Neural Networks for Signal Processing-I

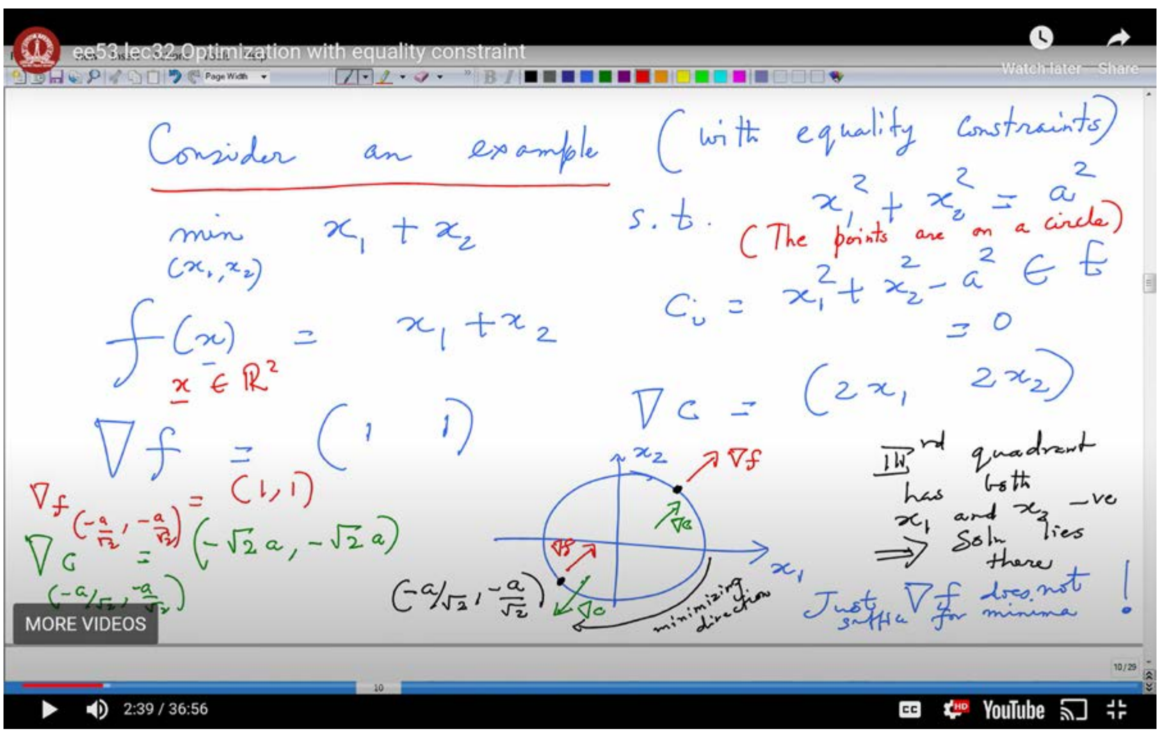
Prof. Shayan Srinivasa Garani
Department of Electronic System Engineering
Indian Institute of Science, Bengaluru

Lecture – 32

Optimization with Equality Constraint

Having reviewed some fundamental terminologies related to constraint optimization, let us now address equality constraints. Specifically, we want to minimize our objective function f(x) subject to constraints that must be satisfied with equality. It is essential to visualize the problem geometrically, focusing on how the constraints intersect with the objective function, and then optimize over these intersecting points.

(Refer Slide Time: 02:39)



To illustrate this, consider the following example where we aim to minimize the function $x_1 + x_2$, where x_1 and x_2 are real numbers. The geometry of this problem is essentially two-dimensional, constrained by x_1 and x_2. Our constraint is $x_1^2 + x_2^2 = a^2$, which represents a circle centered at the origin with radius a.

The objective function f(x), where x is a two-dimensional vector, is given by $f(x) = x_1 + x_2$. The constraint c_i (for simplicity, we omit x in the notation) is $c_i = x_1^2 + x_2^2 - a^2$, which corresponds to the equality constraint $x_1^2 + x_2^2 - a^2 = 0$.

To gain an intuitive understanding of this problem, let us examine the geometry. The circle, defined by $x_1^2 + x_2^2 = a^2$, is centered at the origin with radius a. By analyzing $f(x) = x_1 + x_2$ in different quadrants:

- In the first quadrant, both x_1 and x_2 are positive, so f(x) will be positive.
- In the third quadrant, both x_1 and x_2 are negative, so f(x) will be negative.
- In the second quadrant, x_1 is negative and x_2 is positive, while in the fourth quadrant, x_1 is positive and x_2 is negative. Thus, f(x) can vary between positive and negative values depending on the specific values of x_1 and x_2.

To minimize the objective function $x_1 + x_2$ subject to the constraint that the points lie on the circle, the optimal solution will be found in the third quadrant, where both x_1 and x_2 are negative, resulting in the minimum value of $x_1 + x_2$.

Conversely, in the first quadrant, where both coordinates are positive, the objective function will achieve its maximum value.

Next, we compute the gradient of the objective function and the gradient of the constraint to further analyze this problem. Taking the partial derivatives will help us understand how these gradients interact and contribute to finding the optimal solution.

