What does "differentiable" mean?

These notes are intended as a supplement (not a textbook-replacement) for a class at the level of Calculus 3, but can be used in a higher-level class as well. Instructors may find some of the material useful as well. Students seeing this material for the first time should not worry about all the relationships diagrammed on pp. 10 and 11, but these diagrams may be useful for instructors and very interested students.

1 Differentiability at a point

In Calculus 1, we learn that a real-valued function f is called *differentiable* at a if

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a} \tag{1}$$

exists. If the limit exists, we call it *the derivative of f at a*, denoted $f'(a)$. Thinking of a as a point on the real line, we call the notions above *differentiability at a point* and *the derivative at a point.*

One of the first things done in Calculus 1, after defining the derivative of f at a point a, is to let a vary over all numbers for which the limit (1) exists, obtaining the function f' (less precisely, "$f'(x)$") that we call the derivative of f.

Once we move into multivariable calculus (usually in Calculus 3), however, the definition of "differentiable at a point" becomes more complicated, and the usual definition—i.e. the one you'll find in most Calculus 3 textbooks—looks very different from the one-variable definition and can appear unmotivated[1]. However, there is a unified definition of "differentiable at a point" for a function of n variables that is equivalent to the usual definition for $n = 1$, and also equivalent to usual definition for $n > 1$. The purpose of these notes is to give that definition, see how it reduces to the usual one when $n = 1$, and to explore the definition a bit.

First, we define some terms we will use. Throughout these notes, when we are talking about functions on $\mathbf{R}^n$ and n is not specified to be 1, 2, or 3, we will use notation such as $\mathbf{x} = (x_1, x_2, \ldots x_n)$ and $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ for points of $\mathbf{R}^n$.[2] For notational simplicity, we will also implicitly identify points with their position vectors. Thus, for example, $\mathbf{x} - \mathbf{a} = (x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$, and

$$(\text{distance from } \mathbf{x} \text{ to } \mathbf{a}) = \|\mathbf{x}-\mathbf{a}\| = \sqrt{(x_1-a_1)^2 + (x_2-a_2)^2 + \cdots + (x_n-a_n)^2}\ .$$

---

[1] The textbook by Bona and Shabanov, *Concepts in Calculus*, currently being used in MAC 2313 at UF, is an exception. The definition in *Concepts of Calculus* is the same as the definition in these notes.

[2] Students who are seeing the material for the first time should think "$n = 1, 2,$ or 3" in all references to n and $\mathbf{R}^n$. After thoroughly understanding those cases, the student can re-read portions of these notes to understand that there is no conceptual difference for larger n. The fundamental change occurs when passing from $n = 1$ to $n = 2$.

Definition 1.1 Let f be a real-valued function on $\mathbf{R}^n$ and let $\mathbf{a}$ be a point in $\mathbf{R}^n$. Let g also be a real-valued function on $\mathbf{R}^n$.

- We call g an *approximation*[3] *of* f *near* $\mathbf{a}$ if g is continuous at $\mathbf{a}$, and $g(\mathbf{a}) = f(\mathbf{a})$.
- We call g a *good* approximation[4] of f near $\mathbf{a}$ if g is approximation of f near $\mathbf{a}$ and

  $$\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x}-\mathbf{a}\|} = 0. \tag{2}$$

  (If f itself is continuous at $\mathbf{a}$, then the first half of this definition is not needed: condition (2) automatically implies that g is continuous at $\mathbf{a}$ and that $g(\mathbf{a}) = f(\mathbf{a})$; i.e. that g is approximation of f near $\mathbf{a}$. But if f is not continuous at $\mathbf{a}$, then (2) does not not imply that g is continuous at $\mathbf{a}$.)
- We call g *linear*[5], or a *linear function*, if g is a polynomial of degree at most 1 in the coordinates on $\mathbf{R}^n$. (A degree-zero polynomial is a constant function.)

Observe that, in the setting of Definition 1.1, if g is *any* approximation of f near $\mathbf{a}$, and f is continuous at $\mathbf{a}$, then $\lim_{\mathbf{x}\to\mathbf{a}}(f(\mathbf{x}) - g(\mathbf{x})) = 0$. The added feature of a *good* approximation g of a continuous function f, near $\mathbf{a}$, is that as $\mathbf{x} \to \mathbf{a}$, the quantity $f(\mathbf{x}) - g(\mathbf{x})$ approaches zero much faster than the distance from $\mathbf{x}$ to $\mathbf{a}$ does: even when $f(\mathbf{x}) - g(\mathbf{x})$ is divided by the ever-smaller quantity $\|\mathbf{x}-\mathbf{a}\|$, the quotient approaches zero as $\mathbf{x} \to \mathbf{a}$.

Approximations g of f near $\mathbf{a}$ can be good without being linear, and can be linear without being good. (An example of the latter type: writing $c = f(\mathbf{a})$, the constant function g defined by $g(\mathbf{x}) = c$ is a linear approximation of f near $\mathbf{a}$, but usually is not a *good* approximation.)

