optimization problem (programming).

$\min f(x) \Leftrightarrow \max -f(x)$

unconstrained optimization

$\min f(x)$

$x \in S$

S: feasible set.

constrained optimization

$\min f(x)$

$h_i(x)=0 \quad i=1,2,\cdots p$ (equality strains).

$g_i(x) \le 0 \quad i=1,2,\cdots m$ (inequality strains).

$x \in S$.

linear programming

$f(x)$, $h_i(x)$, $g_i(x)$

are linear functions

otherwise, nonlinear programming.

Ex: $\min \quad w_1 x_1 + w_2 x_2$

Subject to $\begin{cases} x_1 + x_2 = 1. \Rightarrow x_1 + x_2 - 1 = 0. \\ x_1 \ge 0, \quad x_2 \ge 0 \end{cases}$

$x_1 + x_2 - 1 = 0 \Rightarrow x_2 = 1 - x_1$

$\min \quad w_1 x_1 + w_2 (1 - x_1)$

s.t. $\quad x_1 \ge 0 \quad 1 - x_1 \ge 0$

Review

- Sets

$$\min_{x \in S} f(x) \iff \min\{f(x).\ x \in S\}.$$

$$\operatorname*{argmin}_{x \in S} f(x) \quad (\text{value of } x).$$

- matrices, vectors

$\vec{a} = [a_1, \cdots a_n]$: row vector

$\vec{a} = (a_1, \cdots a_n)$: column vector.

$$A = \begin{array}{c} \begin{array}{ccc} \vec{a_1} & \cdots & \vec{a_n} \\ \downarrow & & \end{array} \\ \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & & \vdots \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix} \begin{array}{l} \to \vec{a^1} \\ \\ \\ \to \vec{a^m} \end{array} \end{array}$$

$= [\vec{a_1}, \cdots \vec{a_n}]$

$= (\vec{a^1} \cdots \vec{a^m})$

- spaces

$\mathbb{R}^n$

E^n : Euclidean space of dim n.

(Cauchy-Schwarz Inequality : $|X^T y| \le |X||y|$).

$\{X_1, X_2 \cdots X_k\}$ are linearly dependent

if $\sum_{i=1}^{k} C_i X_i = 0$ for some i that $C_i \neq 0$

rank(A) = max number of linearly independent row/column vectors.

$A_{m \times n}$ is called full rank if rank(A) = $\min\{m, n\}$.

Eigenvalues λ of A:

$$Ax = \lambda x \qquad x \neq 0$$

x: eigenvector.

A is symmetric if $A^T = A$. for symmetric matrix

- eigenvalues are all real
- eigenvectors for distinct eigenvalues are orthogonal.
- There is an orthnomal basis for E^n. consisting of eigenvalues of A.

if the basis is is $\{u_1, u_2 \cdots u_n\}$. $\quad Au_i = \lambda u_i$

Let $Q = [u_1, \cdots u_n]$

Then $Q^T Q = I$. $Q^T = Q^{-1}$

$$Q^{-1}AQ = Q^TAQ = Q^T(Au_1 \cdots Au_n) = Q^T(\lambda_1 u_1 \cdots \lambda_n u_n) = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{bmatrix} = D$$

$$A = QDQ^T$$

A symmetric matrix A is called · positive definite if $X^TAx > 0$. $\forall x \neq 0$

· positive semidefinite if $X^TAX \geq 0$. $\forall X$

$$X^TAX = X^TQDQ^TX = (Q^Tx)^T D (Q^Tx)$$
$$= Y^T D Y. = \Sigma \lambda_i y_i^2$$

positive definite $\Leftrightarrow \lambda_i > 0 \quad \forall i$

positive semidefinite $\Leftrightarrow \lambda_i \geq 0 \quad \forall i$

$$A^{\frac{1}{2}} = Q\sqrt{D}\,Q^T$$

· Topology

$\{x_1, x_2, x_3 \cdots\cdots\} = \{x_k\}_{k=1}^{\infty}$

$\lim_{k\to\infty} x_k = x$: $\forall \varepsilon>0, \exists k \in \mathbb{N}$ $|x_k - x| < \varepsilon$ for $k.>k$

$x_k \to x.$

x^* : limit point of $\{x_k\}_{k=0}^{\infty}$

if x^* is the limit of some subsequence of $\{x_k\}_{k=0}^{\infty}$

open Ball $B_R : \{|x|<R\}$, $B_R(x_0) = \{|x-x_0|<R\}$.

