The Derivative of a Function from $\mathbb{R}^n$ to $\mathbb{R}^m$

Def. A **linear transformation,** $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$, is a function such that for all $u, v \in \mathbb{R}^n$ and $c \in \mathbb{R}$:

a. $T(u+v) = T(u) + T(v)$

b. $T(cu) = cT(u)$

A linear transformation $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be represented with respect to the usual basis in $\mathbb{R}^n$ and $\mathbb{R}^m$ by an $m \times n$ matrix.

$$T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

where $T(e_i) = \sum_{j=1}^{m} a_{ji} e_j$, $e_j = (0, 0, \dots, 1, 0, 0, \dots, 0)$ and the 1 is in the j^{th} place.

The coefficients of $T(e_i)$ appear in the i^{th} column of the matrix.

$$T(e_i) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix}.$$

Ex. Let $T: \mathbb{R}^2 \to \mathbb{R}^4$ and $S: \mathbb{R}^4 \to \mathbb{R}^3$ be linear transformations. Suppose:

$$T(1,0) = (0,2,3,1) \qquad S(1,0,0,0) = (1,2,3)$$
$$T(0,1) = (2,-1,-1,2) \qquad S(0,1,0,0) = (-1,3,1)$$
$$S(0,0,1,0) = (2,3,1)$$
$$S(0,0,0,1) = (0,1,2).$$

Find a matrix representation of S and T with respect to the standard basis, then find a matrix representation of $S \circ T: \mathbb{R}^2 \to \mathbb{R}^3$.

$$T = \begin{pmatrix} 0 & 2 \\ 2 & -1 \\ 3 & -1 \\ 1 & 2 \end{pmatrix}; \qquad S = \begin{pmatrix} 1 & -1 & 2 & 0 \\ 2 & 3 & 3 & 1 \\ 3 & 1 & 1 & 2 \end{pmatrix}.$$

The matrix representation of the composition, $S \circ T$, is gotten by matrix multiplication.

$$S \circ T = \begin{pmatrix} 1 & -1 & 2 & 0 \\ 2 & 3 & 3 & 1 \\ 3 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 2 & -1 \\ 3 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 16 & 0 \\ 7 & 8 \end{pmatrix}.$$

Prop. Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a number, M, such that: $|T(h)| \leq M|h|$ for all $h \in \mathbb{R}^n$.

Proof: Let $h = (h_1, h_2, \dots, h_n)$ and $T = \begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix}$ then,

$$|T(h)| = \left\| \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} \right\|$$

$$= \begin{vmatrix} a_1 \cdot h \\ a_2 \cdot h \\ \vdots \\ a_m \cdot h \end{vmatrix} \qquad \text{where } a_i = (a_{i1}, a_{i2}, \dots, a_{in})$$

$$= \sqrt{(a_1 \cdot h)^2 + (a_2 \cdot h)^2 + \cdots + (a_m \cdot h)^2}$$

$$\leq \sqrt{(|a_1||h|)^2 + (|a_2||h|)^2 + \cdots + (|a_m||h|)^2} \quad \text{Cauchy-Schwarz Inequality}$$

$$= \left(\sqrt{|a_1|^2 + |a_2|^2 + \cdots + |a_m|^2}\right)|h|.$$

Thus:

$$|T(h)| \leq \left(\sqrt{|a_1|^2 + |a_2|^2 + \cdots + |a_m|^2}\right)|h|.$$

So take $$M = \sqrt{|a_1|^2 + |a_2|^2 + \cdots + |a_m|^2}.$$

If $f: \mathbb{R} \to \mathbb{R}$, we say that f is differentiable at $a \in \mathbb{R}$ if

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = f'(a)$$

This definition doesn't make any sense for a function $f: \mathbb{R}^n \to \mathbb{R}^m$ (in that case, $h \in \mathbb{R}^n$ and dividing by a vector is not defined).

