Differentiation

note that for functions on $\mathbb{R}$ $f'(x)$ is defined so that

$$\cancel{f(x) =} \quad f(x+h) = f(x) + f'(x)h + r(h)$$

with remainder $r(h)$ small compared to h

$$\lim_{h \to 0} \frac{r(h)}{h} = 0$$

here $f'(x) \in \mathbb{R}$ which can be identified w/ $L(\mathbb{R}^1)$.

Def Suppose $E \subset \mathbb{R}^n$ open and $f: E \to \mathbb{R}^m$, $x \in E$.
If there exists $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$f(x+h) = f(x) + Ah + r(h) \quad \text{w/} \lim_{h \to 0} \frac{r(h)}{|h|} = 0$$

or equivalently

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

Then we say f is diff at x and $f'(x) = A$.

~~As usual we are not being precise w/~~

note here $h \in \mathbb{R}^n$ and

$f(x+h) - f(x) - Ah \in \mathbb{R}^m$. norms are

on $\mathbb{R}^n$ or $\mathbb{R}^m$ respectively.

Thm Suppose $f: E \subset \mathbb{R}^n \to \mathbb{R}^m$ diff at $x \in E$ and

$f'(x) = A_1$ and A_2 then $A_1 = A_2$

proof: Call $B = A_1 - A_2$ then

$$|Bh| \leq |f(x+h) - f(x) - A_1 h| + |f(x+h) - f(x) - A_2 h|$$

so that $\lim_{h \to 0} \frac{|Bh|}{|h|} = 0$

And for a fixed $h \in \mathbb{R}^n$ ~~$t \in \mathbb{R}$~~

$$0 = \lim_{t \to 0} \frac{|Bth|}{|th|} = \frac{|Bh|}{|h|}$$

so $Bh = 0 \quad \forall \; h \in \mathbb{R}^n$ $\square$

Note that (1) $f' : E \to L(\mathbb{R}^n, \mathbb{R}^m)$

(2) f differentiable $\Rightarrow$ f cts

(3) often I will call $f'(x) = Df(x)$

<u>Example</u> $f(x) = Ax$ for $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, $x \in \mathbb{R}^n$

then $f'(x) = A$ for all $x \in \mathbb{R}^n$.

just note that $A(x+h) - Ax = Ah$.

<u>Thm</u> (Chain Rule) Suppose E open in $\mathbb{R}^n$, $f : E \to \mathbb{R}^m$, f is diff at $x_0 \in E$, g maps an open set containing $f(E)$ into $\mathbb{R}^k$, g diff. at $f(x_0)$.

Then $F : E \to \mathbb{R}^k$ defined by

$F(x) = g(f(x))$ is diff. at x_0

and $F'(x_0) = g'(f(x_0)) f'(x_0)$

<u>Note</u> $f'(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m)$
$g'(f(x_0)) \in L(\mathbb{R}^m, \mathbb{R}^k)$ so $g'(f(x_0)) f'(x_0) \in L(\mathbb{R}^n, \mathbb{R}^k)$

proof: Call $y_0 = f(x_0)$, $A = f'(x_0)$, $B = g'(f(x_0))$

$$u(h) = f(x_0+h) - f(x_0) - Ah \qquad h \in \mathbb{R}^n$$

~~$v(h) = g(y_0+k)$~~

$$v(\delta) = g(y_0+\delta) - g(y_0) - B\delta \qquad \delta \in \mathbb{R}^m$$

s.t. f, g resp are defined at x_0+h, $y_0+\delta$.

then
$$\lim_{h\to 0} \frac{|u(h)|}{|h|} = 0 \qquad \lim_{\delta\to 0} \frac{|v(\delta)|}{|\delta|} = 0$$

put $\delta = f(x_0+h) - f(x_0)$

$$|\delta| = |Ah + u(h)| \le \left[\|A\| + \frac{|u(h)|}{|h|}\right]|h|$$

$$\begin{aligned} F(x_0+h) - F(x_0) - BAh &= g(y_0+\delta) - g(y_0) - BAh \\ &= B\delta - BAh + v(\delta) \\ &= B(\delta - Ah) + v(\delta) \\ &= Bu(h) + v(\delta) \end{aligned}$$

so
$$\frac{|F(x_0+h) - F(x_0) - BAh|}{|h|} \le \|B\| \frac{|u(h)|}{|h|} + \frac{|v(\delta)|}{|\delta|}\frac{|\delta|}{|h|} \to 0 \text{ as } h \to 0$$

$\square$

Partial Derivatives

E open $\subset \mathbb{R}^n$, $f: E \to \mathbb{R}^m$

$e_1, \dots, e_n$ and $u_1, \dots, u_m$ resp. ~~bases of~~ orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$

(standard bases)

