a crucial aspect, as it relates to how we compute the gradients of both the objective function and the constraints, and how these gradients guide us toward finding a feasible solution.

(Refer Slide Time: 03:49)



Basics of constrained optimization

Motivation: We are interested in minimizing/maximizing functions subject to constraints over the variables

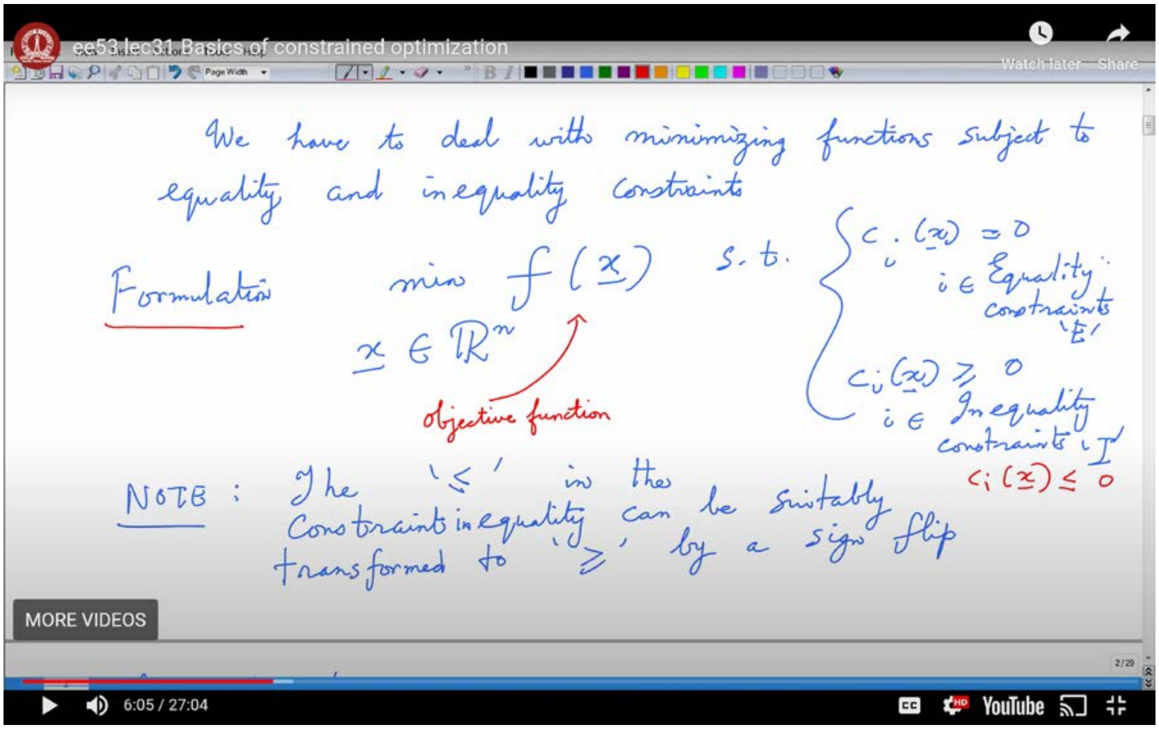
Example: 1) Optimize the path from point A to point B, subject to traffic conditions over all connecting roads.

2) Optimize the square error in the MLP subject to a sparsity constraint in the network connectivity across layers etc

While optimization courses typically cover these fundamentals, my goal here is to ensure that you have a solid grasp of these basics, which will provide continuity in your understanding as we progress.

The motivation behind constrained optimization is to either minimize or maximize an objective function while adhering to constraints imposed on the variables. For example, consider optimizing a route from point A to point B in a city with varying traffic conditions. Here, you might want to find the shortest path, or perhaps a route that avoids heavy traffic, depending on your specific objective. This requires optimizing over parameters such as distance and traffic conditions.

(Refer Slide Time: 06:05)



We have to deal with minimizing functions subject to equality and inequality constraints

Formulation

$$\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x}) \quad \text{s.t.} \begin{cases} c_i(\underline{x}) = 0 & i \in \text{Equality constraints `E'} \\ c_i(\underline{x}) \geq 0 & i \in \text{Inequality constraints `I'} \end{cases}$$

