Optimization

Books name:

- Methods of Optimization:
- Theory of optimization:
- Optimization theory of methods
- An introduction to optimization

Optimization:

Optimization is a subject that is widely and increasingly used in scien
engineering, industry, management and others areas.

It deals with ~~possible~~ ~~st~~ selecting the best of many possible decisio
n real-life environment, constructing computational methods to find
ptimal solution exploring the theoritical properties and studing th
omputational performance of numerical algorithms implimented
ased on compatational methods.

ome applications of optimization:

Optimization in networks:

This topic covers recent advances in linear and non-linear
programing problems and mixed integer programing tecniques
and their novel applications in solving problems from water/
electricity/~~priet~~ pipeline networks.

Optimization in information sciences:

his topics covers various recent ~~optizati~~ optimization tecniq

CamScanner

and their applications in information sciences (including machine learning, speech /image detection processing, wireless communications information theory, estimation and detection theory and so on.

3. Optimization in Big data:

This topic addresses new challenges and opportunities in analyzing large scale data. where sparse optimization, global optimization and stochastic optimization plays a vital rule.

(4) The general form of optimization problem can be written as

$$\min_{x \in \Omega} f(x) \longrightarrow \text{minimize } f(x)$$

Subject to the constants $x \in \Omega$, where $x \in \Omega$ is called decision variable, $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is an objective function which needs to be optimized, Ω is called the feasible region.

$x \in \mathbb{R}^n$ $\quad x = (x_1, x_2 \cdots x_n)^T$

$x \in \mathbb{R}^2$ $\quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$

(4) Unconstants optimization:

$$\min f(x)$$

s.t. $c_i(x) = 0, \quad i \in E \quad i = 1, 2 \cdots m$

$c_i(x) \geq 0, \quad i \in I \quad i = m+1, m+2 \cdots n$

E and I are index set

$i \in E \cup I$

$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

$c_i : \mathbb{R}^n \rightarrow \mathbb{R}^m$

CS CamScanner

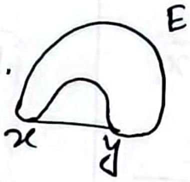
$x \in \mathbb{R}^5$

$f(x) = x^3 + 5x + 3$

$f(x) = 0$

$\boxed{\text{convex set}}$

$t \in (0,1)$



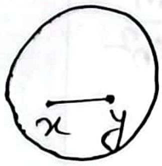
non convex

$tx + (1-t)y \in E$ $\quad [x, y \in E]$

$f : \mathbb{R}^n \to \mathbb{R}^m$

$\Omega \subseteq \mathbb{R}^n$

$\hookrightarrow$ convex



convex

$$f(tx + (1-t)y) \le t f(x) + (1-t) f(y)$$

$f : \mathbb{R}^n \to \mathbb{R}^m$

$f = (f_1, f_2 \cdots f_m)^T$

$f(x) \in \mathbb{R}^m$

$\dfrac{\partial f(x_0)}{\partial x_j} \qquad j = 1, 2 \cdots\cdots n$

$$= \left[\frac{\partial f_1(x_0)}{\partial x_j}, \frac{\partial f_2(x_0)}{\partial x_j}, \cdots\cdots, \frac{\partial f_m(x_0)}{\partial x_j} \right]$$

$$(x_0) = \begin{vmatrix} \frac{\partial f_1(x_0)}{\partial x_1} & \frac{\partial f_1(x_0)}{\partial x_2} & \cdots & \frac{\partial f_1(x_0)}{\partial x_n} \\ \frac{\partial f_2(x_0)}{\partial x_1} & \frac{\partial f_2(x_0)}{\partial x_2} & \cdots & \frac{\partial f_2(x_0)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m(x_0)}{\partial x_1} & \frac{\partial f_m(x_0)}{\partial x_2} & \cdots & \frac{\partial f_m(x_0)}{\partial x_n} \end{vmatrix}$$

This matrix is called Jacobian matrix or derivatives matrix.

CamScanner

Transpose of the Jacobian matrix is called the gradiant of f at x is denoted by $\nabla f(x) \to$

$$Df(x)^T = \nabla f(x)$$

$$D^2 f(x_0) = \begin{vmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f_1}{\partial x_n \partial x_1} \\ \frac{\partial^2 f_2}{\partial x_1 \partial x_2} & \frac{\partial^2 f_2}{\partial x_2^2} & \cdots & \frac{\partial f_2}{\partial x_n \partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f_m}{\partial x_1 \partial x_n} & \frac{\partial^2 f_m}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f_m}{\partial x_n^2} \end{vmatrix}$$

Hession matrix and it is denoted by $F(x)$.

big O
little o

Product rule for defferentiability

$h : \mathbb{R}^n \to \mathbb{R}$ defined by $h(x) = f(x)^T g(x)$

$$Dh(x) = f(x)^T Dg(x) + g(x)^T Df(x)$$

- $D(y^T Ax)$

$= y^T D(Ax) + (Ax)^T D(y)$

$= y^T (ADx + x^T DA) + x^T A^T . 0$

$= y^T A + 0$

$= y^T A$

- $D(x^T Ax)$

$= x^T D(Ax) + (Ax)^T Dx$

$= x^T A + x^T A^T$

$= x^T (A + A^T)$

$D(x^T x)$
$D(y^T x)$

Let,

(iv) $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$ be a given function

Find $\nabla f(x)$ and $F(x)$.

