ARE201-Simon, Fall2015

CALCULUS2: THU, SEP 10, 2015 PRINTED: SEPTEMBER 9, 2015 (LEC# 5)

Contents

2. Univariate and Multivariate Differentiation (cont)

Key Points:

(1) Defn of a directional derivative: $f_h(\mathbf{x}) = \lim_{|k|\to\infty} \frac{\left(f(\mathbf{x}_0+\mathbf{h}/k)-f(\mathbf{x}_0)\right)}{||\mathbf{h}||/k}$ (equation (1) on p. 3).

(2) Defn of a differentiable function: $f : \mathbb{R}^n \to \mathbb{R}$ is *differentiable* at $\mathbf{x}_0$ if $\bigtriangledown f(\mathbf{x}_0)$ exists and if for all $\mathbf{h} \in \mathbb{R}^n$,

$$\lim_{|k|\to\infty} \frac{\left(f(\mathbf{x}_0+\mathbf{h}/k)-f(\mathbf{x}_0)\right) - \bigtriangledown f(\mathbf{x}_0)\cdot\mathbf{h}/k}{||\mathbf{h}||/k} = 0 \qquad \text{(equation (3) on p. 4)}$$

or equivalently

$$\lim_{|k|\to\infty} \frac{\left(f(\mathbf{x}_0+\mathbf{h}/k)-f(\mathbf{x}_0)\right)}{||\mathbf{h}||/k} = \bigtriangledown f(\mathbf{x}_0)\cdot\frac{\mathbf{h}}{||\mathbf{h}||} \qquad \text{(equation (4) on p. 4).}$$

(3) A sufficient condition for differentiability of a function is that each of its partial derivatives is a continuous function.

(4) Relationship between directional derivatives and total derivatives.

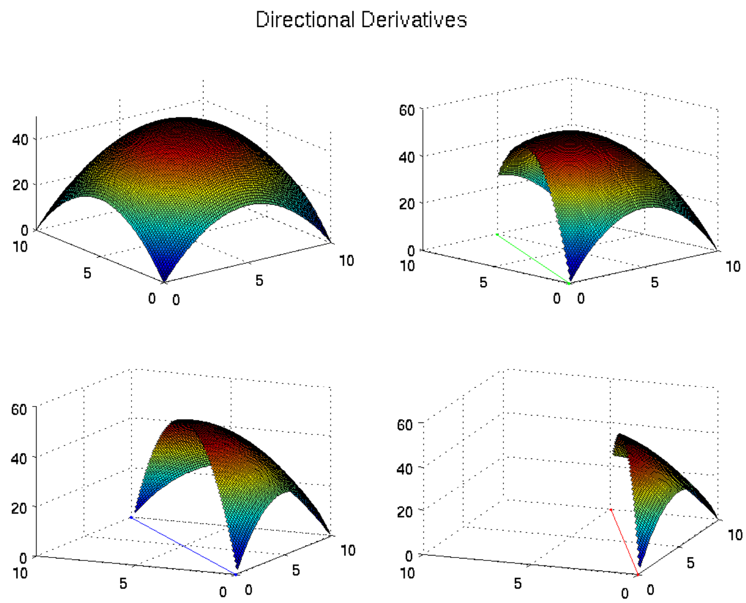

FIGURE 1. Multiple slices of the "cake" in the top left panel

2.3. **Partial, Directional and Total derivatives**

2.3.1. *Directional derivatives.* Suppose we have a function $f : \mathbb{R}^n \to \mathbb{R}$ and we want to know how it behaves as you move out in some direction starting from some given point in the domain $\mathbf{x}^0$. In particular, want to know the slope of f in this direction.

Example: The directional derivative of f in the direction e_i—i.e., the vector whose i'th component is 1, and others are zeros—is just the i'th partial derivative of f.

Intuitively, imagine the graph of f is in fact a cake. Now "cut" the cake in the direction we're interested in, so that the cut passes through $\mathbf{x}^0$. Look at the cross-section of the cake you've obtained. It's just like the graph of a function from $\mathbb{R}$ to $\mathbb{R}$, and, if the graph is smooth, it has a slope. That slope is going to be the directional derivative in the direction you've selected. Fig. 1 illustrates: consider the edge of each "slice" to be the graph of a one-dimensional function. Having said that, it's not so straightforward to compute it. For computational purposes, it's useful to extrapolate from the following facts:

- if $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then the slope of f at x equal to the differential of f at x, evaluated at 1, i.e., $f'(x) = df^x(1)$.
- if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then the i'th partial derivative of f at $\mathbf{x}$ is equal to the differential of f at $\mathbf{x}$, evaluated at $\mathbf{e}^i$, the i'th unit vector, i.e., $e^i_k = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}$, i.e., $f_i(x) = df^x(\mathbf{e}^i)$.
- by analogy if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then the directional derivative of f at $\mathbf{x}$ in an arbitrary direction h had better be equal to the differential of f at $\mathbf{x}$, evaluated at the (unique) *unit length* vector that points in the direction h. If this weren't true, then partial derivatives would not be special cases of directional derivatives!

This observation motivates why there's an $||\mathbf{h}||$ in the denominator of the definition below: note that if the h in the numerator is of unit length, then that term disappears, if not, the norm makes the necessary adjustment.

Definition: Given $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{h} \in \mathbb{R}^n$, the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{h}$ is given by[1]

$$\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0 + \mathbf{h}/k) - f(\mathbf{x}_0)\big)}{||\mathbf{h}||/k} \tag{1}$$

Notice that the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{h}$ has the same magnitude but the opposite sign from the directional derivative of f at $\mathbf{x}_0$ in the direction $-\mathbf{h}$.

