Optimization Theory and Applications

Kun Zhu (zhukun@nuaa.edu.cn)

November 6, 2018

Introduction

- We consider the optimization problem

$$\begin{array}{ll} \textit{minimize} & f(\mathbf{x}) \\ \textit{subject to} & \mathbf{x} \in \Omega \end{array}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is the ***objective function***
  - Cost function
  - Utility function
- $\mathbf{x}$ is an n-vector of independent ***decision variables***: $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}^T \in \mathbb{R}^n$
- The set $\Omega \subset \mathbb{R}^n$ is the ***constraint set*** or ***feasible set***
- Question: What if $\Omega = \mathbb{R}^n$?

Introduction

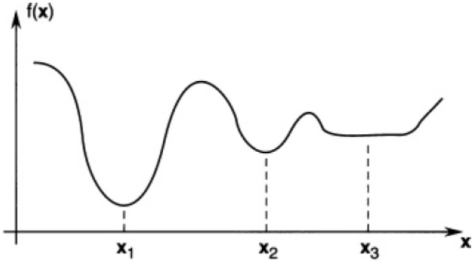
- The optimization problem is a decision problem for finding the "best" $\mathbf{x}$ over all possible vectors in Ω
- The "best" $\mathbf{x}$ which results in the smallest value of the objective function is the ***minimizer***
- Possible multiple minimizers
- Maximization problem

Types of Minimizers

- ***Definition:*** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $\mathbf{x}^*$ is a ***global minimizer*** of f over Ω if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \backslash \{\mathbf{x}^*\}$
- A point $\mathbf{x}^* \in \Omega$ is a ***local minimizer*** of f over Ω if there exists $\varepsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \backslash \{\mathbf{x}^*\}$ and $\|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$
- If we replace $\geq$ with $>$
  - Strict local minimizer
  - Strict global minimizer

Types of Minimizers

- An example for $n = 1$
  - x_1: strict global minimizer
  - x_2: strict local minimizer
  - x_3: local (not strict) minimizer



Introduction

- If $\mathbf{x}^*$ is a global minimizer of f over Ω, we write
  - $f(\mathbf{x}^*) = \min_{\mathbf{x}\in\Omega} f(\mathbf{x})$
  - $\mathbf{x}^* = \arg\ \min_{\mathbf{x}\in\Omega} f(\mathbf{x})$
- Example: $f(x) = (x+1)^2 + 3$
  - $\arg\ \min f(\mathbf{x}) = -1$ for $\Omega = \mathbb{R}^n$
  - $\arg\ \min f(\mathbf{x}) = 0$ for $\Omega = [0, \infty)$
- Strictly speaking, an optimization problem is solved only when a global minimizer is found
- Global minimizer *vs* local minimizer

Existence of Minimizers

- Theorem of Weierstrass: If f is continuous and Ω is closed and bounded, then a global minimizer exists

Conditions: Necessary and Sufficient

- We seek conditions that characterize minimizers
- Two types of conditions: necessary and sufficient
- Necessary condition: If $\mathbf{x}^*$ is a minimizer, then $\mathbf{x}^*$ satisfies this particular condition
- Sufficient condition: If $\mathbf{x}^*$ satisfies this particular condition, then $\mathbf{x}^*$ is a minimizer.

Conditions: Necessary and Sufficient

- A necessary condition limits the set of candidates for minimizers
- A sufficient condition guarantees that a point is a minimizer
- We consider conditions that are based on gradients and Hessians. These conditions apply to local minimizers

Conditions for Local Minimizers

- Consider the totally unconstrained problem

$$\begin{array}{ll} \textit{minimize} & f(\mathbf{x}) \\ \textit{subject to} & \mathbf{x} \in \mathbb{R}^n \end{array}$$

- Assume $f \in C^1$
- Theorem: If $\mathbf{x}^*$ is a local minimizer, then

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

- For $n = 1$, this theorem should be very familiar (“slope=0”)
- First order necessary condition (FONC)

Conditions for Local Minimizers

- Consider the general set constrained problem

$$\begin{array}{ll} \textit{minimize} & f(\mathbf{x}) \\ \textit{subject to} & \mathbf{x} \in \Omega \end{array}$$

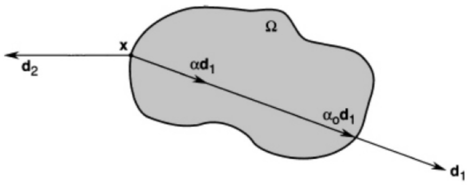
- If $\mathbf{x}^*$ is a local minimizer and an interior point of Ω, then

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

- Need to consider the boundary case

Feasible Direction

- A minimizer could be an interior point or a boundary point
- For discussing the conditions for a boundary minimizer, we introduce the concept of feasible direction
- **Definition:** A vector $\mathbf{d} \in \mathbb{R}^n, \mathbf{d} \neq 0$, is a ***feasible direction*** at $\mathbf{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\mathbf{x} + \alpha\mathbf{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$