Set S is compact if it's closed and bounded

Weirestrass Theorem

if f is continous on a compact set. then it achieves max and min.

if f is continous on S, write $f \in C(S)$

if $\frac{\partial f}{\partial x_i}$ $\forall x_i$ are continous in S, $f \in C^1(S)$.

if $\frac{\partial^\alpha f}{\partial x_i^\alpha}$ $|\alpha|=k$ are continous on S. $f \in C^k(S)$.

Gradient of f

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n}\right]$$

$$F = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

if $f \in C^2$, then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

$F^T = F$

Taylor's Theorem

- if $f \in C^1(S)$,

  $[x_1, x_2]$: line segment in S)

  then $\exists \theta \in [0,1]$ s.t.

  $$f(x_2) = f(x_1) + \nabla f(\underbrace{\theta x_1 + (1-\theta) x_2}_{x_0})(x_2 - x_1)$$

- if $f \in C^2(S)$

  $$f(x_2) = f(x_1) + \nabla f(x_1)(x_2 - x_1). + \frac{1}{2}(x_2 - x_1)^T F(\theta x_1 + (1-\theta) x_2)(x_2 - x_1).$$

other forms

$$f(x_2) = f(x_1) + \nabla f(x_1)(x_2 - x_1) + o(|x_2 - x_1|)$$

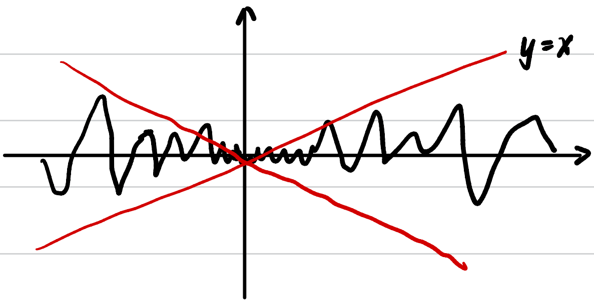
$$f(x_2) = f(x_1) + \nabla f(x_1)(x_2 - x_1) + \frac{1}{2}(x_2 - x_1)^T F(x_1)(x_2 - x_1) + o(|x_2 - x_1|^2).$$

Big-O, small-O notations.

$f(x) = O(x)$ as $x \to 0$ means $\left|\frac{f(x)}{x}\right| < k$ as $x \to 0$

$\exists k, \delta$ s.t. $\left|\frac{f(x)}{x}\right| < k$ if $|x| < \delta$.

e.g. $f(x) = x \sin x$



$\left|\frac{f(x)}{x}\right| = |\sin x| \le 1. \Rightarrow f(x) = O(x).$

$f(x) = o(x)$ if $\left|\frac{f(x)}{x}\right| \to 0$ as $x \to 0$.

Part I: unconstrained optimization

Chapter 1: Basic properties.

$$\begin{cases} \min f(x) \\ x \in \Omega. \end{cases}$$

Definition: $x^* \in \Omega$ is called a relative (local) minium point. of f over Ω if $\exists \varepsilon > 0$, s.t. $f(x^*) \leq f(x)$ $\forall x \in B_\varepsilon(x^*) \cap \Omega$.

Strict relative minium point if $\exists \varepsilon > 0$, s.t. $f(x^*) \leq f(x)$ $\forall x \in B_\varepsilon(x^*) \cap \Omega$.

global minium point if $f(x^*) \leq f(x)$, $\forall x \in \Omega$

Definition: Given $x \in \Omega \subset E^n$. $d \in E^n$ is called a feasible direction at x if $\exists \bar{\alpha} > 0$ s.t. $x + \alpha d \in \Omega$. $\forall 0 \leq \alpha \leq \bar{\alpha}$

proposition (First-order necessary optimality Conditions)

$f \in C'(s)$

if x^* is a relative minium point. then $\forall d$ feasible at x^* we have $\nabla f(x^*) d \geq 0$.

$f(x_1, x_2)$ partial derivatives

$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1+h, x_2) - f(x_1, x_2)}{h}$$

$$\frac{\partial f}{\partial x_2} = \lim_{h \to 0} \frac{f(x_1, x_2+h) - f(x_1, x_2)}{h}$$

$f(x_1, x_2, x_3 \cdots x_n)$

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n}\right]$$

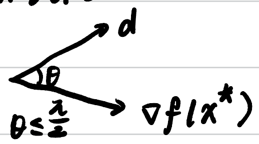
Def: $x \in \Omega$ is an interior point if $\exists \varepsilon > 0$ s.t $B_\varepsilon(x) \subset \Omega$

A set Ω is open if every $\Lambda \in \Omega$ is an interior point.