The notation in (1) is shorthand for the following idea: there exists some scalar $c \in \mathbb{R}$ such that for *any* sequence, $\{k(n)\}$ with the property that $|k(n)|$ increases without bound with n, the expression in (1) must converge to c. To determine whether a directional derivative exists, it is instructive *but not sufficient* to consider whether the limit (1) exists for the particular sequence $k(n) = n$. You also have to worry about sequences that approach $\mathbf{x}_0$ from the opposite direction, e.g., $k(n) = -n$.

Almost everybody freaks out because they don't see what the $||\mathbf{h}||$ is doing in the denominator. It's easy to see why it's there if you compare this definition to that of a partial derivative, i.e.,

$$\frac{df(\mathbf{x})}{dx_1} = \lim_{|k|\to\infty} \frac{\big(f(x_1 + 1/k,\ x_2, ..., x_n) - f(\mathbf{x})\big)}{1/k} \tag{2}$$

This is just the familiar: take the limit of rises over runs. Notice in both numerator and denominator, however, you are working with $1/k$. Compare this to the definition of directional derivative: in the numerator, you have an $\mathbf{h}$ *of unspecified length*. So you have to compensate for this in the denominator. An alternative definition of directional derivative would have been to start with a unit length vector $\mathbf{h}$ in the numerator, then in the denominator you could have written $1/k$, (which would of course been in this case the same as $||\mathbf{h}||/k$).

Here's a computed example of why, if you *didn't* divide by $||\mathbf{h}||$ you would get the wrong answer. Let $f = x_1x_2$, $\mathbf{x}_0 = (1, 1)$, First, let's move in the direction $\mathbf{h} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, i.e., $\mathbf{h}$ is already a *unit length* vector, so that in the denominator $||\mathbf{h}||/k = 1/k$. We have

$$\begin{aligned}
\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0 + \mathbf{h}/k) - f(\mathbf{x}_0)\big)}{1/k} &= \lim_{|k|\to\infty} \frac{\left(\left(1 + \frac{1}{k\sqrt{2}}\right)^2 - 1^2\right)}{1/k} \\
&= \lim_{|k|\to\infty} \frac{\left(1 + \frac{\sqrt{2}}{k} + \frac{1}{2k^2} - 1\right)}{1/k} = \lim_{|k|\to\infty} (\sqrt{2} + 0.5/k) = \sqrt{2}
\end{aligned}$$

[1] An alternative defn is $\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0 + \mathbf{h}/(k||\mathbf{h}||) - f(\mathbf{x}_0)\big)}{1/k}$. The difference is that in this alternative defn, the the norm of $\mathbf{h}$ is moved into the numerator. For a proof that the two definitions are equivalent, see /home/simon/teaching/classes/mathLectures/AltDirectionalDerivDefn.tex

This calculation is done in the correct way, so $\sqrt{2}$ is the correct answer for the derivative of f at $\mathbf{x}_0$ in the direction h. Next we're going to take $\boldsymbol{\eta}$ to be the *non*-unit vector $(1,1)$ pointing in the same direction as $\mathbf{h}$, but we're going to divide by $1/k$ instead of $||\boldsymbol{\eta}||/k$. I emphasize AGAIN this this is the WRONG thing to do, I just am doing it to convince you of why you have to divide by the norm of the η vector.

$$\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0+\boldsymbol{\eta}/k)-f(\mathbf{x}_0)\big)}{1/k} = \lim_{|k|\to\infty} \frac{\Big(\big(1+\frac{1}{k}\big)^2-1^2\Big)}{1/k}$$
$$= \lim_{|k|\to\infty} \frac{\big(1+\frac{2}{k}+\frac{1}{k^2}-1\big)}{1/k} = \lim_{|k|\to\infty}(2+1/k) = 2$$

I.e., I get the WRONG answer, by a factor of $\frac{\sqrt{2}}{2}$, which is exactly the norm of $\boldsymbol{\eta}$. The point is that if you don't have the right denominator throughout the sequence, you get a sequence that's too big—in our example this sequence is $\{2+1/k\}$ instead of $\{\sqrt{2}+0.5/k\}$, and hence converges to something that's too big, i.e., 2 instead of $\sqrt{2}$.

2.3.2. *Computing Directional Derivatives from Partial Derivatives.* Computing directional derivatives can be a huge pain. It turns out, however, that provided the function f is *differentiable*, you can infer any directional derivative just by knowing the partial derivatives of f. Specifically:

Definition: a function $f:\mathbb{R}^n\to\mathbb{R}$ is said to be *differentiable* at $\mathbf{x}_0$ if $\bigtriangledown f(\mathbf{x}_0)$ exists and if for all $\mathbf{h}\in\mathbb{R}^n$,

$$\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0+\mathbf{h}/k)-f(\mathbf{x}_0)\big) - \bigtriangledown f(\mathbf{x}_0)\cdot\mathbf{h}/k}{||\mathbf{h}||/k} = 0 \tag{3}$$

Note that $(\mathbf{h}/k)_{k=1}^{\infty}$ is a sequence of vectors, all pointing in the same direction, whose lengths shrink to zero. Also, the sequence of k's can change sign, just as in the primitive definition of the derivative of a function defined on $\mathbb{R}$.

