

Neural Networks for Signal Processing-I
Prof. Shayan Srinivasa Garani
Department of Electronic System Engineering
Indian Institute of Science, Bengaluru

Lecture – 33
Optimization with Inequality Constraint

Alright, let's dive into the concept of inequality constraints and how they differ from equality constraints.

We've already discussed the conditions for achieving an optimal solution with equality constraints. Now, let's shift our focus to inequality constraints. Suppose we have an inequality constraint of the form $c(x) \geq 0$. When considering a direction d around a point x , we need to examine how $c(x) + \Delta c^T d$ behaves.

(Refer Slide Time: 02:46)

Single inequality constraint $c(\underline{x}) \geq 0$

$$0 \leq c(\underline{x} + \underline{d}) \approx c(\underline{x}) + \nabla^T c(\underline{x}) \underline{d}$$

Feasibility of \underline{d} is retained while still improving the objective if

$$c(\underline{x}) + \nabla^T c(\underline{x}) \underline{d} \geq 0 \quad \text{--- (C)}$$

Observe this as against in equality constraints

MORE VIDEOS

2:46 / 28:05

Specifically, we need to ensure that:

$$c(x) + \Delta c^T d \geq 0$$

Here, we need to be mindful of the change in inequality compared to the strict equality constraints we previously examined. In the case of a perturbation around x , if we expand $c(x) + \mathbf{d}$ using a Taylor series approximation, we get:

$$c(x + d) \approx c(x) + \Delta c^T d$$

Since $c(x)$ is already greater than or equal to zero, and for the inequality to hold, $c(x) + \Delta c^T d$ must also be greater than or equal to zero. This simplifies to:

$$c(x) + \Delta c^T d \geq 0$$

which we can denote as inequality C. This inequality ensures that the perturbed point $x + \mathbf{d}$ remains feasible under the constraint $c(x) \geq 0$.

(Refer Slide Time: 05:02)

ee53. lec33. Optimization with inequality constraint

Watch later Share

Considering the example from the circular constraints with inequality conditions, we are optimizing over all points lying on & inside the circle.

$c(x) \geq 0$

$x_1^2 + x_2^2 \leq a^2$

$-(x_1^2 + x_2^2) \geq -a^2$

$a^2 - x_1^2 - x_2^2 \geq 0$

MORE VIDEOS

5:02 / 28:05

CC YouTube

Let's consider an example with circular constraints. Suppose we have a constraint of the form $x_1^2 + x_2^2 \leq a^2$. This inequality describes all points within and on the boundary of a

circle of radius a . If we negate the inequality, changing it to $-x_1^2 - x_2^2 \geq -a^2$, and then rearrange it to $a^2 - x_1^2 - x_2^2 \geq 0$, this still represents the constraint where we are optimizing points inside and on the boundary of the circle.

So, for the constraint $c(x) \geq 0$, we can either be optimizing over points that lie strictly inside the circle (if $c(x) > 0$) or on the boundary of the circle (if $c(x) = 0$).

(Refer Slide Time: 08:57)

We have 2 cases

Case A : The strict inequality holds i.e., $c(\underline{x}) > 0$

Whenever $\nabla f(\underline{x}) \neq 0$ i.e., when we have not yet reached optimum points

(II) $\nabla f(\underline{x})^T \underline{d} < 0$ ———— (∵ (B))

$c(\underline{x}) + \nabla c(\underline{x})^T \underline{d} \geq 0$ ———— (∵ (C))

That satisfies the constraints is

$\underline{d} = -\frac{c(\underline{x}) \nabla f(\underline{x})}{\|\nabla f(\underline{x})\| \|\nabla c(\underline{x})\|}$ (D)

We encounter two cases in this context:

- Case A: The strict inequality $c(x) > 0$ holds, meaning the constraint is satisfied strictly inside the feasible region.
- Case B: The constraint is satisfied exactly, i.e., $c(x) = 0$, which places us on the boundary of the feasible region.

It's crucial to carefully distinguish between these two cases to accurately assess and solve optimization problems involving inequality constraints.

It's important to understand that when $c(x) > 0$, it doesn't always imply that we are dealing with points outside the circle. For instance, if we take $c(x) \geq 0$ and transform the constraint into $c(x) = a^2 - x_1^2 - x_2^2 \geq 0$, this still describes points on or inside the circle. Flipping the sign to $-x_1^2 - x_2^2 \geq -a^2$, which rearranges to $a^2 - x_1^2 - x_2^2 \geq 0$, is equivalent to describing points within or on the boundary of the circle. Thus, $c(x) \geq 0$ does not exclusively imply being outside; it also includes the interior and boundary of the circle.