objective function

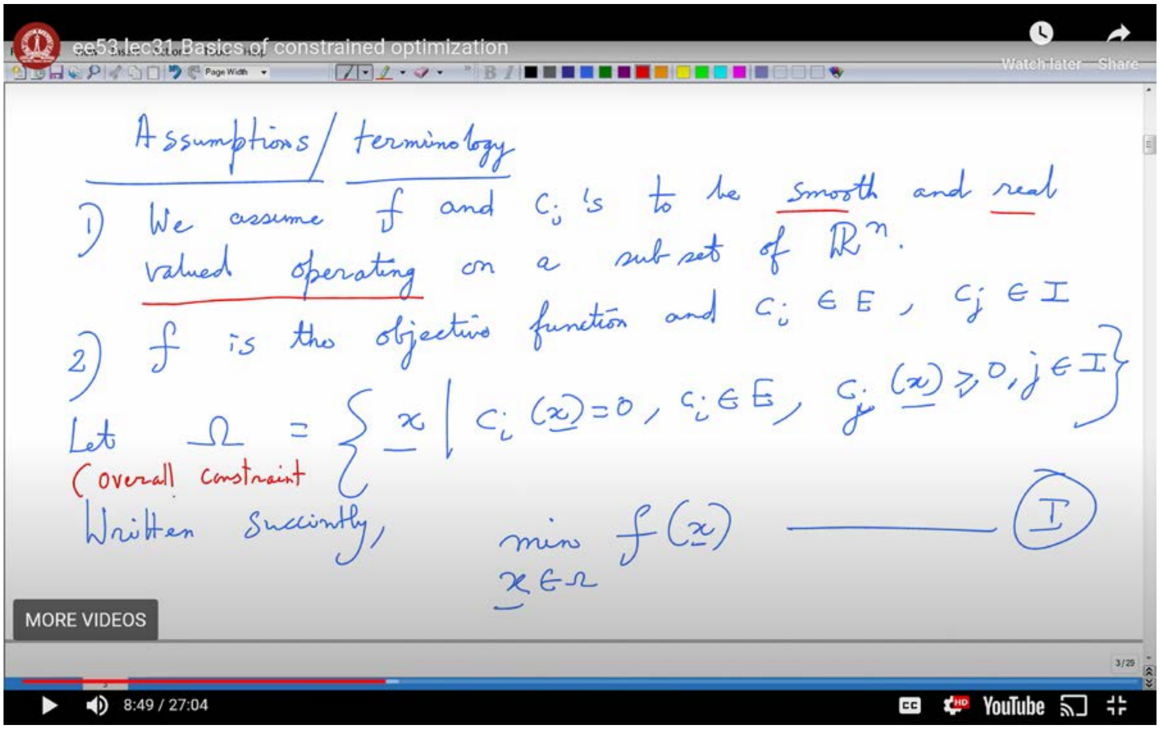
$c_i(\underline{x}) \leq 0$

NOTE : The '$\leq$' in the constraint inequality can be suitably transformed to '$\geq$' by a sign flip

Another example is related to neural networks. Suppose we want to minimize the squared error in a multilayer perceptron while maintaining a sparsity constraint on the network's connectivity across layers. In this case, we aim to retain only the synapses with significant strength, effectively eliminating connections that fall below a certain threshold. This optimization process helps to remove redundant connections or neurons, leading to a more efficient network.

In general, when formulating constrained optimization problems, we have an objective function f(x) that we wish to minimize or maximize, where x is a vector in an n-dimensional space. This minimization or maximization is subject to certain equality and inequality constraints. By understanding how to properly set up and solve these problems, we can handle a variety of practical scenarios where optimizing an objective function under constraints is required.

(Refer Slide Time: 08:49)



Assumptions / terminology

1) We assume f and c_i's to be smooth and real valued operating on a subset of $\mathbb{R}^n$.

2) f is the objective function and $c_i \in E$, $c_j \in I$

Let $\Omega = \{ \underline{x} \mid c_i(\underline{x}) = 0,\ c_i \in E,\ c_j(\underline{x}) \geq 0, j \in I \}$