The component functions of f are $f_1, \dots, f_m$ defined by

$$f(x) = \sum_{i=1}^{m} f_i(x) u_i \qquad x \in E.$$

(or since $u_1, \dots, u_m$ an ONB $f_i(x) = f(x) \cdot u_i$)

for $x \in E$, $1 \leq i \leq m$, $1 \leq j \leq n$ define the partial derivative in direction e_j

$$(D_j f_i)(x) = \lim_{t \to 0} \frac{f_i(x + t e_j) - f_i(x)}{t}$$

provided that the limit exists

we may also use the notation

$$\frac{\partial f_i}{\partial x_j} \quad \text{corresponding to notation} \quad f(x) = f(x_1, \dots, x_n)$$

33

Thm f diff at $x \in E$, then $(D_j f_i)$ ~~exists~~ (x) exist

and
$$f'(x) e_j = \sum_{i=1}^{m} (D_j f_i)(x)\, u_i \qquad (1 \le j \le n)$$

(i.e. $f'(x)$ can be written as matrix

with entries $(D_j f_i)(x)$

$$[f'(x)] = \begin{bmatrix} D_1 f_1 & \cdots & D_n f_1 \\ D_1 f_2 & & \vdots \\ \vdots & & \vdots \\ D_1 f_m & & D_n f_m \end{bmatrix}$$
)

41

proof since f is diff at x

$$f(x + te_j) = f(x) + f'(x)(te_j) + r(te_j)$$

so by linearity of $f'(x)$, $\frac{r(te_j)}{t} \to 0$ as $t \to 0$

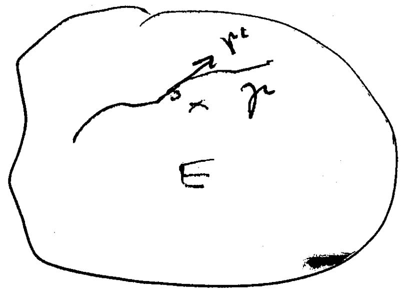
$$\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x) e_j$$

$\Rightarrow$ same limit holds for components

$$\lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t} = u_i \cdot f'(x) e_j = D_j f_i(x)$$

Directional Derivatives and Gradient

function $f: E \subset \mathbb{R}^n \to \mathbb{R}$



curve $\gamma: (a,b) \to E$

f, γ differentiable

then $f(\gamma(t)) : (a,b) \to \mathbb{R}$ diff.

$g(t) = f(\gamma(t))$, $g'(t) = f'(\gamma(t))\,\gamma'(t)$

where $\gamma'(t) \in L(\mathbb{R}, \mathbb{R}^n)$

$f'(\gamma(t)) \in L(\mathbb{R}^n, \mathbb{R})$

so $g'(t) \in L(\mathbb{R}, \mathbb{R})$ i.e. it is a real number

W.r.t. the standard basis on $\mathbb{R}^n$ $e_1, \dots, e_n$

$$[\gamma'(t)] = \begin{pmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{pmatrix} \qquad [f'(\gamma(t))] = (D_1 f, D_2 f, \dots, D_n f)\Big|_{\gamma(t)}$$

so $$g'(t) = \sum_{i=1}^{n} D_i f(\gamma(t))\,\gamma_i'(t)$$

This motivates us to define the gradient of f at x

$$(\nabla f)(x) = \sum_{i=1}^{n} (D_i f)(x) e_i \quad \text{a vector in } \mathbb{R}^n$$

(~~although~~ ~~note~~ see HW ~~#~~5)

Then $g'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t)$.

i.e., for any diff curve γ through x ~~at~~

at time 0

$$\frac{d}{dt} f(\gamma(t))\Big|_{t=0} = \nabla f(x) \cdot \gamma'(0)$$

depends only on the velocity vector of the curve.

given $u \in \mathbb{R}^n$, $x \in E$ and $\gamma(t) = x + tu$

$$g'(0) = \nabla f(x) \cdot u$$

but $g(t) - g(0) = f(x+tu) - f(x)$

so $$\lim_{t\to 0} \frac{f(x+tu) - f(x)}{t} = \nabla f(x) \cdot u$$

We call this limit the directional derivative of f at x in dir u, called $D_u f(x)$

When f is diff at x all the directional derivatives exist and have value given by $\nabla f(x) \cdot u$.

note: $D_u f(x) = \nabla f(x) \cdot u \leq |\nabla f(x)| \, |u|$

letting u vary over unit sphere we see

$$\max_{|u|=1} D_u f(x) = |\nabla f(x)|$$

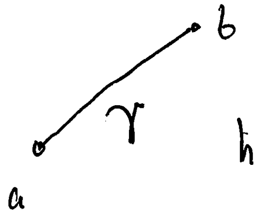
is attained when $u = \dfrac{\nabla f(x)}{|\nabla f(x)|}$

The gradient is the direction of fastest increase for f at x.