If g is a linear function on $\mathbf{R}^n$ then

$$g(\mathbf{x}) = c + m_1x_1 + m_2x_2 + \cdots + m_nx_n \tag{3}$$

for some real numbers c and $m_1, \ldots, m_n$. If, in addition, g is an approximation of f near $\mathbf{a}$, then

$$g(\mathbf{a}) = f(\mathbf{a}) = c + m_1a_1 + m_2a_2 + \cdots + m_na_n. \tag{4}$$

Solving (4) for c, substituting into (3), and regrouping terms, we find

---

[3] Our definition of "approximation" is special to these notes. There is no standard definition of "approximation". In principle, *any* function on $\mathbf{R}^n$ can be considered an approximation of any other function; the only question is how good the approximation is.

[4] The definition of the term "*good* approximation" used in these notes is also not universal; however, many authors use it exactly way we are using it here.

[5] The term "linear function", as we are using it here, does not mean the same thing as "linear transformation", as used in a linear algebra course.

$$\begin{aligned} g(\mathbf{x}) &= f(\mathbf{a}) + m_1(x_1 - a_1) + m_2(x_2 - a_2) + \cdots + m_n(x_n - a_n) \\ &= f(\mathbf{a}) + \mathbf{m}\boldsymbol{\cdot}(\mathbf{x} - \mathbf{a}), \end{aligned} \tag{5}$$

where $\mathbf{m} = (m_1, m_2, \ldots m_n)$ and "$\boldsymbol{\cdot}$" is the usual dot-product on $\mathbf{R}^n$.

Lemma 1.2 *Let f be a real-valued function on $\mathbf{R}^n$ and let $\mathbf{a} \in \mathbf{R}^n$. Then f has* ***at most one*** *good linear approximation near $\mathbf{a}$.*

Proof: If g and h are linear approximations of f near $\mathbf{a}$, then $g(\mathbf{x}) = f(\mathbf{a}) + \mathbf{m}\boldsymbol{\cdot}(\mathbf{x} - \mathbf{a})$ and $h(\mathbf{x}) = f(\mathbf{a}) + \mathbf{w}\boldsymbol{\cdot}(\mathbf{x} - \mathbf{a})$ for some vectors $\mathbf{m}, \mathbf{w} \in \mathbf{R}^n$. If both g and h are *good* approximations of f near $\mathbf{a}$, then

$$\begin{aligned} \lim_{\mathbf{x}\to\mathbf{a}} \frac{h(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} &= \lim_{\mathbf{x}\to\mathbf{a}} \frac{(f(\mathbf{x}) - g(\mathbf{x})) - (f(\mathbf{x}) - h(\mathbf{x}))}{\|\mathbf{x} - \mathbf{a}\|} \\ &= \lim_{\mathbf{x}\to\mathbf{a}} \left(\frac{f(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} - \frac{f(\mathbf{x}) - h(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} \right) \\ &= \lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} - \lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - h(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} \qquad (6) \\ &= 0 - 0 \\ &= 0. \end{aligned}$$

Here, in passing from the second line to the third, we have used the fact that "The limit of a difference exists and is the difference of the limits, **provided that** both of the limits whose difference is being taken exist." In the application above, both limits on the right-hand side of (6) exist *by hypothesis* because we assumed that both g and h are good approximations of f near $\mathbf{a}$, and the definition of "good approximation" says that these limits are zero (hence exist). Therefore if both g and h are good linear approximations of f near $\mathbf{a}$, and the vectors $\mathbf{m}$ and $\mathbf{w}$ are as above, then

$$\begin{aligned} 0 &= \lim_{\mathbf{x}\to\mathbf{a}} \frac{h(\mathbf{x}) - g(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} \\ &= \lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{a}) + \mathbf{w}\boldsymbol{\cdot}(\mathbf{x} - \mathbf{a}) - (f(\mathbf{a}) + \mathbf{m}\boldsymbol{\cdot}(\mathbf{x} - \mathbf{a}))}{\|\mathbf{x} - \mathbf{a}\|} \\ &= \lim_{\mathbf{x}\to\mathbf{a}} \frac{(\mathbf{w} - \mathbf{m})\boldsymbol{\cdot}(\mathbf{x} - \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|} \\ &= \lim_{\mathbf{x}\to\mathbf{a}} (\mathbf{w} - \mathbf{m})\boldsymbol{\cdot}\frac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|}\ . \end{aligned}$$

Restating the result of this calculation in one line:

$$\lim_{\mathbf{x}\to\mathbf{a}} (\mathbf{w} - \mathbf{m})\boldsymbol{\cdot}\frac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|} = 0. \tag{7}$$

Recall that (7) means that if $\mathbf{x}$ approaches $\mathbf{a}$ along *any path whatsoever*, then the limit taken along this path must be 0. If $\mathbf{w}-\mathbf{m}\neq\mathbf{0}$, we can let $\mathbf{x}$ approach $\mathbf{a}$ along the ray (half-line) through $\mathbf{a}$ that extends in the direction of $\mathbf{w}-\mathbf{m}$. For every $\mathbf{x}$ on this ray, other than $\mathbf{a}$ itself, we have

$$\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|}=\frac{\mathbf{w}-\mathbf{m}}{\|\mathbf{w}-\mathbf{m}\|}\ ,$$

so

$$(\mathbf{w}-\mathbf{m})\boldsymbol{\cdot}\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|}=\frac{(\mathbf{w}-\mathbf{m})\boldsymbol{\cdot}(\mathbf{w}-\mathbf{m})}{\|\mathbf{w}-\mathbf{m}\|}=\frac{\|\mathbf{w}-\mathbf{m}\|^2}{\|\mathbf{w}-\mathbf{m}\|}=\|\mathbf{w}-\mathbf{m}\|\ .$$