$$Df(x) = \left[\frac{\partial f(x)}{\partial x_1}, \quad \frac{\partial f(x)}{\partial x_2}\right]$$

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

$$\frac{\partial f(x)}{\partial x_1} = 5 + x_2 - 2x_1$$

$$\frac{\partial f(x)}{\partial x_2} = 8 + x_1 - 4x_2$$

$$\begin{bmatrix} 5 - 2x_1 + x_2 \\ 8 + x_1 - 4x_2 \end{bmatrix} = \nabla f(x)$$

$$\begin{bmatrix} 5 - 2x_1 + x_2 & 8 + x_1 - 4x_2 \end{bmatrix} = Df(x)$$

$$F(x) = D^2 f(x)$$

$$F(x) = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}$$

norm $= \|x\|$

$\exists^s$

$f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

$g : \mathbb{R}^n \longrightarrow \mathbb{R}$

CamScanner

! $\boxed{\text{norm}}$

$$\frac{\|f(x)\|}{|g(x)|} \le K;\ K>0 \qquad g(x)\neq 0$$

$f(x) = O(g(x))$ [big oh) $(g(x))$

$f(x) = o(g(x))$ (little oh) $(g(x))$

↳ $$\lim_{x\to 0} \frac{\|f(x)\|}{|g(x)|} = 0$$

$x^2 = o(x)$

$$\frac{\|x\|}{|x|} = 1$$

$$\underset{x\to 0}{Lt} \frac{\|x^2\|}{|x|} = \underset{x\to 0}{Lt}\, x = 0$$

$$\|x\| = \sqrt{x_1^2 + x_2^2}$$

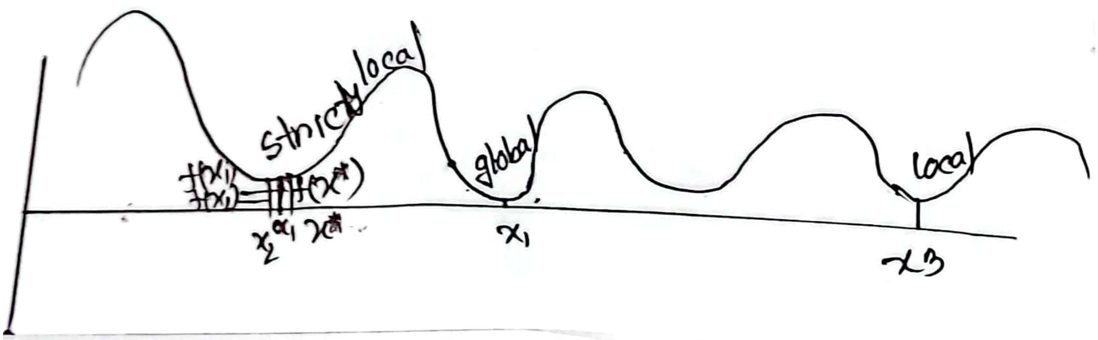
Ⓔ $$\begin{bmatrix} x^3 \\ 2x^2+3x^4 \end{bmatrix} = O(x^2)$$

$$= \frac{x^6 + 4x^4 + 12x^6 + 9x^8}{x^2}$$

$\Rightarrow x = O(1)$

Ⓔ. Local minimizer:

Global minimizer:



$f: \mathbb{R}^n \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$

$x^* \in \Omega \longrightarrow$ local minimizer if $\exists^s\ \varepsilon > 0$, s.t. $f(x) \geq f(x^*), \forall x \in \Omega$ and $|x - x^*| < \varepsilon$

☐ arg min (argument of minimizer)

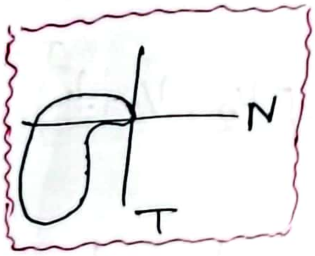
arg min $f(x)$; $x \geq 0$

$f(x) = (x+1)^2 + 3$

$x = 0$; $f(0) = 4$

$x = -1$; $f(-1) = 3$

$x = 1$, $f(1) = 7$



☐ Feasiable direction:

A vector $d \in \mathbb{R}^n$, $d \neq 0$ is said to be a feasiable direction of a vector $x \in \Omega$ if $\exists^s$, $\alpha_0 > 0$ s.t.

$x + \alpha d \in \Omega$ for all $\alpha \in (0, \alpha_0)$

$f: \mathbb{R}^n \to \mathbb{R}$

$$\frac{\partial f(x)}{\partial d} = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} \quad \hookrightarrow \boxed{\text{directional derivative}}$$

$f: \mathbb{R}^n \to \mathbb{R}$ directional derivative in the feasiable direction

$d \in \mathbb{R}^n$

$$\text{if } \frac{\partial f}{\partial d}(x) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} \quad \alpha > 0 ;$$

$$\frac{\partial f(x)}{\partial d} = \frac{d f}{d \alpha}(x + \alpha d) \Big|_{\alpha = 0}$$

$$\frac{\partial f(x+\alpha d)}{\partial \alpha} = f'(x+\alpha d)\cdot d \Big|_{\alpha=0}$$

$$= f'(x)\cdot d$$

$$= \nabla f(x)^T d$$

$$= \langle \nabla f(x), d \rangle$$

$$\frac{\partial f(x)}{\partial d} = \frac{df(x+\alpha d)}{d\alpha}\Big|_{\alpha \to 0}$$

This is called the rate of increase of f at x in the direction d if $\|d\| = 1$

$x \in \mathbb{R}^3$

- $f: \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x) = x_1 x_2 x_3$ and let $d = [1/2, 1/2, 1/\sqrt{2}]^T$

Find the directional derivative of f in the direction d, Also, show that the directional derivative is the rate of increase of f at x in the direction d.