To determine the direction of the gradient of the objective function Δ f, we first calculate the partial derivatives with respect to x_1 and x_2. Specifically, Δ f is given by the partial derivative of f with respect to x_1 and the partial derivative of f with respect to x_2. Mathematically, this is represented as:

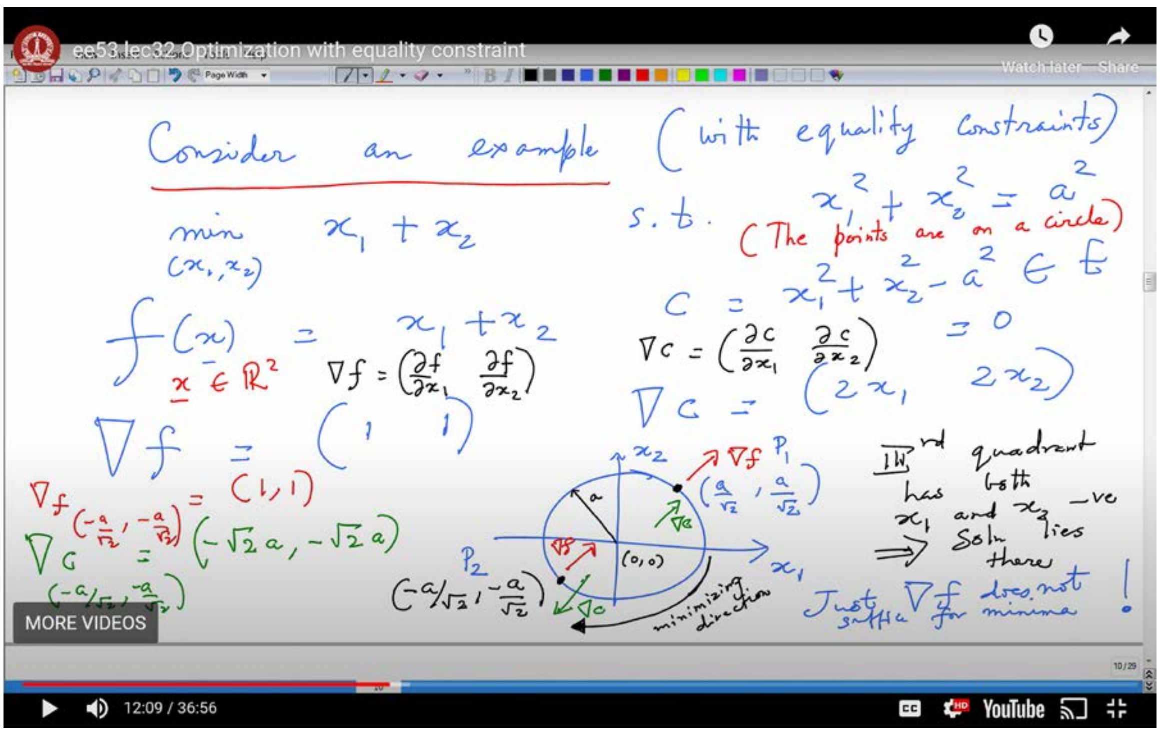
$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right).$$

In this case, the computation straightforwardly results in the vector $\nabla f = (1{,}1)$. This vector points at a 45-degree angle relative to the origin. Consequently, the direction of Δ f is always at 45 degrees, and its unit vector representation is:

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

This direction is indicated by the red arrows.

(Refer Slide Time: 12:09)



Now, let's analyze the gradient of the constraint Δ C. This gradient is obtained by taking the partial derivatives of the constraint function with respect to x_1 and x_2. For the constraint function $C = x_1^2 + x_2^2 - a^2$, the gradient is:

$$\Delta C = \left(\frac{\partial C}{\partial x_1}, \frac{\partial C}{\partial x_2}\right) = (2x_1, 2x_2).$$

Thus, the gradient of the constraint depends on the coordinates x_1 and x_2.

Consider the specific points where we want to evaluate the gradients:

- Point P_1 with coordinates $\left(\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right)$
- Point P_2 with coordinates $\left(-\frac{a}{\sqrt{2}}, -\frac{a}{\sqrt{2}}\right)$

At P_2, which is the point of minimum value, the gradient of the constraint evaluates to:

$$\Delta C = \left(-\sqrt{2}a, -\sqrt{2}a\right).$$

At P_1, which is the point of maximum value, the gradient of the constraint is:

$$\Delta C = \left(\sqrt{2}a, \sqrt{2}a\right).$$

An important observation here is that the gradients at P_2 and P_1 are in exactly opposite directions. The green arrows indicate that at P_2, the gradient is directed downward, while at P_1, it is directed upward.

To find the minimizing direction, suppose we start at a point on the x_1-axis. The black arrow shows the clockwise direction, which indicates the path we should follow to move toward the minimum.

