APM462

Lecture 01 Jan 6th

R. Jerrard 250 Huron, 1001B

Office hrs T, Th 3:30-4:30 (by appointment)

rjerrard@math.toronto.ca

Text:

Nonlinear Opt. Linear and nonlinear Programming 3rd ed, by D. Luenberger & Y. Ye.

- unconstrained optimization (Ch 7, 8, 9, 11, 12, 13, 15)
- constrained

got it

online lecture notes: L.C. Evans (Ch 1-5)

optimal control theory (pdf) got it.

				①	②
marking scheme	1 midterm	March 3rd	2h.	~~32%~~	10%
	1 final		3h	~~48%~~	70%
	assignments	every 2 weeks		20%	20%

Review of some calculus ~~one~~ with maybe new notations.

lin. alge.

"little oh" notation. If h is a variable,

$o(h)$ denotes a quantity that (depending on h)

is negligible compared to h as $h \to 0$.

This means

$$\left|\frac{\text{quantity}}{h}\right| \to 0 \text{ as } h \to 0.$$

E.g. Spo g is a C^1 function of a single variable.

Definition of derivative: $g(x+h) - g(x) = h g'(x) + o(h)$ (*)

To see this, ~~divide by h~~ rewrite:

$$\cancel{\frac{g(x+h)-g(x)}{h} - g'(x) = \frac{o(h)}{h}}$$

$$g(x+h)-g(x)-hg'(x) = o(h)$$

This means

$$\lim_{h\to 0}\left|\frac{---}{h}\right| = 0$$

$$\lim_{h\to 0}\left|\frac{g(x+h)-g(x)}{h} - g'(x)\right| = 0$$

Can also write (*) as

$$g(x+h) = \underbrace{g(x)+hg'(x)}_{\text{linear function of } h} + \underbrace{o(h)}_{\text{"negligible compared with } h\text{"}}$$

Another true statement: "Taylor's Theorem" "2nd order Taylor Series"

If g is C^2 then

$$g(x+h) = g(x) + hg'(x) + \frac{1}{2}h^2 g''(x) + o(\cancel{h^2})$$

$$\lim_{h\to 0}\frac{g(x+h)-[g(x)+hg'(x)+\frac{1}{2}h^2g''(x)]}{h^2} = 0$$

$$\frac{h^2}{2}\left(g''(x)-g''(x+\theta h)\right)$$

some $\theta \in (0,1)$

~~Hatt Not~~ <u>Notation in Textbook</u>:

E^n = <u>column</u> vector w/ n components

E_n = <u>row</u> vector w/ n components

E^n = n-dim Euclidean space

$(v_1, \ldots, v_n)$ column vector $\rightarrow$ i.e. $= \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$

$[v_1, \ldots, v_n]$ row ~~colo~~ vector.

m

Multivariable Taylor expansions
(1st order and 2nd order)

Sps f is a C^1 function on E^n, x is a point in E^n, and $v \in E^n$

Claim:
$$f(x+v) = f(x) + \nabla f(x) v + o(|v|) \quad (**)$$

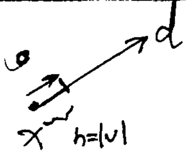
Notation: ∇f is always a row vector $[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}]$

Idea of (**)

Also $|v| = (v^T \cdot v)^{\frac{1}{2}} = (\sum_{i=1}^{n} v_i^2)^{\frac{1}{2}}$

Idea of (**). Since I'm interested in small v, let's ~~make~~ write $v = hd$, h is a number
d is a vector, $|d| = 1$

i.e. $h = |v|$



~~defite~~ define $g(h) = f(x+hd)$

g is a function of a single variable, so we can use earlier Taylor expansion

$$\boxed{g(h) = g(0) = h \cdot g'(0) + o(h)}$$

I want to rewrite this to get (**)

$g(h) = f(x+hd) = f(x+v)$

$g(0) = f(x)$,

since $h = |v|$, $o(h) = o(|v|)$

Also $g'(0) = \frac{d}{dh} f(x+hd)\Big|_{h=0} = \frac{d}{dh} f(x_1 + hd_1, \cdots, x_n + hd_n)$

$$= \frac{\partial f}{\partial x_1}(x+hd)\, d_1 + \cdots \frac{\partial f}{\partial x_n}(x+hd)\, d_n$$

set $h=0$, then

$$g'(0) = [\frac{\partial f}{\partial x_1}, \cdots \frac{\partial f}{\partial x_n}] \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = \nabla f(x) v$$

Reading week : Feb. 17-21

2nd order Taylor expansion in E^n.

x fixed, v small.

$$\boxed{f(x+v) = f(x) + \nabla f(x) v + \frac{1}{2} v^T \nabla^2 f(x) v + o(|v|^2)}$$

where $\nabla^2 f$ = matrix of 2nd derivatives.

