

I pause to outline a few formal ideas that can help bring more clarity to our calculations. Read Colley §2.3, 2.4 and 2.5 for the whole story. Colley uses matrix notation so she can say far more general things efficiently. This material is not all req^d, I include it for breadth, so that you can be "well-rounded"

Defⁿ Let \mathcal{X} be open in \mathbb{R}^2 and $f: \mathcal{X} \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ then f is differentiable at $(a,b) \in \mathcal{X}$ if the partial derivatives $f_x(a,b)$ and $f_y(a,b)$ exist and if the function

$$h(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

is a good linear approx. to f near (a,b) . That is if

$$\lim_{(x,y) \rightarrow (a,b)} \frac{f(x,y) - h(x,y)}{|(x,y) - (a,b)|} = 0$$

Moreover, if f is differentiable at (a,b) then the eqⁿ $z = h(x,y)$ defines the tangent plane to the graph of f at $(a,b, f(a,b))$. If f is diff. $\forall (a,b) \in \text{dom}(f)$ then we say f is differentiable and write $f \in C^1(\mathcal{X}, \mathbb{R})$

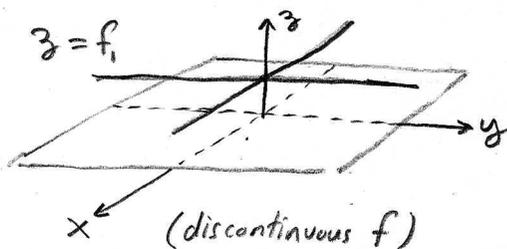
You can compare this to Stewart's comments on p. 930 where he defines the tangent plane and Defⁿ (7) on p. 931 where he defines differentiability of f . This is simply a precise way of saying the same things.

Th^m Suppose \mathcal{X} is open in \mathbb{R}^2 . If $f: \mathcal{X} \rightarrow \mathbb{R}$ has continuous partial derivatives in a neighborhood of (a,b) in \mathcal{X} then f is differentiable at (a,b) .

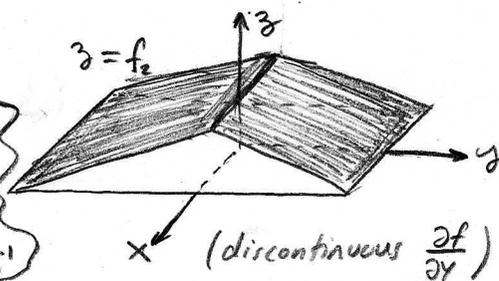
A function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ may fail to be differentiable due to discontinuity or discontinuous partials.

$$f_1(x,y) = \begin{cases} 0 & xy \neq 0 \\ 1 & xy = 0 \end{cases}$$

$$f_2(x,y) = 1 - |x|$$



NO WELL-DEFINED TANGENT PLANE BECAUSE $f_1 \notin C^1$



Just as in the case $f: \mathbb{R} \rightarrow \mathbb{R}$ (calc. I) we can prove that,

Thⁿ/ If $f: \mathcal{X} \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ is differentiable at (a, b) then f is continuous at (a, b)

We may generalize the idea of differentiability to $f: \mathcal{X} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. But first let me look ahead a bit and define the gradient

$$\nabla f = [\partial_1 f, \partial_2 f, \dots, \partial_n f] \quad \text{where } \partial_i f \equiv \frac{\partial f}{\partial x_i}$$

$$(\nabla f)(\vec{a}) = [\partial_1 f(\vec{a}), \partial_2 f(\vec{a}), \dots, \partial_n f(\vec{a})] \quad \text{where } \vec{a} = (a_1, a_2, \dots, a_n)^T$$

$$\underbrace{\nabla f(\vec{a}) \cdot (\vec{x} - \vec{a})}_{\text{dot-product in } \mathbb{R}^n} = [\partial_1 f(\vec{a}), \dots, \partial_n f(\vec{a})] \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{bmatrix} = \partial_1 f(\vec{a})(x_1 - a_1) + \dots + \partial_n f(\vec{a})(x_n - a_n).$$

dot-product in \mathbb{R}^n , can be encoded by matrix multiplication $\vec{x} \cdot \vec{y} = x^T y$