This leads to two crucial questions:

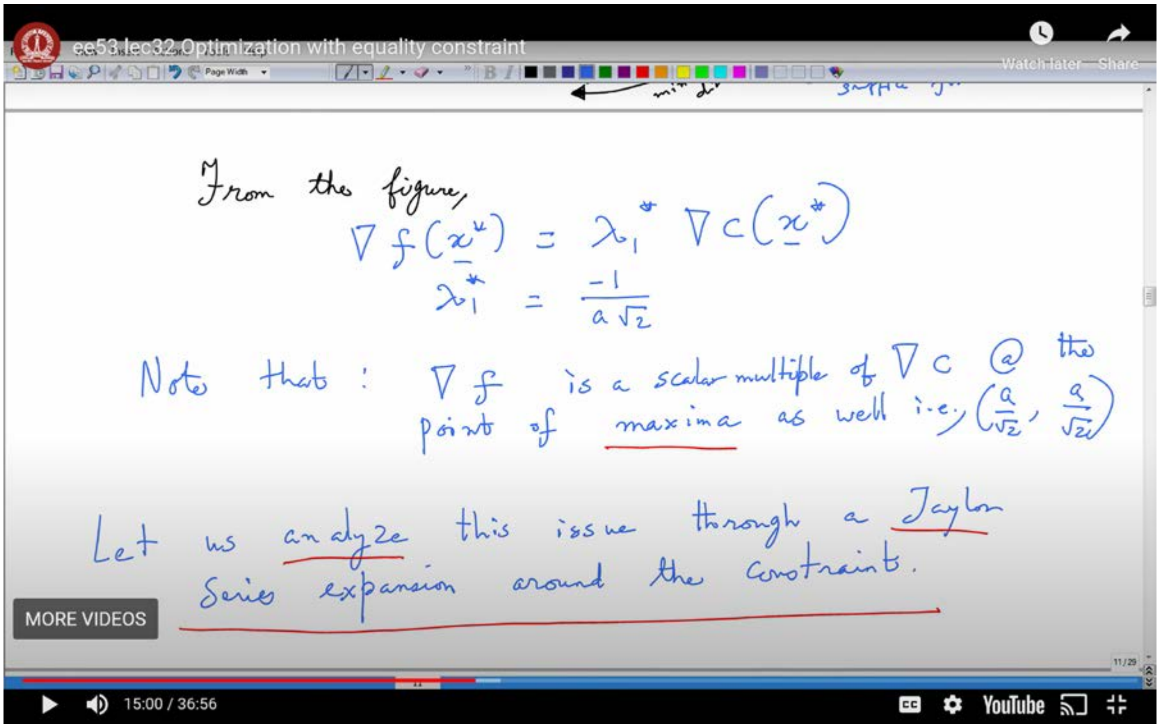
1. How should we choose the minimizing direction?

2. Is there a relationship between the minimizing direction and the gradients of the constraint and the objective function?

To address these questions, we need to explore the relationships between the gradients of the constraint and the objective function, and how they guide us toward finding the optimal minimum. Understanding these relationships will help us determine the appropriate direction for minimizing the objective function effectively.

It's important to note that simply computing Δ f is not enough to determine whether we have a minimum. To truly understand the nature of a point, you need to examine Δ^2 f, which is the Hessian matrix of second-order partial derivatives. While Δ f might indicate potential extremum points, the Hessian will help you determine whether these points are minima, maxima, or saddle points.

(Refer Slide Time: 15:00)