~~(?)~~ i,j entry is $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$

Compare, for g function of a single variable.

$$g(x+h) = g(x) + hg'(x) + \frac{1}{2}h^2 g''(x) + o(h^2)$$

idea: as before, $v = h \cdot d$, d = unit vector, $h = |v|$

$g(h) = f(x+hd)$

write down 2nd ~~or~~ order expansion for g and then translate to f.

Only new part: $h^2 g''(0)$.

can check that in fact: $h^2 g''(0) = v^T \nabla^2 f(x) v$

Final calculus fact:

we saw that for C^1 for f on E^n.

$f(x+v) = f(x) + \nabla f(x) v + o(|v|)$

Conversely, x is any point in E^n, and $p \in E_n$

(✱) $\left\{ \begin{array}{l} \text{s.t. if } f(x+v) = f(x) + pv + o(|v|) \text{ then in fact} \\ p = \nabla f(x) \end{array} \right.$

True because:

we want to show (✱) $\Rightarrow p = \nabla f(x)$

equivalent to show $p \neq \nabla f(x) \Rightarrow$ (✱) not true

Let's try to do this: if $p \neq \nabla f(x)$, then $f(x+v) - [f(x) + pv]$

$$= f(x) + \nabla f(x) v + o(|v|)$$
$$= (\nabla f(x) - p)v + o(|v|)$$

Is it true that this ~~expansion~~ expression is $o(|v|)$?

No, because

$$\lim_{|v|\to 0} \frac{(\nabla f(x)-p)v + o(|v|)}{|v|} = \lim_{|v|\to 0}\left[(\nabla f(x)-p)\frac{v}{|v|}\right]$$

, depends only on direction of v, not on its size.

i.e. if we write $v = hd$, with $|d| = 1$, & $h = |v|$,
this is $\lim_{h\to 0} (\nabla f(x) - p)d$, ind. of h.

If $\nabla f(x) - p \neq 0$, then there is a d such that

column vector

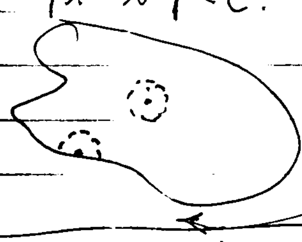
$(\nabla f(x) - p)d \neq 0$.

~~Uncons~~ Unconstrained optimization.

Basic problem

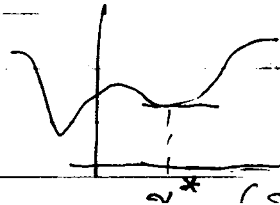
given function f on E^n & $\Omega \subseteq E^n$
minimize f in Ω.

Definition: x^* is a local or relative minimum point for f over Ω if there exist~~s~~ some number $\varepsilon > 0$, such that $f(x^*) \le f(x)$ for all $x \in \Omega$ such that $|x - x^*| < \varepsilon$.



If x^* in interior

Strict local minimum if strict inequality where $x \neq x^*$



x^* (strict local min)

x^*

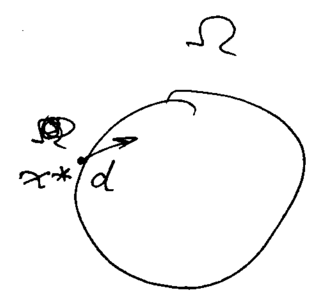
x^* local min, but not strict

x^* is global min of f over Ω if $f(x^*) \le f(x)\ \forall x \in \Omega$.

strict global min if $f(x^*) < f(x)$ for $\forall x \in \Omega,\ x = x^*$.

Basic problem, rewrite n:
find x^*, a global minimum of f over Ω.

We will usually consider $\Omega = E^n$.