(overall constraint (

Written succinctly,

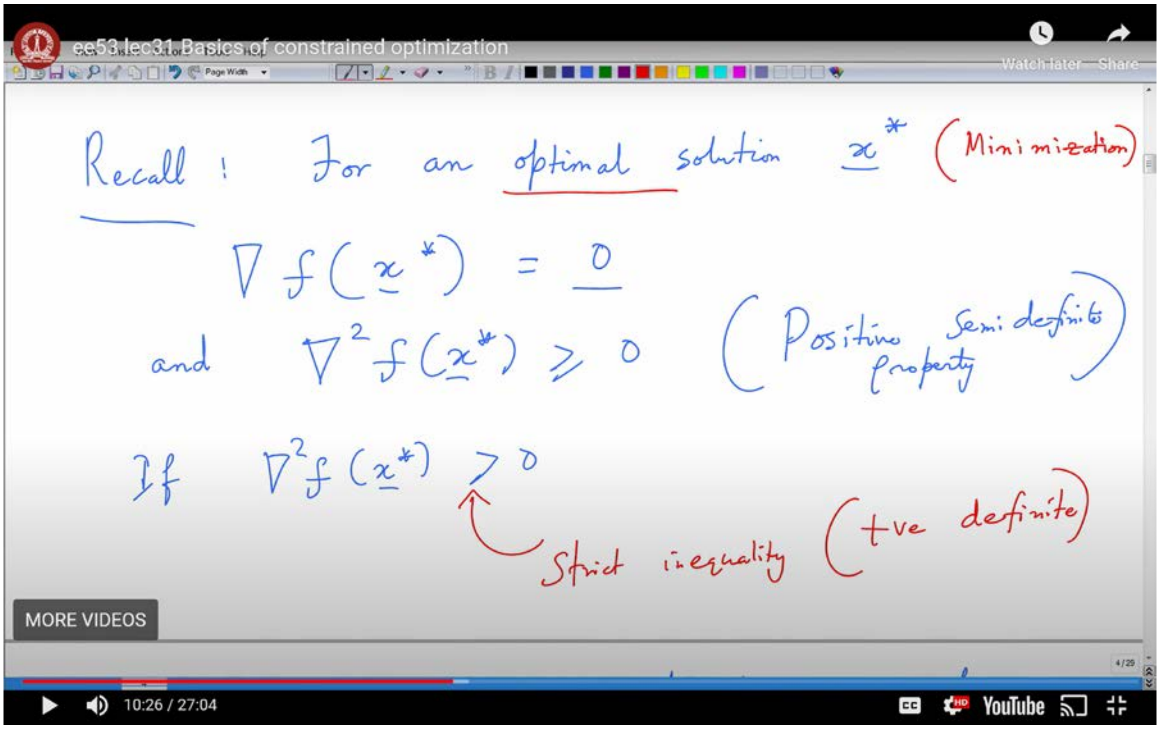
$$\min_{\underline{x} \in \Omega} f(\underline{x}) \quad \text{———} \quad \textcircled{I}$$

Let's delve into the details of equality and inequality constraints. Equality constraints are represented as $c_i(x) = 0$, where each $c_i(x)$ is a function of x. On the other hand, inequality constraints can take the form $c_i(x) \geq 0$ or $c_i(x) \leq 0$. Both forms are valid; essentially, if you have an inequality constraint $c_i(x) \geq 0$, you can flip the sign to convert it into $c_i(x) \leq 0$ by simply taking the negative of the function. This principle allows for flexibility in formulating constraints: minimizing a function can be approached by negating it and then maximizing, depending on the context.

For instance, consider the objective of minimizing the distance subject to an acceleration constraint. If you need to travel from point A to point B, you can adjust your speed—accelerate or decelerate. If there's a constraint on acceleration (i.e., you cannot exceed a certain acceleration limit), this becomes a practical constraint that must be considered.

Before we proceed, let’s establish some foundational assumptions and terminology. We assume that both the objective function f and the constraints c_i are smooth and real-valued, operating in an n-dimensional real space. Smoothness is important because we need to compute derivatives and perform analysis, so we require the functions to be differentiable up to a certain degree. While f and c_i could potentially be complex-valued, for simplicity in our analysis, we will consider them as real-valued.

(Refer Slide Time: 10:26)



Recall : For an optimal solution $\underline{x}^*$ (Minimization)

$$\nabla f(\underline{x}^*) = \underline{0}$$

and $\nabla^2 f(\underline{x}^*) \geq 0$ (Positive Semi definite property)

If $\nabla^2 f(\underline{x}^*) > 0$

Strict inequality (+ve definite)

The objective function is f, and c_i represents the equality constraints, while c_j represents the inequality constraints. We can combine all constraints into a single set Ω, defined as:

$$\Omega = \{x \mid c_i(x) = 0 \text{ for equality constraints } c_i \text{ and } c_j(x) \geq 0 \text{ for inequality constraints } c_j\}.$$

In this compact notation, our goal is to minimize the objective function f where x belongs to the set Ω.

To revisit basic calculus principles, if we solve an optimization problem without constraints, the derivative of the objective function must be zero, and the Hessian matrix (second derivative) of f should be positive semi-definite for minimization problems. This positive semi-definite property ensures that we are dealing with a minimization problem. For a minimization problem, if the Hessian is positive definite (i.e., all eigenvalues are positive), then the solution is a strict local minimum.

