# MATH 3742

We just saw that the derivative of $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ can be represented by the Jacobian $Df$, when it exists.

The formula for the Jacobian also reinforces something we already discovered, namely that

$$f'(\vec{c};\vec{u}) = \vec{u}_1 \frac{\partial f}{\partial x_1}(\vec{c}) + \cdots + \vec{u}_n \frac{\partial f}{\partial x_n}(\vec{c}) \quad \text{if } \vec{u} = (u_1, \dots, u_n),$$

when $f: \mathbb{R}^n \longrightarrow \mathbb{R}$.

That's because when $m=1$ (so $f$ is real-valued), then $Df$ is a row:

$$Df(\vec{c}) = [D_1 f(\vec{c}) \;\; D_2 f(\vec{c}) \cdots\cdots D_n f(\vec{c})]$$

and the directional derivative in the direction of $\vec{u}$ is

$$[Df(\vec{c})]\vec{u} = [D_1 f(\vec{c}) \cdots D_n f(\vec{c})] \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \sum_{i=1}^{n} u_i D_i f(\vec{c}).$$

Or, using the gradient $\nabla f$ of $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, this can be written as

$$Df(\vec{c}) = \nabla f(\vec{c}) \text{ (again, only if } m=1)$$

and for a vector-valued function $f = (f_1, \dots, f_m)$

$$(*) \quad Df(\vec{c})(\vec{v}) = \underbrace{f'(\vec{c};\vec{v}) = \sum_{k=1}^{m} f_k'(\vec{c};\vec{v})\vec{e}_k = \sum_{k=1}^{m} (\nabla f_k(\vec{c}) \cdot \vec{v})\vec{e}_k}_{\text{this is just from the definition of matrix multiplication, and the}}$$

definition of the notation

$$\nabla f_k(\vec{c}) = (D_1 f_k(\vec{c}), \ldots, D_n f_k(\vec{c})).$$

We need this equation for a lemma.

Lemma: With the setup above,

$$\|Df(\vec{c})(\vec{v})\| \le \|\vec{v}\| \sum_{k=1}^{m} \|\nabla f_k(\vec{c})\|, \quad \text{and} \quad \lim_{\vec{v}\to\vec{0}} Df(\vec{c})(\vec{v}) = \vec{0}.$$

$$= M\|\vec{v}\|$$

Proof: From equation (*), the inequality we want follows from the Cauchy Schwartz inequality since

$$\boxed{\sum_{k=1}^{m} (\nabla f_k(\vec{c})\cdot\vec{v})\vec{e}_k \le \|\vec{v}\| \sum_{k=1}^{m} \|\nabla f_k(\vec{c})\|}$$

$$\text{VI}$$

$$\left\|\sum_{k=1}^{m} f_k'(\vec{c};\vec{v})\vec{e}_k\right\|$$

$$\|$$

$$\|Df(\vec{c})(\vec{v})\|$$

This is Cauchy-Schwartz, recall it says

$$|\vec{u}\cdot\vec{v}| \le \|\vec{u}\|\|\vec{v}\|.$$

Second, $\lim_{\vec{v}\to\vec{0}} \|Df(\vec{c})(\vec{v})\| \le \lim_{\vec{v}\to\vec{0}} \|\vec{v}\| \underbrace{\sum_{k=1}^{m} \|\nabla f_k(\vec{c})\|}_{\text{constant wrt } \vec{v}} = 0,$

and thus $Df(\vec{c})(\vec{v}) \to 0$ as $\vec{v}\to\vec{0}$.