Defⁿ/ Let \mathcal{X} be open in \mathbb{R}^n and $f: \mathcal{X} \rightarrow \mathbb{R}$ and take a point $\vec{a} = (a_1, a_2, \dots, a_n)^T \in \mathcal{X}$. We say f is differentiable at \vec{a} if all the partials $\partial_i f(\vec{a})$ exist and if $h: \mathbb{R}^n \rightarrow \mathbb{R}$ defⁿ by
$$h(\vec{x}) = f(\vec{a}) + [\nabla f(\vec{a})] \cdot [\vec{x} - \vec{a}]$$
 is a good linear approximation to f near \vec{a} , meaning,
$$\lim_{\vec{x} \rightarrow \vec{a}} \frac{f(\vec{x}) - h(\vec{x})}{|\vec{x} - \vec{a}|} = 0$$

Again the condition of differentiability amounts to the existence of a best linear approximation ($h(\vec{x})$). This is in turn equivalent to the existence of a well-defined hyperplane to the graph $x_{n+1} = f(x_1, x_2, \dots, x_n)$ (a hypersurface in \mathbb{R}^{n+1}). The eqⁿ to the tangent hyperplane at $(\vec{a}, f(\vec{a}))$ is,

$$x_{n+1} = h(\vec{x}) = f(\vec{a}) + (\nabla f)(\vec{a}) \cdot (\vec{x} - \vec{a})$$

Remark: $z = f(a, b) + (\partial_x f)(a, b)(x-a) + (\partial_y f)(a, b)(y-b)$ and $y = f(a) + f'(a)(x-a)$ are the cases $n=2$ and $n=1$ where the "tangent hyperplanes" are actually an ordinary plane and a line.

The Jacobian Matrix

308

We bring our tour of the theory of multivariate differentiation to the most general case. We consider $f: \mathcal{X} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, this is an m -dim'l vector-valued function of n -variables x_1, x_2, \dots, x_n , we can express f in terms of its component functions f_j ,

$$f(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_m(\vec{x}))^T \in \mathbb{R}^m$$

Then the matrix $Df(x_1, x_2, \dots, x_n)$ is the Jacobian Matrix of f ,

$$Df(x_1, x_2, \dots, x_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} (\nabla f_1)^T \\ (\nabla f_2)^T \\ \vdots \\ (\nabla f_m)^T \end{bmatrix}$$

the "T" is for transpose it makes the column ∇f into a row $(\nabla f)^T$.

Defⁿ/ Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open and let $f: \mathcal{X} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Pick $\vec{a} \in \mathcal{X}$ then f is differentiable at \vec{a} if $Df(\vec{a})$ exists and if the function $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ defined by

$$h(\vec{x}) = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a})$$

is a good linear approx. to f near \vec{a} . That is

$$\lim_{\vec{x} \rightarrow \vec{a}} \frac{|f(\vec{x}) - h(\vec{x})|}{|\vec{x} - \vec{a}|} = \lim_{\vec{x} \rightarrow \vec{a}} \frac{|f(\vec{x}) - f(\vec{a}) - Df(\vec{a})(\vec{x} - \vec{a})|}{|\vec{x} - \vec{a}|} = 0$$

So we finally arrive at the general idea of differentiation. The derivative of a function is the best linear approximation.

That is if $f: \mathcal{X} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ then f is diff. at $a \in \mathcal{X}$ if $\exists L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ which is a linear mapping ($L(\alpha x + \gamma) = \alpha L(x) + L(\gamma)$) such that

$$\lim_{x \rightarrow a} \frac{|f(x) - f(a) + L(x-a)|}{|x-a|} = 0$$

Remark: If you find these generalities interesting you should take advanced calc' where you'll prove many of these results. I include them because of the examples I'm about to give. You are only expected to get the big ideas here,

PROPERTIES OF THE "DERIVATIVE"

These results are quite general. We consider $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ then if f is differentiable then we say $Df: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the derivative of f . Supposing $f \& g$ are diff. from $\mathbb{R}^n \rightarrow \mathbb{R}^m$ ($f, g \in C^1(\mathbb{R}^n, \mathbb{R}^m)$) then for $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

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

$$D(cf)(a) = c Df(a).$$

I don't find these results particularly surprising, however the next property, the Generalized Chain Rule, I found surprisingly simple.