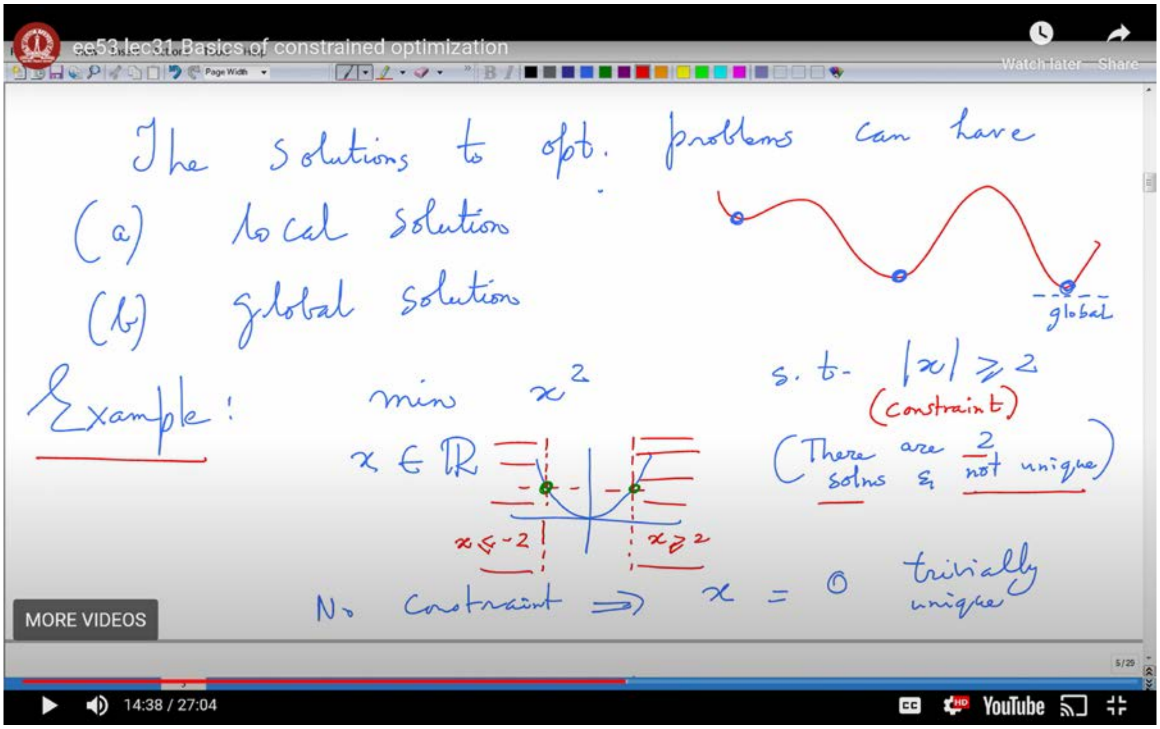
Conversely, for maximization problems, the Hessian should be negative semi-definite. If the Hessian is negative definite (i.e., all eigenvalues are negative), the solution is a strict local maximum.

These conditions are essential for ensuring that we have valid solutions for optimization problems, whether minimizing or maximizing the objective function.

Let's explore the concepts of local and global solutions in optimization with an illustrative example. Consider the simple quadratic function $y = x^2$. To find the minimum value, note that x^2 is always positive, regardless of whether x is positive or negative. The function $y = x^2$ achieves its minimum value when x = 0. This is a clear example of a global minimum: there is only one minimum point in this case.

Now, imagine a more complex surface, where you might have multiple valleys and peaks. In such a scenario, there could be several local minima, depending on where you are in the surface. The global minimum is the absolute lowest point on the surface, while the local minima are points that are lower than their immediate surroundings but not necessarily the lowest overall. Finding the global optimum in such cases can be challenging, as no algorithm guarantees finding it precisely. Instead, approximation algorithms are often used to get reasonably good solutions within the search space.

(Refer Slide Time: 14:38)



The solutions to opt. problems can have

(a) local solution

(b) global solution

Example: min x^2 s.t. $|x| \geq 2$ (constraint)

$x \in \mathbb{R}$

(There are 2 solns & not unique)

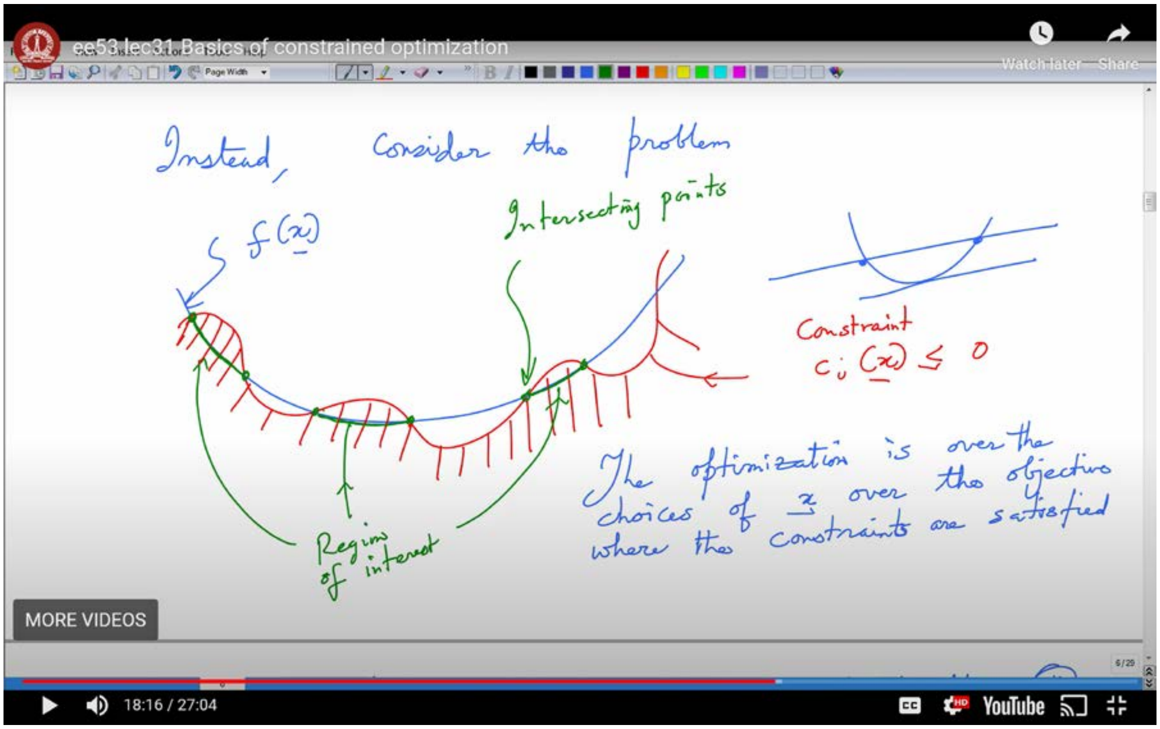
No constraint $\Rightarrow$ $x = 0$ trivially unique

