Optimization Methods (CS1.404), Spring 2024
Lecture 18

Naresh Manwani

Machine Learning Lab, IIIT-H

March 18th, 2024

Characterization of Local Minima in Constrained Optimization

Theorem

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

- $\mathbf{x}^* \in \mathcal{X}$ is a local minima if $\Rightarrow \mathcal{F}(\mathbf{x}^*) \cap \mathcal{D}(\mathbf{x}^*) = \phi$.
- Consider any $\mathbf{x} \in \mathcal{X}$ and assume $f \in \mathbb{C}^2$. Then $\nabla f(\mathbf{x})^T \mathbf{d} < 0$ implies there exists $\delta > 0$ such that $f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x})$, $\forall \alpha \in (0, \delta)$. Such $\mathbf{d}$ is a descent direction ($\mathbf{d} \in \mathcal{D}(\mathbf{x})$).
- Let $\tilde{\mathcal{D}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla f(\mathbf{x})^T \mathbf{d} < 0\}$, then $\tilde{\mathcal{D}}(\mathbf{x}) \subseteq \mathcal{D}(\mathbf{x})$.

Active Constraint

- Consider the problem $\min\ f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0,\ j = 1 \ldots l$, $\mathbf{x} \in \mathbb{R}^n$.
- Assume $h_j \in \mathbb{C}^2,\ j = 1 \ldots l$.
- Let $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0,\ j = 1 \ldots l\}$.
- An inequality constraint $h_j(\mathbf{x}) \leq 0$ is said to be active at $\mathbf{x}^*$ if $h_j(\mathbf{x}^*) = 0$. It is inactive if $h_j(\mathbf{x}^*) < 0$.
- Set of active constraints $\mathcal{A}(\mathbf{x}) = \{j \in \{1, \ldots, l\} \mid h_j(\mathbf{x}) = 0\}$.

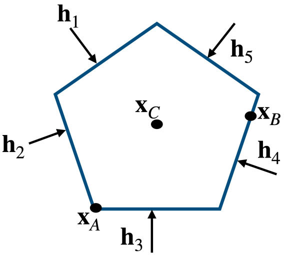

Figure: At $\mathbf{x}_A$, h_2 and h_3 are active. At $\mathbf{x}_B$, h_4 is active. At $\mathbf{x}_C$, no constraint is active.

Characterization of Feasible Directions

Lemma

For any $\mathbf{x} \in \mathcal{X}$, we have

$$\tilde{\mathcal{F}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x})^T \mathbf{d} < 0,\ j \in \mathcal{A}(\mathbf{x})\} \subseteq \mathcal{F}(\mathbf{x}).$$

Necessary Condition for Local Minima

- Consider the problem $\min \ f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0, \ j = 1 \dots l$, $\mathbf{x} \in \mathbb{R}^n$.
- Let $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0, \ j = 1 \dots l\}$.
- $\tilde{\mathcal{D}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla f(\mathbf{x})^T \mathbf{d} < 0\} \subseteq D(\mathbf{x})$.
- $\tilde{\mathcal{F}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x})^T \mathbf{d} < 0, \ j \in \mathcal{A}(\mathbf{x})\} \subseteq \mathcal{F}(\mathbf{x})$.
- $\mathbf{x}^* \in \mathcal{X}$ is a local minima if $\Rightarrow \mathcal{F}(\mathbf{x}^*) \cup \mathcal{D}(\mathbf{x}^*) = \phi$. Which implies $\tilde{\mathcal{F}}(\mathbf{x}) \cap \tilde{\mathcal{D}}(\mathbf{x}) = \phi$.
- $\tilde{\mathcal{F}}(\mathbf{x}) \cap \tilde{\mathcal{D}}(\mathbf{x}) = \phi$ is only necessary. If this condition is satisfied for $\mathbf{x}^*$, then it does not mean that $\mathbf{x}^*$ is a local minima.