Directional Derivative

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function
- Let $\mathbf{d}$ be a feasible direction at $\mathbf{x} \in \Omega$
- The ***directional derivative*** of f in the direction $\mathbf{d}$, denoted by $\partial f / \partial \mathbf{d}$, is the ***real-valued function*** defined by

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha \to 0} \frac{f(\mathbf{x}+\alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

- If $\|\mathbf{d}\| = 1$, $\partial f / \partial \mathbf{d}$ is the ***rate of increase*** of f at $\mathbf{x}$ in the direction $\mathbf{d}$

Directional Derivative

- Suppose $\mathbf{x}$ and $\mathbf{d}$ are given, then $f(\mathbf{x}+\alpha\mathbf{d})$ is ***a function of*** α

$$\phi(\alpha) = f(\mathbf{x}+\alpha d)$$

- The directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{d}$ is

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \lim_{\alpha\to 0}\frac{f(\mathbf{x}+\alpha\mathbf{d})-f(\mathbf{x})}{\alpha} = \frac{d}{d\alpha}f(\mathbf{x}+\alpha\mathbf{d})|_{\alpha=0}$$

- By the chain rule

$$\frac{d}{d\alpha}f(\mathbf{x}+\alpha\mathbf{d})|_{\alpha=0} = \nabla f(\mathbf{x})^T\mathbf{d} = \mathbf{d}^T\nabla f(\mathbf{x})$$

- Hence

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \mathbf{d}^T\nabla f(\mathbf{x})$$

- When $\|\mathbf{d}\| = 1$, the rate of increase of f in direction $\mathbf{d}$ can be represented by $< \nabla f(\mathbf{x}), \mathbf{d} >$

Directional Derivative

- Example: Define $f(\mathbf{x}) = x_1 x_2 x_3$ and let $\mathbf{d} = [1/2, 1/2, 1/\sqrt{2}]^T$. Calculate the directional derivative of f in the direction $\mathbf{d}$

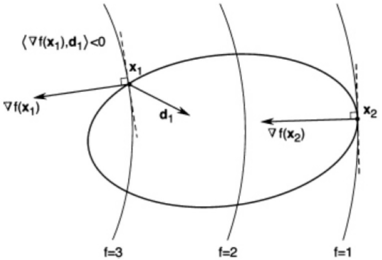
Directional Derivative

- Example: Define $f(\mathbf{x}) = x_1x_2x_3$ and let $\mathbf{d} = [1/2, 1/2, 1/\sqrt{2}]^T$. The directional derivative of f in the direction $\mathbf{d}$ is

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^\top \boldsymbol{d} = [x_2x_3, x_1x_3, x_1x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2x_3 + x_1x_3 + \sqrt{2}x_1x_2}{2}.$$

Conditions for Local Minimizers

- **Theorem: First-Order Necessary Condition** (**FONC**). Let $f \in \mathcal{C}^1$ a real-valued function on $\Omega \subset \mathbb{R}^n$. If $\mathbf{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\mathbf{d}$ at $\mathbf{x}^*$, we have $\mathbf{d}^T \nabla f(\mathbf{x}^*) \geq 0$
  - Alternatively, the directional derivative $\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}^*) \geq 0$ for all feasible directions $\mathbf{d}$
  - ***The rate of increase*** of f at $\mathbf{x}^*$ in ***any feasible direction*** $\mathbf{d} \in \Omega$ is nonnegative



Conditions for Local Minimizers

- **Corollary: Interior Case** (**FONC**). Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\mathbf{x}^*$ is a local minimizer of f over Ω and $\mathbf{x}^*$ is an interior point, then for any feasible direction $\mathbf{d}$ at $\mathbf{x}^*$, we have

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

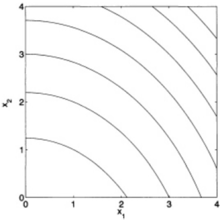
- We can use ***FONC*** to find candidate optimal solutons

Conditions for Local Minimizers

- Example 1:

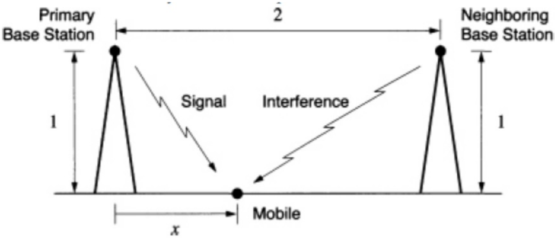
$$\begin{aligned} \textit{minimize}: &\quad x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ \textit{subject to}: &\quad x_1, x_2 \geq 0 \end{aligned}$$

- Questions:
  - Is the first-order necessary condition (FONC) for a local minimizer satisfied at $\mathbf{x} = [1, 3]^T$?
  - Is the FONC for a local minimizer satisfied at $\mathbf{x} = [0, 3]^T$?
  - Is the FONC for a local minimizer satisfied at $\mathbf{x} = [1, 0]^T$?
  - Is the FONC for a local minimizer satisfied at $\mathbf{x} = [0, 0]^T$?
- The plot of level sets of example 1