We will use this lemma to prove:

## §12.9 The chain rule.

Theorem 12.7. Let $f: \mathbb{R}^n \to \mathbb{R}^m$ and $g: \mathbb{R}^p \to \mathbb{R}^n$ be functions such that their composition $h = f\circ g$

is defined in a neighbourhood of $\vec{a} \in \mathbb{R}^p$. Suppose $g$ is differentiable at $\vec{a}$ with total derivative $g'(\vec{a})$, and $f$ is differentiable at $\vec{b} = g(\vec{a})$ with total derivative $f'(\vec{b})$. Then $h = f \circ g$ is differentiable at $\vec{a} \in \mathbb{R}^p$ and

$$h'(\vec{a}) = \underbrace{f'(\vec{b}) \cdot g'(a)}$$

matrix multiplication, or composition of linear functions, depending on how you are thinking of total derivatives.

Proof: We consider $h(\vec{a}+\vec{y}) - h(\vec{a})$ for small $\|\vec{y}\|$, and show

$$h(\vec{a}+\vec{y}) - h(\vec{a}) = \left[f'(b) \cdot g'(a)\right]\vec{y} + \|\vec{y}\| \, E_a(y).$$

for an appropriate error function $E_e(\vec{y})$.

First

$$h(\vec{a}+\vec{y}) - h(\vec{a}) = (f \circ g)(\vec{a}+\vec{y}) - (f \circ g)(\vec{a}) = f(\vec{b}+\vec{v}) - f(\vec{b}),$$

if we take $\vec{b} = g(\vec{a})$ and $\vec{v} = g(\vec{a}+\vec{y}) - \vec{b}$.

Then since $g$ obeys a Taylor formula we get

$$\vec{v} = g'(\vec{a})(\vec{y}) + \|\vec{y}\| \, E_a(\vec{y}), \quad E_a(\vec{y}) \to \vec{0} \text{ as } \vec{y} \to \vec{0};$$

and since $f$ obeys a Taylor formula we get

$$f(\vec{b}+\vec{v}) - f(\vec{b}) = f'(\vec{b})(\vec{v}) + \|\vec{v}\| \, E_b(\vec{v}), \text{ where}$$

$$E_b(\vec{v}) \to \vec{0} \text{ as } \vec{v} \to \vec{0}.$$

Combining these, we get:

$$f(\vec{b}+\vec{v})-f(\vec{b}) = f'(b)\left(g'(\vec{a})(\vec{y}) + \|\vec{y}\|\, E_a(\vec{y})\right) + \|\vec{v}\|\, E_b(\vec{v})$$

$$= f'(\vec{b})\cdot g'(\vec{a})(\vec{y}) + E(\vec{y})\,\|\vec{y}\|, \text{ where}$$

$$\|\vec{y}\|\, E(\vec{y}) = f'(\vec{b})\,\|\vec{y}\|\, E_a(\vec{y}) + \|\vec{v}\|\, E_b(\vec{v})$$

$$\Rightarrow E(\vec{y}) = f'(\vec{b})\, E_a(\vec{y}) + \frac{\|\vec{v}\|}{\|\vec{y}\|} E_b(\vec{v}) \quad \text{if } \vec{y} \neq 0.$$

Thus the proof is done if we can show $E(\vec{y}) \to 0$ as $\vec{y} \to \vec{0}$.

Now as $\vec{y} \to 0$ the formula for $\vec{v}$

$$\vec{v} = g'(\vec{a})(y) + \|\vec{y}\|\, E_a(\vec{y})$$

shows that $\vec{v} \to 0$ as well. Thus the term $f'(\vec{b})E_a(\vec{y}) \to \vec{0}$ as $\vec{y} \to 0$, but $\frac{\|\vec{v}\|}{\|\vec{y}\|} E_b(\vec{v})$ may or may not. It depends on "how fast" $\vec{v} \to 0$ as $\vec{y} \to 0$. We'll show $\frac{\|\vec{v}\|}{\|\vec{y}\|}$ is bounded as $\|\vec{y}\| \to 0$, completing the proof. We estimate:

$$\|\vec{v}\| \leq \|g'(\vec{a})(\vec{y})\| + \|\vec{y}\|\,\|E_a(\vec{y})\|$$

$$\leq \|\vec{y}\|\left(M + \|E_a(\vec{y})\|\right)$$

This is the M provided by our previous lemma

Thus

$$\frac{\|\vec{v}\|}{\|\vec{y}\|} \leq M + \|E_a(\vec{y})\|, \text{ so as } \vec{y} \to 0 \text{ this term}$$

remains bounded.