Examples: Necessary Condition for Local Minima

- **Example 1:** Consider optimization problem min $x_1^2 + x_2^2$ such that $(x_1 + x_2 - 1)^3 \leq 0$ and $x_1, x_2 \geq 0$.
  - Here, $h_1(x_1, x_2) = (x_1 + x_2 - 1)^3$, $h_2(x_1, x_2) = -x_1$, $h_3(x_1, x_2) = -x_2$
  - Consider a point $\mathbf{x}_A = (a, b) \in \mathbb{R}^2$ such that $a + b = 1$ and $a > 0, b > 0$.
  - We can see that $\nabla h_1(\mathbf{x}_A) = \mathbf{0}$. Thus, $\tilde{\mathcal{F}}(\mathbf{x}_A) = \{\mathbf{d} \mid \nabla h_1(\mathbf{x}_A)^T \mathbf{d} < 0\} = \phi$.
  - Here, $\mathbf{x}_A$ is not a local minima but $\tilde{\mathcal{F}}(\mathbf{x}_A) \cup \tilde{\mathcal{D}}(\mathbf{x}_A) = \phi$.
- **Example 2:** Consider optimization problem min $x_1^2 + x_2^2$ such that $(x_1 + x_2 - 1) \leq 0$ and $x_1, x_2 \geq 0$.
  - Here, $h_1(x_1, x_2) = (x_1 + x_2 - 1)$, $h_2(x_1, x_2) = -x_1$, $h_3(x_1, x_2) = -x_2$
  - Consider a point $\mathbf{x}_A = (a, b) \in \mathbb{R}^2$ such that $a + b = 1$ and $a > 0, b > 0$.
  - We can see that $\nabla h_1(\mathbf{x}_A) = [1\ 1]^T$. Thus, $\tilde{\mathcal{F}}(\mathbf{x}_A) = \{\mathbf{d} \mid \nabla h_1(\mathbf{x}_A)^T \mathbf{d} < 0\} = \{\mathbf{d} \mid d_1 + d_2 < 0\} \neq \phi$.
  - Here, $\mathbf{x}_A$ is not a local minima because $\tilde{\mathcal{F}}(\mathbf{x}_A) \cup \tilde{\mathcal{D}}(\mathbf{x}_A) \neq \phi$.

Example 3: Consider optimization problem min $x_1^2 + x_2^2$ such that $(x_1 + x_2 - 1) = 0$.

- Here, $h_1(x_1, x_2) = (x_1 + x_2 - 1)$, $h_2(x_1, x_2) = -x_1 - x_2 + 1$
- Consider a point $\mathbf{x}_A = (a, b) \in \mathbb{R}^2$ such that $a + b = 1$.
- Thus, $\tilde{\mathcal{F}}(\mathbf{x}_A) = \{\mathbf{d} \mid \nabla h_1(\mathbf{x}_A)^T \mathbf{d} < 0,\ \nabla h_2(\mathbf{x}_A)^T \mathbf{d} < 0\} = \phi$.
- This does not guarantee that $\mathbf{x}_A$ is a local minima.
- **The above Necessary Condition for Local Minima is only applicable when there are inequality constraints.**
- **It is not applicable when there are equality constraints.**

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Necessary Condition for Local Minima

- Consider the problem $\min\ f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0,\ j = 1 \dots l$, $\mathbf{x} \in \mathbb{R}^n$.
- Let $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0,\ j = 1 \dots l\}$.
- $\tilde{\mathcal{D}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla f(\mathbf{x})^T \mathbf{d} < 0\} \subseteq D(\mathbf{x})$.
- $\tilde{\mathcal{F}}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x})^T \mathbf{d} < 0,\ j \in \mathcal{A}(\mathbf{x})\} \subseteq \mathcal{F}(\mathbf{x})$.
- $\mathbf{x}^* \in \mathcal{X}$ is a local minima if $\tilde{\mathcal{F}}(\mathbf{x}) \cap \tilde{\mathcal{D}}(\mathbf{x}) = \phi$.
- Let $A = \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ \nabla h_{j_1}(\mathbf{x}^*) \\ \vdots \\ \nabla h_{j_k}(\mathbf{x}^*) \end{bmatrix}$ assuming that there are k active constraints $h_{j_1}, \dots, h_{j_k}$.
- Then $\mathbf{x}^* \in \mathcal{X}$ is a local minima if $\{\mathbf{d} \in \mathbb{R}^n \mid A\mathbf{d} < \mathbf{0}\} = \phi$.

