01. Introduction

최적화 이론 발전과정

1) Pythagoras

: 수학의 원리는 모든 것이다.

→ 수학을 통해 우리가 살고 있는 세상을 모형화 할 수 있다.

2) Plato / Aristole

: 추론을 이용한 사회의 최적화

→ 개인의 생활 방식을 최적화하고, 상태를 기능화하면

최상의 삶을 구현할 수 있다.

3) Euclid

: 수학적 추상의 최적화 (기하학의 최적화)

→ 기하학의 최적화 문제 : 한 점에서 원주에 이르는 거리 최적화.

4) Zenodorous

: Dido's 문제

→ 쇠가죽 채찍으로 만들 수 있는 최대 면적

5) Heron

: 자연 현상 최적화

→ 빛은 두 지점을 지나는 가장 짧은 길이로 지난다.

6) Pappus

: 자연 현상의 최적화

→ 육각형 : 최소 재료를 이용해 이차원 평면상의 단위격자 구성

7) al-khwarizmi

: 대수와 알고리즘 → 수학 기호를 다루기 위한 규칙

8) Descartes

: 이차원 공간의 좌표 표현해 곡선의 접선

→ 공간 (좌표 집합) 탐색을 통한 최적화 해 함수의 최소/최대.

9) Fermat

: 도함수 정리 → 도함수를 이용한 최적점 식별

10) Leibniz / Newton

: 미적분학 → 무한 수열의 수렴성.

$\boxed{\text{컴퓨터 등장}}$: 최적화를 위한 계산 알고리즘

- kantorovich : 선형 계획법
- Dantzing : 단체법 (심플렉스 법)
- Bellman : 동적 프로그래밍 (동적 계획)
- IBM Deep Blue : "Chess"
- IBM Watson : Jeopardy
- Google AlphaGo : Go (+ 의사 결정 문제)
- AI : Deep Neural Network

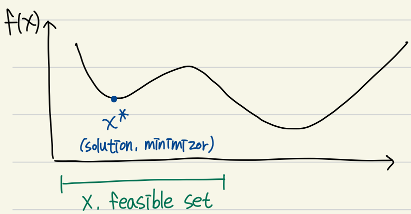
* 최적화 방법 학습 위한 필요 지식

: 미적분학, 행렬 이론 , Python

최적화 문제의 근본

$$\boxed{\underset{x}{\text{minimize}}\ f(x) \qquad \text{subject to } x \in X}$$

- x : 계획점 (design point) ← 모수, vector
- 계획 변수 : $[x_1, x_2, \cdots, x_n]$
  - └ x_i는 i번째 계획 변수, scalar
- f : 목적함수 (objective function)
- X : 해가 될 수 있는 모든 점들의 집합 (공간), feasible set
  - ex) 거리, 자연수
- x^* : 해, minimizer, 목적 함수를 최소화 하는 점.

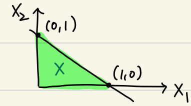


$\underset{x}{\text{maximize}}\ f(x)$, subject to $x \in X$

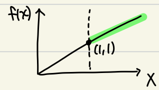
$\Leftrightarrow \underset{x}{\text{minimize}}\ -f(x)$, subject to $x \in X$ (자주 사용)

제약 조건

$\underset{x_1, x_2}{\text{minimize}}\ f(x_1, x_2)$, subject to $x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \geq 0$



$\underset{x}{\text{minimize}}\ x$, subject to $x > 0$



⇒ 0에 가까운 값이면 전부 해. (유일 X), 제약 조건 하에서

목적함수 f를 최소화 하는 minimize X는 존재X

↓

feasible set 어떻게 설정하느냐에 따라

해 존재 / 여러 해 존재 / 해 X

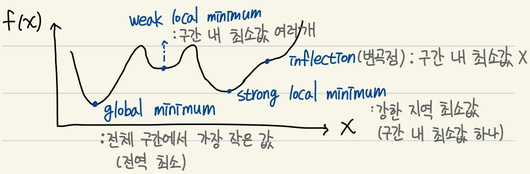
Unconstraints Optimization Problem 문제만 공부!

(제약 조건 X)

임계점

- critical points (임계점)

: 도함수 (derivative) 가 '0' 인 점.



→ 최적화 과정 : design point 찾기.

최적화 문제 : 목적 함수의 임계점 찾기

cf. gradient (기울기, 경사도)

: 변수 여러개 일 때의 도함수

⇒ 최적화

: 주어진 feasible set 에서 strong local min 찾는것.

- 전역 최소 (Global minimum)

: 목적 함수 최소

- 국소 최소 (Local minimum)

: 특정 구간, 공간, 집합으로 제한

$\exists\, \delta > 0 \;:\; f(x^*) \le f(x) \quad \forall\, |x - x^*| < \delta$

dim : n

$\exists\, \delta > 0 \;:\; f(\mathbf{x}^*) \le f(\mathbf{x}) \quad \forall\, \|\mathbf{x} - \mathbf{x}^*\| < \delta$ ⟨다변량⟩

→ ∃ : there exists, : such that, ∀ : for all

- 순 국소 최소 (strict local minimum)

: 유일하게 하나 존재

$\exists\, \delta > 0 \;:\; f(x^*) < f(x)\ ,\ x^* \neq x$ and $|x - x^*| < \delta$

- 변곡점 (Inflection point)

: 도함수는 '0' 이지만 국소 최소가 아닌 점.