Therefore, if we approach $\mathbf{a}$ along this straight line from one side of $\mathbf{a}$, the quantity $(\mathbf{w}-\mathbf{m})\boldsymbol{\cdot}\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|}$ is *constant* it is always equal to $\|\mathbf{w}-\mathbf{m}\|$), and its limit is $\|\mathbf{w}-\mathbf{m}\|$, which is not zero. So, under the assumption that $\mathbf{w}-\mathbf{m}\neq\mathbf{0}$, the limit (7) cannot be zero, a contradiction. Hence $\mathbf{w}-\mathbf{m}=\mathbf{0}$, so $\mathbf{w}=\mathbf{m}$ and the functions h and g are identical. Thus, any two good linear approximations of f near $\mathbf{a}$ are identical; i.e. there is at most one such approximation. ■

Now we're ready to define what "differentiable at $\mathbf{a}$" means.

Definition 1.3 A real-valued function f on $\mathbf{R}^n$ is *differentiable at* $\mathbf{a}$ if f has a good linear approximation near $\mathbf{a}$ (i.e. if a good linear approximation of f near $\mathbf{a}$ exists).

So, a corollary of Lemma 1.2 is: if f is differentiable at $\mathbf{a}$, then f has *exactly one* good linear approximation near $\mathbf{a}$; there is *exactly one* vector $\mathbf{m}$ such that the function g defined by $g(\mathbf{x})=f(\mathbf{a})+\mathbf{m}\boldsymbol{\cdot}(\mathbf{x}-\mathbf{a})$ is a good approximation of f near $\mathbf{a}$.

A good linear approximation of f near $\mathbf{a}$ is also called a **linearization** *of f at* $\mathbf{a}$. Since a function f, differentiable at $\mathbf{a}$, has exactly one such approximation near $\mathbf{a}$, we can unambiguously refer to this approximation as **the** linearization of f at $\mathbf{a}$.

The case $n=1$

For $n=1$, "vectors in $\mathbf{R}^n$" and "points in $\mathbf{R}^n$" are just real numbers, so we will dispense with boldface-notation and subscripts.

Suppose that a real-valued function f on $\mathbf{R}$ is differentiable at a, *as defined in Definition 1.3* rather than as in Calculus 1. Then f has a linearization at a: there is a real number m such that the function g defined by $g(x)=f(a)+m(x-a)$ is a good approximation of f near a. According to (2), this means that

$$\lim_{x\to a}\frac{f(x)-g(x)}{|x-a|}=0. \tag{8}$$

But since $\frac{f(x)-g(x)}{|x-a|}=\pm\frac{f(x)-g(x)}{x-a}$ for all $x\neq a$, (8) is equivalent to the statement that

$$\lim_{x\to a}\frac{f(x)-g(x)}{x-a}=0.$$

Therefore

$$\begin{aligned}0 &= \lim_{x\to a}\frac{f(x)-g(x)}{x-a}\\ &= \lim_{x\to a}\frac{f(x)-f(a)-m(x-a)}{x-a}\\ &= \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a}-m\right). \qquad (9)\end{aligned}$$

Note that $\frac{f(x)-f(a)}{x-a}=\left(\frac{f(x)-f(a)}{x-a}-m\right)+m$. Since the limit on the right-hand side of (9) exists, and so does the limit as $x\to a$ of m (a constant function of x),

$$\begin{aligned}\lim_{x\to a}\frac{f(x)-f(a)}{x-a} &= \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a}-m\right)+\lim_{x\to a}m\\ &= 0+m\\ &= m. \qquad (10)\end{aligned}$$

In other words, the limit in (1) exists—so f is differentiable at a, *as this terminology is defined in Calculus 1*—and $f'(a)=m$. The linearization g of f at a can therefore be rewritten as

$$g(x)=f(a)+f'(a)(x-a). \qquad (11)$$

Note that the graph of g—the graph of the equation $y=f(a)+f'(a)(x-a)$—is exactly *the straight line tangent to the graph of $y=f(x)$ at $(a,f(a))$*. This is the source of the terminology "linear approximation" and "linearization". (Although we use the same terminology for functions of more than one variable, only for functions of a single variable is the graph of the linearization a *line*.) Our "good linear approximation" is exactly the tangent-line approximation.

Conversely, let us assume from the start that our function f is differentiable at a as defined in Calculus 1, rather than as defined in Definition 1.3. Define $g(x)=f(a)+f'(a)(x-a)$. Then g is a linear approximation of x near a, but is the approximation *good*? Well, we have

$$\begin{aligned}\lim_{x\to a}\frac{f(x)-g(x)}{x-a} &= \lim_{x\to a}\frac{f(x)-f(a)-f'(a)(x-a)}{x-a}\\ &= \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a}-f'(a)\right) \qquad (12)\\ &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}-\lim_{x\to a}f'(a) \qquad (13)\\ &= f'(a)-f'(a)\\ &= 0.\end{aligned}$$

(In passing from (12) to (13), we have again used the fact that "The limit of a difference exists and is the difference of the limits, **provided that** both of the limits whose difference is being taken exist.") So $\lim_{x\to a}\frac{f(x)-g(x)}{x-a}=0$, implying that $\lim_{x\to a}\frac{f(x)-g(x)}{|x-a|}=0$—which is exactly the $n=1$ case of the definition of "good approximation". Therefore f is differentiable at a, *as this terminology is defined in Definition 1.3.*

Thus, the two definitions of "differentiable at a point" are entirely equivalent for functions of a single variable (i.e., in the case $n=1$).