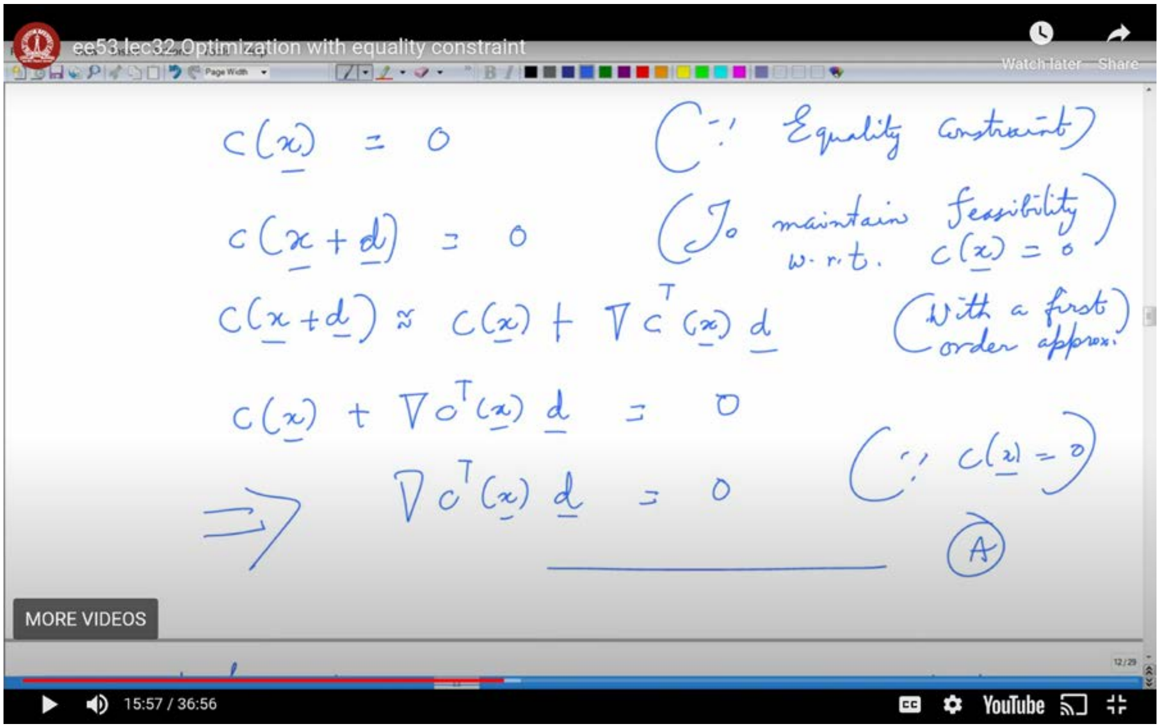


From the figure, we can see that Δf^* is some scalar multiple of ΔC evaluated at x^*. If x^* is our saddle point, the scalar λ_1^* is straightforward to compute. For instance, if λ_1^* *is* $-\frac{1}{a\sqrt{2}}$, it is easy to verify this by examining the relationship between Δf and ΔC. The scalar λ_1 links Δf and ΔC since Δf is a constant and ΔC has symmetry.

Importantly, Δf is also a scalar multiple of ΔC at the point of maximum. However, the sign of the scalar will differ; for instance, it may be $\frac{1}{\sqrt{2}}$ at a maximum, while it is $-\frac{1}{\sqrt{2}}$ at a minimum. This sign difference is crucial and implies that merely identifying a scalar multiple is not sufficient for determining if a point is a maximum or minimum. Further analysis is necessary to classify the saddle point correctly.

To proceed with this analysis, we can use a Taylor series expansion around the constraints. Since we're dealing with equality constraints, C(x) = 0, we need to maintain feasibility near x. We choose a direction ***d*** such that $C(x + d) = 0$.

(Refer Slide Time: 15:57)



Expanding C(x + ***d***) using a first-order Taylor series approximation, we get:

$$C(x + d) \approx C(x) + \Delta C^T \cdot d.$$

Since C(x) = 0, this simplifies to:

$$C(x + d) \approx \Delta C^T \cdot d.$$

Thus, the condition C(x + ***d***) = 0 leads to the equation:

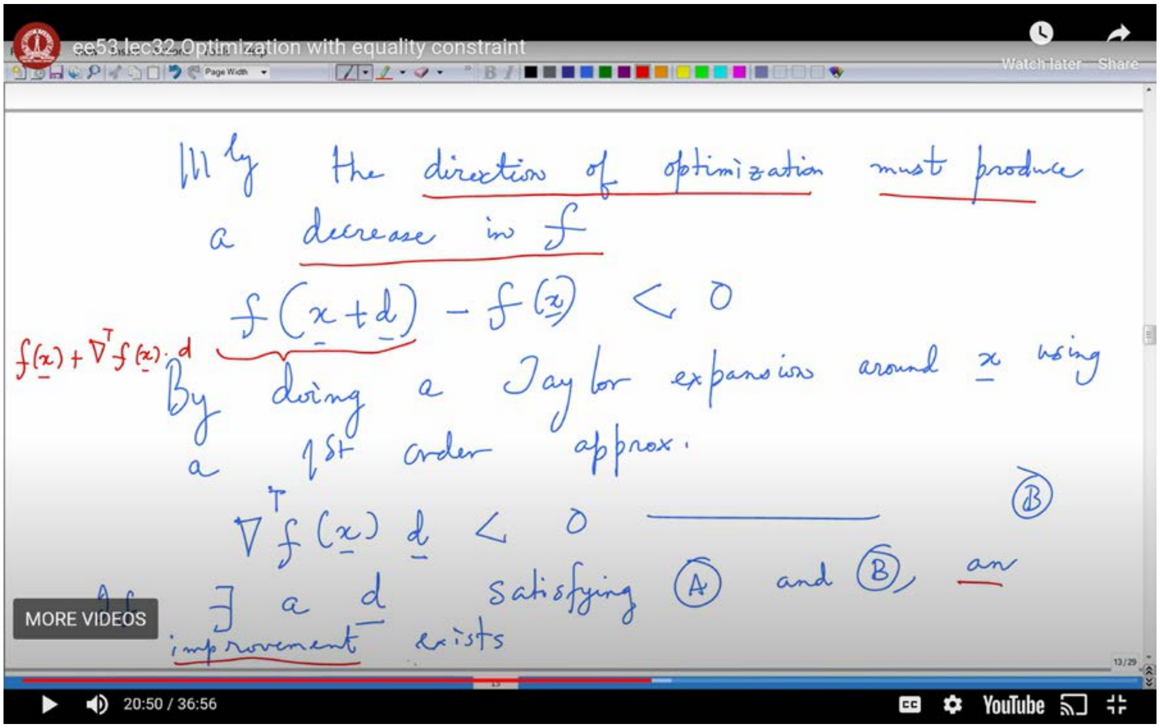
$$\Delta C^T \cdot d = 0.$$

This equation implies that the gradient Δ C at x and the direction ***d*** must be orthogonal.