$f \in C^1(\Omega)$, $x^* \in \Omega$ a local minium of f

d: any feasible direction

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



proof: d is a feasible direction

$\Rightarrow \exists \bar{\alpha}$ s.t.

$$x^* + \alpha d \in \Omega, \ \forall \alpha \in [0, \bar{\alpha}]$$

let $g(\alpha) = f(x^* + \alpha d)$

Taylor's theorem

$$g(\alpha) = g(0) + g'(0)\alpha + O(\alpha)$$

$$= g(0) + [\nabla f(x^*) d]\alpha + O(\alpha)$$

$$g'(\alpha) = \nabla f(x^* + \alpha d) d$$

$$g'(\alpha) = \lim_{h \to 0} \frac{g(\alpha + h) - g(\alpha)}{h}$$

$$= \lim_{h \to 0} \frac{f(x^* + (\alpha + h) \cdot d) - f(x^*)}{h}$$

$$= \lim_{h \to 0} \frac{f(x_1^* + (\alpha + h) d_1, \ x_2^* + (\alpha + h) d_2, \ \cdots \ x_n^* + (\alpha + h) d_n) - f(x^*)}{h}$$

$$= \frac{\partial f}{\partial x_1}(x^*) d_1 + \frac{\partial f}{\partial x_2}(x^*) d_2 + \cdots \frac{\partial f}{\partial x_n}(x^*) d_n$$

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

Suppose $\nabla f(x^*) d < 0$, then $g(\alpha) < g(0)$ for sufficiently small α.

$$[\nabla f(x^*) d]\alpha + M(\alpha)\cdot\alpha$$
$$= \alpha[\nabla f(x^*) d + M(\alpha)] < 0 \quad \text{as } \alpha \to 0$$

Taylor's theorem.

Peano

if f is k-times differentiable at x_0

Then $f(x) = f(x_0) + f'(x_0)(x-x_0)^1 + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + r_k(x)$.

where $r_k(x) = o(|x-x_0|)^k$,

Ex: $f(x) = e^x \quad x_0 = 0$

$$e^x = 1 + x + \frac{1}{2}x^2 + \cdots \frac{1}{k!}x^k + o(x^k).$$

Lagrange form

If $f \in C^{k+1}(I)$, $x_0 \in I$, $x \in I$

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \cdots + \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + r_k(x).$$

$$r_k(x) = \frac{f^{(k+1)}(\bar{x})}{(k+1)!}(x-x_0)^{k+1}$$

$$\bar{x} \in (x_0, x).$$

$r_k(x) = o(|x-x_0|^k)$.

proof: $\lim_{x \to x_0} \frac{r_k(x)}{|x-x_0|^k} = \frac{f^{(k+1)}(\bar{x})}{(k+1)!}(x-x_0) = 0.$

$f: \mathbb{R}^n \to \mathbb{R}$, differentiable at x_0. (e.g. $f \in C^1$)

$$f(x) = f(x_0) + \nabla f(x_0)(x-x_0) + o(\|x-x_0\|).$$

if $f \in C^2$, then

$$f(x) = f(x_0) + \nabla f(x_0)(x-x_0) + \frac{1}{2}(x-x_0)^T \nabla^2 f(x_0)(x-x_0) + o(\|x-x_0\|^2).$$

Ex: $f(x,y)$ at (x_0, y_0)

$$f(x,y) = f(x_0,y_0) + \nabla f(x_0,y_0)\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}$$

$$= f(x_0, y_0) + \left[\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0)\right] \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + o(\|x - x_0\|).$$

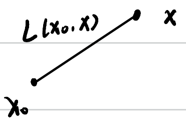
$$= f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) + o\left(\sqrt{(x-x_0)^2 + (y-y_0)^2}\right).$$

Hession : $\nabla^2 f(x_0, y_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} (x_0, y_0).$

$$\frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0),$$

$$= \frac{1}{2} \cdot [x - x_0, \ y - y_0] \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

$$= \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}(x_0, y_0)(x - x_0)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x - x_0)(y - y_0) + \frac{\partial^2 f}{\partial y^2}(y - y_0)^2.$$



define the line segment $L[x_0, x] = \{x = t x_0 + (1-t)x\}$.