Directional derivatives

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

$$= [x_2 x_3, x_1 x_3, x_1 x_2] \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \frac{x_2 x_3 + x_1 x_3 + \sqrt{2}\, x_1 x_2}{2}$$

$$\|d\| = \sqrt{\frac{1}{4}+\frac{1}{4}+\frac{1}{2}} = \sqrt{\frac{1+1+2}{4}} = \sqrt{\frac{4}{4}} = \sqrt{1} = 1$$

$x \in \mathbb{R}^n$

$d \neq 0$

Theorem 6.1

First order necessary condition (FONC):

Let, $f: \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be a differentiable real valued function on Ω. If x^* is a local minimizer of f over Ω, then for any feasible direction d at x^*,

we have $d^T \nabla f(x) \geq 0$

$f \in C' = \nabla f(x)^T d \geq 0$

$f(x^* + \alpha d) - f(x^*)$

$= f(x^*) + \alpha \nabla f(x^*)^T d + \frac{\alpha^2}{2!} d^T f(x) d + O(\alpha)$

$= \alpha \nabla f(x^*)^T d + O(\alpha)$

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \cdots$$

Since x^* is a local minimizer then,

$f(x^* + \alpha d) \geq f(x^*)$

$\Rightarrow \nabla f(x^*)^T d + \lim_{d \to 0} \frac{O(\alpha)}{\alpha} \geq 0$

$\Rightarrow \nabla f(x^*)^T d \geq 0$

$\Rightarrow d^T \nabla f(x^*) \geq 0$

Example:

minimize $x_1^2 + \frac{1}{2} x_2^2 + 3x_2 + \frac{9}{2}$

$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$

subject to $x_1, x_2 \geq 0$

CamScanner

Test whether the FONC is satisfied or not for a local minimizer at $x = [0, 3]^T$

$$d = (d_1, d_2)^T \quad d \neq 0$$

$$\nabla f(x) = [2x_1, x_2 + 3]^T$$

$$= [0, 6]^T \quad \text{at } x = [0, 3]^T$$

But $x + \alpha d \in \mathbb{R}^2$

$$\Omega = \{x : x_1 \geq 0; \ x_2 \geq 0\}$$

$$[0, 3]^T + \alpha [d_1, d_2]^T \in \mathbb{R}^2$$

$$\Rightarrow (0 + \alpha d_1, \ 3 + \alpha d_2)^T \in \Omega$$

$$\Rightarrow \alpha d_1 \geq 0; \quad 3 + \alpha d_2 \geq 0$$

$$d_1 \geq 0, \quad 3 + \alpha d_2 \geq 0$$

$$d^T \nabla f(x) = [d_1 \ d_2] [0 \quad 6]^T$$

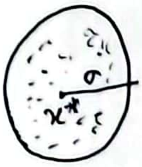
$$= 6 d_2$$

$\boxed{\text{Ex :}}$ $\quad d = [1 \quad -1]$

$$d^T \nabla f(x) = [1 \quad -1][0 \quad 6]$$

$$= -6 \leq 0$$

This is not satisfied FONC condition.



$N_\delta(x^*) = \{ y : \|y - x^*\| \leq \delta \} \ \forall y$

$\bar{N}_\delta(x^*) = \{ y : N_\delta(y) \subseteq N_\delta(x^*) \}$

□ Neibougr hood interiorpoint.

SONC : Second order necessary condition:

1. $f \in C^2$

$F(x)$ – Hession matrix and $d^T \nabla f(x^*) = 0$, then $d^T F(x^*) d \geq 0$

$x^* \to$ local minimizer

and $d \to$ fessible direction vector

$$\alpha \nabla f(x)^T d + \frac{\alpha^2}{2!} d^T F(x) d + \cdots\cdots$$

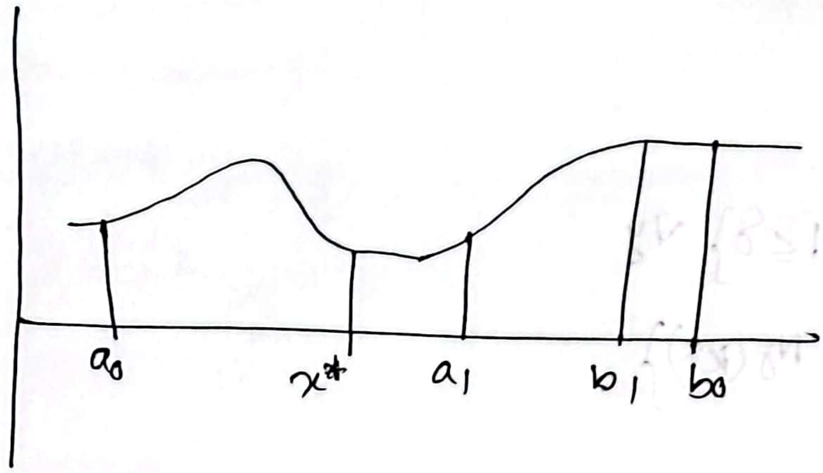
$$f(x^* + \alpha d) - f(x^*)$$

$$f(x^* + \alpha d) - f(x^*) = f(x^*) + \alpha \nabla f(x^*)^T d + \frac{\alpha^2}{2!} d^T F(x^*) d + O(\alpha) - f(x^*)$$

$$= \alpha \nabla f(x^*)^T d + \frac{\alpha^2}{2!} d^T F(x^*) d$$

□ Uni model function (one minimizer)