~~반드시 도함수가 '0' 이 아니어도 됨~~ why?

- 수치 최적화 (numerical optimization)

- 가정과 제한점

① 목적 함수는 미분 가능 함수

② 제약식이 없는 최적화 방법

국소 최소 조건 : 일변량

- 계획점이 순 국소 최소가 되기 위한 필요조건

① $f'(x^*) = 0$: first - order necessary condition

② $f''(x^*) > 0$: Second - order necessary condition

→ x^* : 해, minimizor

- FONC : first - order necessary condition.

→ 첫번째 필요조건 만족 증명

$$f(x^*+h) = \underbrace{f(x^*) + h \cdot f'(x^*)}_{x^* \text{에 대한 테일러 전개}} + \boxed{O(h^2)}$$

2, 3, 4, … 차 (빅오) h^2보다 더 빨라지거나 작아지지 않는다.

$\rightarrow f(x^*+h) \geq f(x^*) \quad \because h \cdot f'(x^*) \geq 0$

$f(x^*-h) = f(x^*) - h \cdot f'(x^*) + O(h^2)$

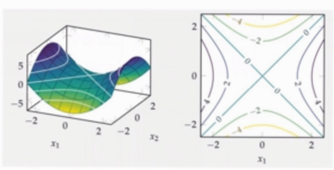
$\rightarrow f(x^*-h) \geq f(x^*) \quad \because h \cdot f'(x^*) \leq 0 , \ f'(x^*) = 0$

- SONC : Second - order necessary condition

$$f(x^*+h) = \underbrace{f(x^*) + hf'(x^*) + \frac{h^2}{2} f''(x^*)}_{\text{2차항까지 테일러 전개}} + O(h^2)$$

$\rightarrow f(x^*+h) \geq f(x^*) \quad \because \frac{h^2}{2} f''(x^*) \geq 0$

\+ $f(x_1, x_2) = x_1^2 - x_2^2$ **등고선**



→ Saddle , local minimum 존재 X

등고선 : 같은 f값 갖는 선들

국소 최소 조건 : 다변량

$\nabla f(x)$: 기울기 또는 경사도 (gradient)

$\nabla^2 f(x)$: 헷세 행렬 (Hessian) , 편미분 값 나타냄

- 계획점이 순 국소최소가 되기 위한 조건

① $\nabla f(x) = 0$: first - order necessary condition

② $\nabla^2 f(x)$는 양의 준정 부호 (positive semidefinite)

: Second - order necessary condition

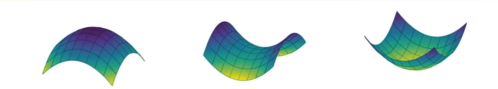
- FONC : first - order necessary condition

$f(x^*) \leq f(x+hy) \Leftrightarrow f(x+hy) - f(x^*) \geq 0$

- SONC : Second-Order necessary condition

$f(x^*+hy) = f(x^*) + h \nabla f(x^*)^T y + \frac{1}{2} h^2 \cdot y^T \nabla^2 f(x^*) \cdot y + O(h^3)$

$\frac{1}{2} h^2 y^T \nabla^2 f(x^*) \cdot y = f(x+hy) - f(x^*) \geq 0$

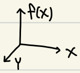


A *local maximum*. The gradient at the center is zero, but the Hessian is negative definite.

A *saddle*. The gradient at the center is zero, but it is not a local minimum.

A *bowl*. The gradient at the center is zero and the Hessian is positive definite. It is a local minimum.

→ 변수가 2개일 때, 3차원 공간에 표현.



: gradient = 0 이 되는 값이 minimum 아닐 수도 있음

변곡점과 비슷한 개념 → saddle

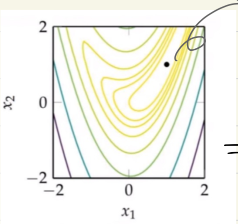
예제

Rosenbrock banana function.

: $f(x) = (1-x_1)^2 + 5(x_2 - x_1^2)^2$

→ minimizer : (1,1)

local minimum.



⇒ 2차원 contour plot

다변량 이니까 기울기 (gradient) 이용해 (1, 1) 찾자!

- Strong local minimum 확인

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2(10x_1^3 - 10x_1x_2 + x_1 - 1) \\ 10(x_2 - x_1^2) \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 x_1} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2 x_2} \end{bmatrix}$$

$$= \begin{bmatrix} -20(x_2 - x_1^2) + 40x_1^2 + 2 & -20x_1 \\ -20x_1 & 10 \end{bmatrix}$$