However, we can think of any number, $f'(a)$, as defining a linear transformation λ of $\mathbb{R}$ into $\mathbb{R}$ by:

$$\lambda: \mathbb{R} \to \mathbb{R}$$
$$\lambda(h) = \big(f'(a)\big)h.$$

So we could reformulate our definition of the derivative, $f'(a)$, by saying:

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0$$

Or

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda(h)}{h} = 0$$

Thus, we say a function $f: \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if there is a linear transformation $\lambda: \mathbb{R} \to \mathbb{R}$ such that:

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda(h)}{h} = 0$$

Note: Any linear transformation, λ, of $\mathbb{R}$ into $\mathbb{R}$, $\lambda: \mathbb{R} \to \mathbb{R}$, is just multiplication by a fixed number; $\lambda(h) = ph;\ p \in \mathbb{R}$.

Now we can generalize this definition to: $f: \mathbb{R}^n \to \mathbb{R}^m$.

Def. A function $f: \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $a \in \mathbb{R}^n$ if there is a linear transformation, $\lambda: \mathbb{R}^n \to \mathbb{R}^m$, such that:

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0.$$

Notice that $\big(f(a+h) - f(a) - \lambda(h)\big) \in \mathbb{R}^m$ and $h \in \mathbb{R}^n$.

If this limit is 0, then we say: $Df(a) = \lambda$.

Theorem: If $f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$, then there is a unique linear transformation, $\lambda: \mathbb{R}^n \to \mathbb{R}^m$, such that:

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0.$$

Proof: Suppose $\tau: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation that also satisfies

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \tau(h)|}{|h|} = 0.$$

Then we have:

$$0 \le \lim_{h\to 0} \frac{|\lambda(h) - \tau(h)|}{|h|}$$

$$= \lim_{h\to 0} \frac{\big|\big(\lambda(h) - f(a+h) + f(a)\big) + \big(f(a+h) - f(a) - \tau(h)\big)\big|}{|h|}$$

$$\le \lim_{h\to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} + \lim_{h\to 0} \frac{|f(a+h) - f(a) - \tau(h)|}{|h|}$$

$$= 0 + 0 = 0. \qquad \Rightarrow \lambda(h) = \tau(h).$$

Ex. Let $f: \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x, y) = (xy, x + 2y)$. Using the definition of the derivative, show that:

$$Df(0,0) = \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}.$$

We must show that $\lim_{h \to 0} \frac{\left|f\left(\vec{0}+h\right) - f\left(\vec{0}\right) - \lambda(h)\right|}{|h|} = 0$, where $\lambda = \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}$.
If we let $h = (h_1, h_2)$ then:

$$\lim_{h \to 0} \frac{\left|f\left(\vec{0}+h\right) - f\left(\vec{0}\right) - \lambda(h)\right|}{|h|} = \lim_{h \to 0} \frac{\left|(h_1 h_2, h_1 + 2h_2) - \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\right|}{|h|}$$

$$= \lim_{h \to 0} \frac{|(h_1 h_2, h_1 + 2h_2) - (0, h_1 + 2h_2)|}{|h|} = \lim_{h \to 0} \frac{|h_1 h_2|}{\sqrt{h_1^2 + h_2^2}}$$

Notice that $(h_1 + h_2)^2 = h_1^2 + 2h_1 h_2 + h_2^2 \geq 0$

$$h_1^2 + h_2^2 \geq -2h_1 h_2$$

$$\frac{h_1^2 + h_2^2}{2} \geq |h_1 h_2| .$$

So:

$$0 \leq \lim_{h \to 0} \frac{|h_1 h_2|}{\sqrt{h_1^2 + h_2^2}} \leq \lim_{h \to 0} \frac{\frac{h_1^2 + h_2^2}{2}}{\sqrt{h_1^2 + h_2^2}} = \lim_{h \to 0} \frac{1}{2} \sqrt{h_1^2 + h_2^2} = 0$$

Thus $\lim_{h \to 0} \frac{\left|f\left(\vec{0}+h\right) - f\left(\vec{0}\right) - \lambda(h)\right|}{|h|} = 0$ and:

$$Df(0,0) = \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}.$$

Ex. Let $f: \mathbb{R}^2 \to \mathbb{R}$ by:

$$f(x,y) = \frac{xy}{\sqrt{x^2+y^2}} \quad ; \quad (x,y) \neq (0,0)$$
$$= 0 \quad ; \quad (x,y) = (0,0).$$

Show f is not differentiable at $(0,0)$.