CS CamScanner



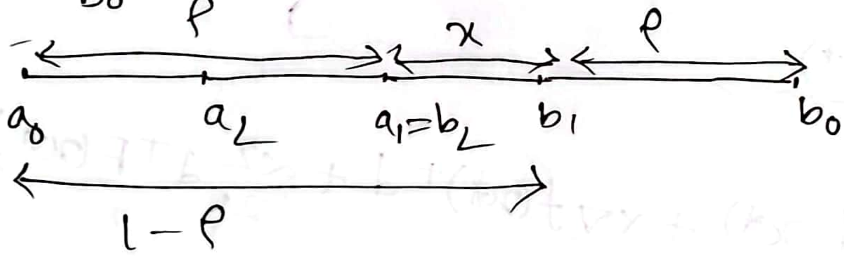
$$a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0), \quad \rho < \tfrac{1}{2}$$

if $f(a_1) < f(b_1)$, the minimizer lies in the interval $[a_0, b_1]$

On the other hand if $f(a_1) > f(b_1)$ then the minimizer lies interval $[a_1, b_0]$

$b_0 - a_0 = 1$

$b_0 - a_0 = 1$



$$a_2 - a_0 = b_1 - b_2 = \rho(b_1 - a_0)$$

$\Rightarrow$ But $b_1 - a_0 = 1 - \rho$

$\Rightarrow b_1 - b_1 = 1 - 2\rho$

$\Rightarrow 1 - 2\rho = \rho(1 - \rho)$

$\Rightarrow \rho - \rho^2 = 1 - 2\rho$

$\rho^2 - 3\rho + 1 = 0$

$\rho = \frac{3-\sqrt{5}}{2}, \frac{3 \pm \sqrt{5}}{2} \sim 0.382$

$1-\rho = \frac{\sqrt{5}-1}{2}$

$\therefore \frac{\rho}{1-\rho} = \frac{1-\rho}{1}$

$b_1 = b_0 + 1 - 2\rho$

$= b_0 + b_0 - a_0 - 2a_1$

$= 2b_0 - 2a_1$

$= 2(b_0 - a_1)$

$= 2(a_1 - b_0)$

$(1-\rho)^2 \simeq (0.6183)^2$

$\simeq (0.6183)^2 \leq$ Least interval $(b_0 - a_0)$

$a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$

$a_1 = a_0 + \rho(b_0 - a_0)$

Use Golden section search to find the minimise

$f(x) = x^4 - 14x^3 + 60x^2 - 70x$

in the range $[0,2]$ Locate this value for x to within a range 0.3. (H.W. Algorithm)

$(1-\rho)^N \leq \frac{0.3}{2-0}$

$\Rightarrow N = 4$

$a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$

$a_1 = a_0 + \rho(b_0 - a_0) = 0 + 0.382 \times 2 = 0 + 0.61$

$= 1.236$

$$f(a_1) = f(0.7639)$$

$$= (0.7639)^4 - 14(0.7639)^3 + 60 \times (0.7639)^2 - 70(0.7639)$$

$$= -24.36$$

$$f(b_1) = -18.96$$

$$f(a_1) < f(b_1)$$

$$[0, b_1] = [0, 1.236]$$

$$b_2 = 0.7639$$

$$a_2 = a_0 + \rho(b_1 - a_0)$$

$$= 0.4721$$

$$f(a_2) = (0.4721)^4 - 14(0.4721)^3 + 60(0.4721)^2 - 70(0.4721)^3$$

$$= -21.10$$

$$f(b_2) = f(a_1) = -24.36$$

$$f(a_2) > f(b_2)$$

$$[a_2, b_1] = [0.4721, 1.236]$$

$$b_3 = a_2 + \rho(b_1 - a_2)$$

$$= 0.9443$$

~~Net~~ Newton method

1) $f(x)=0$

(ii) differentiable হতে হবে

(iii) $f'(x) \neq 0$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$f(x) = x^6 - 9x^5 - 3$

$f(x) = \frac{1}{2}x^2 - \sin x = 0$

$f'(x) = x - \cos x$

K=0;

$$x_1 = x_0 - \frac{f(x_1)}{f(x_0)} = \frac{1}{2} - \frac{0.116}{-0.3775}$$

initial guess $x_0 = \frac{1}{2}$

$f(\frac{1}{2}) = \frac{1}{2}\cdot\frac{1}{4} = \sin\frac{1}{2} = 0 \quad = 0.116$