Goal: necessary conditions for minima:

Definition:

If $x^* \in \Omega$, a vector d is a feasible direction at x^* if $\exists$ some number $\bar{\alpha} > 0$ such that

$$x^* + \alpha d \in \Omega \text{ whenever } 0 < \alpha < \bar{\alpha}$$

Note: if $x^* \in$ interior, every direction is feasible.

Proposition (1st order ~~necessary conditions~~)
If x^* is a relative minimum point for f over Ω. and if f is C^1, then $\nabla f(x^*) d \ge 0$ for ~~every fe~~ all feasible direction d.

Corollary: if $\Omega = E^n$, and x^* is a local min. pt for f, then $\nabla f(x^*) = 0$

Proof of proposition: for any feasible d,

$$f(x^* + hd) \ge f(x^*) \text{ if } 0 < h < \bar{\alpha}, \text{ and } h < \frac{\varepsilon}{|d|}$$

where ε comes from def of local minimum.

But $f(x^* + hd) = f(x^*) + \nabla f(x^*) \cdot hd + \cancel{o(|hd|)}$

rewrite $|d|\, o(h)$

rewrite $\underbrace{f(x^* + hd) - f(x^*)}_{\text{nonnegative}} = \nabla f(x^*) hd + |d|\, o(h)$

255 college street.
printorium

so $\lim_{h \to 0} \left(\frac{\text{left hand side}}{h} \right) = \nabla f(x^*) d \geq 0$

Point is $f(x^* + hd) = f(x^*) + h(\nabla f(x^*) \cdot d) + o(h)|d|$

if negative then near x^*, fcn decreases as h increases
IMPOSSIBLE

Proof of corollary:
If $\Omega \in E^n$, then every d is feasible, so
$\nabla f(x^*) d \geq 0$ for all d.
Also for every d, $-d$ is feasible direction.
$\nabla f(x^*) d \leq 0$ for all d.
Thus $\nabla f(x^*) d = 0$. for all d.
so $\nabla f(x^*) = 0$.

Proposition 2nd order necessary conditions
Sps f is a C^2 function on E^n, and x^* is a local min point for f. then $d^T \nabla^2 f(x^*) d \geq 0$ for all $d \in \mathbb{R}^n$

of the form c

Pf. $f(x^* + hd) = f(x^*) + h[\nabla f(x^*) d] + \frac{1}{2} h^2 d^T \nabla^2 f(x^*) d + o(h^3)$

=0 by the 1st order conditions

This $\frac{1}{2} h^2 d^T \nabla^2 f(x^*) d + o(h^2) \geq 0$ for all sufficiently small h.
divide by h^2 & let $h \to 0$ to find
$d^T \nabla^2 f(x^*) d \geq 0$

[redacted]

Recall 1st year calculus.

① local min $\Rightarrow$ $f'(x^*) = 0$

$f''(x^*) \geq 0$

② converse is false: can happen that

$f'(x^*)=0, f''(x^*) \geq 0$ but x^* not a local min.

However, $\left.\begin{array}{l} f'(x^*)=0 \\ f''(x^*) > 0 \end{array}\right\} \Rightarrow x^*$ is a local min.

For ②, e.g. $f(x) = x^3$

$f'(0) = f''(0) = 0$, but not a local min.

We have shown, for f function on E^n, x^* local min $\Rightarrow$
$\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ positive semi-definite

by definition, this means

$d^T \nabla^2 f(x^*) d \geq 0$ for all d.

Also true that converse is false.

Can happen that

$\left.\begin{array}{l} \nabla f(x^*)=0 \\ \nabla^2 f(x^*) \text{ pos. semi-definite} \end{array}\right\}$ but x^* not a local min.

e.g. $f(x_1, \cdots, x_n) = x_1^3 + \cdots + x_n^3$

$\nabla f(0) = 0, \nabla^2 f(0) = 0, \cdots, \nabla^n f(0) = 0$.

but 0 not a local min.

Analog of 3rd fact also holds.

$\left.\begin{array}{l} \nabla f(x^*)=0 \\ \nabla^2 f(x^*) \text{ positive definite} \end{array}\right\} \Rightarrow x^*$ is a local min for f.