Let's assume f is differentiable at $(0,0)$ and derive a contradiction.

If f is differentiable at $(0,0)$, then there is a linear transformation: $\lambda: \mathbb{R}^2 \to \mathbb{R}$ such that:

$$\lim_{h\to 0} \frac{|f(\vec{0}+h) - f(\vec{0}) - \lambda(h)|}{|h|} = 0$$

where $h = (h_1, h_2)$.

Let $\lambda = (a_{11} \;\; a_{12})$ so if $(x,y) \in \mathbb{R}^2$, then:

$$\lambda(x,y) = (a_{11} \;\; a_{12}) \binom{x}{y} = a_{11}x + a_{12}y, \quad \text{and}$$

$$\lim_{h\to 0} \frac{\left| \frac{h_1 h_2}{\sqrt{h_1^2+h_2^2}} - (a_{11}h_1 + a_{12}h_2) \right|}{\sqrt{h_1^2+h_2^2}} = 0.$$

For this limit to exist we must get the same value, 0, no matter which direction h approaches $(0,0)$.

Suppose $h = (h_1, 0)$; i.e. we approach $(0,0)$ along the x-axis.

$$\lim_{h_1 \to 0} \frac{|-a_{11}h_1|}{\sqrt{h_1^2}} = \lim_{h_1 \to 0} \frac{|a_{11}||h_1|}{|h_1|} = |a_{11}| = 0$$

so $a_{11} = 0$.

Now approach $(0,0)$ along the y-axis, i.e. $h_1 = 0$.

$$\lim_{h_2 \to 0} \frac{|-a_{12}h_2|}{\sqrt{h_2^2}} = \lim_{h_2 \to 0} \frac{|a_{12}||h_2|}{|h_2|} = |a_{12}| = 0$$

so $a_{12} = 0$.

Thus, $\lambda = (a_{11} \quad a_{12}) = (0 \quad 0)$ maps all vectors in $\mathbb{R}^2$ to 0.

Knowing $\lambda = (0 \quad 0)$, let's approach $(0,0)$ by
$h = (h_1, h_1)$, i.e. $h_2 = h_1$.

$$\lim_{h_1 \to 0} \frac{\left|\frac{h_1^2}{\sqrt{h_1^2 + h_1^2}} - 0\right|}{\sqrt{h_1^2 + h_1^2}} = \lim_{h_1 \to 0} \frac{h_1^2}{(h_1^2 + h_2^2)} = \lim_{h_1 \to 0} \frac{h_1^2}{2h_1^2} = \frac{1}{2}.$$

But then:

$$\lim_{h \to 0} \frac{\left|f(\vec{0} + h) - f(\vec{0}) - \lambda(h)\right|}{|h|} \neq 0$$

thus, $f(x, y)$ does not have a derivative at $(0,0)$.

Ex. Let $f: \mathbb{R}^2 \to \mathbb{R}$ by:

$$f(x,y) = \frac{x^2y^4}{x^4+6y^8} \qquad (x,y) \neq (0,0)$$
$$= 0 \qquad (x,y) = (0,0).$$

Show that $f(x,y)$ is not differentiable at $(0,0)$.

Let's assume $f(x,y)$ is differentiable at $(0,0)$ and derive a contradiction.

If f is differentiable at $(0,0)$, then there is a linear transformation, $\lambda: \mathbb{R}^2 \to \mathbb{R}$, where:

$$\lambda(x,y) = (a_{11} \quad a_{12})\binom{x}{y} = a_{11}x + a_{12}y$$

and

$$\lim_{h\to 0} \frac{|f(\vec{0}+h) - f(\vec{0}) - \lambda(h)|}{|h|} = 0.$$

If we let $h = (h_1, h_2)$ then:

$$\lim_{h\to 0} \frac{\left|\dfrac{h_1^2h_2^4}{h_1^4 + 6h_2^8} - (a_{11}h_1 + a_{12}h_2)\right|}{\sqrt{h_1^2 + h_2^2}} = 0.$$

For this limit to exist, we must get the same value, 0, when approaching $(0,0)$ from any direction. In particular, suppose $h = (h_1, 0)$ (i.e. we approach $(0,0)$ along the x-axis).

$$\lim_{h_1\to 0} \frac{|-a_{11}h_1|}{\sqrt{h_1^2}} = 0 \quad \Rightarrow \quad a_{11} = 0.$$

Now approach $(0,0)$ along the y-axis (i.e. $h_1 = 0$):

$$\lim_{h_2 \to 0} \frac{|-a_{12}h_2|}{\sqrt{h_2^2}} = 0 \quad \Rightarrow \quad a_{12} = 0$$

Thus, $\lambda = (a_{11} \quad a_{12}) = (0 \quad 0)$.