Conditions for Local Minimizers

- Example 2: Two base stations transmitting with equal power. We are interested in finding the position of the mobile that maximizes the signal-to-interference ratio



Conditions for Local Minimizers

- The signal-to-interference ratio is
$$f(x) = \frac{1+(2-x)^2}{1+x^2}$$
and we have
$$\begin{aligned} f'(x) &= \frac{-2(2-x)(1+x^2) - 2x(1+(2-x)^2)}{(1+x^2)^2} \\ &= \frac{4(x^2-2x-1)}{(1+x^2)^2}. \end{aligned}$$

- By the FONC, at the optimal position x^*, we have $f'(x^*) = 0$. Accordingly, we have $x^* = 1-\sqrt{2}$ or $x^* = 1+\sqrt{2}$

Conditions for Local Minimizers

- **Theorem: Second-Order Necessary Conditions (SONC).** Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, $\mathbf{x}^*$ a local minimizer of f over Ω, and $\mathbf{d}$ a feasible direction at $\mathbf{x}^*$. If $\mathbf{d}^T \nabla f(\mathbf{x}^*) = 0$ then

  $$\mathbf{d}^T F(\mathbf{x}^*)\mathbf{d} \geq 0,$$

  where F is the Hessain of f

- Question: Why do we need the SONC?

Conditions for Local Minimizers

- **Corollary: Interior Case.** Let $\mathbf{x}^*$ be an interior point of $\Omega \subset \mathbb{R}^n$. If $\mathbf{x}^*$ is a local minimizer of f over Ω,then

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

and $F(\mathbf{x}^*)$ is positive semidefinite $F(\mathbf{x}^*) \geq 0$; that is, for all $\mathbf{d} \in \mathbb{R}^n$,

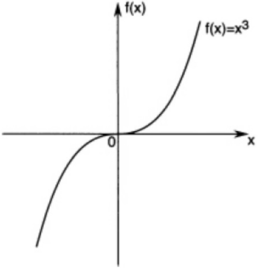
$$\mathbf{d}^T F(\mathbf{x}^*)\mathbf{d} \geq 0$$

Summary: Second Order Necessary Condition

- Interior case: $F(\mathbf{x}^*) \geq 0$
- General case: If $\mathbf{d}^T \nabla f(\mathbf{x}^*) = 0$ for a feasible direction $\mathbf{d}$, then $\mathbf{d}^T F(\mathbf{x}^*) \mathbf{d} \geq 0$
- Note: the above conditions are ***necessary conditions*** and not ***sufficient conditions***

Conditions for Local Minimizers

- Example: Consider a function $f(x) = x^3$. We have $f'(0) = 0$ and $f''(0) = 0$, the point $x = 0$ satisfies both the FONC and SONC. However, $x = 0$ is not a minimizer

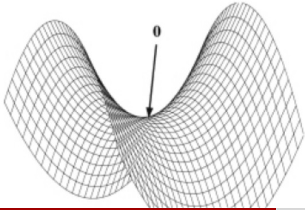


Conditions for Local Minimizers

- Example: Consider a function $f(x) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(\mathbf{x}) = [2x_1, -2x_2]^T = 0$. Thus $\mathbf{x} = [0, 0]^T$ satisfies the FONC. The Hessian matrix of f is

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

- The Hessian matrix is indefinite
  - For some $\mathbf{d_1}$, we have $\mathbf{d_1}^T F \mathbf{d_1} > 0$, (e.g., $\mathbf{d_1} = [1, 0]^T$)
  - For some $\mathbf{d_2}$, we have $\mathbf{d_2}^T F \mathbf{d_2} < 0$, (e.g., $\mathbf{d_2} = [0, 1]^T$).
  - Thus $\mathbf{x} = [0, 0]^T$ does not satisfy the SONC, and is not a minimizer



Conditions for Local Minimizers

- **Theorem: Second-Order Sufficient Condition (SOSC), Interior Case.** Let $\Omega \subset \mathcal{C}^2$ be defined on a region in which $\mathbf{x}^*$ is an interior point. Suppose that

  1. $\nabla f(\mathbf{x}^*) = 0$
  2. $F(\mathbf{x}^*) > 0$

  Then, $\mathbf{x}^*$ is a strict local minimizer of f

Conditions for Local Minimizers

- Example: Let $f(\mathbf{x}) = x_1^2 + x_2^2$. We have $\nabla f(\mathbf{x}) = [2x_1, 2x_2]^T = 0$ if and only if $\mathbf{x} = [0, 0]^T$. For all $\mathbf{x} \in \mathbb{R}^2$, we have

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0$$

- The point $\mathbf{x} = [0, 0]$ satisfies the FONC, SONC, Aand SOSC. It is a strict local minimizer