$f'(\frac{1}{2}) = \frac{1}{2} - \cos\frac{1}{2}$

$f'(\frac{1}{2}) = -0.3775$

$$= \frac{1}{2} + \frac{0.116}{0.3775}$$

$|f'(x_k)| < 0.001$

1. $x_0 \in \mathbb{R}$, $k=0$ $\quad \varepsilon = 0.001$

2. If $\|f'(x'_k)\| \le \varepsilon$, then stop;
   otherwise go to steps.

3. $$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

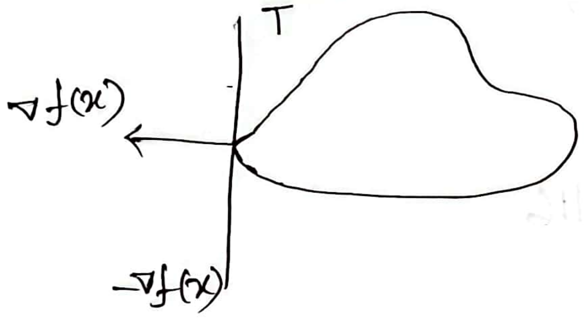
4. Set $k = k+1$

5. go to step 2

Apply Newton method, find the minimizer of

$f(x) = x^3 - 12x^2 + 7x + 42 = 0$

Gradiant method:



x_0

$f(x_{k+1}) < f(x_k)$

x_0

$\downarrow$ improvement

$x_0 - \alpha \nabla f(x_0)$

x_0
$\downarrow$
x_1
$\downarrow$
$x_2 = x_k = \dfrac{f(x_k)}{f(x_k+1)}$
$\downarrow$
x_{k+1}

$f(x_0 - \alpha \nabla f(x_0))$

$= f(x_0) - \alpha \|\nabla f(x_0)\|^2 + o(\alpha)$

$f(x+h) = f(x) + h f'(x) + \dfrac{h^2}{2!} f''(x)$ ~~$-\alpha \nabla f(x_0) \cdot f'(x) \nabla f(x)$~~

$+ \cdots$

$-\alpha \nabla f(x_0) \cdot f'(x) \nabla f(x) \qquad \langle u, v \rangle$

CamScanner

If $\nabla f(x_0) \neq 0$, then

$\underbrace{f(x_0 - \alpha \nabla f(x_0))}_{x_1} < f(x_0)$

$f(x_1) < f(x_0)$

step 1: Given initial point $x_0 \in \mathbb{R}^n$; $k := 0$, $\varepsilon = 0.0001$

step 2: If $\nabla f(x_k) = 0$, then stop, otherwise go to. If

$\|\nabla f(x_k)\| < \varepsilon$ go to step 3

step 3: $x_{k+1} = x_k - \alpha k \nabla f(x_k)$

step 4: Ste $k := k+1$ and go to step 2

$$\boxed{\begin{aligned} &\phi_k(\alpha) := f(x_k - \alpha \nabla f(x_k)) \\ &\alpha_k = \arg\min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k)) \end{aligned}}$$

Stepest descent algorithm

- It is one kind of gradiant method

Use stepest descent method to minimize

$f(x) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$ with initial point

$x_0 = [4, 2, -1]^T$

Sol^n next page

Sol^n:

$$\nabla f(x) = \left[4(x_1-4)^3,\ 2(x_2-3),\ 16(x_3+5)^3\right]^T$$

$$\nabla f(x_0) = [0, -2, 1024]^T$$

$$x_0 - \alpha \nabla f(x_0)$$

$$= [4, 2, -1]^T - [0, -2\alpha, 1024\alpha]^T$$

$$= [4,\ 2+2\alpha,\ -1-1024\alpha]^T$$

$$f(x_0 - \alpha \nabla f(x_0))$$

$$= 0 + (2+2\alpha-3)^2 + 4(4-1024\alpha)^4$$

$$\frac{df}{d\alpha} = 0$$

$$\Rightarrow 2(2+2\alpha-3)\cdot 2 + 16(4-1024\alpha)^3(-1024) = 0$$

$$\Rightarrow 2+2\alpha-3 - 4096(4-1024)^3 = 0$$

→ solve this equation and assume that

$$\alpha_0 = 3.967\times 10^{-3}$$

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$\therefore x_1 = x_0 - \alpha_0 \nabla f(x_0)$$

$$= [4, 2, -1]^T - 3.967\times 10^{-3}[0, -2, 1024]^T$$

$$= [4, 2.008, -5.062]^T$$

$x_{k+1} - x_k \xrightarrow{\text{tends to}} 0$ $\quad \|\nabla f(x_k)\| \le \epsilon$

জিরোর কাছাকাছি value পেলে সেটাই answer

Chapter 19

□ Consider the optimization problem minimize $f(x)$

subject to $h(x) = 0$

$g(x) \leq 0$

where $f: \mathbb{R}^n \to \mathbb{R}$, $h: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^n \to \mathbb{R}^p$

$\Omega = \mathbb{R}^n$

minimize $f(x)$

st: $x \in \Omega$

chapter 19

chapter 18

$f, h \in C'$

$h = [h_1, h_2 \ldots\ldots h_m]$

$\nabla h = \nabla [h_1, h_2 \ldots\ldots h_m]$

$= \nabla h_1, \nabla h_2 \ldots\ldots \nabla h_m$

$h(x) = 0$

A vector x^* is called a regular point of the given optimization problem if $h(x^*) = 0$ and $\nabla h_1(x^*) . \nabla h_2(x^*) \ldots \nabla h_m(x^*)$ are linearly independent.