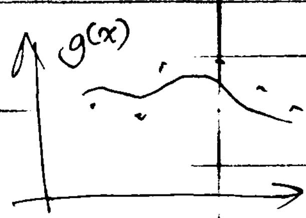
(i.e. $d^T \nabla^2 f(x^*) d > 0$ whenever d nonzero vector)

Why true?

$$f(x^*+hd) = f(x^*) + h\underbrace{\nabla f(x^*)}_{0}d + \frac{h^2}{2}\overbrace{d^T\nabla f(x^*)d}^{>0} + \underbrace{o(h^2)}_{\downarrow}$$

negligible compared to h^2 for h small

~~for h small~~



Example: Sps g is an unknown function of a single variable and suppose we measure g at points $x_1 \cdots x_m$

Goal: find polynomial function $p(x)$ of degree n which is a good fit for g.

What is a "good fit"?

Let's say, we want to minimize

$$\sum_{k=1}^{m}(p(x_k)-g(x_k))^2$$

Here the polynomial p has the form $p(x) = a_0 + a_1x + \cdots + a_nx^n$

We must make best choice of coefficient $a_0, \ldots, a_n$.

So we minimize. $f(a_0, \ldots, a_n) = \sum_{k=1}^{n}(p(x_k)-g(x_k))^2$

$$= \sum_{k=1}^{m}[(a_0 + a_1x_k + \cdots + a_nx_k^n) - g(x_k)]^2$$

Let's rewrite f:

introduce notation: $a \in \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} \in E^{n+1}$

column vector

For $k=1, m$, let $w_k = (1, x_k, \ldots, x_k^n) \in E^{n+1}$

Then $f = \sum_{k=1}^{m}[a^Tw_k - g(x_k)]^2$

where $Q = \sum_k w_k w_k^T$

continue to rewrite:

$$f = \sum_{k=1}^{n}(a^Tw_k)^2 - 2a^Tw_kg(x_k) + g(x_k)^2$$

Note: $\sum_{k=1}^{n}(a^Tw_k)^2 = \sum_{k=1}^{n}(a^Tw_k)(w_k^Ta) = \sum a^Tw_kw_k^Ta = a^TQa$

~~Proceed~~ Proceeding in this way:

$$\boxed{f = a^T Q a - 2b^T a + c}$$

Q as above, $b = 2\sum w_k g(x_k)$, $c = \sum_k g(x_k)^2$

So finally we want to minimize f.

$\boxed{\text{① first order condition : need } \nabla f}$

$\boxed{\text{Claim: } \nabla f(Q) = 2a^T Q - 2b^T}$

If the claim is true, then every candidate a^* for a minimum must satisfy $0 = \nabla f(a^*) = 2(a^{*T} Q - b^T)$

i.e. $Qa^* = b$

(Q is a symmetric matrix).

Why is claim true?

$$f(a+v) = (a+v)^T Q (a+v) - 2b^T(a+v) + c$$

$$= a^T Q a + a^T Q v + v^T Q a + v^T Q v - 2b^T a - 2b^T v + c$$

"3 terms are $f(a)$"

$$= f(a) + \left(a^T Q v + v^T Q a - 2b^T v\right) + v^T Q v$$

this can be rewritten $f(a+v) = f(a) + [2a^T Q - 2b]v + O(|v|)$

(using $Q^T = Q$, which follows from definition)

so this gives best linear approximation hence equals $\nabla f(a)$.

Lecture 02 Jan 13th

Tutorial

+ Office hour W Th
5-6 5-6

Today: convex functions.

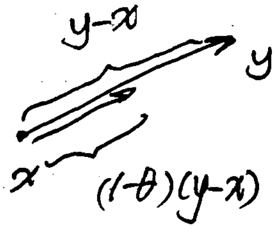
We are interested b/c for a convex function, every local minimum is a global minimum.

Def: A set $\Omega \subseteq E^n$ is <u>convex</u> if $\forall\ x, y \in \Omega$ and $\theta \in [0,1]$.