This definition says, literally, that a function is differentiable at $\mathbf{x}_0$ if you can put a flat board up against the graph of the function, so that it touches the graph vertically above/below $\mathbf{x}_0$, *and if nearby*, the board is very close to the graph (cf what happens when you put a board on top of a Hershey Kiss, or to the side of a pin-wheel). Fig. 2 illustrates how you put a flat board against a graph. Note however, that the relationship between the graph and the board may be less pretty than in this figure: the board could be above the graph on one side of $\mathbf{x}_0$ and below it on the other side.[2] The important thing is that the board and the graph are *very close* in a neighborhood of $\mathbf{x}_0$. To get a feel for what it means for the board to be close to the graph, try putting a board near the graph in Fig. 3 below, vertically above zero. You just can't do it.

Since the second term on the left hand side of (3), i.e., $\frac{\bigtriangledown f(\mathbf{x}_0)\cdot\mathbf{h}/k}{||\mathbf{h}||/k}$, clearly does not depend on k, we can rewrite definition (3) as:

$$\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0+\mathbf{h}/k)-f(\mathbf{x}_0)\big)}{||\mathbf{h}||/k} = \bigtriangledown f(\mathbf{x}_0)\cdot\frac{\mathbf{h}}{||\mathbf{h}||} \tag{4}$$

[2] The board is a first order approximation to the function. Whether the board is above the graph, below it, or cuts through it depends on the definiteness of the Hessian. If the Hessian is positive (negative) definite, the board will be above (below) the graph. If it is indefinite, then the board will actually cut through the graph. It if is semi-definite, you have to check the third-order terms in the Taylor expansion.

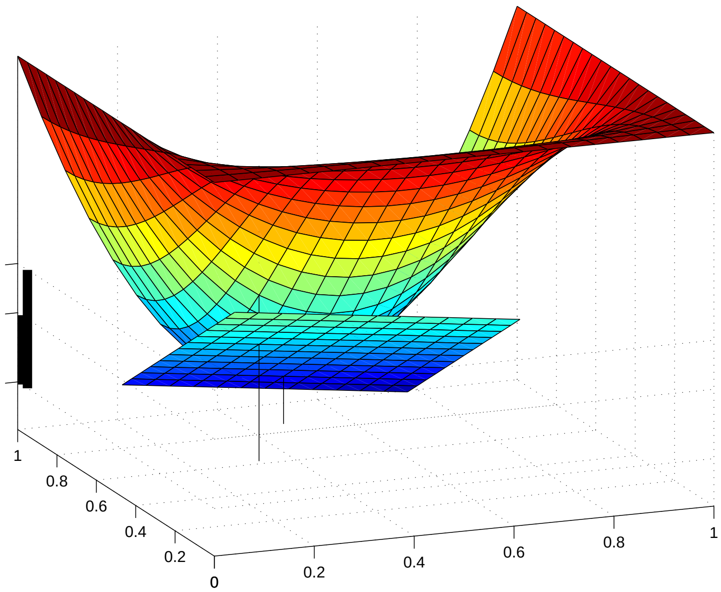

FIGURE 2. Flat board against the graph

Since the length of the vector $\frac{\mathbf{h}}{||\mathbf{h}||}$ is unity, a verbal version of this definition is: f is differentiable at x_0 if for every direction $\mathbf{h}$, the directional derivative of f at x_0 in the direction $\mathbf{h}$, i.e., $\lim_{k\to\infty} \frac{\left(f(\mathbf{x}_0+\mathbf{h}/k)-f(\mathbf{x}_0)\right)}{||\mathbf{h}||}$ is equal to the value of the differential function, when evaluated at *the unique unit length* vector pointing in the direction $\mathbf{h}$. Some times in class I have sloppily talked as if the vectors $\mathbf{h}$ and $-\mathbf{h}$ "point in the same direction," i.e., that all that matters is the angle of the line thru the origin on which $\mathbf{h}$ lies, and this line stays the same when you flip $\mathbf{h}$ through 180 degrees. This was bad sloppy talk, as this definition clearly indicates. From now on, I'll try to consistently say that $\mathbf{h}$ and $-\mathbf{h}$ "point in opposite directions."

There is yet another way of saying what differentiability means, this time in terms of linear algebra. Loosely, a function is differentiable at $\mathbf{x}_0$ if the directional derivatives of f at $\mathbf{x}_0$ live in the *vector space* spanned by the partial derivatives. This statement is not precise: the following theorem makes it correct. Recalling the definition above of $\mathbf{e}^i$, the unit vector pointing in the i'th direction, we have

Theorem: a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$ iff $\bigtriangledown f(\mathbf{x}_0)$ exists and for every $\mathbf{h} \in \mathbb{R}$ with $||\mathbf{h}|| = 1$, the pair $(\mathbf{h}, f_{\mathbf{h}}(\mathbf{x}_0))$ belongs to the n dimensional vector subspace of $\mathbb{R}^{n+1}$ spanned by the set of vectors $\{\left(\mathbf{e}^i, f_i(\mathbf{x}_0)\right) : i = 1, ...n\}$.

Note that this is exactly the same as the following:

Theorem: a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$ iff $\bigtriangledown f(\mathbf{x}_0)$ exists and for every $\mathbf{h} \in \mathbb{R}$ with $||\mathbf{h}|| = 1$, the pair $(\mathbf{h}, f_{\mathbf{h}}(\mathbf{x}_0))$ can be written as a linear combination of the set of vectors $\{\left(\mathbf{e}^i, f_i(\mathbf{x}_0)\right) : i = 1, ...n\}$.