(Refer Slide Time: 12:43)

(ii) Consider

$$c(x) + \nabla^T c(x) d$$

$$= c(x) + \nabla^T c(x) \left[-\frac{c(x) \nabla f(x)}{\|\nabla f(x)\| \|\nabla c(x)\|} \right]$$

$$= c(x) - c(z) \frac{\nabla^T c(z) \nabla f(z)}{\|\nabla f(z)\| \|\nabla c(z)\|} < 1$$

Unless $\nabla f(x) \neq \lambda \nabla c(x)$,

Moving on, if the gradient Δf is not zero, it indicates that we have not yet reached the optimal point. In this context, we have two key inequalities to consider:

1. $\Delta f^T \mathbf{d} < 0$ (which we denote as inequality B)
2. For the inequality constraint, $c(x) + \Delta c^T \mathbf{d} \geq 0$

When $c(x)$ is zero, the condition simplifies, but for general inequality constraints, $c(x) + \Delta c^T \mathbf{d} \geq 0$ reflects a stricter requirement. The inequality constraint $c(x) + \Delta c^T \mathbf{d} \geq 0$ differs from the equality constraint where $\Delta c^T \mathbf{d} = 0$ is required for orthogonality.

We denote the set of these inequalities as "inequality set number 2". A feasible direction \mathbf{d} that satisfies these constraints is given by:

$$\mathbf{d} = -\frac{c(x)\Delta f(x)}{|\Delta f||\Delta c|}$$

To verify this direction, we need to check that it satisfies both constraints in inequality set number 2.

(Refer Slide Time: 16:22)

ee53 lec33 Optimization with inequality constraint

$$\nabla c^T(\underline{x}) \nabla f(\underline{x}) < \|\nabla f(\underline{x})\| \|\nabla c(\underline{x})\|$$

We have $c(\underline{x}) + \nabla c^T(\underline{x}) \underline{d} \geq 0$

$$\Rightarrow c(\underline{x}) - c(\underline{x}) \underline{d} \geq 0$$

$$c(\underline{x})(1 - \underline{d}) \geq 0$$

\underline{d} can be +ve or -ve

(The equality is only over the case when $\underline{d} = 1$)

MORE VIDEOS

16:22 / 28:05

YouTube

First, we test the condition $\Delta f^T \mathbf{d}$. Substituting \mathbf{d} into this condition gives:

$$\Delta f^T \mathbf{d} = -\frac{c(x)\Delta f^T \Delta f}{|\Delta f||\Delta c|}$$

Here, $c(x)$ is a scalar, and $\Delta f^T \Delta f$ evaluates to $|\Delta f|^2$. Simplifying this expression:

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

Since $|\Delta f|$ and $|\Delta c|$ are non-zero positive quantities, $\Delta f^T \mathbf{d}$ is strictly negative, satisfying the first constraint.

Next, we check the second condition $c(x) + \Delta c^T \mathbf{d} \geq 0$. Substituting \mathbf{d} into this condition:

$$c(x) + \Delta c^T \mathbf{d} = c(x) - \frac{(\Delta c^T \Delta f)}{|\Delta f| |\Delta c|}$$

Here, $\Delta c^T \Delta f$ is a scalar quantity. We must verify that:

$$\frac{\Delta c^T \Delta f}{|\Delta f| |\Delta c|}$$

is less than or equal to 1. By the Cauchy-Schwarz inequality, this term will always be less than or equal to 1, and if $\Delta f \neq \lambda \Delta c$, the inequality is strict.

Thus, the direction \mathbf{d} derived satisfies the constraints set in inequality set number 2. This insight is crucial for effectively solving optimization problems with inequality constraints.

(Refer Slide Time: 21:00)

Case B: When \underline{x} is on the boundary of the constraint eqn i.e., $c(\underline{x}) = 0$

We have

- $\nabla f^T(\underline{x}) \underline{d} < 0$ — III.A
- $\nabla c^T(\underline{x}) \underline{d} \geq 0$ — III.B

(III) boundary case
 $\because c(\underline{x}) = 0$
 \therefore plug into (III)

FEASIBLE SOLV REGION (GEOMETRY)

Region where the (III) are constraints in mat.

$\nabla f^T(\underline{x}) \underline{d} < 0$ (open half plane)

MORE VIDEOS

21:00 / 28:05

YouTube

Let's analyze the situation where $\Delta C^T \Delta F$ is strictly less than $|\Delta f| \cdot |\Delta c|$. This inequality holds because we are not considering the case where Δf is equal to $\lambda \Delta c$, and thus we disregard that specific condition. Consequently, the quantity we are analyzing, which I've highlighted in red, evaluates to some value α that can be positive or negative, but in absolute terms, it is less than 1. Hence, the absolute value of this term is strictly less than 1.