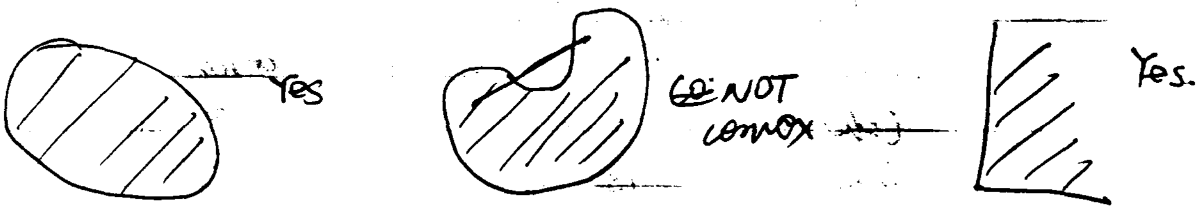
$\boxed{\theta x + (1-\theta) y \in \Omega}$

This means that if $x, y \in \Omega$, then the line segment joining x to y is contained in Ω.

$$\theta x + (1-\theta) y = x + \underbrace{(1-\theta)}_{\text{between } 0 \,\&\, 1}(y-x)$$



since as picture shows, the set of points $\{\theta x + (1-\theta) y : 0 \le \theta \le 1\}$ is exactly the line segment from x to y.



x-axis in x-y coordinate is convex.

A function f on a convex set $\Omega \in E^n$ is convex if for any $x, y \in \Omega$ and $\theta \in [0,1]$.

So $\boxed{f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)}$.

<u>strictly convex</u> if $f(\theta x + (1-\theta) y) < \theta f(x) + (1-\theta) f(y)$

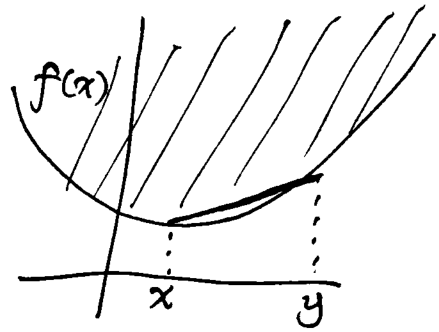
(note strict ineq)

~~MAT3??~~
MAT327
MAT357
MAT315

(STA300 level +)×(3 or 2)?
LIN203? F evening
STA304? F

- f is concave if $-f$ is convex.
  strictly concave if $-f$ is strictly convex.

e.g



θ

~~$f(x)$~~ f on the line segment joining x ~~and~~ and y
$\leq$ linear function on line segment that equals f at end pts.

i.e. is below

~~De~~ Remark: f is a convex function iff
$\{(x,z)\in E^{n+1} : z \geq f(x)\}$ is a convex set.

To prove ① Assume the function is convex and show that the set is convex. (using defn of convex ~~set &~~ convex function)
② Assume set is convex, deduce that function is convex.

Properties of convex functions:
① If f_1 and f_2 are convex functions
(always understood to be defined on a convex set Ω)
Then $f_1 + f_2$ is convex.

I must show that for any x,y in Ω, any $\theta \in [0,1]$
$$(f_1+f_2)(\theta x+(1-\theta)y) \leq \theta(f_1+f_2)(x) + (1-\theta)(f_1+f_2)(y)$$

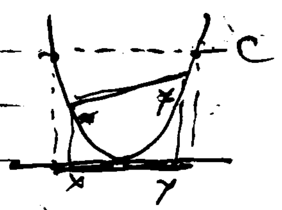
~~f(x)~~
$* \; f_1(x)+f_2(x) = (f_1+f_2)(x)$
why?

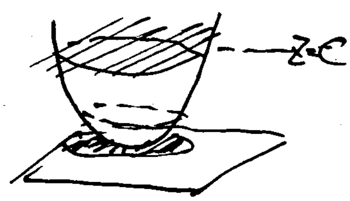
True because:
$$(f_1+f_2)(\theta x+(1-\theta)y) = f_1(\theta x+(1-\theta)y) + f_2(\theta x+(1-\theta)y)$$
$$\leq \theta f_1(x) + (1-\theta) f_1(y) + \theta f_2(x) + (1-\theta) f_2(y)$$
$$= \theta(f_1+f_2)(x) + (1-\theta)(f_1+f_2)(y)$$

② If f is convex and $a>0$, then $a \cdot f > 0$ is convex.
Pf: similar but easier.

<u>Note</u>: by combining the above:
if $f_1, \dots, f_k$ are convex and $a_1, \dots, a_k \geq 0$
then $a_1 f_1 + \cdots + a_k f_k$ is convex.