Furthermore, for the direction of optimization to result in a decrease in f, we need:

$$f(x + d) - f(x) < 0.$$

(Refer Slide Time: 20:50)



Thus, the difference f(x) - f(x + $\boldsymbol{d}$) must be positive. To analyze this, we expand f(x + $\boldsymbol{d}$) using a first-order Taylor series:

$$f(x + d) \approx f(x) + \Delta f^T \cdot d.$$

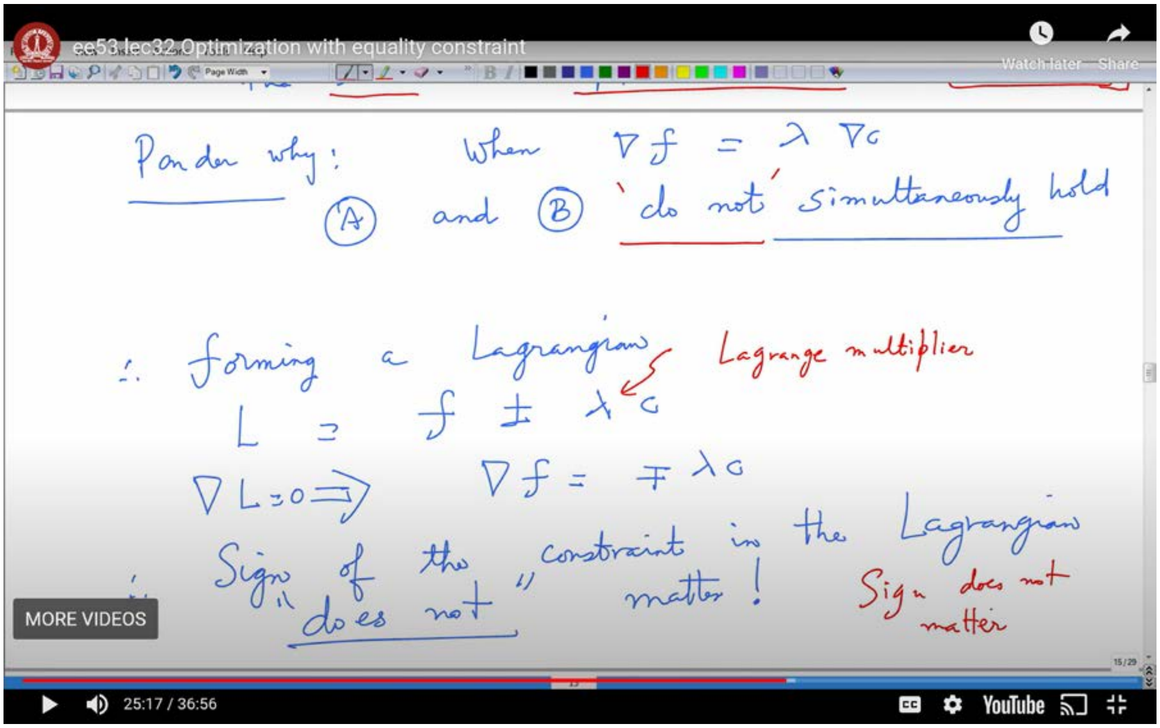
Subtracting f(x), we get:

$$f(x + d) - f(x) \approx \Delta f^T \cdot d.$$

This expansion confirms that to achieve a decrease in the function, $\Delta f \cdot \boldsymbol{d}$ must be negative.

It's crucial to understand that simply calculating $\Delta f \cdot d$, which represents the inner product of the gradient of the objective function and the direction vector $\boldsymbol{d}$, must be strictly less than zero for an optimal solution. We need to find a direction $\boldsymbol{d}$ that satisfies this condition, which I will refer to as inequality B, while also ensuring that $\Delta C \cdot d = 0$, which I will call equation A. If such a direction $\boldsymbol{d}$ exists, then we are improving our solution, meaning we're moving toward a better outcome.

(Refer Slide Time: 25:17)