(Note that $(\mathbf{h}, f_{\mathbf{h}}(\mathbf{x}_0))$ and $\left(\mathbf{e}^i, f_i(\mathbf{x}_0)\right)$ both live in $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$.)

Letting $I = [\mathbf{e}^1, ..., \mathbf{e}^i, ..., \mathbf{e}^n]$ denote the $n \times n$ identity matrix, the above theorem is saying that f is differentiable at $\mathbf{x}_0$ iff for every $\mathbf{h} \in \mathbb{R}^n$ with $||\mathbf{h}|| = 1$,

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \\ f_{\mathbf{h}}(\mathbf{x}_0) \end{bmatrix} = h_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ f_1(\mathbf{x}_0) \end{bmatrix} + h_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ f_2(\mathbf{x}_0) \end{bmatrix} + \cdots + h_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ f_n(\mathbf{x}_0) \end{bmatrix}$$

Thus, this theorem simplys restate the definition of differentiability in vector space language. The span of the set of vectors $\{(\mathbf{e}^i, f_i(\mathbf{x}_0)) : i = 1, ...n\} \subset \mathbb{R}^{n+1}$ is precisely the tangent plane translated back to the origin. The theorem says that a function is differentiable iff for each direction $\mathbf{h}$ in the unit circle, the directional derivative $f_h(\mathbf{x}_0)$ is the height of the tangent plane vertically above $\mathbf{h}$.

Take an example: $f(\mathbf{x}) = x_1 x_2$, $\bar{\mathbf{x}} = (1,2)$, $\bigtriangledown f(\mathbf{x}) = (x_2, x_1)$, $\bigtriangledown f(\bar{\mathbf{x}}) = (2,1)$, $\mathbf{h} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Now

$$\begin{aligned} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ f_{\mathbf{h}}(\bar{\mathbf{x}}) \end{bmatrix} &= h_1 \begin{bmatrix} 1 \\ 0 \\ f_1(\bar{\mathbf{x}}) \end{bmatrix} + h_2 \begin{bmatrix} 0 \\ 1 \\ f_2(\mathbf{x}_0) \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} \end{bmatrix} \end{aligned}$$

On the other hand, consider the function graphed in Fig. 3 below. At the origin, both partials for this function are zero, so that span of $\{(\mathbf{e}^i, f_i(\mathbf{0})) : i = 1, 2\}$ is the horizontal plane. However, almost none of the $(\mathbf{h}, f_{\mathbf{h}}(\mathbf{0}))$'s live in this plane.

So far, we have been talking about differentiability *at a point*. We now define what it means for a *function* to be differentiable.

Definition: a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *differentiable* if it is differentiable at every point in its domain.

Simon & Blume make a particularly egregious mistake when they say what differentiability is. They don't have a formal definition, but their very sloppy choice of words imply that a function is *differentiable* at $\mathbf{x}_0$ if $\bigtriangledown f(\mathbf{x}_0)$ exists. They rarely make egregious mistakes, but in this case... This illustrates the perils of using friendly wordy definitions (something I, of course, would *never* do) rather than writing out the math.

Obviously, it could be a big pain to check that for every unit length vector $\mathbf{h}$, the differential of f at $\mathbf{x}_0$, evaluated at $\mathbf{h}$, coincides with the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{h}$. Fortunately, we don't have to do this, because of the following sufficiency theorem.

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that all partial derivatives of f exist and are continuous in a neighborhood of $\mathbf{x}_0 \in \mathbb{R}^n$. Then f is differentiable at $\mathbf{x}_0$.

A function f defined on an open set U is said to be *continuously differentiable* if $\bigtriangledown f(\cdot)$ is a continuous function on U. (Notice that there's an abuse of terminology here: the whole point of the definition of differentiability is that *it's about much more than just the partials*. Yet continuous differentiability, a more stringent concept, concerns *only* the partials!) The above theorem establishes that continuous

