Note that for functions on \( \mathbb{R}^2 \) \( f' \) is defined so that
\[
f(x + h) = f(x) + f'(x) h + r(h)
\]
with remainder \( r(h) \) small compared to \( h \)
\[
\lim_{h \to 0} \frac{r(h)}{h} = 0
\]

Here \( f' : \mathbb{R}^2 \to \mathbb{R} \) which can be identified with
\[
L(\mathbb{R}^2, \mathbb{R}).
\]

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
\[
f(x + h) = f(x) + Ah + r(h)
\]
with
\[
\lim_{h \to 0} \frac{r(h)}{h} = 0
\]
or equivalently
\[
\lim_{h \to 0} \left| \frac{f(x + h) - f(x) - Ah}{h} \right| = 0
\]

Then we say \( f \) is dif. at \( x \) and \( f'(x) = A \).
The usual we are not being precise.

Note here the \( \mathbb{R}^n \) and \( \mathbb{R}^m \), norms are on \( \mathbb{R}^n \) or \( \mathbb{R}^m \) respectively.

Thus suppose \( f: E \subset \mathbb{R}^n \to \mathbb{R}^m \) differ at \( x \in E \) and

\[
f'(x_1) = A_1 \quad \text{and} \quad A_2.
\]

Then \( A_1 = A_2 \).

Proof: Call \( B = A_1 - A_2 \) then

\[
|Bh| \leq |f(x_1) - f(x_2) - A_1h| + |Ax_1 - f(x_1)| - A_2h
\]

so that

\[
\lim_{h \to 0} \frac{|Bh|}{|h|} = 0.
\]

And for a fixed \( h \in \mathbb{R}^n \),

\[
0 = \lim_{h \to 0} \frac{|Bh|}{|lh|} = \frac{|Bh|}{|lh|}.
\]

So \( Bh = 0 \) \( \forall h \in \mathbb{R}^n \) \( \Box \).
Note that (1) \( f' : E \to L(\mathbb{R}^n, \mathbb{R}^m) \)

(2) \( f \) differentiable \( \implies \) \( f \) continuous

(3) Often I will call \( f'(x) = Df(x) \)

Example

\[ f(x) = Ax \quad \text{for } A \in \mathbb{L}(\mathbb{R}^n, \mathbb{R}^m), \ x \in \mathbb{R}^n \]

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

Note that \( A(x+h) - Ax = Ah \).

The Chain Rule: Suppose \( E \) open in \( \mathbb{R}^n \), \( f : E \to \mathbb{R}^k \), \( f \) is differentiable at \( x_0 \in E \), \( g \) maps an open set containing \( f(x_0) \) into \( \mathbb{R}^m \), \( g \) differentiable at \( f(x_0) \).

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

\[ F(x) = g(f(x)) \]

is differentiable at \( x_0 \)

and

\[ F'(x_0) = g'(f(x_0)) f'(x_0) \]

Note:

\[ f'(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m) \]

\[ g'(f(x_0)) \in L(\mathbb{R}^m, \mathbb{R}^k) \]

so \( g'(f(x_0)) f'(x_0) \in L(\mathbb{R}^n, \mathbb{R}^k) \)
Proof: \( y_0 = f(x_0), \ A = f'(x_0), \ B = g'(f(x_0)) \)

\[
\begin{align*}
\mathbf{u(h)} &= f(x_0 + h) - f(x_0) - Ah \\
\mathbf{v(s)} &= g(y_0 + s) - g(y_0) - BS & s \in \mathbb{R}^m
\end{align*}
\]

s.t. \( f, g \) are defined at \( x_0 + h, y_0 + s \)

Then

\[
\lim_{h \to 0} \frac{\mathbf{u(h)}}{\mathbf{l}_{\mathbf{h}}} = 0 \quad \lim_{s \to 0} \frac{\mathbf{v(s)}}{\mathbf{l}_{\mathbf{s}}} = 0
\]