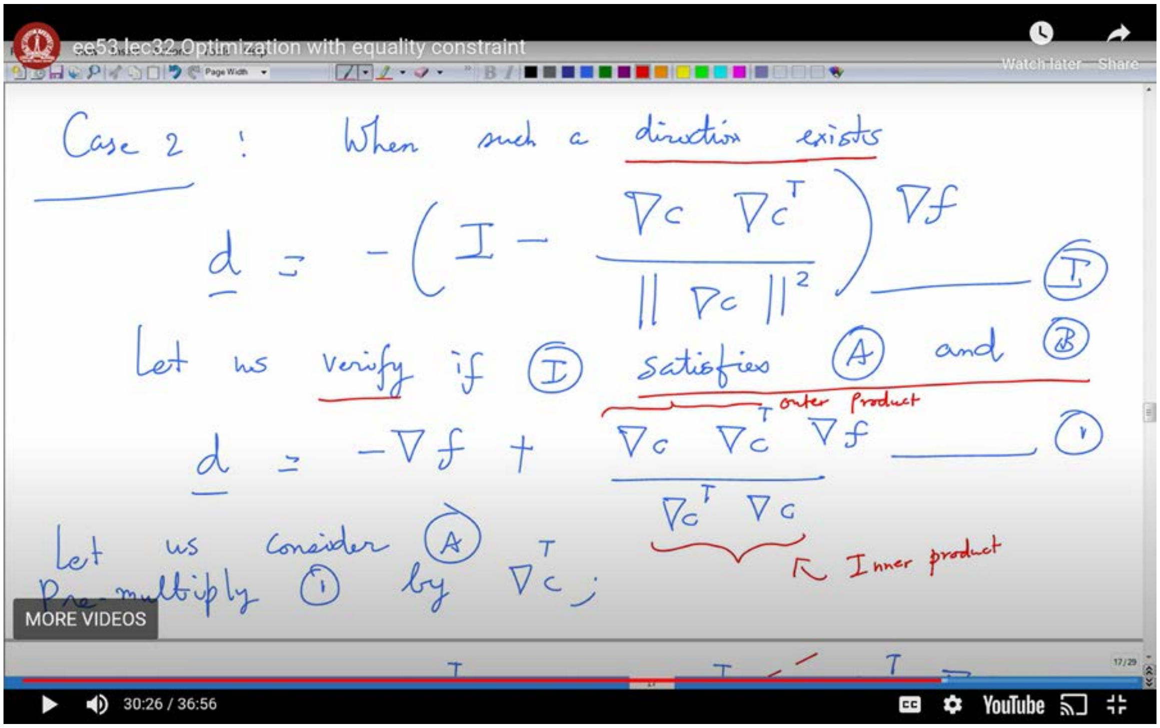


We need to consider two scenarios here:

1. Case 1: Such a Direction Does Not Exist

- If no such direction exists, it implies that we are at a saddle point where Δf and ΔC are scalar multiples of each other. This means that Δf could be aligned with or oppose ΔC. Mathematically, this is represented as $\Delta f = \lambda \Delta C$, where λ is a scalar.
- This situation was evident when we minimized $x_1 + x_2$ subject to x_1 and x_2 lying on a circle. The critical insight is that if Δf equals $\lambda \Delta C$, equations A and inequality B cannot both hold simultaneously. This is because equation A implies $\Delta C \cdot \boldsymbol{d} = 0$, while inequality B requires $\Delta f \cdot \boldsymbol{d} < 0$. Therefore, if $\Delta f = \lambda \Delta C$, substituting this into equations A and B shows that they cannot be satisfied at the same time.

(Refer Slide Time: 30:26)



2. Case 2: Such a Direction Exists

- If such a direction $\boldsymbol{d}$ does exist, then we are not at a saddle point. We need to solve for $\boldsymbol{d}$ analytically. The direction $\boldsymbol{d}$ can be obtained using the following formula:

$$d = -\left(I - \frac{\Delta C \Delta C^T}{|\Delta C|^2}\right) \Delta f$$

- Here, I is the identity matrix, and $\Delta C \Delta C^T$ is the outer product of the gradient of the constraint with itself. The term $|\Delta C|^2$ is a scalar representing the norm squared of ΔC. This formula essentially projects Δf onto the space orthogonal to ΔC.