The case $n = 2$

For this case we will write (x, y) instead of (x_1, x_2), and (a, b) instead of (a_1, a_2).

Suppose that f is differentiable at (a, b). Then f has a good linear approximation g near (a, b), and from equation (5), there are numbers m_1, m_2 such that

$$g(x,y) = f(a,b) + m_1(x-a) + m_2(y-b).$$

Because g is a *good* approximation of f near (a, b), we have

$$\lim_{(x,y)\to(a,b)} \frac{f(x,y) - f(a,b) - m_1(x-a) - m_2(y-b)}{\sqrt{(x-a)^2+(y-b)^2}} = 0. \tag{14}$$

Recall that (14) means that if (x, y) approaches (a, b) along *any path whatsoever*, then the limit taken along this path must be 0. We consider two special straight-line paths, one parallel to the x-axis and one parallel to the y-axis.

On the path parallel to the x-axis, we have $y \equiv b$ and $x \neq a$. The expression whose limit is taken in (14) then simplifies to $\frac{f(x,b)-f(a,b)-m_1(x-a)}{|x-a|}$, and, on this path, the condition "$(x, y) \to (a, b)$" simplifies to "$x \to a$". Therefore equation (14) implies that

$$\lim_{x\to a} \frac{f(x,b) - f(a,b) - m_1(x-a)}{|x-a|} = 0,$$

which in turn implies that, if we erase the absolute-value symbols in the denominator, the limit is still zero:

$$\lim_{x\to a} \frac{f(x,b) - f(a,b) - m_1(x-a)}{x-a} = 0.$$

Then, by the same reasoning that in the $n = 1$ case led us to (9), and then to (10), we find that

$$\lim_{x\to a} \frac{f(x,b) - f(a,b)}{x-a} = m_1. \tag{15}$$

But the left-hand side of (15) is exactly the definition of $\frac{\partial f}{\partial x}(a,b)$ (also denoted $f_x(a,b)$). Therefore we conclude that this partial derivative exists, and has the value m_1.

Completely analogously, by approaching (a, b) along the straight line parallel to the y-axis, we conclude $\frac{\partial f}{\partial y}(a,b)$ (also denoted $f_y(a,b)$) exists, and has the value m_2. Thus

$$(m_1, m_2) = (f_x(a,b), f_y(a,b)) \tag{16}$$

Definition 1.4 Let f be a real-valued function on $\mathbf{R}^2$ that is differentable at the point (a,b). The *gradient* of f at (a,b), denoted $(\text{grad } f)|_{(a,b)}$ or $\nabla f|_{(a,b)}$, is defined to be the vector $(f_x(a,b), f_y(a,b))$. [6]

Thus if f is differentiable at $\mathbf{a} = (a,b)$, the linearization g of f at $\mathbf{a}$ can be written as

$$g(\mathbf{x}) = f(\mathbf{a}) + (\text{grad } f)|_{\mathbf{a}} \cdot (\mathbf{x} - \mathbf{a}), \tag{17}$$

or equivalently as

$$g(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b). \tag{18}$$

Thus, using variables x, y, z on $\mathbf{R}^3$, the graph of g is the graph of the equation

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b). \tag{19}$$

This graph is a plane—specifically, the plane tangent to the graph of $z = f(x,y)$ at the point $(a,b,f(a,b))$ (this is the *definition* of "tangent plane"). The good linear approximation of f near (a,b) is exactly the *tangent-plane approximation* defined in any Calculus 3 textbook. Geometrically, the tangent plane can be characterized as the best planar approximation to the (generally curved) surface $z = f(x,y)$ near the point $(a,b,f(a,b))$; differentiability of f at (a,b) can be characterized geometrically as the condition that this surface *has* a good planar approximation near (a,b).

The case $n > 2$

The case $n > 2$ is essentially no different from the case $n = 2$; there are just more variables, making our formulas longer unless we use vector notation. The same argument used in the $n = 2$ case shows that, if f is differentiable at $\mathbf{a}$, then all the first partial derivatives of f exist at $\mathbf{a}$. Therefore we can define the *gradient* of f at $\mathbf{a}$ completely analogously to the dimension-two definition:

---

[6] **Note to instructors.** Care should be taken not to give students the impression that the gradient is defined as long as $f_x(a,b)$ and $f_y(a,b)$ exist. While that's all that's needed to define the vector $f_x(a,b)\mathbf{i} + f_y(a,b)\mathbf{j}$, this vector is not called the *gradient* unless f is differentiable at (a,b). The vector $f_x(a,b)\mathbf{i} + f_y(a,b)\mathbf{j}$ has no useful properties unless all directional derivatives of f at (a,b) exist, and the generalized-directional-derivative map $\mathbf{u} \mapsto (D_{\mathbf{u}}f)(a,b) := \frac{d}{dt}f((a,b)+t\mathbf{u})|_{t=0}$ is linear in $\mathbf{u}$. It would be unwise to call a vector "the gradient of f at (a,b)" if its existence or value varied under rotation of the coordinate axes. This is why, at a higher level of sophistication, the gradient is defined in a coordinate-independent way determined solely by an inner product—*not* by equation (20)—and only at points of differentiability: if $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable at $\mathbf{a}$, then $(\text{grad } f)|_{\mathbf{a}}$ is the unique vector $\mathbf{m} \in \mathbf{R}^n$ such that $df|_{\mathbf{a}}(\mathbf{u}) = \mathbf{m} \cdot \mathbf{u}$ for all $\mathbf{u} \in \mathbf{R}^n$. (Here the differential $df|_{\mathbf{a}}$ is viewed as a linear map $\mathbf{R}^n \to \mathbf{R}$.) Equivalently, if $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable at $\mathbf{a}$, then $(\text{grad } f)|_{\mathbf{a}}$ is the unique vector $\mathbf{m} \in \mathbf{R}^n$ such that all the directional derivatives of f at $\mathbf{a}$ are given by $D_{\mathbf{u}}f(\mathbf{a}) = \mathbf{m} \cdot \mathbf{u}$.

For similar reasons, one should take care not to define "*the* linearization of f at $\mathbf{a}$" (or "*the* tangent plane approximation at $\mathbf{a}$", in dimension two) if it is not stated that f is differentiable at $\mathbf{a}$, even if all the first partials of f at $\mathbf{a}$ exist. In the non-differentiable case, there may infinitely many such "linearizations", none of them a good approximation to f near $\mathbf{a}$; or there may be a unique such "linearization" that still is not a good approximation. "Pseudo-linearization" would be a better term in these cases; these approximations are not true linearizations.

$$(\text{grad } f)|_{\mathbf{a}} = (f_{x_1}(\mathbf{a}), f_{x_2}(\mathbf{a}), \ldots, f_{x_n}(\mathbf{a})) \tag{20}$$

(where, of course, f_{x_i} means $\frac{\partial f}{\partial x_i}$). If f is differentiable at $\mathbf{a}$, then the linearization of f at $\mathbf{a}$—i.e. the good linear approximation of f near $\mathbf{a}$—is given by (17), regardless of the dimension n. For $n = 3$, choosing non-subscripted names for coordinates, we can write (17) as

$$g(x, y, z) = f(a, b, c) + f_x(a, b, c)(x - a) + f_y(a, b, c)(y - b) + f_z(a, b, c)(z - c). \tag{21}$$

For $n \geq 3$ there *is* a slight difference in terminology, and a greater difference in visualizability. The graph of an n-variable real-valued function lives in $\mathbf{R}^{n+1}$ (n variables for the domain, plus one for the range). Already for $n = 3$, we cannot graph functions as meaningfully as we can for $n \leq 2$. The graph of the function g in (21) is a three-dimensional set in $\mathbf{R}^4$, so we do not use the term "tangent plane" for it (by definition, planes are two-dimensional). Similarly, we do not have terminology analagous to "tangent-line approximation" and "tangent-plane approximate" for $n \geq 3$. Instead, we usually just use the term "linear approximation" or "linearization". However, the graph of (21) or, more generally, (17), is still an "n-dimensional flat object in $\mathbf{R}^{n+1}$," and it is the best "n-dimensional flat-object approximation" of the graph of f (which is usually a curved n-dimensional object) near the point $(\mathbf{a}, f(\mathbf{a})) = (a_1, \ldots, a_n, f(a_1, \ldots, a_n))$ in $\mathbf{R}^{n+1}$.

You may have noticed that, while we have given a definition of the terminology "f is differentiable at $\mathbf{a}$" (for a real-valued function f on $\mathbf{R}^n$) that applies for all n, have talked about *partial* derivatives for $n > 1$, and you have learned (or will soon learn) about *directional* derivatives for $n > 1$, we have *not* defined the terminology "the derivative of a differentiable function f at $\mathbf{a}$" for $n > 1$. This omission is intentional! A definition of "the derivative of f at $\mathbf{a}$" that applies for all n is best postponed until you have had (or are taking) a course in linear algebra.[7]

2 Differentiability on a set

To avoid the distracting the student with longer hypotheses, so far we have assumed our functions f to be defined on all of $\mathbf{R}^n$. However, so large a domain of f is not needed for the concept of "differentiability at a point $\mathbf{a}$" to be meaningful. All that is needed is that f be defined on a *neighborhood of* $\mathbf{a}$, defined below.