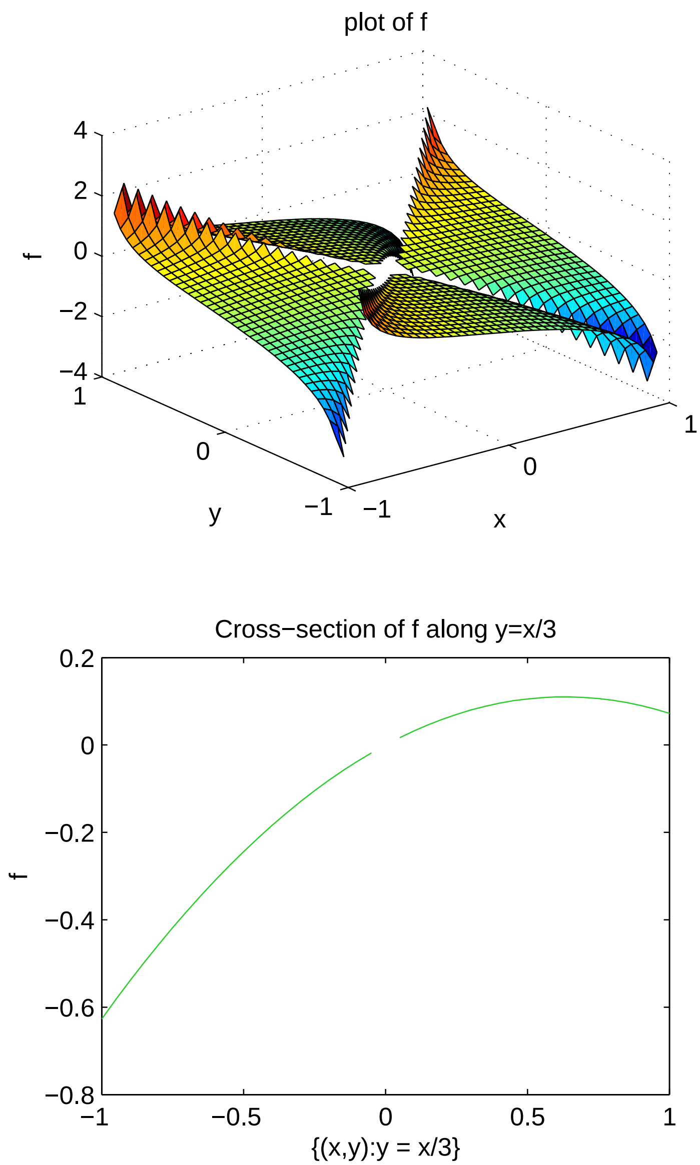

FIGURE 3. Graph of f: partials convey no information about other directional derivatives

differentiability is a sufficient condition for differentiability. It is not, however, a necessary condition for differentiability. To see this, consider the function $f(\mathbf{x}) = \begin{cases} x^2\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$. As Fig. 4

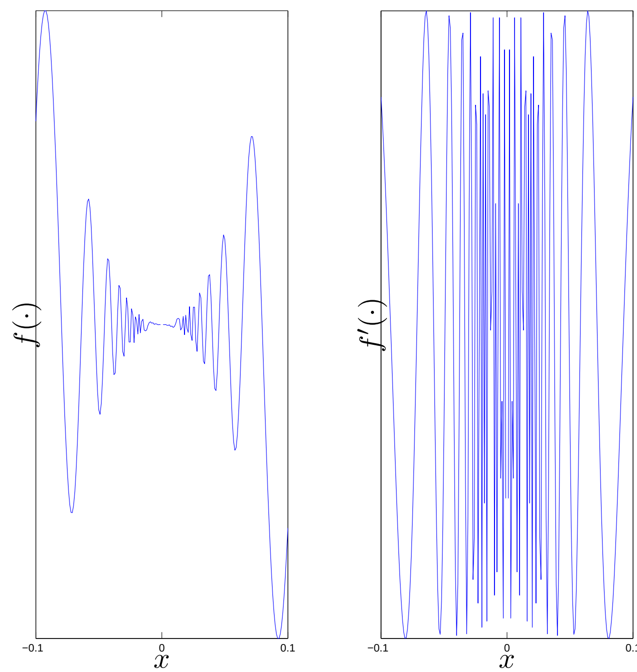

FIGURE 4. A differentiable function that is not continuously differentiable

suggests, the graph of $f(\cdot)$ (plotted the left panel) has a tangent plane at zero—you have to use your imagination a bit to visualize it—but the derivative $f(\cdot)$ (given by $f(x) = 2x\sin(1/x) - \cos(1/x)$ and plotted in the right panel) oscillates increasingly wildly as you approach zero, and so is not a continuous function at zero. The set of continuously differentiable functions is denoted by $\mathbb{C}^1$. More generally, the set of i-times continuously differentiable functions (i.e., each of the first partials are continuously differentable, etc., etc) is denoted by $\mathbb{C}^i$.

Returning to the relationship between directional derivatives and the gradient for well-behaved functions, the following example illustrates that provided a function is differentiable, then computing a directional derivative in the primitive way—i.e., the left hand side of (4)—will yield the same answer as computing it as a linear combination of partials, i.e., the right hand side of (4).

Example: Let $f = x_1x_2$, $\mathbf{x}_0 = \mathbf{h} = (1,1)$. We'll compute the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{h}$, both ways:

(1) <u>The direct way</u>:

$$\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0 + \mathbf{h}/k) - f(\mathbf{x}_0)\big)}{||h||/k} = \lim_{|k|\to\infty} \frac{\Big(\big(1+\frac{1}{k}\big)^2 - 1^2\Big)}{\sqrt{2}/k}$$

$$= \lim_{|k|\to\infty} \frac{\big(1 + \frac{2}{k} + \frac{1}{k^2} - 1\big)}{\sqrt{2}/k} = \lim_{|k|\to\infty} \frac{2+1/k}{\sqrt{2}} = \sqrt{2}$$

(2) <u>The indirect way</u>: (right hand side of (4))

$$\triangledown f(\mathbf{x}_0)\cdot\frac{\mathbf{h}}{||\mathbf{h}||} = \frac{\begin{bmatrix}1 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 1\end{bmatrix}'}{\sqrt{2}} = \frac{2}{\sqrt{2}} = \sqrt{2}$$