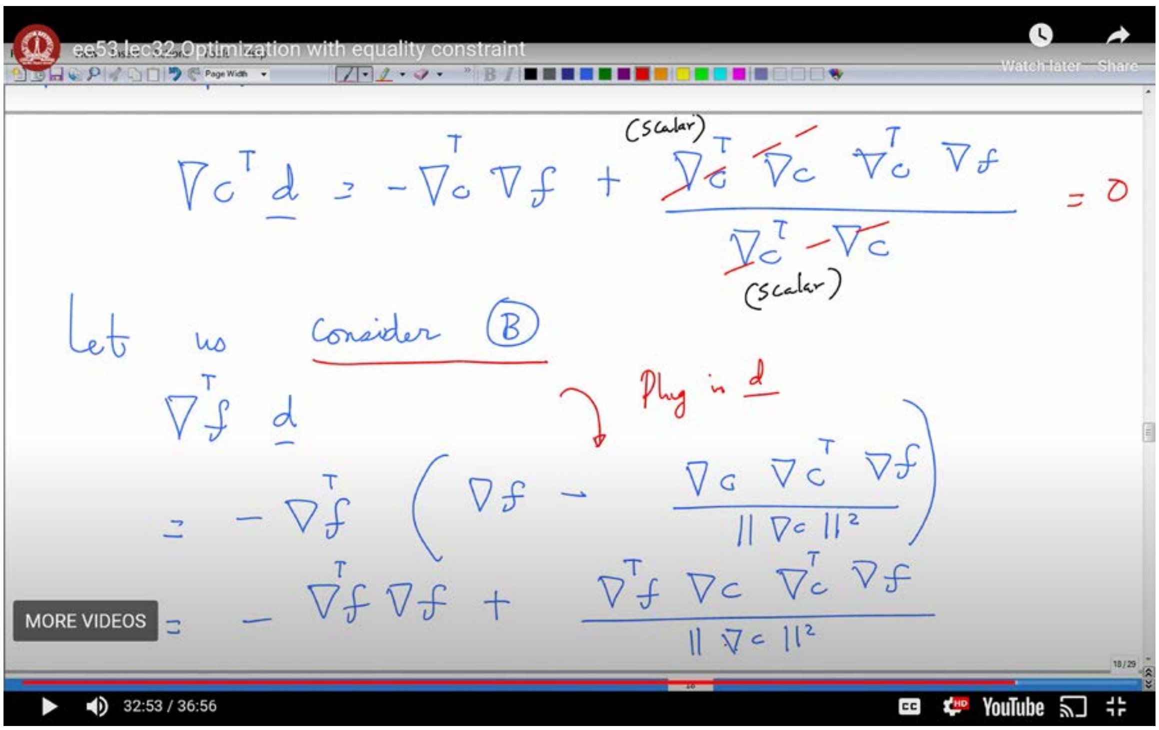
Thus,

- Case 1: If Δf and ΔC are aligned (i.e., $\Delta f = \lambda \Delta C$), equations A and inequality B cannot both be satisfied. This indicates a saddle point rather than an extremum.

- Case 2: If a suitable direction $\boldsymbol{d}$ is found, it indicates that we have not yet reached a saddle point, and we can use this direction to improve our solution. The derived direction $\boldsymbol{d}$ must be verified to ensure it satisfies equations A and inequality B.

By ensuring that $\boldsymbol{d}$ satisfies these conditions, we can correctly identify whether we're moving towards a minimum or a maximum.

(Refer Slide Time: 32:53)



Let's break down the analysis and calculations involved here with clarity.

Firstly, when dealing with gradients and directions, the equation

$$d = -\left(I - \frac{\Delta C \Delta C^T}{|\Delta C|^2}\right)\Delta f$$

provides a direction vector $\boldsymbol{d}$ that should be analyzed. In this context, we note that the product of two negatives results in a positive. Hence, the term simplifies to

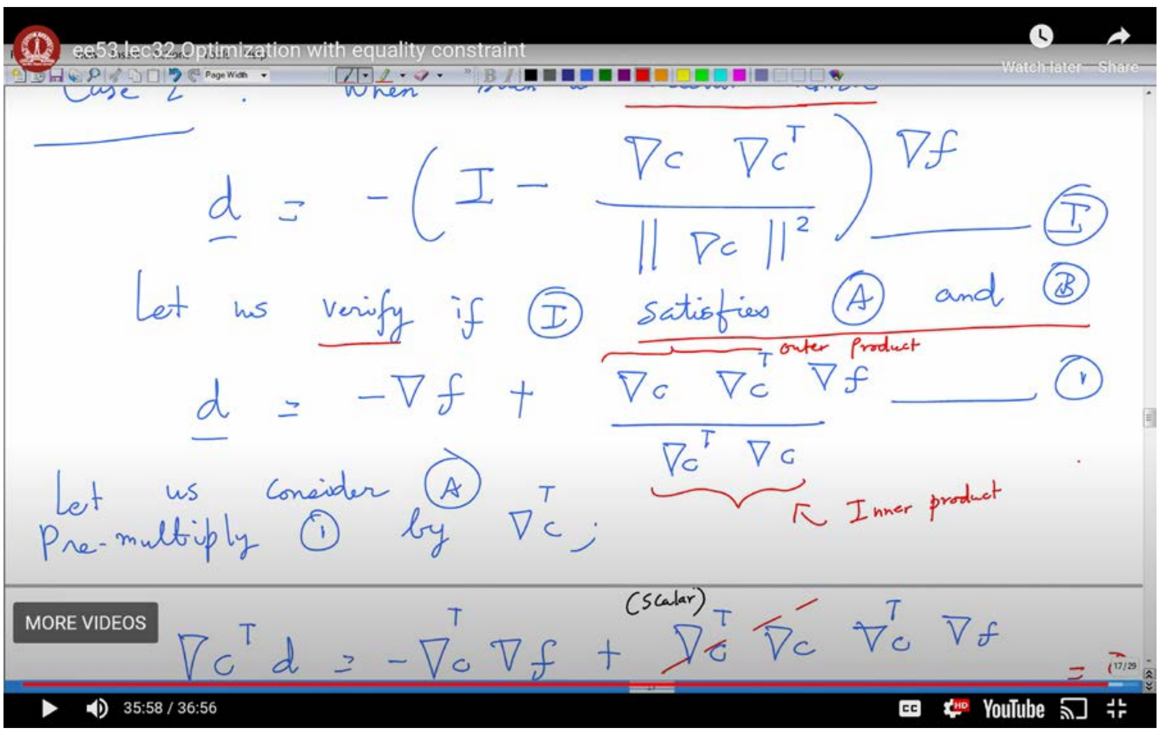
$$d = \frac{-(\Delta C \Delta C^T)\Delta f}{|\Delta C|^2}$$

where $|\Delta C|^2$ can be expressed as the inner product $\Delta\ C^T\ \Delta\ C$, representing the norm squared of Δ C.