Consider now adding constraints to the problem. Take the same function y = x^2, but this time we minimize it subject to the constraint $|x| \geq 2$. This constraint limits the feasible region to $x \geq 2$ and $x \leq -2$, which corresponds to the shaded red regions in this example.

Under this constraint, the function y = x^2 has minimum values at x = 2 and x = -2. These are the points where the constraint intersects with the curve, and in this case, there are two solutions. Unlike the unconstrained case where x = 0 is the unique minimum, the constrained problem has multiple minimum points.

Introducing constraints can lead to more complex situations where multiple solutions might exist, or where you might need additional regulations to ensure a unique solution. This nuance highlights the importance of understanding the nature of solutions—whether they are local or global—and how constraints impact the optimization process.

(Refer Slide Time: 18:16)



Instead, consider the problem

The optimization is over the choices of $\underline{x}$ over the objective where the constraints are satisfied

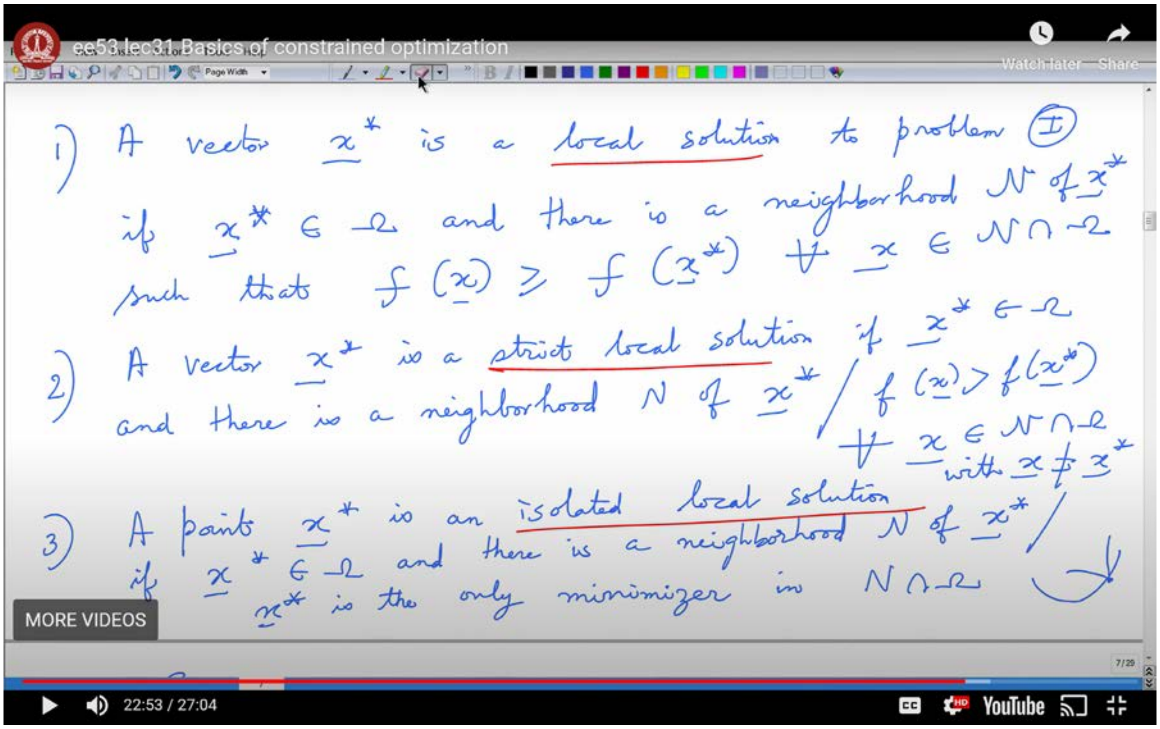
Finally, let’s consider a more general problem. Suppose you have a curve f(x) (depicted as the blue curve) and a constraint equation represented by another curve (the red curve). If the constraint is an equality, $c_i(x) = 0$, you need to look for intersection points between the constraint curve and the objective function.