Now, let's consider $c(x) - c(x) \cdot \alpha$, where α can be positive or negative, but its absolute value is less than 1. Simplifying this, we get $c(x) \cdot (1 - \alpha)$. Since $1 - \alpha$ remains positive whether α is positive or negative (as long as its absolute value is less than 1), $c(x) \cdot (1 - \alpha)$ is indeed positive and satisfies the condition ≥ 0 . The only exception is when $\alpha = 1$, where the quantity becomes zero. This specific case occurs if and only if $\Delta f = \lambda \Delta c$, which we've already ruled out. Therefore, we conclude that $c(x) \cdot (1 - \alpha)$ is strictly greater than or equal to 0, except in the case where $\alpha = 1$, where it exactly equals 0.

Now, turning to the scenario where x is on the boundary of the constraint. In Case A, we considered x inside the curve. However, when x is on the boundary, we need to take $c(x) = 0$ into account. This doesn't change the essence of the analysis; it just places x on the boundary of the feasible region.

We have two key constraints to consider in this boundary case:

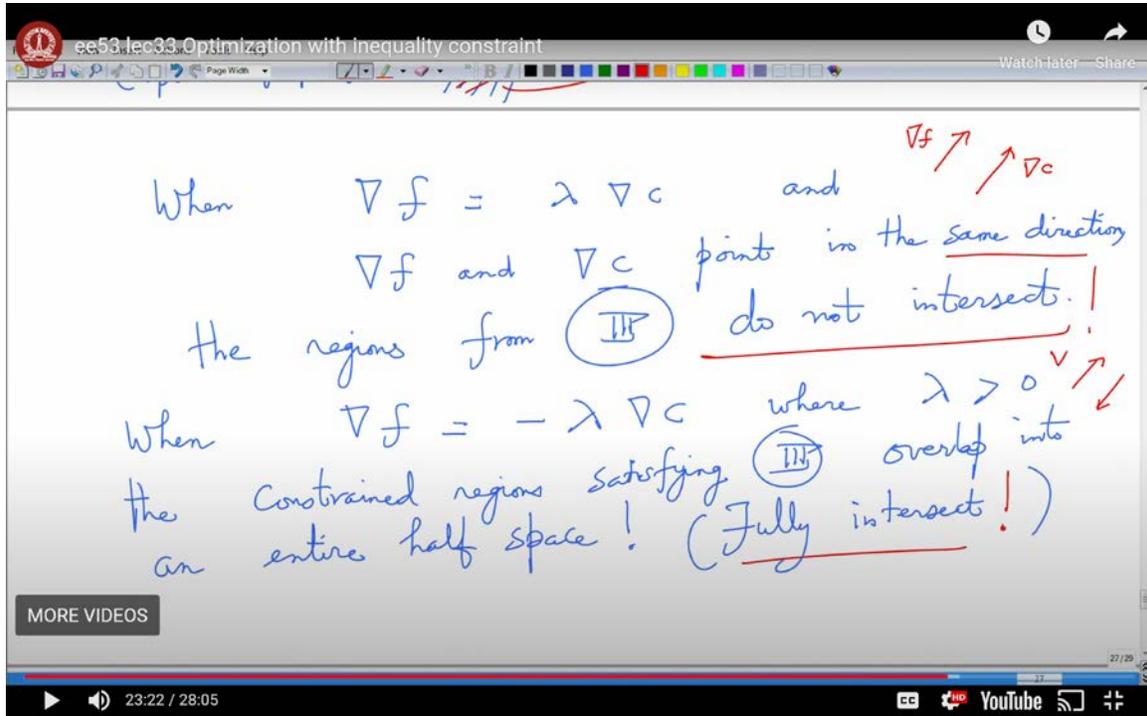
1. $\Delta F^T \mathbf{d} < 0$, which we denote as constraint B.
2. $c(x) + \Delta C^T \mathbf{d} \geq 0$. Given that $c(x) = 0$ on the boundary, this simplifies to $\Delta C^T \mathbf{d} \geq 0$.

To visualize the feasible region, let's label these constraints as follows:

- Constraint B: $\Delta F^T \mathbf{d} < 0$ represents an open half-space. The dotted line in this case represents the hyperplane where $\Delta f \cdot \mathbf{d} = 0$. Everything below this dotted line, shaded in blue, is part of this open half-space.
- Constraint C: $\Delta C^T \mathbf{d} \geq 0$ describes a region shaded in red, where this inequality holds.

The feasible region is the intersection of these two shaded areas: the blue region from B and the red region from C. This overlapped shaded region represents the feasible solution space. If this overlapped region is non-null, it indicates that a solution exists. Conversely, if there is no overlap, then no feasible solution is available.