$v_1, v_2 \ldots\ldots v_n \in V$

$a_1 v_1 + a_2 v_2 + \ldots\ldots + a_n v_n = 0$

if $\exists^s$ scalars $a_1 = a_2 = \ldots\ldots = a_n = 0$

□ Equility Constant

$f(x)$ - objective function

$h(x)$ - constant function

CamScanner

Lagrange theorem: (Chapter 19, theorem 19.3)

Let $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function with equality constants $h(x)=0$. Let x^* be a regular point for which x^* is a local minimizer. Then for every $\lambda^* \in \mathbb{R}^n$, we have

$$Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T.$$

Lagrange function:

A function $l : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ be defined by ~~$l(x,\lambda)$~~

$l(x,\lambda) := f(x) + \lambda^T h(x)$ is called a language function.

$D_x l(x,\lambda) = Df(x) + \lambda^T Dh(x)$

$D_\lambda l(x,\lambda) = 0 + h^T(x)$

$D_\lambda l(x,\lambda) = h^T(x)$

if x^* is a regular point and x^* is a local minimizer, then

$D_x L(x^*, \lambda^*) = Df(x^*) + \lambda^{*T} Dh(x^*) = 0$

and $D_\lambda l(x^*, \lambda^*) = h^T(x^*) = 0^T$

$$\begin{cases} D_x l(x^*, \lambda^*) = 0^T \\ D_\lambda l(x^*, \lambda^*) = 0^T \end{cases}$$

are called Lagrange equations or Lagrange conditions.

Find the extremisers of the objective function ~~$f(x) = x_1^2$~~

$f(x) = x_1^2 + x_2^2$ .... $x = (x_1, x_2)^T$

subject to the constraints on Ω = ~~{~~

$$\Omega = \{x : h(x) = x_1^2 + 2x_2^2 - 1 = 0\}$$

$\therefore \nabla f(x) = [2x_1, 2x_2]^T$

$\because \nabla h(x) = [2x_1, 4x_2]^T$

$$D_x l(x, \lambda) = Df(x) + \lambda^T Dh(x)$$
$$= (2x_1 + 2\lambda x_1, 2x_2 + 4\lambda x_2) = 0$$

$f : \mathbb{R}^n \to \mathbb{R}$
$h : \mathbb{R}^n \to \mathbb{R}^m$
$\lambda \in \mathbb{R}^m$
$f : \mathbb{R}^2 \to \mathbb{R}$
$\lambda : \mathbb{R}^2 \to \mathbb{R}$

$D_\lambda l(x, \lambda) = 0$

$h(x) = 0$

$\Rightarrow x_1^2 + 2x_2^2 = 1$

$2x_1 + 2\lambda x_1 = 0$
$2x_2 + 4\lambda x_2 = 0$
$x_1^2 + 2x_2^2 = 1$

$x_1 = 0$; $\lambda = -1$

then $x_2 = \pm\sqrt{\frac{1}{2}}$

Then 2nd equation given $\lambda = -\frac{1}{2}$

Now, put $\lambda = -1$, then

$$x_2 = 0$$
$$x_1 = \pm 1$$

$$x_1^* = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$x_1^{**} = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$x_2^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$x_2^{**} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

minimize $f(x)$

s.t. $h(x) = 0$
$g(x) \leq 0$

$f : \mathbb{R}^n \to \mathbb{R}$
$h : \mathbb{R}^n \to \mathbb{R}^m, \forall \lambda \in \mathbb{R}^m$
$g : \mathbb{R}^n \to \mathbb{R}^p, \forall, \mu \in \mathbb{R}^p$
$f, h, g \in C'$
$\nabla h, \nabla g$

1. $\mu^*(x^*) \geq 0$
2. $Df(x^*) + \lambda^{*T} \nabla h(x^*) + \mu^{*T} \nabla g(x^*) = 0^T$
3. $\mu^{*T} g(x^*) = 0$
4. $h(x^*) = 0$
5. $g(x^*) \leq 0$

KKT $\to$ condition

$$\bigcup_{\Omega}$$

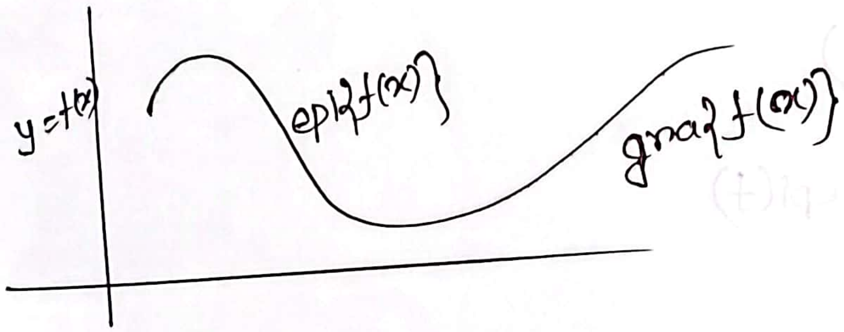
$$\bigcap$$

$\forall x, y \in \Omega$

$t \in (0,1)$

$tx + (1-t)y \in \Omega$

- If every point of a set element A is include in Ω set it's called convex set.
- If is not it's concave set.



$f : \mathbb{R}^n \to \mathbb{R}$

$gph(f) = \{(x, y)^T : y = f(x)\}$

graph of f

epigraph of f

$epi(f) = \{(x, \beta)^T : f(x) \leq \beta\}$

Convexity

Def^n: A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex on $\Omega \subseteq \mathbb{R}^n$ if its epigraph is convex.