A set U in $\mathbf{R}^n$ is called *open* if for every point $\mathbf{a}$ of U, there is some $r > 0$ such that every point of $\mathbf{R}^n$ within a distance r of $\mathbf{a}$ is contained in U. To say this more succinctly, for $\mathbf{a} \in \mathbf{R}^n$ and $r > 0$, we define the *open ball of radius r centered at* $\mathbf{a}$, denoted $B_r(\mathbf{a})$, to be the set of points in $\mathbf{R}^n$ a distance less than r from $\mathbf{a}$. Then the definition of "open set" above simply says: a set U in $\mathbf{R}^n$ is open if for every $\mathbf{a} \in U$, there is some $r > 0$ such

---

[7] **Note to instructors.** There is a natural temptation to tell students that for $n > 1$, $(\text{grad } f)|_{\mathbf{a}}$ is the generalization of "the derivative of f at $\mathbf{a}$". For students who take no calculus beyond Calculus 3, this may do no harm. But students who go on to take advanced differential calculus, as presented in Dieudonné's *Foundations of Modern Analysis* and many less ambitious texts, will have more to unlearn.

that $B_r(\mathbf{a})$ is entirely contained in U. It is not hard to show that "the open ball of radius r centered at $\mathbf{a}$" is itself an open set, so the terminology "open ball" is not misleading (an open ball is a ball that is open). In this definition, *it does not matter how small* r *is*, as long as r is positive. Typically, mathematicians use the Greek letter ϵ or δ, rather than r, for a positive quantity that may potentially be very small.

A *neighbhorhood* of a point $\mathbf{a}$ is simply an open set that contains $\mathbf{a}$. A statement of the form "X is true on a neighborhood of $\mathbf{a}$" means that there is *some* neighborhood of $\mathbf{a}$ at every point of which X is true. This is equivalent to saying that there is some $r > 0$ (we don't care how small) such that X is true at every point $\mathbf{q}$ a distance less than r from $\mathbf{a}$.

All that is needed for the concept of "good approximation of f near $\mathbf{a}$" to make sense is that f be defined on a neighborhood of $\mathbf{a}$. We generalize Definition 1.1 and Lemma 1.2 by replacing their first sentences with "Let $\mathbf{a} \in \mathbf{R}^n$ and let f be a real-valued function defined on a neighborhood of $\mathbf{a}$." Similarly, we extend Definition 1.3 by starting it with, "A real-valued function f defined on a neighborhood of $\mathbf{a}$ in $\mathbf{R}^n$ is differentiable if ...."

These generalizations allow us to define what "differentiable on a set" means for certain sets U in $\mathbf{R}^n$. What we intend "differentiable on U" to mean is that f is differentiable at every point of U—but for this to make sense, f must be defined on a neighbhorhood of each point of U. This is no problem if U is an open set contained in the domain of f. So, we make the following definition:

Definition 2.1 Let U be an open subset of $\mathbf{R}^n$, and let f be a real-valued function defined on U. We say that f is *differentiable on* U if f is differentiable at $\mathbf{a}$ for every point $\mathbf{a}$ in U.

3 Properties related to differentiability; tests for differentiability

We saw in Section 1, when examining the case $n = 2$, that differentiability of f at (a, b) implies the existence of the first partials $f_x(a, b)$, $f_y(a, b)$. **The converse is false.** Existence of the first partials at (a, b) does *not* imply differentiability of f at (a, b). **Differentiability at a point is a *stronger* condition than existence of the first partials at that point.**

Furthermore, in our derivation of the fact that "differentiability of f at (a, b) implies the existence of $f_x(a, b)$ and $f_y(a, b)$," had we approached (a, b) along a straight line in an arbitrary direction rather than only lines that were parallel to the coordinate axes, we would have found that *differentiability of* f *at* (a, b) *implies the existence of every directional derivative of* f *at* (a, b). The converse of this implication is also false: even if the directional derivative of f at (a, b) exists for every direction, f need not be differentiable at (a, b). **Differentiability at a point is a *stronger* condition than existence of all directional derivatives at that point.** This is true for all $n > 1$.

Returning to the general-n case, if all n first partial derivatives of f not only *exist* at a point $\mathbf{a}$, but are continuous there, then it can be shown that f is differentiable at

a. A seemingly logical name for this continuity condition on the first partials would be "continuous partial differentiability at **a**"; alternatively, we could say that f is continuously partial-differentiable at **a**. However, because "continuous partial differentiability of f at **a**" implies "differentiability of f at **a**"—no "partial"—nobody uses the long, awkward terms "continuous partial differentiability" or "continuously partial-differentiable". Instead, we simply use the terms "continuous differentiability" and "continuously differentiable." Note that "continuously differentiable" would be poor terminology were it not for the fact that "continuous partial-differentiability" implies differentiability: by standard rules of English, it would be misleading to call a function "continuously differentiable" if it isn't differentiable.

"Continuous differentiability at **a** implies differentiability at **a**" is another non-reversible implication, and the same is true for continuous differentiability and differentiability on an open set U. *Continuous differentiability at a point* **a***, or on an open set* U*, is a* **stronger** *condition than differentiability at* **a***, or on* U*.*

There are several properties you've learned that are related to differentiability at a point **a**. Some of them *imply* differentiability at **a**, and some of them are *implied by* differentiability at **a**. But, for $n > 1$, none of them is *equivalent* to differentiability at **a**. Using the mathematician's notation "A ⇒ B" for "A implies B," we give the implication-relations among many of these properties in the diagram below.