These intersection points (indicated by green dots) are potential solutions to the optimization problem under the given constraints. If the constraint is an inequality, such as $c_i(x) \le 0$, the feasible region is determined by the area below the constraint curve.

Understanding the geometry of these constraints and their interaction with the objective function is crucial in solving optimization problems effectively.

Let's delve into how we determine the optimal solution in the context of constraints. When optimizing a function f(x), we focus on points where the curve intersects with the objective function. For equality constraints, this involves examining the points where the constraint curve intersects the objective function. The function f(x) is optimized over these specific intersection points.

(Refer Slide Time: 22:53)

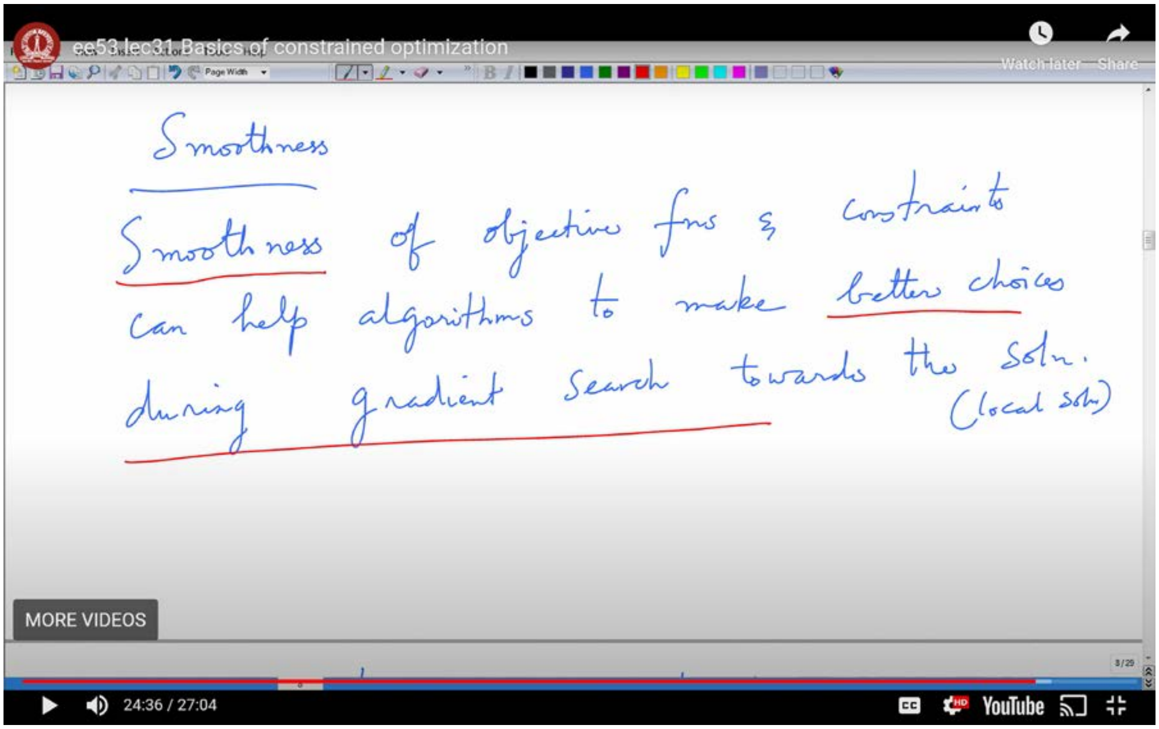


1) A vector $\underline{x}^*$ is a local solution to problem $\textcircled{I}$ if $\underline{x}^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of $\underline{x}^*$ such that $f(\underline{x}) \geq f(\underline{x}^*)$ $\forall \underline{x} \in \mathcal{N} \cap \Omega$

2) A vector $\underline{x}^*$ is a strict local solution if $\underline{x}^* \in \Omega$ and there is a neighborhood N of $\underline{x}^*$ / $f(\underline{x}) > f(\underline{x}^*)$ $\forall \underline{x} \in \mathcal{N} \cap \Omega$ with $\underline{x} \neq \underline{x}^*$

3) A point $\underline{x}^*$ is an isolated local solution if $\underline{x}^* \in \Omega$ and there is a neighborhood N of $\underline{x}^*$ / $\underline{x}^*$ is the only minimizer in $N \cap \Omega$

(Refer Slide Time: 24:36)



Smoothness

Smoothness of objective fns & constraints can help algorithms to make better choices during gradient search towards the soln. (local soln)