Put \( \mathbf{z} = f(x_0 + h) - f(x_0) \)

\[
|\mathbf{z}| = |A\mathbf{h} + \mathbf{u}(h)| \leq |A\mathbf{h}| + |\mathbf{u}(h)| \frac{1}{\mathbf{l}_{\mathbf{h}}}
\]

\[
F(x_0 + h) - F(x_0) - B\mathbf{A}h = g(y_0 + s) - g(y_0) - B\mathbf{A}h
\]

\[
= BS - B\mathbf{A}h + v(s)
\]

\[
= B(s - Ah) + v(s)
\]

\[
= B\mathbf{u}(h) + v(s)
\]

\[
\Rightarrow \quad \left| \frac{F(x_0 + h) - F(x_0) - B\mathbf{A}h}{\mathbf{l}_{\mathbf{h}}} \right| \leq \frac{|B||\mathbf{u}(h)|}{\mathbf{l}_{\mathbf{h}}} + \frac{|v(s)|}{\mathbf{l}_{\mathbf{s}}} \to 0 \quad \text{as} \quad l \to 0
\]
Partial Derivatives

Let \( E \) be an open subset of \( \mathbb{R}^n \), and \( f: E \to \mathbb{R}^m \).

Let \( e_1, \ldots, e_n \) and \( u_1, \ldots, u_m \) be orthonormal bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \), respectively.

The component functions of \( f \) are

\[
 f_1, \ldots, f_m \quad \text{defined by}
\]

\[
 f_i(x) = \sum_{j=1}^{n} f_j(x) u_j, \quad x \in E.
\]

(Or since \( u_1, \ldots, u_m \) is 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 \( e_j \)

\[
 (D_{e_j} f_i)(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t},
\]

provided that the limit exists.
we may also use the notation
\[
\frac{\partial f}{\partial x_j}
\]
corresponding to notation \( f(x_1, \ldots, x_n) \)

Thus if \( f \) is diff at \( x \in E \), then \( (D_j f_i)(x) \) exists and
\[
f'(x)e_j = \frac{1}{h}(D_j f_i)(x)u_i \quad (1 \leq j \leq n)
\]
(i.e., \( f'(x)e_j \) can be written as matrix with entries \( (D_j f_i)(x) \))

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

(since \( f \) is diff at \( x \))

\[
f(x+te_j) = f(x) + f'(x)(te_j) + o(te_j)
\]
so by linearity of \( f'(x) \), \( \frac{f(x+te_j)-f(x)}{t} \) as \( t \to 0 \)

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

⇒ same limit holds for components
\[
\lim_{t \to 0} \frac{f_i(x+te_j)-f_i(x)}{t} = u_i \cdot f'(x)e_j = D_i f(x)
\]
function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \)

curve \( \gamma : (0, b) \rightarrow \mathbb{R}^n \), \( f \), \( \gamma \) differentiable

then \( f(\gamma(t)) : (0, b) \rightarrow \mathbb{R} \), differentiable

\[ g(t) = f(\gamma(t)) \quad \Rightarrow \quad g'(t) = f'(\gamma(t)) \gamma'(t) \]

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

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

so \( g'(t) \in L([0, b], \mathbb{R}) \) i.e. it is a real number

wrt the standard basis on \( \mathbb{R}^n \) e1, ..., en

\[ \left[ \gamma'(t) \right] = \left[ \begin{array}{c} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{array} \right] \quad \Rightarrow \quad \left[ f'(\gamma(t)) \right] = \left[ \begin{array}{c} D_1 f(\gamma(t)) \\ \vdots \\ D_n f(\gamma(t)) \end{array} \right] \]

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 \( x \) as

\[
(\nabla f)(x) = \sum_{i=1}^{n} (\frac{\partial f}{\partial x_i})(x) \cdot e_i
\]

where \( e_i \) is a vector in \( \mathbb{R}^n \).

(Although take some time)

Then

\[ g'(t) = (\nabla f)(y(t)) \cdot \dot{y}(t) \]

i.e., for any differentiable curve \( \gamma \) through \( x \) at \( t = 0 \),

\[
\frac{d}{dt} f(\gamma(t)) \bigg|_{t=0} = (\nabla f)(x) \cdot \dot{\gamma}(0)
\]

depends only on the velocity vector of the curve.

Given \( u \in \mathbb{R}^n \), \( x \in E \) and \( y(t) = x + tu \),

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

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 that the directional derivative of \( f \) at \( x \) in the direction of \( u \), called \( D_u f(x) \), when \( f \) is differentiable at \( x \) all the directional derivatives exist and have value given by \( D_u f(x) \).

Note: \( D_u f(x) = \langle \nabla f(x), u \rangle \leq |\nabla f(x)| |u| \).

Letting \( u \) vary over unit sphere \( \mathbb{S}^1 \), we see

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

is attained when \( u = \frac{\nabla f(x)}{|\nabla f(x)|} \).