$$\begin{tikzcd}
\left\{\begin{array}{c} f\text{ is}\\ \text{continuously}\\ \text{differentiable}\\ \text{on a}\\ \text{neighborhood}\\ \text{of }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \left\{\begin{array}{c} f\text{ is}\\ \text{differentiable}\\ \text{on a}\\ \text{neighborhood}\\ \text{of }\mathbf{a}\end{array}\right\} \arrow[d, Rightarrow] & & & & \\
\left\{\begin{array}{c} f\text{ is}\\ \text{continuously}\\ \text{differentiable}\\ \text{at }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} f\text{ is}\\ \text{differentiable}\\ \text{at }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \left\{\begin{array}{c} \text{(i) the}\\ \text{directional}\\ \text{derivative}\\ D_{\mathbf{u}}f(\mathbf{a})\\ \text{exists for}\\ \text{every unit}\\ \text{vector}\\ \mathbf{u}\in\mathbf{R}^n,\\ \text{and (ii) there}\\ \text{is a vector}\\ \mathbf{m}\in\mathbf{R}^n\\ \text{such that}\\ D_{\mathbf{u}}f(\mathbf{a})\\ =\mathbf{m}\cdot\mathbf{u}\\ \text{for all }\mathbf{u},\\ \text{and}\\ \text{(iii) } f\text{ is}\\ \text{continuous}\\ \text{at }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} \text{(i) the}\\ \text{directional}\\ \text{derivative}\\ D_{\mathbf{u}}f(\mathbf{a})\\ \text{exists for}\\ \text{every unit}\\ \text{vector}\\ \mathbf{u}\in\mathbf{R}^n,\\ \text{and (ii) there}\\ \text{is a vector}\\ \mathbf{m}\in\mathbf{R}^n\\ \text{such that}\\ D_{\mathbf{u}}f(\mathbf{a})\\ =\mathbf{m}\cdot\mathbf{u}\\ \text{for all }\mathbf{u}\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} \text{the}\\ \text{directional}\\ \text{derivative}\\ D_{\mathbf{u}}f(\mathbf{a})\\ \text{exists for}\\ \text{every unit}\\ \text{vector}\\ \mathbf{u}\in\mathbf{R}^n\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} \text{all the}\\ \text{first}\\ \text{partials}\\ \frac{\partial f}{\partial x_i}(\mathbf{a})\\ \text{exist}\end{array}\right\} \\
& \left\{\begin{array}{c} f\text{ is}\\ \text{continuous}\\ \text{at }\mathbf{a}\end{array}\right\} & & & &
\end{tikzcd}$$

For $n > 1$, *not a single one of these implications is reversible.* (For $n = 1$, all the arrows to the right of "f is differentiable at $\mathbf{a}$" are reversible, but none of the others are.) Furthermore, the only condition, or set of conditions, to the right of "f is differentiable at $\mathbf{a}$" in this diagram, that implies "f is continuous at $\mathbf{a}$", is the one directly to the right (where the continuity implication is trivial, since continuity is explicitly assumed). In other words, for multivariable functions, differentiability at a point implies continuity at that point, just as for single-variable functions. But, among the conditions considered above (or any conditions known to the author of these notes), *no differentiability-related condition weaker than differentiability at* $\mathbf{a}$*, implies continuity at* $\mathbf{a}$, other than by explicitly assuming continuity to begin with.

An additional true, non-reversible, implication not shown in the diagram is "$\{f$ is continuously differentiable on a neighborhood of $\mathbf{a}\} \Rightarrow \{f$ is continuous on a neighborhood of $\mathbf{a}\}$." We can display this relation in another diagram if we omit "$\{f$ is continuously differentiable at $\mathbf{a}\}$". We do this in the diagram below, most of which is identical to the previous diagram.

$$\begin{tikzcd}
\left\{\begin{array}{c} f\text{ is}\\ \text{continuously}\\ \text{differentiable}\\ \text{on a}\\ \text{neighborhood}\\ \text{of }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] \arrow[dd, Rightarrow] & \left\{\begin{array}{c} f\text{ is}\\ \text{differentiable}\\ \text{on a}\\ \text{neighborhood}\\ \text{of }\mathbf{a}\end{array}\right\} \arrow[d, Rightarrow] & & & & \\
& \left\{\begin{array}{c} f\text{ is}\\ \text{differentiable}\\ \text{at }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \left\{\begin{array}{c} \text{(i) the}\\ \text{directional}\\ \text{derivative}\\ D_{\mathbf{u}}f(\mathbf{a})\\ \text{exists for}\\ \text{every unit}\\ \text{vector}\\ \mathbf{u}\in\mathbf{R}^n,\\ \text{and (ii) there}\\ \text{is a vector}\\ \mathbf{m}\in\mathbf{R}^n\\ \text{such that}\\ D_{\mathbf{u}}f(\mathbf{a})\\ =\mathbf{m}\cdot\mathbf{u}\\ \text{for all }\mathbf{u},\\ \text{and}\\ \text{(iii) }f\text{ is}\\ \text{continuous}\\ \text{at }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} \text{(i) the}\\ \text{directional}\\ \text{derivative}\\ D_{\mathbf{u}}f(\mathbf{a})\\ \text{exists for}\\ \text{every unit}\\ \text{vector}\\ \mathbf{u}\in\mathbf{R}^n,\\ \text{and (ii) there}\\ \text{is a vector}\\ \mathbf{m}\in\mathbf{R}^n\\ \text{such that}\\ D_{\mathbf{u}}f(\mathbf{a})\\ =\mathbf{m}\cdot\mathbf{u}\\ \text{for all }\mathbf{u}\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} \text{the}\\ \text{directional}\\ \text{derivative}\\ D_{\mathbf{u}}f(\mathbf{a})\\ \text{exists for}\\ \text{every unit}\\ \text{vector}\\ \mathbf{u}\in\mathbf{R}^n\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} \text{all the}\\ \text{first}\\ \text{partials}\\ \frac{\partial f}{\partial x_i}(\mathbf{a})\\ \text{exist}\end{array}\right\} \\
\left\{\begin{array}{c} f\text{ is}\\ \text{continuous}\\ \text{on a}\\ \text{neighborhood}\\ \text{of }\mathbf{a}\end{array}\right\} \arrow[r, Rightarrow] & \left\{\begin{array}{c} f\text{ is}\\ \text{continuous}\\ \text{at }\mathbf{a}\end{array}\right\} & & & &
\end{tikzcd}$$