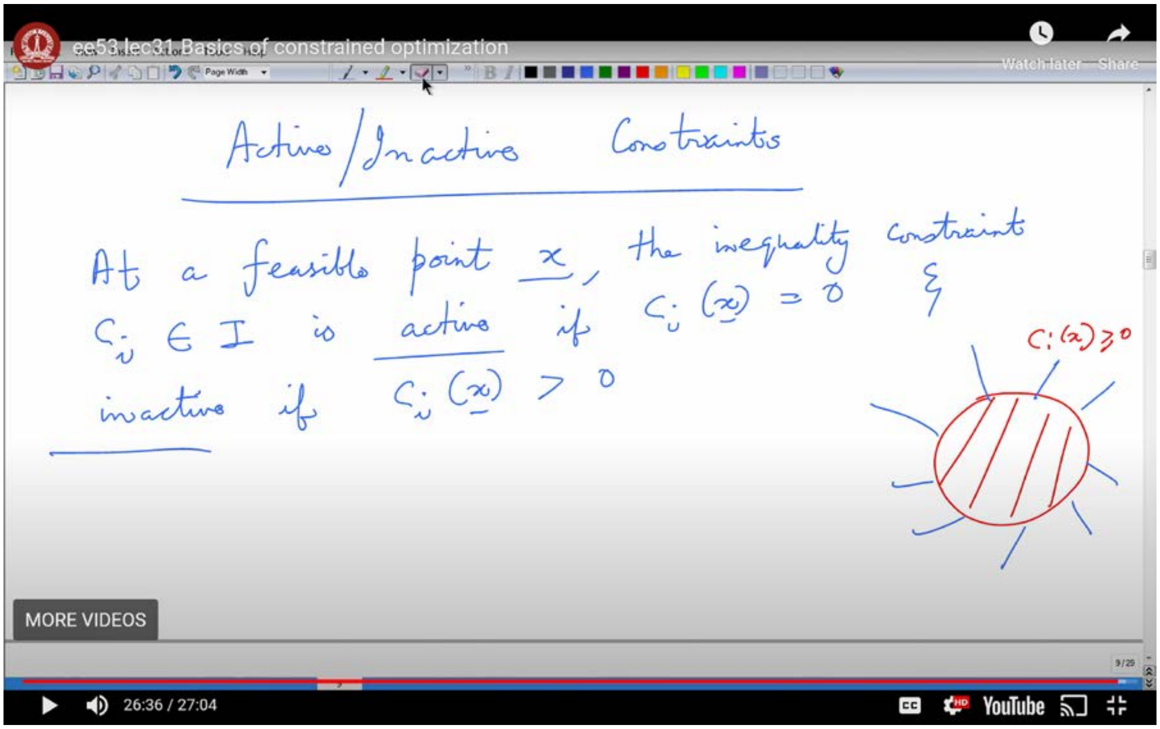
However, when dealing with inequality constraints, such as $c_i(x) \leq 0$, we are interested in the regions below the constraint curve. This shaded region, as illustrated, represents the feasible area where the constraint holds. Optimization is then performed over this feasible region.

Thus, the shaded segments of the curve represent the regions of interest for optimization. This problem can become quite complex, especially with multiple constraints and intersecting curves. Understanding these intersections and feasible regions requires careful consideration of the geometry involved.

Now, let's discuss some basic definitions related to solutions. Suppose we have a minimization problem defined as minimizing f(x) over all x belonging to a set Ω. A vector x^* is considered a local solution to this problem if:

- x^* belongs to Ω
- There exists a neighborhood N around x^* such that $f(x) \geq f(x^*)$ for all x within the intersection of N and Ω

(Refer Slide Time: 26:36)



Active / Inactive Constraints

At a feasible point $\underline{x}$, the inequality constraint $c_i \in I$ is active if $c_i(\underline{x}) = 0$ &

inactive if $c_i(\underline{x}) > 0$

In simpler terms, x^* is a local solution if within some small vicinity around x^*, the value of f(x) does not drop below $f(x^*)$. For a minimization problem, this means x^* provides a local minimum.