③ If f is a convex function on convex set Ω, then for any number C, $\{x \in \Omega : f(x) \leq C\}$ is a convex set.

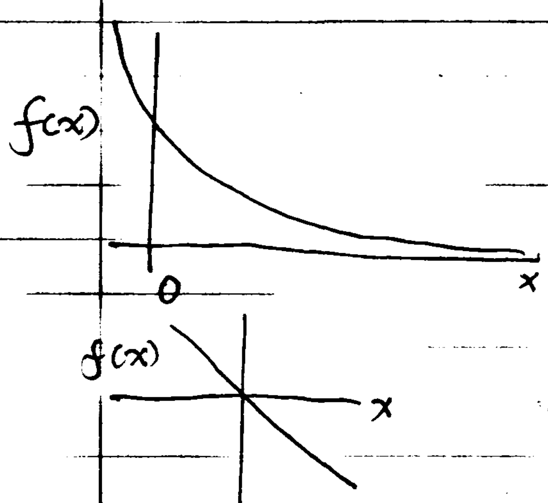




Pf: Sps $x, y \in \{z \in \Omega : f(z) < C\}$
i.e. $f(x) \leq C$ and $f(y) \leq C$

Then for any $\theta \in [0,1]$, by ~~convex theory~~ convexity
$f(\theta x (1-\theta)y)) \leq \theta f(x) + (1-\theta) f(y) \leq \theta \cdot C + (1-\theta) \cdot C = C$

So $\theta x + (1-\theta) y \in \{z \in \Omega : f(z) \leq C\}$.
So this set is convex.



Note:
convex functions need not "curve up" in all directions.

④ If f is a C^1 function on a convex set Ω, then f's ~~convex~~ convex if $\boxed{f(y) \geq f(x) + \nabla f(x)(y-x) \quad \forall x, y \in \Omega}$

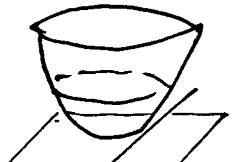
Note: If I write $y = x+v$ then above becomes

(★) $$\boxed{f(x+v) \geq \underbrace{f(x) + \nabla f(x)\, v} \quad \forall x, v \in \Omega}$$

$\Downarrow$ first order Taylor approximation of f near x.

$\Downarrow$

equation for "tangent plane" to f at x (as function of v).



"function $\geq$ every tangent plane"

Proof: of ~~(★)~~

① Assume f is convex

Fix $x, y \in \Omega$, I have to show (★)

By convexity,

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y), \ \forall \theta \in [0,1]$$

$$\| \\ f(x + (1-\theta)(y-x))$$

subtract $f(x)$ from both sides:

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

So $$\frac{f(x+(1-\theta)(y-x)) - f(x)}{(1-\theta)} \leq f(y) - f(x)$$

$$\lim_{\theta \to \cancel{0}1} \frac{f(x+(1-\theta)(y-x)) - f(x)}{(1-\theta)} \leq f(y) - f(x)$$

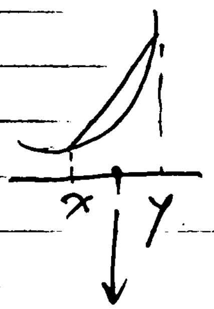
(directional derivative)

Since not hard to check that

$$\lim_{\theta \to 1} \frac{f(x+(1-\theta)(y-x))-f(x)}{1-\theta} = \nabla f(x)(y-x)$$

$$\frac{d}{d\theta} f(x+(1-\theta)(y-x)) \Big|_{\theta=1}$$

(Check it!) this proves ⊛



$X_\theta = \theta x + (1-\theta) y$

Now we assume ⊛ & try to show that f is convex.

Fix x and $y \in \Omega$ and $\theta \in [0,1]$

let $x_\theta = \theta x + (1-\theta) y$.

By condition ⊛:

$$f(x) \geq f(x_\theta) + \nabla f(x_\theta)(x - x_\theta) \qquad (1)$$

and

$$f(y) \geq f(x_\theta) + \nabla f(x_\theta)(y - x_\theta) \qquad (2)$$

Also, $x - x_\theta = (1-\theta)(x-y)$

and $y - x_\theta = \theta(y-x)$

So: $\theta(x - x_\theta) + (1-\theta)(y - x_\theta) = 0$

θ (eqn 1) $+ (1-\theta)$(eqn 2) $\Rightarrow$

$$\boxed{\theta f(x) + (1-\theta) f(y) \geq f(x_\theta)}$$ so f is convex.