It follows that $\lim_{\vec{y}\to 0} \frac{\|\vec{v}\|}{\|\vec{y}\|} E_b(\vec{v}) = \vec{0}$,

so $\lim_{\vec{y}\to 0} E(y) = 0$ and we have the Taylor formula

$$h(\vec{a}+\vec{y}) - h(\vec{a}) = f'(b)\cdot g'(a)(\vec{y}) + \|\vec{y}\| E(\vec{y})$$

as desired. So the total derivative of $h = f\circ g$ at $\vec{a}$ is $f'(b)\cdot g'(a)$.

---

**Start here**

This means that the Jacobians $Dh$, $Df$, $Dg$ obey the law

$$Dh(\vec{a}) = Df(\vec{b})\,Dg(\vec{a}),$$

which our book calls "the matrix form of the chain rule". This ~~&~~ reduces to more familiar-looking forms in low dimensions. E.g.

Example: If $m=1$, so that $h = f\circ g$ is a real-valued function. Then we have, if ~~$h = (h_1, \ldots, h_m)$~~ $h = h(y_1, \ldots, y_n)$

$$\begin{bmatrix} \frac{\partial h}{\partial y_1}(\vec{a}) & \frac{\partial h}{\partial y_2}(\vec{a}) & \cdots & \frac{\partial h}{\partial y_n}(\vec{a}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{b}) & \cdots & \frac{\partial f}{\partial x_p}(\vec{b}) \end{bmatrix}$$

$$\cdot \begin{bmatrix} \frac{\partial g_1}{\partial y_1}(\vec{a}) & \frac{\partial g_1}{\partial y_2}(\vec{a}) & \cdots & \frac{\partial g_1}{\partial y_n}(\vec{a}) \\ \vdots & & & \\ \frac{\partial g_p}{\partial y_1}(\vec{a}) & \cdots & \cdots & \frac{\partial g_p}{\partial y_n}(\vec{a}) \end{bmatrix}$$

So that

$$\frac{\partial h}{\partial y_i}(\vec{a}) = \frac{\partial f\circ g}{\partial y_i}(a) = \sum_{k=1}^{p} \frac{\partial f}{\partial x_k}\frac{\partial g_k}{\partial y_i},$$ which we normally would write as

$$= \sum_{k=1}^{p} \frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial y_i},$$ thinking of the output of the $k^{th}$ coordinate function of $g: \mathbb{R}^n \longrightarrow \mathbb{R}^p$ as the $k^{th}$ argument of $f: \mathbb{R}^p \to \mathbb{R}$.

This gives the familiar chain rule for calculating derivatives of functions $\mathbb{R}^n \longrightarrow \mathbb{R}$.

# MATH 3472

**Example** (Chain rule matrix version vs. case of functions $\mathbb{R}^n \to \mathbb{R}$)

(from Dartmouth College notes)

Define $g: \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ by
$$g(s,t) = (s^2t,\ s+2t^2,\ st)$$
and $f: \mathbb{R}^3 \longrightarrow \mathbb{R}$ by
$$f(x,y,z) = e^{2x-y+z}.$$
We will calculate the derivative of $h = f \circ g$ in two ways at the point $(1,1)$.