How can we use the properties in these diagrams to test for differentiability? As a practical matter, the most useful tests are the "continuity test" and the "continuous first partials" test, and they're usually all that's needed. For simplicity, we will state these tests as they apply to differentiability at a point; we leave the student to figure out the corresponding tests for differentiability on an open set (which may be all of $\mathbf{R}^n$). Below, we assume that f is defined on a neighborhood of a point $\mathbf{a}$ in $\mathbf{R}^n$.

> **Continuity test.** Is f continuous at $\mathbf{a}$? If "no", then f is not differentiable at $\mathbf{a}$. If "yes", the test is inconclusive; try the "continuous first partials" test.
>
> **Continuous first partials test.**
>
> 1. Do all n first partial derivatives of f exist at $\mathbf{a}$? If "no", then f is not differentiable at $\mathbf{a}$. If "yes", go to Step 2.
> 2. Are all n first partial derivatives of f continuous at $\mathbf{a}$? If "yes", then f is differentiable at $\mathbf{a}$. If "no", the test is inconclusive.

Fortunately, the "continuous first partials" test for differentiability is almost always conclusive, in the sense that for *most* functions f we commonly work with, at each point $\mathbf{a}$ in the domain of f, either some first partial of f doesn't exist at $\mathbf{a}$ (implying f is *not* differentiable at a), or all first partials exist *and* are continuous at $\mathbf{a}$ (implying f *is* differentiable at $\mathbf{a}$). Functions that are differentiable at a point, or on a set, but are not *continuously* differentiable there, are rarely encountered—except in examples designed specifically to show that differentiability (either at a point or on an open set) does not guarantee continuous differentiability.

In the relatively rare cases that the "continuous first partials" test is inconclusive, you have more work to do, but don't lose hope. The existence of all the first partial derivatives of f at $\mathbf{a}$ singles out a *candidate* for the linearization of f at $\mathbf{a}$. In the case $n = 2$, $\underline{\text{if}}$ f is differentiable at (a, b), $\underline{\text{then}}$ the linearization of f at (a, b) *must be* the function g in equation (18). For general n, $\underline{\text{if}}$ f is differentiable at $\mathbf{a}$, $\underline{\text{then}}$ the linearization of f at $\mathbf{a}$ *must be* the function g_{cand} defined by

$$g_{\text{cand}}(\mathbf{x}) = f(\mathbf{a}) + f_{x_1}(\mathbf{a})(x_1 - a_1) + f_{x_2}(\mathbf{a})(x_2 - a_2) + \cdots + f_{x_n}(\mathbf{a})(x_n - a_n). \tag{22}$$

We can therefore test for differentiability at $\mathbf{a}$ by examining

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - g_{\text{cand}}(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} \quad . \tag{23}$$

If this limit is zero, then f is differentiable at $\mathbf{a}$. If this limit fails to exist, or exists but is nonzero, then f is not differentiable at $\mathbf{a}$. As a practical matter, since it can be tricky to determine the existence or value of a multivariable limit of "0/0" type such as the one above (there is no multivariable l'Hôpital's Rule!), this final test for differentiability can take some time to carry out. Unfortunately, there are no *completely general*, practical tools for such limits. But fortunately, for many garden-variety functions f, such as *homogeneous rational functions* (polynomials divided by polynomials, where all terms in the numerator have the same degree, and all terms in the denominator have the same degree), if we reach the stage of having to analyze the limit in (23) there are some standard tools that the student has learned (not addressed in these notes) that often lead to a definitive answer.

A remark for the curious student. You may be wondering why we did not write equation (22) more simply as equation (17). The reason is that equation (20), despite looking convenient and understandable whenever all the first partial derivatives $f_{x_i}(\mathbf{a})$ exist, is not a correct definition if f is not differentiable at $\mathbf{a}$. The gradient of f at $\mathbf{a}$ is *not defined* unless f is differentiable at $\mathbf{a}$, *even if all first partial derivatives of f exist at* $\mathbf{a}$. The reason for this is given in a footnote earlier in these notes (footnote 6) addressed to instructors. Understanding the explanation requires more mathematical experience than most Calculus 3 students have, but if you have understood everything else in these notes, and are interested enough to ask, "Why do these notes say that existence of all the first partials isn't enough to define the gradient?" your instructor may be willing to explain the answer.