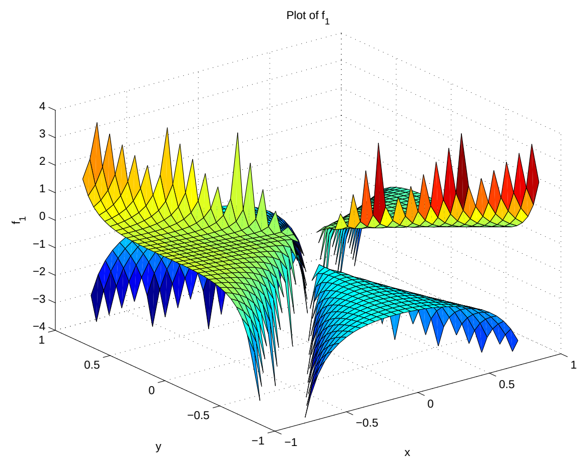

FIGURE 5. Graph of $f_1(\cdot)$

A question came up in class one year: Letting C denote the unit circle, consider the mapping $f_h^{\mathbf{x}} : C \to \mathbb{R}$ that maps each vector in the circle to the corresponding directional derivative, evaluated at $\mathbf{x}$. Is this mapping linear? Answer no, since as we've discussed, the definition of linear only makes sense if both the domain and codomain of the function are vector spaces. What is true is that the graph of f is a subset of the graph of a specific linear function, i.e., the differential of f evaluated at $\mathbf{x}$. In other words, if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then the graph of $f_h : C \to \mathbb{R}$ is contained in a particular n-dimensional vector subspace of of $\mathbb{R}^{n+1}$.

2.3.3. *A nondifferentiable function whose partial derivatives exist.* Consider the function

$$f(x,y) = \; -\gamma(x^2+y^2) \quad + \quad \begin{cases} \frac{xy}{\mathrm{sgn}(x-y)\sqrt{|x^2-y^2|}} & \text{if } |x| \neq |y| \\ 0 & \text{otherwise} \end{cases}$$

where $\gamma > 0$, is graphed in the top panel of Fig. 3. (sgn($\cdot$) is a function that is 1 if the argument is positive, -1 if negative and zero if zero.) Look at the function in a neighborhood of zero. The function is continuous at zero, indeed, has beautifully behaved cross sections in *every* direction, which are strictly concave in every direction (a typical cross-section is graphed in the bottom panel of Fig. 3). In particular, the partials w.r.t. x and y both exist and $\bigtriangledown f(\mathbf{0}) = \mathbf{0}$. To see this, note that for all $\alpha \neq 0$, $f(\alpha,0) = f(0,\alpha) = 0/\alpha = 0$. Hence $\lim_{|x|\to 0} \frac{\left(f(\mathbf{0}+(x,0))-f(\mathbf{0})\right)}{||\mathbf{x}||} = \lim_{|y|\to 0} \frac{\left(f(\mathbf{0}+(0,y))-f(\mathbf{0})\right)}{||\mathbf{y}||} = 0.$ However, as Fig. 3 indicates, these partials provide no information about what the slope of f is when you move in any direction other than parallel to an axis. To see that the requirement for differentiability fails, note first from the graph that you clearly can't put a tangent plane on top of the graph at $(0,0)$. More precisely, note that the directional derivatives *don't* live in the vector space spanned by the the vectors $\{\left(\mathbf{e}^i, f_i(\mathbf{x}_0)\right) : i = 1, ...n\}$. Also, observe that the *sufficient* condition for differentiability fails, i.e., neither partial is a continuous function in a neighborhood of zero. (The graph of the first partial is plotted in Fig. 5.)

2.3.4. *Second Partials vs Definiteness of the Hessian.*

Suppose $f : \mathbb{R}^n \to \mathbb{R}$. The gradient $\bigtriangledown f : \mathbb{R}^n \to \mathbb{R}^n$ is the vector-valued mapping that takes each point in the domain of f to its vector of partial derivatives. If f is *twice differentiable* then each partial derivative f_i itself has a vector of partial derivatives, $(f_{i1}(\cdot), ..., f_{ii}(\cdot), ..., f_{in}(\cdot))$, or,

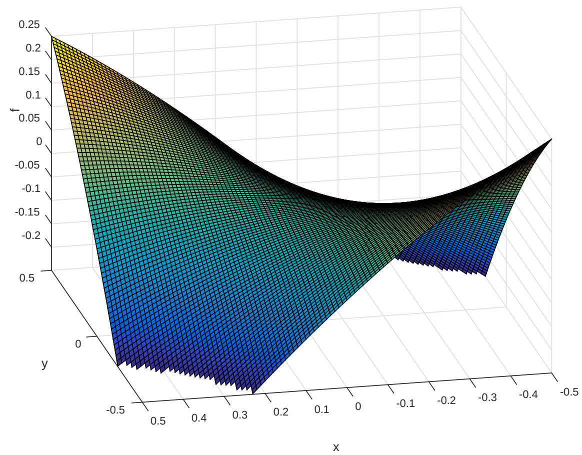

FIGURE 6. Negative second partials don't imply a local maximum

more concisely, $\triangledown f_i(\cdot)$. For now, we'll just mechanically stack all of these n gradient vectors on top of each other to construct the *Hessian* of f, i.e., $\mathrm{Hf}(\mathbf{x}) = \begin{bmatrix} \triangledown f^1(\mathbf{x}) \\ \triangledown f^2(\mathbf{x}) \\ \vdots \\ \triangledown f^m(\mathbf{x}) \end{bmatrix}$. The Hessian of f, or, equivalently, the *derivative of the gradient of* f is an example of the derivative of a function from $\mathbb{R}^n$ to $\mathbb{R}^n$. We'll talk more about these objects in the next lecture. For given $\mathbf{x} \in \mathbb{R}^n$, the matrix $\mathrm{Hf}(\mathbf{x})$ is a point in $\mathbb{R}^n \times \mathbb{R}^n$. Given a twice differentiable function, f, the hessian *mapping* $\mathrm{Hf} : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ maps every point in $\mathbb{R}^n$ to the matrix of its second partial derivatives.