Now $g(1,1) = (1,3,1)$ and so
$$(f \circ g)'(1,1) = f'(1,3,1) \cdot g'(1,1)$$

Compute:
$$g'(s,t) = \begin{bmatrix} D_1g_1(s,t) & D_2g_1(s,t) \\ D_1g_2(s,t) & D_2g_2(s,t) \\ D_1g_3(s,t) & D_1g_3(s,t) \end{bmatrix} = \begin{bmatrix} 2st & s^2 \\ 1 & 4t \\ t & s \end{bmatrix}$$

$$\Rightarrow g'(1,1) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix}$$

and similarly

$$f'(x,y,z) = \left[ D_1f(x,y,z) \ \ D_2f(x,y,z) \ \ D_3f(x,y,z) \right]$$
$$= \left[ 2e^{2x-y+z} \quad -e^{2x-y+z} \quad e^{-2x-y+z} \right]$$

Therefore

$$f'(1,3,1) = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}.$$ Then we compute

$$(f\circ g)'(1,1) = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \end{bmatrix}.$$

Another way of seeing this is to use the notation just discussed, ie the equation $\dfrac{\partial y}{\partial t_j} = \sum_{k=1}^{n} \dfrac{\partial y}{\partial x_k} \dfrac{\partial x_k}{\partial t_j}$. In our case, set $w = f(x,y,z)$ and $(x,y,z) = g(s,t)$ (so $x$ = outputs of coordinate function $g_1$, etc).

Then the above formula in our present setting gives

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial s} + \frac{\partial w}{\partial z} \cdot \frac{\partial z}{\partial s}$$

then compute

$$\frac{\partial w}{\partial s} = 2(2) + (-1)(1) + (1)(1) = 4,$$ which

agrees with the first entry in our matrix. Similarly

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial t} + \frac{\partial w}{\partial z} \cdot \frac{\partial z}{\partial t} = -1.$$

---

Sample application of Chain Rule.

**Theorem:** Suppose $f$ and $D_2 f$ are continuous on $[a,b]\times[c,d]$. Let $p(y), q(y)$ be functions $p,q:[c,d] \longrightarrow [a,b]$. Define

$$F(y) = \int_{p(y)}^{q(y)} f(x,y)\,dx \quad \text{if} \quad y\in[c,d]$$

Then $F'(y)$ exists and for each $y \in [c,d]$ we have

$$F'(y) = \int_{p(y)}^{q(y)} D_2 f(x,y)\,dx + f(q(y),y)q'(y) - f(p(y),y)p'(y).$$

---

Remark: This is viewed as an improvement of a theorem proved in Chapter 7 (Riemann-Stieltjes integration). Namely if $\alpha$ is of bounded variation on $[a,b]$, $f \in R(\alpha)$ and

f Riemann int. on $[a,b]$ wrt $\alpha$.

$$F(x) = \int_a^x f\,d\alpha \quad (x \in [a,b])$$

then

$$F'(x) = f(x)\alpha'(x)$$ (provided $\alpha$ is increasing and $\alpha'(x)$ exists. When $\alpha(x) = x$ is the identity, this reduces to the expected formula:

$$\left(\int_a^x f\,dx\right)' = f(x).$$

---

Proof of theorem: Set

$$G(x_1, x_2, x_3) = \int_{x_1}^{x_2} f(t, x_3)\,dt \text{ whenever } x_1, x_2 \in [a,b]$$

and $x_3 \in [c,d]$. Then we can rewrite $F$ as

$$F = G(p(y), q(y), y)$$

ie the composition

$$\mathbb{R} \xrightarrow{(p(y),q(y),y)} \mathbb{R}^3 \xrightarrow{G} \mathbb{R}$$

so the chain rule gives

$$F'(y) = \begin{bmatrix} D_1 G(p(y), q(y), y) & D_2 G(p(y), q(y), y) & D_3 G(p(y), q(y), y) \end{bmatrix} \begin{bmatrix} p'(y) \\ q'(y) \\ 1 \end{bmatrix}$$

$$\Rightarrow F'(y) = D_1 G(p(y), q(y), y) p'(y) + D_2 G(p(y), q(y), y) q'(y) + D_3 G(p(y), q(y), y)$$

By Theorems from Chapter 7,

$$D_1 G(x_1, x_2, x_3) = -f(x_1, x_3) \text{ and } D_2 G(x_1, x_2, x_3) = f(x_2, x_3).$$

and

$$D_3 G(x_1, x_2, x_3) = \int_{x_1}^{x_2} D_2 f(t, x_3)\, dt.$$

These equations result in the claimed formula for $F'(y)$.