Lagrange form for multi-variable.

$f \in C^1$: $f(x) = f(x_0) + \nabla f(\bar{x})(x - x_0)$, $\bar{x} \in L[x_0, x]$.

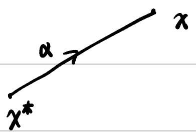
$f \in C^2$: $f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla f(\bar{x})(x - x_0)$

First order necessary condition for local minimum.

$f \in C^1(S)$, $x^* \in S$ local min

$\Rightarrow \nabla f(x^*) d \geq 0$, $\forall d$ feasible.

proof:



$d = x - x^*$

$g(\alpha) = f(x^* + \alpha d)$. , $\alpha \in [0, \bar{\alpha}]$.

$g(0) = f(x^*)$ is a local minimum of g.

$g'(\alpha) = \nabla f(x^* + \alpha d) d$.

$g'(0) = \nabla f(x^*) d$

Taylor :

$g(\alpha) = g(0) + g'(0)\alpha + o(\alpha)$

Suppose $g'(0)<0$

Then $g'(0)\alpha + o(\alpha) = \alpha\underbrace{\left[g'(0)+\frac{o(\alpha)}{\alpha}\right]}_{<0 \text{ for small } \alpha}$, $\alpha>0$

$\Rightarrow g(\alpha)<g(0)$ for small $\alpha>0$.

Contradicting that $g(0)$ is a local minium.

$f(x)=f(x^*)+\nabla f(x^*)(x-x^*)+o(\|x-x^*\|)$

$x-x^*=d$.

$\nabla f(x^*)d<0 \Rightarrow f(x)<f(x^*)$.

Corally

x^* is local minium, x^* is a interior point.

$\Rightarrow \nabla f(x^*)=0$.

proof: x^* is interior $\Rightarrow$ any d is feasible.

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

let $d=-\nabla f(x^*)$

$\Rightarrow (\nabla f(x^*))^2\leq 0$

$\Rightarrow (\nabla f(x^*))=0$

Ex: $f(x)=x^2$, $x\in[-1,1]$

① if x^* is local minium, & interior

$\Rightarrow f'(x^*)=0$ $\quad x^*=0$

② if $x^*=-1$ $\quad d=1$.

$\nabla f(x^*)d=-2<0$

Ex: $f(x)=-x^2$ $\quad x\in[-1,1]$

$f'(0)=0$ exactly not a local minium.

$x^*=-1$, $d=1$, $\nabla f(x^*)d=2>0$.

Ex: $f(x,y)=x^2-y^2$ $\quad (x,y)\in\mathbb{R}^2$.

$\nabla f=[2x,-2y]=0 \Rightarrow \begin{cases}x=0\\y=0\end{cases}$

Second order Condition

If $f \subset C^2(S)$

$x^* \in S$ is a local minium

d is feasible direction

s.t. $\nabla f(x^*) d = 0$.

Then $d^T \nabla^2 f(x^*) d \geq 0$

proof: $f(x) = f(x^*) + d^T \nabla^2 f(x^*) d + o(\|d\|^2)$.

Corally: if x^* is also an interior point, then $d^T f(x^*) d \geq 0$, $\forall d$.

i.e. $\nabla^2 f(x^*)$ is positive semidefinite.

Theorem: 2nd order sufficient Condition

If $f \subset C^2(S)$, x^* is an interior point of S and $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) > 0$

then x^* is a strict local minium of S.

proof: $f(x) = f(x^*) + \frac{1}{2}(x-x^*) \nabla^2 f(x^*)(x-x^*) + o(\|x-x^*\|^2)$.

$> f(x^*)$. for $\|x-x^*\|$ sufficiently small.

$\Rightarrow x^*$ is a strictly local minium.

Ex. $f(x,y) = x^4 + y^4 - 4xy + 1$. $(x,y) \in \mathbb{R}^2$

$\nabla f = [4x^3 - 4y, 4y^3 - 4x]$. let $\nabla f = 0$ $\begin{cases} x^3 - y = 0 \\ y^3 - x = 0 \end{cases}$ $\Rightarrow x^9 = x$ $x(x^8 - 1) = 0$

$\begin{cases} x = 0, 1, -1. \\ y = 0, 1, -1. \end{cases}$

stationary points: $(0,0)$ $(1,1)$ $(-1,-1)$.

$$\nabla^2 f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{bmatrix}$$

① at $(0,0)$, $\nabla^2 f = \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix} := F$

$\det F = -16$. $\Rightarrow F$ is indefinite

$\Rightarrow (0,0)$ isn't a local minium.