Now, consider equation A, which states that

$$\Delta C^T d = 0$$

(Refer Slide Time: 35:58)



We need to verify if this condition is satisfied with our direction $\boldsymbol{d}$. Substituting $\boldsymbol{d}$ into this equation, we compute

$$\Delta C^T d = -\frac{\Delta C^T(\Delta C \Delta C^T \Delta f)}{|\Delta C|^2}$$

Simplify this expression:

$$\Delta C^T d = -\frac{(\Delta C^T \Delta C)(\Delta C^T \Delta f)}{|\Delta C|^2}$$

Since Δ C^T Δ C is a scalar (which is $|\Delta C|^2$), it cancels out the denominator. Thus,

$$\Delta C^T d = -(\Delta C^T \Delta f) + (\Delta C^T \Delta f) = 0$$

So, equation A is satisfied.

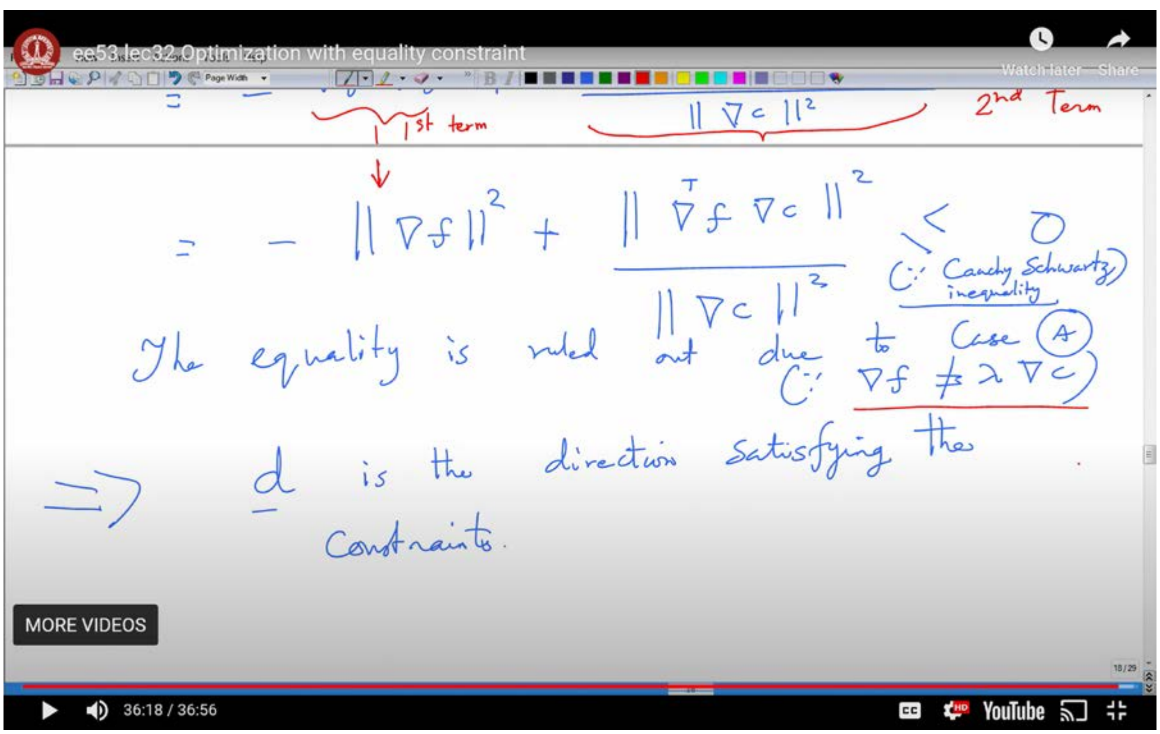
Next, we need to examine the inequality B, which is

$$\Delta f^T d < 0$$

Substitute $\boldsymbol{d}$ into this inequality:

$$\Delta f^T d = -\frac{\Delta f^T (\Delta C \Delta C^T \Delta f)}{|\Delta C|^2}$$

(Refer Slide Time: 36:18)



Simplify it:

$$\Delta f^T d = -\frac{(\Delta f^T \Delta C)(\Delta C^T \Delta f)}{|\Delta C|^2}$$

Here, $\Delta f^T \Delta f$ is the norm squared of Δf, and $\Delta f^T \Delta C \Delta C^T \Delta f$ can be interpreted as the squared term of Δf projected onto ΔC. The expression becomes:

$$-\frac{|\Delta f|^2 |\Delta C|^2}{|\Delta C|^2}$$

Simplify this to:

$$-|\Delta f|^2 + \frac{(\Delta f^T \Delta C)^2}{|\Delta C|^2}$$

According to the Cauchy-Schwarz inequality, this quantity will be less than or equal to zero, with equality when $\Delta f = \lambda \Delta C$. Since equality is ruled out in our analysis, we have a strict inequality:

$$\Delta f^T d < 0$$

Therefore, $\boldsymbol{d}$ is indeed a valid direction for optimization, satisfying the constraints of the problem.

To summarize:

- If Δf and ΔC are scalar multiples, we are at a saddle point. In such cases, equations A and inequality B cannot both be satisfied simultaneously.
- If such a direction d exists, it means we are not at a saddle point. Use the direction provided by the formula to proceed towards optimization, whether it's using gradient descent or another algorithm, to reach an optimal point.

This approach gives valuable insight into navigating the geometry of optimization problems, particularly when dealing with equality constraints. Next, we will explore how to handle inequality constraints.