The gradient is the direction of fastest increase for \( f \) at \( x \).
The support of a mapping \( F \) is open, convex in \( \mathbb{R}^n \) into \( \mathbb{R}^m \).

If \( f \) is diff in \( E \), and \( \|f'(x)\| \leq M \) for every \( x \in E \). Then

\[
|f(b) - f(a)| \leq M|b-a|
\]

for all \( a, b \in E \).

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

\(
\gamma(0) = a, \quad \gamma(1) = b, \quad \gamma(t) \subseteq E \)

Since \( E \) is convex.

Then \( h(t) = f(\gamma(t)) \) and

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

So

\[
|h'(t)| = \|f'(\gamma(t))\| |b-a| \leq M |b-a|
\]

Then the "MVF" for vector valued fun, given

\[
|h(b) - h(a)| \leq E \quad \Delta E \subseteq \gamma(0) \quad \sin
\]

\[
|f(b) - f(a)| \leq |h(b) - h(a)| \leq |h'(t_0)| \leq M|b-a|
\]
Cor. If \( f(y) = 0 \) \( \forall x \in E \) then \( f \) constant.

Def: A mapping \( f : E \subset \mathbb{R}^n \rightarrow \mathbb{R}^k \) is called \( C^1 \) on \( E \) if

\[ f' : E \rightarrow L(\mathbb{R}^n, \mathbb{R}^k) \]

continuously differentiable in \( E \) for \( C^1 \). If \( f' \) is continuous.

more explicitly \( \forall x \in E \) and \( \forall \xi, \eta \in E \).

\[ \| f'(y) - f'(x) \| \leq C \| x - y \| \] for some \( C > 0 \).

Then \( f \in C^1(E) \) iff the partial derivatives \( Df_i, x \) exist and are continuous on \( E \) for all \( i \).

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

Then \( Df_i(x_i) = (f'(x_i)e_i) \cdot u_i \) (in particular \( i = 1 \)).

Hence \( f' \) is \( C^1 \) for \( x, y \in E \).

\[ f(x) - f(y) = (f'(x)e_i) \cdot u_i - (f'(y)e_i) \cdot u_i \]

\[ = \left[ (f'(x) - f'(y))e_i \right] \cdot u_i \]
so by CS
\[ |f(x) - f(y)| \leq \frac{1}{n} |f'(x) - f'(y)| \leq \frac{1}{n} |f(x) - f(y)| \]

since \( |e_j|, \|u_i\| = 1 \).

Thus \( f' \) is \( \text{cts} \) \( \implies \) \( \text{diff} \) \( \text{cts} \)

(\text{truf has the only direction})

Now assume \( \text{Diff} \) exists and \( \text{cts} \) \( \implies \) \( \text{cts} \).

We can assume \( m=1 \) since
\[ \lim_{t \to 0} \frac{|f(x+tu) - f(x)|}{|u|} = 0 \text{ if limit holds for } \]

each component. Similarly \( f'(x) \) \( \text{cts} \) \( \iff \)
\[ f'(x) \text{ is } \text{cts} \text{ for each component.} \]

Let \( x \in \mathbb{E} \) and \( \epsilon > 0 \), \( \exists r > 0 \) s.t.
\[ |f(y) - f(x)| \leq \frac{\epsilon}{n} \text{ for } |y-x| < r \]
by compactness of \( \mathbb{D} \) \( \implies \) \( \text{cts} \).
Now let $h = \sum_{j=1}^{n} \lambda_j v_j$ with $\|h\| < r$

Let $v_0 = 0$ and $v_k = \lambda_1 v_1 + \ldots + \lambda_k v_k$

Then $v_k$ is the $k$th term of the linear transformation $L(\mathbb{R}^n, \mathbb{R})$

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

by the MVT for $f(x+v_{j-1} + b(y_j-v_{j-1}))$ for some $0 < b_j < 1$ such that

$$f(x+v_j) - f(x+v_{j-1}) = Df(x + v_{j-1} + b_j h_j e_j) \cdot h_j$$

by the continuity of $Df$

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

so

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

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

and then $\varepsilon \to 0$ identifies $f'(x)$ as the linear transformation $L(\mathbb{R}^n, \mathbb{R})$.
\[ f'(x) \mathbf{e}_j = \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j} \mathbf{e}_j \]

which (writing the \( \mathbf{e}_i \) basis) has matrix form

\[ \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = [f'(x)] \]

since \( \frac{\partial f}{\partial x} \) one of the functions \( f(x) \) is as well.

\[ \square \]