② at $(1,1)$.

$F = \nabla^2 F = \begin{bmatrix} 12 & -4 \\ -4 & 12 \end{bmatrix}$. $\det F > 0$, $12 > 0$, F is positive definite.

By the 2nd order sufficient conditions, $(1,1)$ is a strict local minium.

③ at $(-1,-1)$.

Def: A set is called convex is $L[a,b] \subset S$, $\forall a,b \in S$.

properities:

① C is convex, $\beta \in \mathbb{R}$

$\Rightarrow \beta C = \{\beta c : c \in C\}$ convex

② C, D convex

$C+D = \{c+d,\ c \in C,\ d \in D\}$ convex

③ $C_1 \cdots C_n$ convex

$\bigcap_{i=1}^{n} C_i$ convex.

Def. Convex hull of S is the smallest convex set containg S.

Def: S convex

$f: S \to \mathbb{R}$ is convex if $\forall x_1, x_2 \in S$, $\forall \alpha \in [0,1]$

$f(\alpha x_1 + (1-\alpha) x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)$.

f is strictly convex if $\forall x_1, x_2 \in S$, $x_1 \neq x_2$, $\forall \alpha \in (0,1)$

$f(\alpha x_1 + (1-\alpha) x_2) < \alpha f(x_1) + (1-\alpha) f(x_2)$.

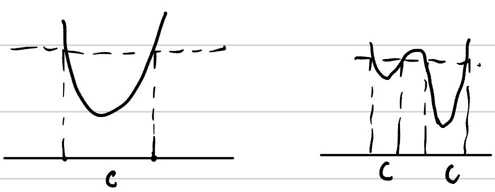
f is concave if $-f$ is convex.

prop: ① if f_1, f_2 convex $\Rightarrow f_1 + f_2$ convex

② f convex, $\alpha > 0 \Rightarrow \alpha f$ convex

f convex over S. $\Rightarrow \forall c \in \mathbb{R}$

$\Gamma_c = \{x \in S \mid f(x) \le c\}$ is convex.



Theorem: f is differentiable over a convex set S, then

$$f \text{ convex} \iff f(y) \geq f(x) + \nabla f(x)(y-x) \quad \forall x, y \in S.$$

1D: $f(y) \geq f(x) + f'(x)(y-x)$, if $y > x$

$$\frac{f(y)-f(x)}{y-x} \geq f'(x).$$

proof: "$\Rightarrow$": Suppose f is convex, let $x, y \in S$, $\forall \alpha \in [0,1]$

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

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

$$\frac{1}{\alpha}[f(x+\alpha(y-x)) - f(x)] \leq f(y) - f(x)$$

let $\alpha \to 0$

$$\nabla f(x)(y-x) \leq f(y) - f(x)$$

* directional derivative

$$D_u f(x) = \lim_{h \to 0} \frac{f(x+hu) - f(x)}{h}$$

Thm: $D_u f(x) = \nabla f(x) \cdot u$

"$\Leftarrow$" Suppose $f(y) \geq f(x) + \nabla f(x)(y-x)$, $\forall x, y \in S$

let $x_1, x_2 \in S$, $\alpha \in [0,1]$

let $\bar{x} = \alpha x_1 + (1-\alpha) x_2$

① $f(x_1) \geq f(\bar{x}) + \nabla f(\bar{x})(x_1 - \bar{x})$

② $f(x_2) \geq f(\bar{x}) + \nabla f(\bar{x})(x_2 - \bar{x})$

$\alpha ① + (1-\alpha) ② \Rightarrow \alpha f(x_1) + (1-\alpha) f(x_2) \geq \alpha[f(\bar{x}) + \nabla f(\bar{x})(x_1 - \bar{x})] + (1-\alpha)[f(\bar{x}) + \nabla f(\bar{x})(x_2 - \bar{x})]$

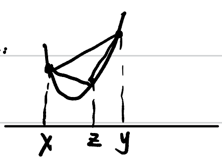
$$= f(\bar{x}) + \nabla f(\bar{x})[\alpha(x_1 - \bar{x}) + (1-\alpha)(x_2 - \bar{x})] = f(\bar{x}).$$

then we have $\alpha f(x_1) + (1-\alpha) f(x_2) \geq f(\alpha x_1 + (1-\alpha) x_2)$.