Knowing $\lambda = (0 \quad 0)$, let's approach $(0,0)$ by $h_1 = h_2^2$ (i.e. $h = (h_2^2, h_2)$).

$$\begin{aligned}\lim_{h \to 0} \frac{\left|f\left(\vec{0}+h\right) - f\left(\vec{0}\right) - \lambda(h)\right|}{|h|} &= \lim_{h_2 \to 0} \frac{h_2^4 h_2^4}{(h_2^8 + 6h_2^8)\sqrt{h_2^4 + h_2^2}} \\ &= \lim_{h_2 \to 0} \frac{h_2^8}{(7h_2^8)\sqrt{h_2^4+h_2^2}} \\ &= \lim_{h_2 \to 0} \frac{1}{(7)\sqrt{h_2^4+h_2^2}} \neq 0.\end{aligned}$$

Thus:

$$\lim_{h \to 0} \frac{\left|f\left(\vec{0}+h\right) - f\left(\vec{0}\right) - \lambda(h)\right|}{|h|} \neq 0$$

and $Df(0,0)$ does not exist.

Theorem: If $f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$, then it's continuous at $a \in \mathbb{R}^n$.

Proof: To show f is continuous at $a \in \mathbb{R}^n$ we need to show:

$\lim_{x \to a} f(x) = f(a)$ or equivalently $\lim_{h \to 0}\big(f(a+h) - f(a)\big) = 0$.

We need to use the fact that $Df(a)$ exists:

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0$$

for some linear transformation $\lambda: \mathbb{R}^n \to \mathbb{R}^m$.

Notice that:

$$0 \leq |f(a+h) - f(a)| = |f(a+h) - f(a) - \lambda(h) + \lambda(h)|$$

$$\leq |f(a+h) - f(a) - \lambda(h)| + |\lambda(h)|$$

$$= \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} \cdot |h| + |\lambda(h)|.$$

For any linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$, we know there is a $M \in \mathbb{R}$ such that:

$$|T(h)| \leq M|(h)|.$$

Thus:

$$0 \leq |f(a+h) - f(a)| \leq \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} |h| + M|h|.$$

Letting $h \to 0$ we know the RHS becomes 0 (why?).

Thus by the squeeze theorem:

$$\lim_{h \to 0} |f(a+h) - f(a)| = 0.$$

Hence:

$$\lim_{h \to 0} (f(a+h) - f(a)) = 0$$

$$\lim_{h \to 0} f(a+h) = f(a).$$

Differentiation Theorems:

The Chain Rule: If $f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a, and $g: \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(a)$, then $g \circ f: \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a, and:

$$D(g \circ f)(a) = Dg\big(f(a)\big) \circ Df(a).$$

1) If $f: \mathbb{R}^n \to \mathbb{R}^m$ is a constant function, then

$$Df(a) = 0$$

2) If $f: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then

$$Df(a) = f$$

3) If $f: \mathbb{R}^n \to \mathbb{R}^m$, then f is differentiable at $a \in \mathbb{R}^n$ if, and only if, f_i is differentiable and $Df(a) = \big(Df_1(a), \dots, Df_m(a)\big)$. Thus, $Df(a)$ is the $m \times n$ matrix whose i^{th} row is $Df_i(a)$

4) If $g: \mathbb{R}^2 \to \mathbb{R}$ is defined by $g(x, y) = x + y$, then

$$Dg(a, b) = g$$

5) If $m: \mathbb{R}^2 \to \mathbb{R}$ is defined by $m(x, y) = xy$, then

$\big(Dm(a, b)\big)(x, y) = bx + ay$, thus $Dm(a, b) = (b \quad a)$.