The best known use of the Hessian is that it enables us to distinguish between the maxima, the minima and the "neither" of a function $f : \mathbb{R}^n \to \mathbb{R}$.

Question: Suppose I know that the slope of f is zero at $\boldsymbol{x}^*$. What additional information about the slope of f at $\boldsymbol{x}^*$ do I need to know that f attains a strict local maximum at $\boldsymbol{x}^*$?

First shot at an answer:[3] A first guess is that the slopes of all of the partial derivatives have to be decreasing, i.e., the second order partials must all be negative. Turns out however that it isn't enough. Reason is that while the *first* partials of a differentiable function provide all the information you need to compute directional derivatives (i.e., directionals live in the vector space spanned by the partials, etc), the second partials give you *no information whatsoever* about the *slopes* of the other directional derivatives. Imagine a graph where if you took *circular* cross-sections, i.e., sliced the graph with a circular cookie-cutter, what you got when you laid out the cut was a sine curve, with zeros corresponding to the points at the axes. Here you might think you had a weak local maximum if you just looked at the second-order partials, but you can't put the board on top. More concretely, consider Fig. 6, which is the graph of the function $f(x, y) = 1.5xy - (y^2 + x^2)/4$.

To see what's going on with this function, we'll plot the *second* derivatives of the diagonal cross-sections as we work our way around the unit circle. More precisely, for each *angle* $\theta \in [0, 360°]$, denote by $f_{\theta\theta}(0, 0)$, the slope of the slope of the cross-section you get when you slice the graph of the function plotted in Fig. 6 in the direction $(\sin(\theta), \cos(\theta))$. Observe that along the directions parallel to the axes, i.e., when $\theta \in \{0°, 90°, 180°, 270°, 360°\}$, the slope of the slope of the diagonal

[3] Many thanks to YoungDong Liu for help with this section.

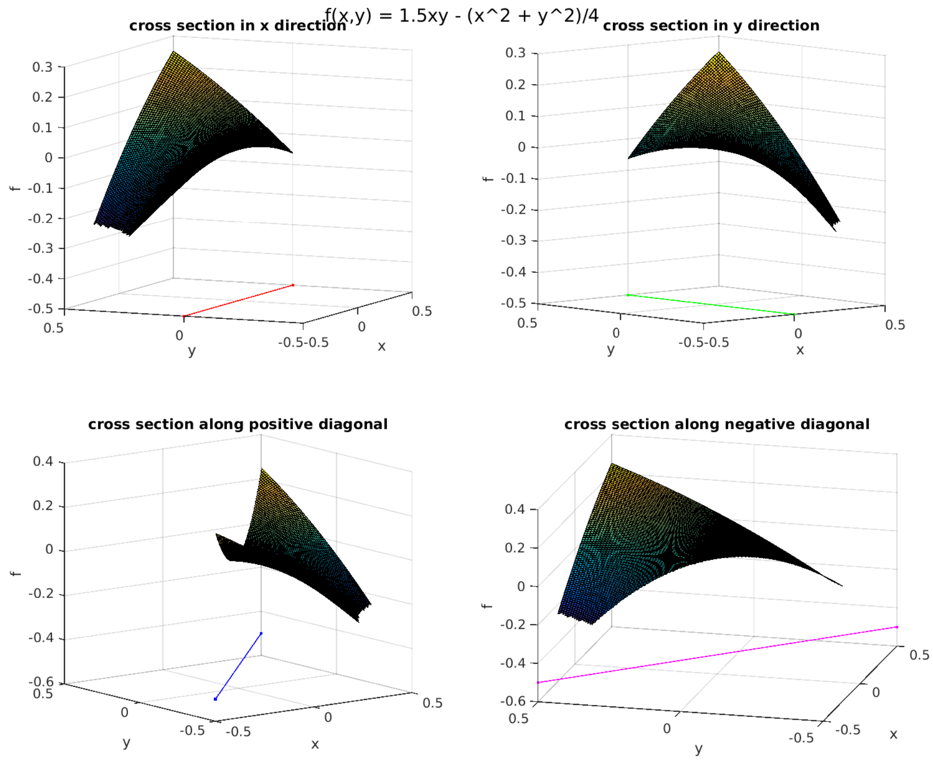

FIGURE 8. Cross sections of f

cross-section is negative, while along the direction parallel to the positive 45 degree line, i.e., when $\theta \in \{45°, 225°, \}$, it is positive. *Note well* that this picture is not a pathological one like Fig. 3. All of the derivatives in sight are nicely continuously differentiable. So, we've established that negativity of the second partials, i.e., $f_{ii} < 0$, for each $i = 1, ...n$, isn't enough to guarantee $f_{hh} < 0$, for each *direction* h, and hence isn't enough to guarantee a maximum.

So what do we need to guarantee a maximum at $\boldsymbol{x}^*$? A sufficient requirement is that the second derivative of the function f evaluated at $\boldsymbol{x}^*$ (which is a *matrix*) is *negative definite*. What does this mean? As we saw in the linear algebra section, it means that as you move $\mathbf{dx}$ in *any* direction away from $\boldsymbol{x}^*$, $\mathbf{dx}$ makes an obtuse angle with its image under the linear mapping $H(\boldsymbol{x}^*)$. Diagramatically it means that you can put a flat board on top of the graph of the function at $\boldsymbol{x}^*$ and this graph will be everywhere below the board. Note that in figure 6, you can't do this. Alternatively, if you were to take a function whose second derivative were negative definite and repeat the exercise we did to produce Fig. **??**, the graph we would obtain would be weakly below the origin for all values of θ.