Thm: f is differentiable on S, then f is strictly convex $\iff f(y) > f(x) + \nabla f(x)(y-x)$, $\forall x, y \in S$.

proof: "$\Leftarrow$": similar.

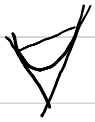
"$\Rightarrow$": Hint:



$$\frac{f(z)-f(x)}{z-x} < \frac{f(y)-f(x)}{y-x} < \frac{f(y)-f(z)}{y-z}$$

Collary: $f: \mathbb{R} \to \mathbb{R}$. differentiable

f convex $\iff f'$ increasing

proof:  $x<y$, $f'(x) \le \frac{f(y)-f(x)}{y-x} \le f'(y)$

Thm: $f \in C^2(S)$

① f Convex $\iff \nabla^2 f(x) \ge 0$

② $\nabla^2 f(x) > 0 \Rightarrow f$ strictly Convex.

proof: (by Taylor's Thm)

① "$\Leftarrow$" $f(y) = f(x) + \nabla f(x)(y-x) + \frac{1}{2}(y-x)^T \underbrace{\nabla^2 f(\bar{x})}_{\ge 0}(y-x)$

$\Rightarrow f(y) \ge f(x) + \nabla f(x)(y-x) \Rightarrow f(x)$ Convex.

"$\Rightarrow$" Suppose $\exists x^*$ s.t.

$\nabla^2 f(x^*)$ isn't positive semidefinite.

② Similar

Thm: f convex over convex set S, then

① the set where f achieves global minium is Convex

② any local minium is a global minium.

proof. let D: set of global minium points

let $x^* \in D$. Then

$D = \{x \in S.\ f(x^*) \le f(x)\}$ is Convex by proposition

② let $x \in S$ be a local minium, Suppose x isn't a local minium, then $\exists y \in S$

s.t. $f(y) < f(x)$ Then $\forall \alpha \in [0,1]$

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

Contradicts with x is local minium.

Theorem: $f \in C^1(S)$, S convex

if $\nabla f(x)(y-x) \geq 0 \quad \forall x, y \in S$, Then x is a global minium of f.

proof: $f(y) \geq f(x) + \nabla f(x)(y-x) \geq f(x) \quad \forall x, y \Rightarrow x$ is global minium.

Corolly: $f \in C^1(S)$: convex

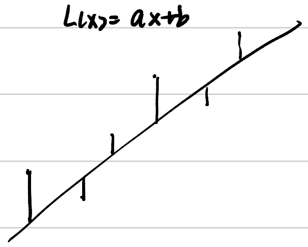
if $\nabla f(x) = 0$. Then x is global minium.

Ex: linear regression $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

Fit (x_i, y_i), $i = 1, 2, \cdots m$

by a straight line

$L(x) = ax + b$



$\min \sum | L(x_i) - y_i |^2$ least square method.

$f(a,b) = \frac{1}{2} \sum_{i=1}^{m} | a x_i + b - y_i |^2$

$\min\limits_{(a,b) \in \mathbb{R}^2} f(a,b)$

$$\begin{cases} \frac{\partial f}{\partial a} = \sum_{i=1}^{m} (a x_i + b - y_i) \cdot x_i \\ \frac{\partial f}{\partial b} = \sum_{i=1}^{m} (a x_i + b - y_i) \end{cases}$$

$$\Rightarrow \begin{cases} \frac{\partial f}{\partial a} = (\sum x_i^2) a + (\sum x_i) b - \sum x_i y_i \\ \frac{\partial f}{\partial b} = mb + (\sum x_i) a - (\sum y_i) \end{cases}$$

$\frac{\partial^2 f}{\partial a^2} = \sum x_i^2 \qquad \frac{\partial^2 f}{\partial a \partial b} = \sum x_i \qquad \frac{\partial^2 f}{\partial b^2} = m$

$$\nabla^2 f = \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & m \end{bmatrix}$$

$\det(\nabla^2 f) = m \sum x_i^2 - (\sum x_i)^2 \geq 0$. $= 0$ if all x_i the same.