Let $h = F \circ G$ where $F: \mathbb{R}^m \rightarrow \mathbb{R}^p$ and $G: \mathbb{R}^n \rightarrow \mathbb{R}^m$ so that $h: \mathbb{R}^n \xrightarrow{G} \mathbb{R}^m \xrightarrow{F} \mathbb{R}^p$ and suppose that $\vec{x} \in \mathbb{R}^n$ such that $h = F \circ G$ is differentiable at \vec{x} then

$$[Dh](\vec{x}) = [D(F \circ G)](\vec{x}) = (DF)(G(\vec{x})) DG(\vec{x})$$

where there are matrix multiplications implicit in the above, if we use operator notation then $D(F \circ G) = DF \circ DG$. This encompasses all the various unconstrained chainrules we've detailed upto now. (See Colley §2.5 for more details)

E69 Suppose $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ is differentiable. Define the polar coordinate change map; $\mathbb{X}(r, \theta) \equiv (r \cos \theta, r \sin \theta)$ this means $\mathbb{X}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ it takes $(r, \theta) \mapsto (x(r, \theta), y(r, \theta))$. Consider $g = f \circ \mathbb{X}$. Then $Dg = Df \circ D\mathbb{X}$ where $\mathbb{X} = (x, y)$

$$Df = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \quad D\mathbb{X} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{bmatrix}$$

Hence if $w = f \circ \mathbb{X} = g$

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

$$\left[\frac{\partial w}{\partial r}, \frac{\partial w}{\partial \theta} \right] = \left[\cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y}, -r \sin \theta \frac{\partial f}{\partial x} + r \cos \theta \frac{\partial f}{\partial y} \right]$$

Thus we find, (see §11.5 # 37, I mean you can find this w/o the matrix approach)

$$\frac{\partial w}{\partial r} = \cos \theta \frac{\partial w}{\partial x} + \sin \theta \frac{\partial w}{\partial y} \longrightarrow$$

$$\frac{\partial w}{\partial \theta} = -r \sin \theta \frac{\partial w}{\partial x} + r \cos \theta \frac{\partial w}{\partial y} \longrightarrow$$

$$\boxed{\begin{aligned} \frac{\partial}{\partial r} &= \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} \\ \frac{\partial}{\partial \theta} &= -r \sin \theta \frac{\partial}{\partial x} + r \cos \theta \frac{\partial}{\partial y} \end{aligned}}$$

this is an operator eqⁿ!

E70 Let $z = f(x, y) = x^2 - 3y^2$ and let $x = uv$ & $y = u + v^2$ calculate $\partial z / \partial u$ and $\partial z / \partial v$.

310

$$\frac{\partial z}{\partial u} = \frac{\partial}{\partial u} [f(x(u, v), y(u, v))] = \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u} = (2x)(v) - 6y(1).$$

$$\frac{\partial z}{\partial v} = \frac{\partial}{\partial v} [f(x(u, v), y(u, v))] = \frac{\partial f}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial v} = (2x)(u) - 6y(2v).$$

This is simple enough, you can use a tree-diagram if you like, but I've never needed them, you just identify the intermediate variables and sort-of "conserve partials". Lets see how this is done in the matrix/Jacobian formalism. We define

$$\Sigma(u, v) \equiv (x(u, v), y(u, v)) = (uv, u + v^2).$$

Thus, notice $x_1 = u$ and $x_2 = v$ while $x = f_1$, $y = f_2$ and $\Sigma = f$ to use (308),

$$D\Sigma = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \quad \text{while} \quad Df = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

Then $z = f \circ \Sigma$ so $z = z(u, v)$

$$\begin{aligned} D_z &= \left[\frac{\partial z}{\partial u}, \frac{\partial z}{\partial v} \right] = (Df)(D\Sigma) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \\ &= \left[\underbrace{\frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u}}_{\frac{\partial z}{\partial u}}, \underbrace{\frac{\partial f}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial v}}_{\frac{\partial z}{\partial v}} \right] \end{aligned}$$

So you see, its just the same formulas, the nice thing about the matrix is you get everything at once. Pragmatically speaking its the same calculation for particular physical problems. But, every view is another tool, it may make certain general arguments much more efficient. I mean just think about the chain rule, it gets all the other chain rules.