<u>Note</u>: $D_1 G$ and $D_2 G$ are relatively straight forward consequences of what is called (sometimes) the fundamental theorem of calculus:

$$\frac{d}{dx}\left(\int_a^x f(t)\, dt\right) = f(x)$$

Application: Multivariable chain rule and change of coordinates.

Recall that cartesian $(x,y)$-coordinates are related to polar coordinates $(r,\theta)$ via

$$x = r\cos\theta, \quad y = r\sin\theta, \quad \text{and}$$
$$r = \sqrt{x^2+y^2}, \quad \theta = \tan^{-1}\frac{x}{y}.$$

Suppose we have a function expressed in $xy$-coordinates, like $g(x,y) = x^2y^3$. Then its derivative (in $xy$ coordinates) is

$$Dg(x,y) = \begin{bmatrix} 2xy^3 & 3x^2y^2 \end{bmatrix}.$$

What is its derivative with respect to $(r,\theta)$ coordinates? One way: plug $x = r\cos\theta$, $y = r\sin\theta$ into $g$, and differentiate. Alternatively, set

$$f: \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad f(r,\theta) = (r\cos\theta, r\sin\theta)$$

Then $g(r,\theta) = g \circ f(r,\theta)$, and the chain rule gives

$$D(g\circ f)(r,\theta) = Dg(f(r,\theta)) \cdot Df(r,\theta)$$

$$= \begin{bmatrix} 2(r\cos\theta)(r\sin\theta)^3 & 3(r\cos\theta)^2(r\sin\theta)^2 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \underbrace{5r^4\cos^2\theta\sin^3\theta}_{= \frac{\partial g}{\partial r}} & \underbrace{r^5\cos\theta\sin^2\theta(3\cos^2\theta - 2\sin^2\theta)}_{= \frac{\partial g}{\partial \theta}} \end{bmatrix}$$

This could have been done by substituting at the beginning $x = r\cos\theta$, $y = r\sin\theta$. But this is <u>not always</u> possible.

<u>Example</u>: Suppose $u : \mathbb{R}^2 \to \mathbb{R}$ satisfies

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \text{(Laplace's equation).}$$

What form would Laplace's equation take in polar coordinates?

From the chain rule,

$$Du(r,\theta) = D(u \circ f)(r,\theta) = Du(f(r,\theta)) \circ Df(r,\theta)$$

$$\begin{bmatrix} \frac{\partial u}{\partial r} & \frac{\partial u}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x}(f(r,\theta)) & \frac{\partial u}{\partial y}(f(r,\theta)) \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

But the matrix on the right is invertible, with inverse

$\begin{bmatrix} \cos\theta & \sin\theta \\ \frac{-\sin\theta}{r} & \frac{\cos\theta}{r} \end{bmatrix}$, so left-multiplying the above eqn by this inverse gives:

$$\begin{bmatrix} \frac{\partial u}{\partial x}(f(r,\theta)) & \frac{\partial u}{\partial y}(f(r,\theta)) \end{bmatrix} = \begin{bmatrix} \cos\theta \frac{\partial u}{\partial r} - \frac{\sin\theta}{r}\frac{\partial u}{\partial \theta} & \sin\theta \frac{\partial u}{\partial r} + \frac{\cos\theta}{r}\frac{\partial u}{\partial \theta} \end{bmatrix}$$

Now we compute second derivatives, and get

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}(v)$$

$$= \cos\theta \frac{\partial v}{\partial r} - \frac{\sin\theta}{r}\frac{\partial v}{\partial \theta}$$

since $u$ was arbitrary, same formula holds for $v$

$$= \cos\theta \frac{\partial}{\partial r}\left(\cos\theta \frac{\partial u}{\partial r} - \frac{\sin\theta}{r}\frac{\partial u}{\partial \theta}\right) - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial u}{\partial r} + \frac{\cos\theta}{r}\frac{\partial u}{\partial \theta}\right)$$

$=$ some equation giving $\frac{\partial^2 u}{\partial x^2}$ in terms of $r, \theta$.