Proof:

1.

$$\lim_{h\to 0}\frac{|f(a+h)-f(a)-0|}{|h|}=0$$

2.

$$\lim_{h\to 0}\frac{|f(a+h)-f(a)-f(h)|}{|h|}=\lim_{h\to 0}\frac{|f(a)+f(h)-f(a)-f(h)|}{|h|}=0$$

3. If f_i is differentiable at a and $\lambda=\big(Df_1(a),\dots,Df_m(a)\big)$, then:

$$f(a+h)-f(a)-\lambda(h)$$
$$=\big(f_1(a+h)-f_1(a)-Df_1(a)(h),\dots,f_m(a+h)-f_m(a)-Df_m(a)(h)\big)$$

So,

$$\lim_{h\to 0}\frac{|f(a+h)-f(a)-\lambda(h)|}{|h|}\le\lim_{h\to 0}\sum_{i=1}^{m}\frac{|f_i(a+h)-f_i(a)-Df_i(a)(h)|}{|h|}=0.$$

Thus f is differentiable at a and $Df(a)=\lambda$.

If f is differentiable at a, then by #2 and the chain rule, $f_i=\pi_i\circ f$ is differentiable at a where $\pi_i(x)=x_i$, and

$$\begin{aligned}Df_i(a)&=D\pi_i(f(a))\circ Df(a)\\&=\pi_i\circ Df(a)\end{aligned}$$

Thus $Df(a)=\big(Df_1(a),\dots,Df_m(a)\big)$.

4. Since $g(x,y)=x+y$ is a linear transformation from $\mathbb{R}^2\to\mathbb{R}$, it follows from #2 that $Dg(a,b)=g$.

5. Let $\lambda(x, y) = bx + ay$, then

$$\lim_{h \to 0} \frac{|m(a + h_1, b + h_2) - m(a, b) - \lambda(h_1, h_2)|}{|(h_1, h_2)|} = \lim_{h \to 0} \frac{|h_1\ h_2|}{\sqrt{h_1^2 + h_2^2}}.$$

Notice that:

$$|h_1\ h_2| \le |h_1\ |^2 \quad \text{if } |h_2| \le |h_1|$$

$$|h_1\ h_2| \le |h_2\ |^2 \quad \text{if } |h_1| \le |h_2|$$

Hence: $\qquad |h_1\ h_2| \le |h_1|^2 + |h_2|^2.$

So we can write:

$$0 \le \frac{|h_1\ h_2|}{\sqrt{h_1^2 + h_2^2}} \le \frac{h_1^2 + h_2^2}{\sqrt{h_1^2 + h_2^2}} = \sqrt{h_1^2 + h_2^2}$$

$$\Rightarrow \qquad \lim_{h \to 0} \frac{|h_1\ h_2|}{\sqrt{h_1^2 + h_2^2}} = 0.$$

Corollary: If $f, g: \mathbb{R}^n \to \mathbb{R}$ are differentiable at a, then

i) $D(f + g)(a) = Df(a) + Dg(a)$

ii) $D(f \cdot g)(a) = g(a)Df(a) + f(a)Dg(a)$

iii) If $g(a) \neq 0$, then:

$$D\left({}^{f}\!/_{g}\right)(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{\left(g(a)\right)^2}$$

Proof of ii:

Let $F: \mathbb{R}^n \to \mathbb{R}^2$ by $F(x) = \big(f(x), g(x)\big)$
$p: \mathbb{R}^2 \to \mathbb{R}$ by $p(x_1, x_2) = x_1 \cdot x_2$
then, $f(x)g(x) = p \circ F(x)$.

$$
\begin{aligned}
D(fg)(a) &= D(p \circ F)(a) \\
&= Dp\big(F(a)\big) \circ DF(a) && \text{Chain Rule} \\
&= Dp\big(f(a), g(a)\big) \circ DF(a) \\
&= \big(g(a)\ \ f(a)\big)\begin{pmatrix} Df(a) \\ Dg(a) \end{pmatrix} && \text{by \#3, \#5} \\
&= g(a)Df(a) + f(a)Dg(a).
\end{aligned}
$$