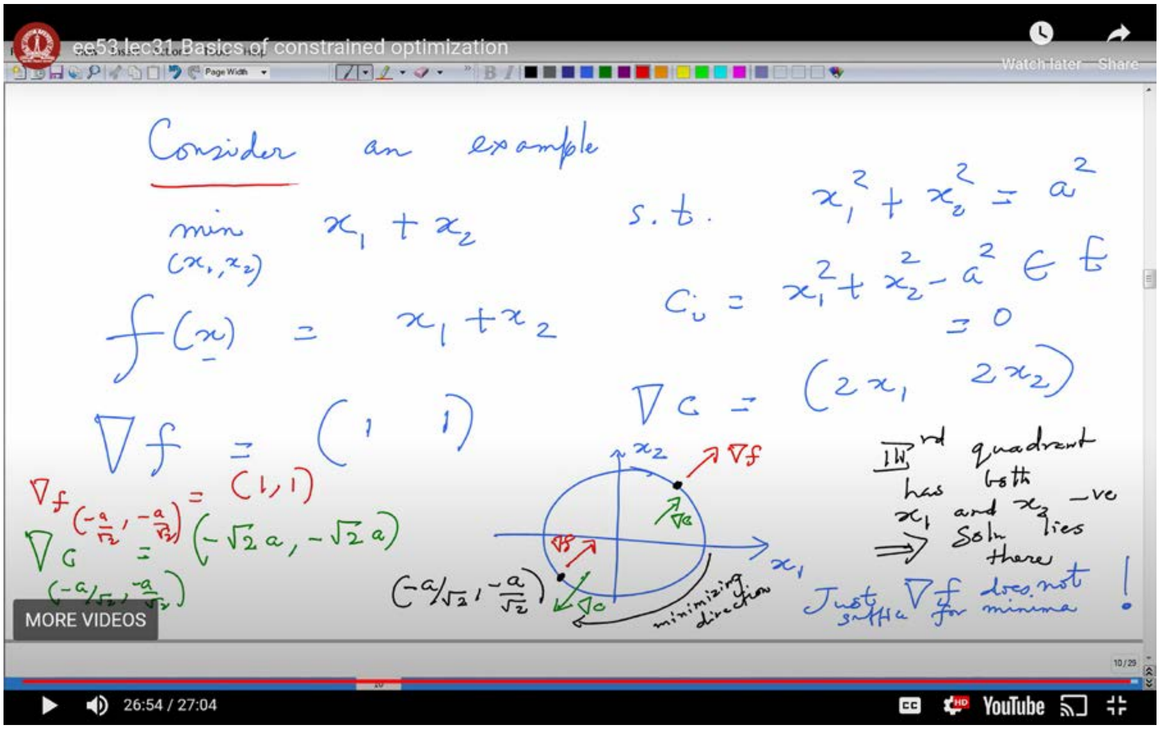
For a stricter definition, x^* is a strict local solution if:

- x^* belongs to Ω
- There exists a neighborhood N around x^* such that $f(x) > f(x^*)$ for all x within the intersection of N and Ω, and $x \neq x^*$

In this case, x^* is not only a local minimum but a strict local minimum. In the neighborhood around x^*, f(x) is strictly greater than $f(x^*)$ for all other points x within the feasible region.

Understanding these definitions helps clarify how solutions are characterized and how constraints influence the optimization process.

(Refer Slide Time: 26:54)



ee53 lec31 Basics of constrained optimization

Consider an example

$\min_{(x_1,x_2)} \; x_1 + x_2 \quad$ s.t. $\quad x_1^2 + x_2^2 = a^2$

$f(\underline{x}) = x_1 + x_2 \qquad c_i = x_1^2 + x_2^2 - a^2 \in \mathcal{E}$

$= 0$

$\nabla f = (1 \quad 1) \qquad \nabla c = (2x_1 \quad 2x_2)$

$\nabla f_{(-\frac{a}{\sqrt{2}}, -\frac{a}{\sqrt{2}})} = (1,1)$

$\nabla c_{(-a/\sqrt{2}, -\frac{a}{\sqrt{2}})} = (-\sqrt{2}a, -\sqrt{2}a)$

$(-a/\sqrt{2}, -\frac{a}{\sqrt{2}})$

minimizing direction

$\overline{\text{IV}}$ rd quadrant has both x_1 and x_2 −ve

$\Rightarrow$ Soln lies there

Just ∇f does not suffice for minima !

Let's delve into the concept of an isolated local solution. A point x^* is considered an isolated local solution if it belongs to the constraint set Ω and there exists a neighborhood N around x^* such that x^* is the only minimizer within the intersection of N and Ω.

To illustrate this, imagine a curve and a constraint, and consider the neighborhood intersecting Ω. If this neighborhood contains only one point where the function achieves its minimum, then x^* is an isolated local solution. In other words, x^* is unique in minimizing the objective function within that region.

Now, let's revisit the importance of smoothness in optimization. We aim for smoothness in both the objective function and the constraints to facilitate gradient-based search techniques. Smooth constraints make it easier to compute gradients and navigate the search space effectively. If the derivatives of the constraints are not well-defined, it complicates the search for optimal solutions. Hence, while smooth constraints are ideal, in practice, we may encounter situations where this is not guaranteed. In such cases, careful handling and definition of derivatives over the regions of interest become essential.

Additionally, we need to understand the notions of active and inactive constraints. At a feasible point x, an inequality constraint c_i is considered active if $c_i(x) = 0$ and inactive if $c_i(x) > 0$. For example, if our constraint is represented by a circle, then points lying on the circle or inside it satisfy the constraint. If we have a strict inequality constraint, say $c_i(x) \geq 0$, points lying exactly on the circle represent the active constraint, while points strictly outside the circle are inactive. This illustration clarifies how to categorize constraints and helps us understand their role in optimization problems.