Lecture 1

Instructor: R. Jerrard rjerrard@math.toronto.ca

Review of some calculus with maybe new notations.
LINEAR ALGEBRA

"little oh" notation :
If h is a variable, $o(h)$ denotes a quantity (depending on h) that is negligible compared to h as $h \to 0$.

This means

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

e.g: Sps g is a C^1 function of a single variable.
Definition of derivative: $g(x+h)-g(x) = hg'(x) + o(h)$ ⊛

To see this, rewrite:

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

This means

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

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

Can also write ⊛ as:

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

Another true statement: "Taylor's thm" "2nd order Taylor Series"
If g is C^2 then

$$g(x+h) = g(x) + hg'(x) + \cancel{\tfrac{1}{2}h^2 g''(x)} + \cancel{o(h^2)} \quad \frac{h^2}{2}\left(g''(x) - g''(x+\theta h)\right)$$

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

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

NOTATIONS IN TEXTBOOK

E^n = column vector w/ n components
E_n = row vector w/ n components

up: column
down: row

E^n = n-dim Euclidean space.

$(v_1,\cdots,v_n)$ column vector $\longrightarrow$ i.e. $\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$

$[v_1,\cdots,v_n]$ row vector.

Multivariable Taylor Expansion
(1st order & 2nd order)

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

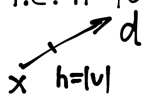
Claim: $f(x+v)=f(x)+\nabla f(x)v+o(|v|)$ (**)

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

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

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

i.e. $h=|v|$



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

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

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

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

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

set $h=0$, then $g'(0)=\left[\frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n}\right]\begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}=\nabla f(x)\cdot v$

2nd order Taylor expansion in E^n.

x fixed, v small.

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

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.

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

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

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

write down 2nd 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, if x is any point in E^n, and $p\in E_n$ s.t. $f(x+v)=f(x)+p\cdot v+o(|v|)$ $(\circledast)$
then in fact $p=\nabla f(x)$

True because: we want to show $(\circledast) \Rightarrow p=\nabla f(x)$
equivalent to show $p\neq\nabla f(x) \Rightarrow (\circledast)$ not true.
let's try to do this: if $p\neq\nabla f(x)$, then $f(x+y)-[f(x)+p\cdot v]=f(x)+\nabla f(x)\cdot v+o(|v|)$
$=(\nabla f(x)-p)v+o(|v|)$

Is it true that this expression is $o(|v|)$?
No, because $\displaystyle\lim_{|v|\to 0}\frac{(\nabla f(x)-p)\cdot v+o(|v|)}{|v|}=\lim_{|v|\to 0}\left[(\nabla f(x)-p)\frac{v}{|v|}\right]$, depends only on direction of v, not on its size.

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

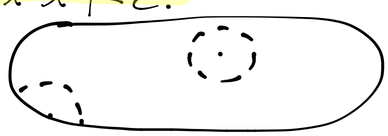
If $\nabla f(x)-p\neq 0$, then there is a column vector d such that $(\nabla f(x)-p)d\neq 0$

UNCONSTRAINED OPTIMIZATION

For the part above the line, see background.pdf.

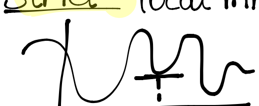
Basic problem: given function f on E^n & $\Omega\subseteq E^n$. minimize f in Ω.

Definition: x^* is a local or relative minimum point for f over Ω if there exist some number $\varepsilon>0$, such that $f(x^*)\leq f(x)$ 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 $f(x^*)\leq f(x)$, $\forall x\in\Omega$
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 $\bar{\alpha}$ s.t. $x^* + \alpha d \in \Omega$ 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 $\nabla f(x^*) d \geq 0$ for 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,

$f(x^* + hd) \geq f(x^*)$ if $0 < h < \bar{\alpha}$, and $h < \frac{\varepsilon}{|d|}$ where ε comes from def of local minimum.

But $f(x^* + hd) = f(x^*) + \nabla f(x^*) hd + \cancel{o(|hd|)}$ rewrite $|d|\, o(h)$

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

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

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

if negative then near x^*, function 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 pt for f, then $d^T \nabla^2 f(x^*) d \geq 0$ for all $d \in \mathbb{R}^n$.

of the form ( )( )( )

Pf: $f(x^* + hd) = f(x^*) + h\underbrace{[\nabla f(x^*) d]}_{=0 \text{ by the 1st order conditions}} + \frac{1}{2} h^2 d^T \nabla^2 f(x^*) d + o(h^2)$

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 and let $h \to 0$ to find $d^T \nabla^2 f(x^*) d \geq 0$

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

$\hookrightarrow$ 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

$\left.\begin{array}{l} \nabla f(x^*)=0 \\ \nabla^2 f(x^*) \text{ pos. semi-definite} \end{array}\right\} \Rightarrow$ but x^* is 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:

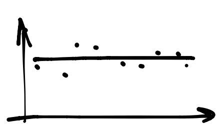
$\left.\begin{array}{l} \nabla f(x^*)=0 \\ \nabla^2 f(x^*) \text{ positive definite} \end{array}\right\} \Rightarrow x^*$ is a local min for f.

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

Why true?

$$f(x^*+hd)=f(x^*)+h\underbrace{\cancel{\nabla f(x^*)}}_{0}d+\overbrace{\frac{h^2}{2}d^T\nabla f(x^*)d}^{>0}+\underbrace{o(h^2)}_{\text{negligible compared to } h^2 \text{ for } h \text{ small}}$$

E.g. Sps g is an unknown function of a single variable and suppose we measure g at pts $x_1,\cdots,x_n$
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 $a_0, \cdots, a_n$ so we minimize:

$$f(a_0, \cdots, a_n) = \sum_{k=1}^{n} (p(x_k) - g(x_k))^2$$

$$= \sum_{k=1}^{n} [(a_0 + a_1 x_k + \cdots + a_n x_k^n) - g(x_k)]^2$$

Let's rewrite f.

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

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

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

<u>continue to rewrite</u>

$$f = \sum_{k=1}^{n} \left((a^T w_k)^2 - 2a^T w_k g(x_k) + g(x_k)^2 \right)$$

Note: $\sum_{k=1}^{n} (a^T w_k)^2 = \sum_{k=1}^{n} (a^T w_k)(w_k^T a) = \sum a^T w_k w_k^T a = a^T Q a$

where $Q = \sum_k w_k w_k^T$

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 we finally want to minimize f.

① First order condition: need ∇f

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

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 symetric matrix)

Why is this <u>claim</u> 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 $f(a)$"

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

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 $\nabla f(a)$.