Similarly we can compute $\frac{\partial^2 u}{\partial y^2}$, and substitute into Laplace's equation. We get:

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{r}\frac{\partial u}{\partial r} = 0.$$

## § 12.11 The Mean Value Theorem.

If $f: \mathbb{R} \longrightarrow \mathbb{R}$, then the Mean-Value theorem says that if $f$ is differentiable on $[a,b]$ then $\exists$ $c$ such that

$$f(b)-f(a)=f'(c)(b-a).$$

When $m>1$, no such formula can hold: Consider $f: \mathbb{R} \longrightarrow \mathbb{R}^2$ given by

$$f(t)=(\cos t, \sin t), \text{ and then}$$

total derivative multiplied by 'vector' $u \in \mathbb{R}$.

$$f'(t)(u)=u[-\sin t, \cos t]^T \text{ for all } u \in \mathbb{R}.$$

Now consider on the interval $[0, 2\pi]$. There we find that the MVT equation gives

$$\begin{array}{c} f(0)-f(2\pi) \\ \parallel \\ (1,0)-(1,0) \\ \parallel \\ \vec{0} \end{array} = \underbrace{u[-\sin t, \cos t]^T \cdot 2\pi}_{\text{a vector of length } 2\pi u.}$$

so the two sides cannot be equal. So the "obvious" generalization from functions $f: \mathbb{R} \longrightarrow \mathbb{R}$ to functions $\mathbb{R} \longrightarrow \mathbb{R}^m$ cannot hold.

The proper generalization is as follows: Use the notation

$$L(\vec{x},\vec{y})=\{t\vec{x}+(1-t)\vec{y} \mid t \in [0,1]\}$$

to denote the line segment connecting $\vec{x}$ and $\vec{y}$.

<u>Theorem</u> (Mean-Value Theorem) Let $S \subset \mathbb{R}^n$ be open and assume $f: S \longrightarrow \mathbb{R}^m$ is differentiable at every $s \in S$. Choose two points $\vec{x}, \vec{y} \in S$ such that $L(\vec{x}, \vec{y}) \in S$. Then for every vector $\vec{a} \in \mathbb{R}^m$ there is a point $z \in L(\vec{x}, \vec{y})$ such that

$$\vec{a} \cdot (f(\vec{y}) - f(\vec{x})) = \vec{a} \cdot \Big(\underbrace{f'(\vec{z})(\vec{y} - \vec{x})}_{\text{matrix multiplication!}}\Big).$$

<u>Proof</u>: Let $\vec{u} = \vec{y} - \vec{x}$. Then $S$ is open and $L(\vec{x}, \vec{y}) \subseteq S$, so $\exists\, \delta > 0$ such that $\vec{x} + t\vec{u} \in S$ for all $t \in (-\delta, 1+\delta)$. Choose $\vec{a} \in \mathbb{R}^m$ and define $F: \mathbb{R} \longrightarrow \mathbb{R}$ by

$$F(t) = \vec{a} \cdot f(\vec{x} + t\vec{u})$$

(actually, $F: (-\delta, 1+\delta) \longrightarrow \mathbb{R}$).