Farkas Lemma

Lemma

Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{c} \in \mathbb{R}^n$. Then, exactly one of the two systems has a solution:

1. $A\mathbf{x} \leq \mathbf{0}, \; \mathbf{c}^T\mathbf{x} > 0$ for some $\mathbf{x} \in \mathbb{R}^n$
2. $A^T\mathbf{y} = \mathbf{c}, \; \mathbf{y} \geq \mathbf{0}$ for some $\mathbf{y} \in \mathbb{R}^m$

Gordon's Theorem

1. $A\mathbf{x} < \mathbf{0}$ for some $\mathbf{x} \in \mathbb{R}^n$
2. $A^T\mathbf{y} = 0, \; \mathbf{y} \geq \mathbf{0}$ for some $\mathbf{y} \in \mathbb{R}^m$

Fritz-John Condition for Local Minima

- $\mathbf{x}^* \in \mathcal{X}$ is a local minima if $\{\mathbf{d} \in \mathbb{R}^n \mid A\mathbf{d} < \mathbf{0}\} = \phi$.
- Using Gordon's Theorem, $\exists \lambda_0 \geq 0$ and $\lambda_j \geq 0,\ j \in \mathcal{A}(\mathbf{x}^*)$, not all λ's zero, such that

$$\lambda_0 \nabla f(\mathbf{x}^*) + \sum_{j \in \mathcal{A}(\mathbf{x}^*)} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

- Define $\lambda_j = 0,\ \forall j \notin \mathcal{A}(\mathbf{x}^*)$. Then the above condition is same as

$$\lambda_0 \nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

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

$$\lambda_j \geq 0; \quad j = 0, 1, \ldots, l$$

Issues with Fritz-John Condition

- A major drawback of the Fritz-John conditions is in the fact that they allows λ_0 to be zero.
- The case $\lambda_0 = 0$ is not particularly informative since condition. In this case, Fritz-John condition becomes

$$\sum_{j \in \mathcal{A}(\mathbf{x}^*)} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

- This means that the gradients of the active constraints are linearly dependent.
- This condition has nothing to do with the objective function.
- This implies that there can be many points which satisfy Fritz-John condition which are not local minima.

Next we will see how KKT condition can overcome this issue.

Regular Point

Definition

A point $\mathbf{x}^* \in \mathcal{X}$ is said to be regular point if the gradient vectors $\nabla h_j(\mathbf{x}^*),\ j \in \mathcal{A}(\mathbf{x}^*)$, are linearly independent. Then, $\sum_{j\in\mathcal{A}(\mathbf{x}^*)} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$ only if $\lambda_j = 0,\ \forall j \in \mathcal{A}(\mathbf{x}^*)$.

Lemma

Consider $\min\ f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0,\ j = 1 \ldots l,\ \mathbf{x} \in \mathbb{R}^n$. If $\mathbf{x}^*$ is a regular point and a local minima, then $\lambda_0 \neq 0$.

Proof:

- If $\mathbf{x}^*$ is a regular point and local minima, then Frtiz-John optimality condition implies,

$$\lambda_0 \nabla f(\mathbf{x}^*) + \sum_{j\in\mathcal{A}(\mathbf{x}^*)} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}.$$

- Let $\lambda_0 = 0$. Then, $\sum_{j\in\mathcal{A}(\mathbf{x}^*)} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$, making gradients of the active constraints linearly dependent.
- However, this contradicts that $\mathbf{x}^*$ is a regular point. Thus, $\lambda_0 \neq 0$.

This is the key idea in KKT conditions taking $\lambda_0 = 1$.

KKT Optimality Conditions: First Order

KKT Optimality Conditions of First Order

Consider the problem $\min \; f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0, \; j = 1 \dots l$, $\mathbf{x} \in \mathbb{R}^n$. Assume that $\mathbf{x}^* \in \mathcal{X}$ to be a regular point and $\mathbf{x}^*$ is a local minima. Then there exist $\lambda_j \; j = 1 \dots l$, such that

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

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

$$\lambda_j \geq 0; \;\; j = 1, \dots, l$$

- These are first order KKT necessary conditions.
- KKT point: $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, where $\boldsymbol{\lambda}^* = [\lambda_1^* \;\; \lambda_2^* \;\; \dots \;\; \lambda_l^*]^T$.