Similarly, to establish that f attains a local *minimum* at $\boldsymbol{x}^*$, need to show that you can put a flat board *below* the graph of the function at $\boldsymbol{x}^*$ and the graph is everywhere *above* the board.

2.3.5. *Total Derivative.* This is a concept that's widely used in economics, but it's mis-named: it's not really a "derivative." Specifically, it is "like" a directional derivative, but it's *not* a directional derivative. The "total" part of the name is also problematic, because people sometimes get confused between the total derivative and the total differential. When people talk about the "total differential," they are really just talking about "the differential".

Often, the arguments of a function depend on each other: we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ which depends on x and w, but w itself depends on x, so we write $f(x, w(x))$. If we want to know how f will change when x changes, we need to take into account that w will change too.

General defn: The *total* derivative of $f : \mathbb{R}^n \to \mathbb{R}$ with respect to x_i is written $\frac{df(x)}{dx_i}$ (cf the partial derivative sign "∂") and is defined by

$$\frac{df(\mathbf{x})}{dx_i} \quad = \quad f_1(\mathbf{x})\frac{\partial x_1}{\partial x_i} + ... + f_i(\mathbf{x}) + ... + f_n(\mathbf{x})\frac{\partial x_n}{\partial x_i}$$

Economic example: the firm's profit function $\pi(p, q(p)) = pq(p)$, where $q(p)$ is the optimal output choice given p. In this case, $\frac{d\pi(p,q(q))}{dp} = \frac{\partial \pi(p,q(p))}{\partial p} + \frac{\partial \pi(p,q(p))}{\partial q}\frac{dq}{dp} = q(p) + pq'(p)$.

Total derivative, directional derivative and differential: What's the relationship between these three concepts? I'll answer this in the context of the example of the firm's profit function $\pi(p, q(p)) = pq(p)$,

- the *total derivative* of $\pi(\bar{p}, q(\bar{p}))$ tells you how much π changes when you increase p by *one unit.*
- the *directional derivative* of $\pi(\bar{p}, q(\bar{p}))$ in the direction $(dp, q'(p)dp)$ tells you how much π changes when you move *one unit in length* from $(\bar{p}, q(\bar{p}))$ in the direction $(dp, q'(p)dp)$.
- but if you increase p by one unit, you *don't* move *one unit* of length in the direction $(dp, q'(p)dp)$; in fact, you move $||\,(1, q'(p)))\,||$ units of length in this direction!
- hence the *total derivative* of $\pi(\bar{p}, q(\bar{p}))$ is the differential of π at $(\bar{p}, q(\bar{p}))$, evaluated at the magnitude of the change, i.e., at $(1, q'(p))$.
- alternatively, the total derivative is the directional derivative in the direction $(1, q'(p))$, *multiplied by the length of the change*, i.e., by $||\,(1, q'(p)))\,||$.

Example: The following example illustrates the above relationships. Consider a firm that supplies the amount $q(p) = \sqrt{p}$ and incurs no costs, so that the firm's profit function is $\pi(p,q) = pq(p)$. Set $p = 1$ and note that $q(p) = 1$ while $q'(p) = \frac{0.5}{\sqrt{p}}$ so that $q'(1) = 0.5$. Note also that $\bigtriangledown\pi(1,1) = (q,p) = (1,1)$. All four of the routes below tell you how much π goes up when you increase p by one unit, and q then increases by $q'(p)$. Happily all four routes give you the same answer.

(1) <u>The derivative of π, viewed as a function of p *only*</u>:

$$\left.\frac{d\pi(p,q(p))}{dp}\right|_{p=1} \quad = \quad \left.\frac{d}{dp}\left(p\sqrt{p}\right)\right|_{p=1} \quad = \quad \left.\left(\sqrt{p} + \frac{0.5p}{\sqrt{p}}\right)\right|_{p=1} \quad = \quad 1.5$$

(2) <u>The total derivative of π, viewed as a function of p and $q(p)$</u>:

$$\frac{d\pi(p,q(p))}{dp} \quad = \quad \frac{\partial\pi(p,q(p))}{\partial p} + \frac{\partial\pi(p,q(p))}{\partial q}\frac{dq}{dp} \quad = \quad 1 + 1 \star 0.5 \quad = \quad 1.5;$$

(3) <u>The differential of $\pi(p,q)$ at $(1,q(1))$, evaluated at $\mathbf{h} = (1,q'(1))$</u>:

$$\bigtriangledown\pi(1,1)\cdot\mathbf{h} \quad = \quad \begin{bmatrix}1 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0.5\end{bmatrix}' \quad = \quad 1.5;$$

(4) <u>The directional derivative of $\pi(p,q)$ in the direction $\mathbf{h} = (1,q'(p))$</u>:

$$\bigtriangledown\pi(1,1)\cdot\frac{\mathbf{h}}{||\mathbf{h}||} \quad = \quad (q(p),p)\cdot\frac{\mathbf{h}}{||\mathbf{h}||} \quad = \quad \frac{\begin{bmatrix}1 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0.5\end{bmatrix}'}{\sqrt{1.25}} \quad = \quad \frac{1.5}{\sqrt{1.25}}$$

Now if you add to $(1,1)$ a vector of length $\sqrt{1.25}$ that points in the direction $\mathbf{h} = (1,q'(p))$ you increase π by

$$\frac{1.5}{\sqrt{1.25}}\sqrt{1.25} \quad = \quad 1.5;$$