Then $F$ is differentiable with derivative

$$F'(t) = \vec{a} \cdot f'(\vec{x} + t\vec{u}; \vec{u}) \longleftarrow \text{directional derivative}$$

$$= \vec{a} \cdot f'(\vec{x} + t\vec{u})(\vec{u}) \longleftarrow \text{total derivative}$$

We can apply the usual mean value theorem to $F$ on $[0, 1]$ to arrive at $\theta \in (0, 1)$ such that

$$F(1) - F(0) = F'(\theta)(1 - 0) = F'(\theta),$$

and then observe

$$F'(\theta) = \vec{a} \cdot f'(\vec{x} + \theta\vec{u})(\vec{u}),$$

If we set $\vec{z} = \vec{x} + \theta\vec{u}$, this becomes

$$\vec{a} \cdot f'(z)(\vec{y} - \vec{x}) \quad (\text{recall } \vec{u} = \vec{y} - \vec{x} \text{ by def})$$

Further, $F(1) - F(0)$

$$= \vec{a} \cdot f(\vec{x} + \vec{u}) - \vec{a} \cdot f(\vec{x})$$

$$= \vec{a} \cdot (f(\vec{y}) - f(\vec{x})),$$

so we arrive at the desired equation.

---

Remarks ① If $f: \mathbb{R} \to \mathbb{R}$ note that this reduces to the original MVT, and the "$\vec{a}$" term becomes redundant (in the sense that it can simply be canceled from each side of the equation, it's just multiplication by a scalar)

② In general, there seems to be some disagreement between texts as to what the "correct" generalization of the MVT to higher dimensions should be.

Eg. Marsden-Hoffman states it as:

(i) If $f: S \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable on $S$ open, then $\forall \vec{x}, \vec{y} \in S$ s.t. $L(\vec{x}, \vec{y}) \subset S$ $\exists$ $c \in L(\vec{x}, \vec{y})$ such that

$$f(\vec{y}) - f(\vec{x}) = Df(\vec{c}) \cdot (\vec{y} - \vec{x})$$

and

(ii) If $f: S \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $S$ open, then $\exists$ $c_1, \dots, c_m$ on $L(\vec{x}, \vec{y})$ such that

$$f_i(\vec{y}) - f_i(\vec{x}) = Df_i(\vec{c})(\vec{y} - \vec{x}), \text{ for } i = 1, \dots, m$$

where $f = (f_1, \dots, f_m)$.

Certainly, wikipedia supports the notion that no generalization to $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is widely accepted as "correct".

The Apostol version of the MVT does allow for the following simple proof, however, which is analagous to the development of calculus for functions $\mathbb{R} \longrightarrow \mathbb{R}$.

Theorem: Let $S$ be an open, connected subset of $\mathbb{R}^n$, and $f: S \longrightarrow \mathbb{R}^m$ differentiable at all points in $S$. If $f'(\vec{c}) = 0$ for all $\vec{c} \in S$, then $f$ is constant on $S$.

Proof:

Since $S$ is open and connected, every pair of points $\vec{x}$ and $\vec{y}$ in $S$ can be connected by a polygonal arc (see Chapter 4). Denote the vertices of this arc by $\vec{x} = \vec{p}_1, \vec{p}_2, \vec{p}_3, \dots, \vec{p}_r = \vec{y}$.

Since each segment $L(\vec{p}_{i+1}, \vec{p}_i) \subseteq S$, the MVT gives $\vec{a}\cdot(f(\vec{p}_{i+1})-f(\vec{p}_i))=\vec{0}$

for every $\vec{a}$, assuming the derivative is everywhere $\vec{0}$.

If we ~~take~~ add together all these equations for $i=1,\dots,r-1$, we get

$$\vec{a}\cdot(f(\vec{x})-f(\vec{y}))=\vec{0}.$$

But $\vec{a}$ can be anything! So in particular, when $\vec{a}$ is $f(\vec{x})-f(\vec{y})$ we get

$$\|f(\vec{x})-f(\vec{y})\|=0$$

So $f(\vec{x})=f(\vec{y})$ and $f$ is constant.