Lagrangian Function

- Lagrangian function is represented as:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \nabla f(\mathbf{x}) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x})$$

- KKT Conditions imply

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$
$$\lambda_j^* \geq 0; \;\; j = 1 \ldots l \quad (\lambda's \text{ are called Lagrange multipliers.})$$
$$\lambda_j^* h_j(\mathbf{x}^*) = 0; \;\; j = 1 \ldots l \quad (\text{Complementary slackness conditions.})$$
$$\lambda_j^* = 0; \;\; \forall j \in \mathcal{A}(\mathbf{x}^*)$$

- Note that for active constraints, $\lambda_j^* h_j(\mathbf{x}^*) = 0$ because $h_j(\mathbf{x}^*) = 0$. Thus, λ_j^* can be zero or greater than zero.
- For non-active constraints, $h_j(\mathbf{x}^*) < 0$. Thus, $\lambda_j^* h_j(\mathbf{x}^*) = 0$ implies $\lambda_j^* = 0$.

$$\min_{x_1, x_2} \; x_1^2 + x_2^2$$
$$s.t. \;\; h_1: \; x_1 + x_2 \geq 1$$
$$h_2: \; x_2 \leq 1$$

Case 1:

- Let h_1 and h_2 both are active constraints. $\mathbf{z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
- At this point both h_1 and h_2 are active. Thus, $\lambda_1^* \geq 0$ and $\lambda_2^* \geq 0$.
- $\mathcal{L} = x_1^2 + x_2^2 + \lambda_1^*(x_2 - 1) + \lambda_2^*(1 - x_1 - x_2)$
- $\nabla f(\mathbf{z}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, $\nabla h_2(\mathbf{z}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\nabla h_1(\mathbf{z}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$.
- $\nabla \mathcal{L} = \mathbf{0}$ implies $\begin{pmatrix} 0 \\ 2 \end{pmatrix} + \lambda_1^* \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
- This results in $\lambda_1^* = 0$ and $\lambda_2^* = -2$. Thus, $\mathbf{z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is not a local minima.

Case 2:

- Assume constraint h_1 is active and h_2 is inactive. Thus, $x_1 + x_2 = 1$ and $x_2 < 1$ at the solution.
- This implies $\lambda_2^* = 0$.
- KKT condition results $\nabla f(\mathbf{x}^*) + \lambda_1^* \nabla h_1(\mathbf{x}^*) = \mathbf{0}$.
- Which results in $\begin{pmatrix} 2x_1^* \\ 2x_2^* \end{pmatrix} + \lambda_1^* \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
- This means $2x_1^* - \lambda_1^* = 0$, $2x_2^* - \lambda_1^* = 0$ and $x_1^* + x_2^* = 1$.
- This results in $x_1^* = x_2^* = 0.5$ and $\lambda_1^* = 1$ and $\lambda_2^* = 0$.
- $x_1^* = x_2^* = 0.5$ is a KKT point at which $f(\mathbf{x}^*) = 0.5$.

Example 1: Continue

Case 3:

- Assume constraint h_1 is inactive and h_2 is active. Thus, $x_1 + x_2 > 1$ and $x_2 = 1$ at the solution.
- This implies $\lambda_1^* = 0$.
- KKT condition results $\nabla f(\mathbf{x}^*) + \lambda_2^* \nabla h_2(\mathbf{x}^*) = \mathbf{0}$.
- Which results in $\begin{pmatrix} 0 \\ 2 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
- This means $\lambda_2^* = -2$, which is not a feasible solution.

Thus, $x_1^* = x_2^* = 0.5$ **is the minima.**

Finding KKT Point

Consider the minimization problem.

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ s.t. \quad & h_j(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

- At a local minimum, active set is unknown.
- Need to investigate all possible active sets for finding KKT points.

Is KKT Point Always Optimal?

KKT Point may not be optimal always. See the example below.

- Consider min $-x^2$ such that $x \leq 0$.
- $\mathcal{L}(x, \lambda) = -x^2 + \lambda x$
- $\frac{\partial \mathcal{L}}{\partial x} = 0 \Rightarrow -2x + \lambda = 0$
- At x^*, the constraint is active. Thus, $x^* = 0$.
- $\frac{\partial \mathcal{L}(x^*, \lambda^*)}{\partial x} = 0 \Rightarrow \lambda^* = 0$.
- (0,0) is a KKT point.
- However, $-x^2$ is unbounded in $x \leq 0$ and $x^* = 0$ is not a local minimum.