(5) If f is a C^2 function on a convex set Ω, then f is convex iff $\boxed{v^T \nabla^2 f(x) v \geq 0, \ \forall x \in \Omega, \text{ and } v \in E^n}$ ⊛⊛

i.e. $\nabla^2 f$ positive semi-definite.

Proof: ① convexity $\Rightarrow$ ⊛⊛

sufficient to show that

if ⊛⊛ not true $\Rightarrow f$ not convex (contrapositive)

If ⊛⊛ not true then there is $x \in \Omega$ and $v \in E^n$.

such that $v^T \nabla^2 f(x) v < 0$.

By Taylor's Theorem

~~This~~ $f(x+hv) = f(x) + h \nabla f(x) v + \frac{1}{2} h^2 v^T \nabla^2 f(x) \cdot v + o(h^2 |v|^2)$

This implies that if h is small enough,

$\frac{1}{2} h^2 v^T \nabla^2 f(x) v + o(h^2 |v|^2) < 0$

so $f(x+hv) < f(x) + \nabla f(x)(hv)$

so f is non-convex by property (4)

Now assume (**). We'll show that property (4) holds.

Fix x, y. Goal: $f(y) \geq f(x) + \nabla f(x)(y-x)$.

Let $g(s) = f(x + s(y-x))$

Then $f(y) = g(1)$

$f(x) = g(0)$

By MVT, $\frac{g(1) - g(0)}{1-0} = g'(\theta)$ for some $\theta \in (0,1)$

~~ie $g(1) - g(0) = g'(\theta)(1-0)$~~

$g(1) - g(0) = g'(\theta) = g'(0) + (g'(\theta) - g'(0))$

$= g'(0) + \theta g''(s)$ for some $s \in (0, \theta)$

then one can check that

$g'(0) = \nabla f(x)(y-x)$ and $g''(\theta) = (y-x)^T \nabla^2 f(x + s(y-x))(y-x)$.

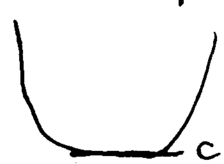
has the form

$v^T \nabla^2 f(z) v$ so ≥ 0

Put together to get $f(y)-f(x) \geq \nabla f(x)(y-x)$

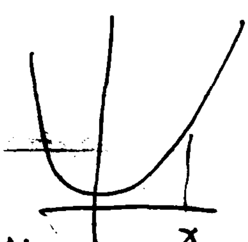
Thm 1: If f is a convex function on a convex set Ω, then
① $\Gamma = \{x \in \Omega : f(x) = \inf_{\Omega} f\}$ is convex, and
② every local min is a global min.

Pf: ① Let $c = \inf_{\Omega} f$
Then $\Gamma = \{x \in \Omega : f(x) \leq c\}$



and this is convex by property ③

Next: ② ~~sho~~ sufficient to show that
if x is not a global min, then
x is not a local min.



If x is not a global min, then $\exists\, y$
s.t. $f(y) < f(x)$.
so $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) < f(x)$, $\forall \theta \in [0,1]$
which implies that x is not a local min.

Thm 2: If f is convex on Ω and x^* is a point where $\boxed{\nabla f(x^*)(y-x^*) \geq 0}$ for all $y \in \Omega$, then x^* is a global min.

Pf: If this holds, by property ④
$f(y) \geq f(x^*) + \nabla f(x^*)(y - x^*) \geq f(x^*)$, $\forall y$.

Corollary: if $\nabla f(x^*) = 0 \Rightarrow x^*$ global min.

Review of linear algebra.

1) transpose of matrix.

$$(AB)^T = B^T A^T$$

2). linear dependent/independent

basis

3). determinants, eigenvectors, eigenvalues.

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Property.

A not invertible $\iff \det A = 0$

$\iff$ there ~~are~~ is some nonzero vector v s.t. $Av = 0$.