Thm Suppose f maps E open convex in $\mathbb{R}^n$, into $\mathbb{R}^m$, f is diff in E, and $\|f'(x)\| \le M$ for every $x \in E$. Then

$$|f(b) - f(a)| \le M|b-a| \quad \text{for all } a, b \in E.$$

proof: Define $\gamma(t) = a + t(b-a)$ path from a to b for $t \in [0,1]$



$\gamma([0,1]) \subset E$ since E convex

$h(t) = f(\gamma(t)) : [0,1] \to \mathbb{R}^m$

by the chain rule

$h'(t) = f'(\gamma(t))\gamma'(t) = f'(\gamma(t))(b-a)$

so $|h'(t)| \le \|f'(\gamma(t))\| \, |b-a| \le M|b-a|$

Then the "MVT" for vector valued fcns gives

$\cancel{|h(b) - h(a)| \le}$ $\exists\, t_0 \in (0,1)$ s.t.

$|f(b) - f(a)| = |h(1) - h(0)| \le |h'(t_0)| \le M|b-a|$ $\square$

~~$\bigcirc$~~

Cor If $f'(x) = 0$ $\forall\, x \in E$ then f constant.

Def:
A mapping $f : E \subset \mathbb{R}^n \to \mathbb{R}^m$ (E open) is called (or f is C^1 or $f \in C^1(E)$)
continuously differentiable in E ~~(or C^1)~~ if

$$f' : E \to L(\mathbb{R}^n, \mathbb{R}^m) \text{ is continuous.}$$

more explicitly $\forall\, x \in E$ and $\forall\, \varepsilon > 0$ $\exists\, \delta > 0$

s.t. $\|f'(y) - f'(x)\| \le \varepsilon$ when $|x - y| \le \delta$.

Thm $f \in C^1(E)$ iff the partial derivatives $D_j f_i$ exist and are continuous on E for $1 \le i \le m$, $1 \le j \le n$.

proof: Assume $f \in C^1(E)$.

Then $D_j f_i(x) = (f'(x) e_j) \cdot u_i$ (in particular it exists)

~~since f' is cts~~ for $x, y \in E$

$$D_j f_i(x) - D_j f_i(y) = (f'(x) e_j) \cdot u_i - (f'(y) e_j) \cdot u_i$$
$$= \left[(f'(x) - f'(y)) e_j\right] \cdot u_i$$

so by C-S $|D_j f_i(x) - D_j f_i(y)| \le |(f'(x) - f'(y))e_j| |u_i|$

$$\le \|f'(x) - f'(y)\|$$

since $|e_j|, |u_i| = 1$.

thus f' cts $\Rightarrow$ $D_j f_i$ cts

(that was the easy direction)

<u>Now assume</u> $D_j f_i$ exist and are cts on E.

We can assume $m = 1$ since

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|} = 0$$ iff limit holds for

each component. Similarly $f'(x)$ cts iff

$f_i'(x)$ is cts for each $1 \le i \le m$.

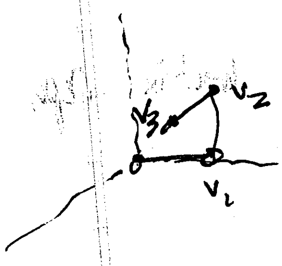
Let $x \in E$ and $\varepsilon > 0$, $\exists$ $r > 0$ s.t.

$|(D_j f)(y) - (D_j f)(x)| \le \frac{\varepsilon}{n}$ for $1 \le j \le n$ if $|y - x| < r$

by continuity of $D_j f$ for $1 \le j \le n$.

Now let $h = \sum_{j=1}^{n} h_j e_j$ w/ $|h| < r$

let $v_0 = 0$ and $v_k = h_1 e_1 + \dots + h_k e_k$



then

$$f(x+h) - f(x) = \sum_{j=1}^{n} f(x+v_j) - f(x+v_{j-1})$$

by the MVT for $f(x + v_{j-1} + t\underbrace{(v_j - v_{j-1})}_{= h_j e_j})$

$\exists$ $\theta_j \in (0,1)$ s.t.

$$f(x+v_j) - f(x+v_{j-1}) = D_j f(x + v_{j-1} + \theta_j h_j e_j)\, h_j$$

by continuity of $D_j f$

$$\left| D_j f(x + v_{j-1} + \theta_j h_j e_j) - D_j f(x) \right| \leq \frac{\varepsilon}{n}$$

so $\left| f(x+h) - f(x) - \sum_{j=1}^{n} D_j f(x) h_j \right| \leq \sum_{j=1}^{n} |h_j| \frac{\varepsilon}{n} \leq |h| \varepsilon$

dividing by $|h|$ and sending $|h| \to 0$

and then $\varepsilon \to 0$ identifies $f'(x)$

as the linear transformation $L(\mathbb{R}^n, \mathbb{R})$

$$[f'(x)](h) = \sum_{j=1}^{n} D_j f(x) h_j \cdot e_j$$

which (wrt the e_i basis) has matrix repn

$$[D_1 f(x), \ldots, D_n f(x)] = [f'(x)]$$

Since $D_j f$ are cts functions $f'(x)$ is as well $\square$