Theorem: Prove that a function $f:\mathbb{R}^n \to \mathbb{R}$ is a convex function on $\Omega \subseteq \mathbb{R}$ if Ω is convex.

The proof will be completed by contraposition.

Suppose that Ω is not a convex set. Then $\forall x, y \in \Omega$ $t \in (0,1)$

$$z = tx + (1-t)y \notin \Omega$$

Let, $f(x) = a$, $f(y) = b$

then

$(x,a)^T, (y,b)^T \in gph(f)$

Also, $(x,a)^T, (y,b)^T \in epi(f)$

$$\boxed{f(x) \leq \beta}$$

$$w = t\begin{pmatrix} x \\ a \end{pmatrix} + (1-t)\begin{pmatrix} y \\ b \end{pmatrix}$$

$$= \begin{pmatrix} tx \\ ta \end{pmatrix} + \begin{pmatrix} (1-t)y \\ (1-t)b \end{pmatrix}$$

$$= \begin{pmatrix} tx + (1-t)y \\ ta + (1-t)b \end{pmatrix}$$

~~$\notin$ epi(f), since~~

$\notin epi(f)$, since $tx + (1-t)y \notin \Omega$

CS CamScanner

$\{(x,a)^T : f(x) \leq a, x \in \Omega\}$

· This shows that epi (f) is not a convex set hence f is not a convex function.

$a = b$
$a^2 = ab$
$a^2 - b^2 = ab - b^2$
$(a+b)(a-b) = b(a-b)$

testing
true
false

$\boxed{\text{}}$ A function $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex on a convex set Ω

iff $\forall x, y \in \Omega, \alpha \in (0,1)$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Soln.

First suppose that if f is a convex function on a convex set Ω. We need to show that,

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y), \quad \forall x, y \in \Omega, \alpha \in (0,1)$$

Let,

$$\begin{pmatrix} x \\ a \end{pmatrix}, \begin{pmatrix} y \\ b \end{pmatrix} \in \operatorname{epi}(f)$$

Then $f(x) \leq a, \quad f(y) \leq b$

$$\therefore \alpha \begin{pmatrix} x \\ a \end{pmatrix} + (1-\alpha) \begin{pmatrix} y \\ b \end{pmatrix} = \begin{pmatrix} \alpha x + (1-\alpha) y \\ \alpha a + (1-\alpha) b \end{pmatrix}$$

CamScanner

This simplies that

$$f(\alpha x+(1-\alpha)y) \le a\alpha+(1-\alpha)b \le \alpha f(x)+(1-\alpha)f(y)$$

$$\Downarrow$$

i.e. $f(\alpha x+(1-\alpha)y) \le \alpha f(x)+(1-\alpha)f(y)$.

II Conversely, suppose that

$f(\alpha x+(1-\alpha)y) \le \alpha f(x)+(1-\alpha)f(y) \quad \forall x,y \in \Omega,\ \alpha\in(0,1)$

we have to show to f is convex on a convex set Ω.

Ex:

$f(x) = x_1 x_2 \qquad f: \mathbb{R}^2 \to \mathbb{R} \qquad x=(x_1,x_2)$

Is f convex on $\Omega = \{x : x_1 \ge 0,\ x_2 \ge 0\}$?

$x = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$$\alpha x+(1-\alpha)y = \begin{pmatrix} 2\alpha \\ \alpha \end{pmatrix} + \begin{pmatrix} (1-\alpha) \\ 2-2\alpha \end{pmatrix}$$

$$= \begin{pmatrix} 2\alpha+1-\alpha \\ \alpha+2-2\alpha \end{pmatrix}$$

$$= \begin{pmatrix} \alpha+1 \\ 2-\alpha \end{pmatrix}$$

$$f(\alpha x+(1-\alpha)y) = (\alpha+1)\cdot(2-\alpha)$$

$$= 2\alpha-\alpha+2-\alpha^2$$

$$= 2+\alpha-\alpha^2$$

CS CamScanner

$\alpha f(x) = x \cdot 2 \cdot 1 = 2\alpha$

$(1-\alpha)f(y) = (1-\alpha) \cdot 1 \cdot 2 = 2(1-\alpha)$

$\alpha f(x) + (1-\alpha) f(y) = 2$

$\therefore f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha) f(y)$

This function f is not convex function

H.W

$f : R^2 \longrightarrow R$ defined by $f(x) = 2x_1x_2 - x_1^2 - x_2^2$ on

$\Omega = \{x ; x_1 \geq 0 ; x_2 > 1\}$ Is f convex on Ω?

Exam

$\boxed{\text{m}}$ Let $f : \Omega \subseteq R^n \rightarrow R$ be convex function defined by $f(x = x^T Q x)$ on a convex set Ω, where ~~convex~~. $Q = Q^T$, $Q_{n\times n}$ matrix. Then f is convex on Ω if and only if $(x-y)^T Q (x-y) \geq 0$, $\forall x, y \in \Omega$

quadratic form

$f(x) = \frac{1}{2} x^T Q x - b^T x$ $\{x \in R^n$, $Q \rightarrow n \times n$, $K \rightarrow n\times 1\}$ $\hookrightarrow R$

proof : First suppose that f is a convex function on a convex set Ω we need to prove that

$(x-y)^T Q (x-y) \geq 0$, $\forall x, y \in \Omega$

CamScanner

Since f is a convex function then $\forall x, y \in \Omega, \alpha \in (0,1)$ we have