A number λ and vector v are eigenvalue and eigenvector if $Av = \lambda v$.

non-zero

equivalently $(A - \lambda I)v = 0$ where I = Identity matrix $= \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$

Combining the above:

λ is an eigenvalue $\iff \det(A - \lambda I) = 0$

nth-order polynomial in variable λ.

Strategy for finding e-values/vectors.

① compute polynomial $\det(A - \lambda I)$

② find roots $\lambda_1, \cdots, \lambda_k$ these are eigenvalues

③ For each λ_i, find all solutions of $Av_i = \lambda_i v$

Def: A matrix S is symmetric if $S^T = S$. as $S = \nabla^2 f(x)$

important fact.

if S symmetric, then • all eigenvalues are real.

• there is an orthonormal basis of ~~eigenvect~~ eigenvectors.

E.g. $S = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$

Find eigenvalues & eigenvectors.

① ~~st~~ eigenvalues. $\det(S - \lambda I) = \begin{vmatrix} 5-\lambda & -1 & -1 \\ -1 & 5-\lambda & -1 \\ -1 & -1 & 5-\lambda \end{vmatrix}$

$= (5-\lambda)^3 - (5-\lambda) \ldots$

$= -(\lambda-3)(\lambda-6)^2$

Find eigenvectors:

i. $\lambda = 3$. must solve $(S - 3I)v = 0$

i.e. $\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$v_1 = v_2 = v_3 = 1$.

so $v = c\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ where c is a constant.

ii. $\lambda = 6$

$S - 6I = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} = -\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ is an eigenvector if

$(S - 6I)v = 0$ i.e.

$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Rightarrow v_1 + v_2 + v_3 = 0$

i.e. v orthogonal to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

So any vector v s.t. $v_1+v_2+v_3=0$ is an eigenvector w/ $\lambda=6$.
But I have to think ~~about~~ a bit to find orthogonal eigenvectors.

e.g. $v=\begin{pmatrix}1\\-1\\0\end{pmatrix}$ and $\begin{pmatrix}1\\0\\-1\end{pmatrix}$ ~~are~~ are e.-vectors.

but not orthogonal.

But $\begin{pmatrix}1\\-1\\0\end{pmatrix}$ & $\begin{pmatrix}1\\1\\-2\end{pmatrix}$ are orthogonal.

Orthonormal basis of eigenvectors

$\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\\0\end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\-2\end{pmatrix}$

Def: A symmetric matrix S is positive semidefinite if $v^TSv \geq 0$ for all $v \in E^n$
positive definite if $v^TSv > 0$ for all nonzero $v \in E^n$.

Fact:
- S ~~is nonnegative~~ positive semi definite $\Leftrightarrow$ all eigenvalues ≥ 0
- S positive definite $\Leftrightarrow$ all eigenvalues > 0.

Why? Fix S symmetric and let $\lambda_1, \dots, \lambda_n$ eigenvalues, and $w_1, \dots, w_n$ = orthonormal basis of eigenvectors.
i.e. $w^T w_k = \delta_{jk} = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if not} \end{cases}$

- Any vector v = a linear combination of $w_1, \dots, w_n$.
 e.g. $v = a_1 w_1 + \dots + a_n w_n$.
 so $v^Tv = (a_1 w_1^T + \dots + a_n w_n^T)(a_1 w_1 + \dots + a_n w_n)$
 $= a_1^2 + \dots + a_n^2$

and

$$v^TSv = (a_1w_1^T + \cdots + a_nw_n^T)\, S\, (a_1w_1 + \cdots + a_nw_n)$$
$$= (\quad '' \quad)(a_1Sw_1 + \cdots + a_nSw_n)$$
$$= (\quad '' \quad)(a_1\lambda_1w_1 + \cdots + a_n\lambda_nw_n)$$

$$\boxed{v^TSv = \lambda_1a_1^2 + \cdots + \lambda_na_n^2}$$

Recall last week:

To find optimal polynomial approximation to a function in g, given $g(x_1), \cdots, g(x_m)$. we tried to minimize

$$f(a) = \sum_{k=1}^{n} (a_0 + a_1x_k + \cdots + a_nx_k^n - g(x_k))^2$$

we ~~we~~ rewrite as

$$f(a) = a^TQa - 2b^Ta - c$$ for some matrix Q, vector b, number c

Q is this function convex?