(Refer Slide Time: 23:22)



When ΔF equals $\lambda \Delta C$, and ΔF and ΔC point in the same direction, this alignment is crucial. Specifically, ΔF and ΔC are parallel, meaning they are pointing in the same direction. In this scenario, the regions described by the constraints do not intersect. To understand this, consider the following:

For the constraints, we have:

- $\Delta F \cdot \mathbf{d} < 0$
- $\Delta C \cdot \mathbf{d} \geq 0$

If ΔC can be expressed as $\frac{1}{\lambda} \Delta F$, substituting this into the second condition yields:

$$\Delta C \cdot d = \frac{1}{\lambda} \Delta F \cdot d$$

Since $\frac{1}{\lambda} \Delta F \cdot d$ essentially translates to $\Delta F \cdot d$, we see that:

- If $\Delta F \cdot d < 0$, then $\Delta C \cdot d \geq 0$ doesn't hold because this implies $\Delta F \cdot d$ should be greater than or equal to 0, which is contradictory.

Therefore, when ΔF and ΔC point in the same direction, the constraint regions described do not intersect.

(Refer Slide Time: 26:33)

ee53.lec33.Optimization with inequality constraint

Watch later Share

Forming the Lagrangian for $\lambda > 0$ ($\lambda > 0$
 \Rightarrow No
Constraint)

If $L = f - \lambda c$ When $\lambda > 0$

$\nabla L = \nabla f - \lambda \nabla c = 0$

$\Rightarrow \nabla f = + \lambda \nabla c \Rightarrow$ The search stops since constraints are not met

With $c(x) \geq 0$ While forming the "Lagrangian" with inequality constraint, have a -1 sign before the constraint scaled by ' $\lambda > 0$ '!

MORE VIDEOS

26:33 / 28:05

YouTube

In contrast, consider the case where $\Delta F = -\lambda \Delta C$, meaning ΔF and ΔC point in opposite directions. Visualize this with the cone concept:

- The first constraint $\Delta F \cdot d < 0$
- The second constraint translates to $\Delta C \cdot d \leq 0$, which, when expressed with ΔF , implies $\Delta F \cdot d \leq 0$

In this situation, when ΔF and ΔC are oppositely directed, the constraint regions fully overlap, forming a half-space. Essentially, the plane defined by $\Delta C^T \mathbf{d} \geq 0$ aligns with the constraint space, leading to a full overlap.

Thus, if ΔF and ΔC point in opposite directions, the constraint regions intersect completely. If they point in the same direction, there is no intersection.

Now, considering the implications for the Lagrangian function L :

$$L = f - \lambda c$$

where $\lambda > 0$ (as $\lambda = 0$ would imply no constraint). To find the saddle point, we compute the gradient of L :

$$\Delta L = \Delta f - \lambda \Delta c$$

Setting this gradient to zero:

$$\Delta f - \lambda \Delta c = 0$$

implies:

$$\Delta f = \lambda \Delta c$$

If a negative sign were mistakenly introduced, we would incorrectly find:

$$\Delta f = -\lambda \Delta c$$

which highlights the critical role of sign consistency in gradient calculations.

In our discussion, we observed that when Δf and Δc point in the same direction, there is no intersection between the feasible regions. This implies that the search should terminate at this point, as the constraints are not being satisfied. It is crucial to recognize this because if Δf and Δc are aligned, it indicates that no further feasible solutions exist, and the optimization process should halt.

When formulating the Lagrangian with inequality constraints, it is essential to include a negative sign before the constraint, scaled by some Lagrange multiplier λ , which must be greater than or equal to zero. This negative sign is crucial; omitting it can render the entire search process ineffective. This detail, though often taken for granted, is vital for correct formulation.

In optimization courses, this aspect is carefully addressed. However, if you have not encountered this in such courses, you might question whether to use a plus or minus sign, and this choice makes a significant difference. Using a plus sign would incorrectly suggest that you are still searching within a half-plane, which ultimately makes the search meaningless.

(Refer Slide Time: 27:19)

If the inequality was $c(\underline{x}) \leq 0$,
we can form a $g(\underline{x}) \geq 0$ such that
 $g(\underline{x}) = -c(\underline{x}) \geq 0$

As previously mentioned, a constraint of the form $c(x) \leq 0$ can be transformed into $g(x) \geq 0$ by negating the sign of $c(x)$. The analysis remains valid regardless of whether the original inequality is ≤ 0 or ≥ 0 . By transforming the inequality and applying a negative sign to the Lagrangian, the formulation and subsequent analysis are appropriately adjusted.

This detail is particularly important when studying support vector machines under various equality and inequality constraints. Understanding this result is crucial for correctly formulating and solving optimization problems.