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

$$\Rightarrow \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha)y) \geq 0$$

$$\Rightarrow \cancel{\alpha x^T Q x + (1-\alpha) y^T Q y - (\alpha x + (1-\alpha)y)^T Q \alpha x}$$

$$\Rightarrow \alpha x^T Q x + (1-\alpha) y^T Q y - (\alpha x + (1-\alpha)y)^T Q (\alpha x + (1-\alpha)y) \geq 0$$

$$\Rightarrow \alpha x^T Q x + (1-\alpha) y^T Q y - \alpha^2 x^T Q x - \alpha(1-\alpha) y^T Q x - \alpha(1-\alpha) x^T Q y - (1-\alpha)^2 y^T Q y \geq 0$$

$$\Rightarrow \alpha(1-\alpha) x^T Q x + \alpha(1-\alpha) y^T Q y - 2\alpha(1-\alpha) x^T Q y \geq 0$$

$$\boxed{(y^T Q x)^T = x^T Q y}$$

$$\Rightarrow \alpha(1-\alpha)(x-y)^T Q (x-y) \geq 0$$

$$\alpha(1-\alpha) x^T Q x - \alpha(1-\alpha) y^T Q x - \alpha(1-\alpha) x^T Q y + \alpha(1-\alpha) y^T Q$$

$$\Rightarrow (x-y)^T Q (x-y) \geq 0$$

(4) $f(x) = x_1 x_2$

$\Omega = \{x : x_1, x_2 \geq 0\}$ If f convex on Ω?

$$f(x) = x^T Q x = \frac{1}{2} x^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} x = \frac{1}{2} (x_1 \;\; x_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \frac{1}{2} (x_2 \;\; x_1) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 x_2$$

CamScanner

Solⁿ:

$f(x) = x_1 x_2 = x^T Q(x)$

$= (x_1, x_2) \begin{pmatrix} a_1 & a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$= \begin{pmatrix} a_1 x_1 + a_2 x_2 & a_2 x_1 + a_1 x_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$= a_1 x_1^2 + a_2 x_1 x_2 + a_2 x_1 x_2 + a_1 x_2^2$

$= a_1 x_1^2 + 2a_2 x_1 x_2 + a_1 x_2^2$

$a_1 = 0 \quad , a_1 = \frac{1}{2}$

$Q = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$(x-y)^T Q(-x-y) \geq 0$

$x = (2, 2)^T$

$y = (1, 3)^T$

$(x-y) = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$(x-y)^T Q(x-y) = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$= \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$= -\frac{1}{2} + \left(-\frac{1}{2}\right) = -1 < 0$

CamScanner

$f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function defined on a con set Ω and $f \in C$ then f is convex on Ω if and only if $f(y) \geq f(x) (-1) f(x) (y-x)$, $\forall x, y \in \Omega$

$$\boxed{\begin{aligned} x_{k+1} &= x_k - \frac{f(x_k)}{Df(x_k)} \\ \Rightarrow f(x_k) &+ Df(x_k)(x_{k+1} - x_k) = 0 \\ \Rightarrow f(x) &+ Df(y)(y-x) = 0 \end{aligned}}$$

Let $f : \Omega \subseteq \mathbb{R}^n$ be a convex function defined on a convex set Ω. Then a point is a global minimize of f over Ω iff (if and only if) it is a local minimizer.

Sol^n:

x^* is not a global minimizer.

$$f(y) < f(x^*)$$

$$\boxed{\begin{gathered} f(x^*) \leq f(x) \\ \forall x \in \Omega \qquad x \neq x^* \end{gathered}}$$

CamScanner

$$\Rightarrow f(\alpha x+(1-\alpha)y) \le \alpha f(x)+(1-\alpha)f(y) \quad ; x,y \in \Omega, y, x^* \in$$

$$\Rightarrow f(\alpha y+(1-\alpha)x^*) \le \alpha f(y)+(1-\alpha)f(x^*)$$

$$\Rightarrow f(\alpha y+(1-\alpha)x^*) \le \alpha(f(y)-f(x^*))+f(x^*)$$

$$\Rightarrow f(\alpha y+(1-\alpha)x^*) < f(x^*)$$

$$y_n = \frac{1}{n} y + \left(1-\frac{1}{n}\right)x^*$$

$$\lim_{n\to\infty} y_n = x^*$$

$$f(y_n) < f(x^*)$$

local minimizer is not a global minimizer.

$\boxed{\text{四}}$ Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function defined on a convex set Ω. Then for all $x \in \Omega$, $x \neq x^*$ s.t.

$$Df(x^*)(x-x^*) \ge 0$$

Then x^* is a global minimizer

$$f(y) > f(x^*) + Df(x^*)(y-x^*)$$

$$\Rightarrow f(y) > f(x^*).$$

CamScanner

if $Df(x^*)(y-x^*) \geq 0$

Stephen decent gardian method

SBG SGD nottorm Ex

Pauly function, newton method এর সাহায্যে প্রমাণ কর। $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

$$f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2, \quad x = (x_1, x_2)^T \; f: \mathbb{R}^n \rightarrow$$

Rosenbroks function (Banana function

Newton method solve

July 10, Assignment submit date.

2.4, 3.4, 4.3, $\boxed{4.4}$, 5.3, 5.5, 6.2, $\boxed{7.1}$, 7.3, 8.2

8.3, 9.2, 9.3, 9.4, 11.2, 11.3, 21.2, 21.3