**ARE211, Fall 2004**

LECTURE #17: THU, NOV 4, 2004 PRINT DATE: NOVEMBER 3, 2004 (CALCULUS2)

## CONTENTS



## 4. UNIVARIATE AND MULTIVARIATE DIFFERENTIATION (CONT)

### 4.3. **Multivariate calculus: functions from $\mathbb{R}^n$ to $\mathbb{R}$**

4.3.1. *Partial Derivative.* . Cookbook approach: If $f : \mathbb{R}^n \to \mathbb{R}$, then $f_i(\cdot) \equiv \frac{\partial f(\cdot)}{\partial x_i}$ is the (usual) derivative of the single variable function you get by treating all other variables as constant.

Example:

$$\begin{aligned} f(\mathbf{x}) &= x_1^2 + 2x_1x_2 + x_2^2 \\ &= x_1^2 + 2x_1\alpha + \beta \end{aligned}$$

$$
\begin{aligned}
f_1(\mathbf{x}) &= 2x_1 + 2\alpha + 0 \\
&= 2x_1 + 2x_2
\end{aligned}
$$

A more graphical view of partial derivatives. Take a cross-section of the graph of the function along the $i$'th axis, and look at the slope of the single-dimensional function you obtain in this way. This slope is the $i$'th partial derivative.

Similarly, you could take a diagonal cross-section, and get a different one-dimensional slope. Generally, you get what is called a *directional derivative*. Partial derivatives are just special kinds of directional derivatives: in particular, they tell you the slopes you obtain when you take cross-sections along the various axes.

4.3.2. *The Gradient.* Given a function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient of $f$ is just the function from $\mathbb{R}^n$ to $\mathbb{R}^n$ that maps each $\mathbf{x} \in \mathbb{R}^n$ to the vector of partial derivatives of $f$ at $\mathbf{x}$. That is,

$$
\nabla f(\cdot) = \begin{bmatrix} f_1(\cdot) \\ \vdots \\ f_i(\cdot) \\ \vdots \\ f_n(\cdot) \end{bmatrix}
$$

Emphasize that $\nabla f(\cdot)$ is the exact analog for $f : \mathbb{R}^n \to \mathbb{R}$. as $f'(\cdot)$ is for $f : \mathbb{R} \to \mathbb{R}$. In fact the words "slope of f," "gradient of f" and "derivative of f" are synonomous.

Emphasize the important distinction between the "gradient of $f$" which is a *function*, written $\nabla f$ or $\nabla f(\cdot)$ and the "gradient of $f$ at $\mathbf{x}$," which is a vector.

4.3.3. *Crosspartial Derivative.* A cross-partial derivative is just an entry in the matrix which is the derivative of the derivative. That is, take a function $f : \mathbb{R}^n \to \mathbb{R}$. The $i$'th partial derivative of this function $f_i(\cdot)$ is a function, like any other, and if the function is differentiable, it has derivatives:

$f_{ij}(\cdot) \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}(\cdot)$ is the $j$'th partial derivative of the function $f_i(\cdot) \equiv \frac{\partial f}{\partial x_i}(\cdot)$.

Thus, the Hessian of $f$ just consists of the matrix of crosspartials of $f$.

4.3.4. *The differential for functions from $\mathbb{R}^n$ to $\mathbb{R}$.* Recall from earlier: the most important thing to keep in mind about calculus is that when you do calculus on $f$ at $x$, you are, ALWAYS, approximating a small change in $f$, starting from $x$, by evaluating *the differential of $f$ at $x$* at the magnitude of the change.

Exactly the same thing for functions from $\mathbb{R}^n$ to $\mathbb{R}$. Evaluate the difference $f(\mathbf{x}+\mathbf{dx}) - f(\mathbf{x})$, where here $\mathbf{dx}$ and $\mathbf{x}$'s are both vectors, by the value of the differential, evaluated at the difference, $\mathbf{dx}$.

What's the differential, in terms that we've used before? Recall that it is a linear function from $\mathbb{R}^n$ to $\mathbb{R}$; so it is characterized by a vector of coefficients. What is the vector? Answer: the gradient, evaluated at $x$.

Thus: $df = \nabla f(\mathbf{x}) \cdot \mathbf{dx}$. See Fig. 1 below. As the figure indicates, it plays exactly the same role as the differential played for functions from $\mathbb{R}$ to $\mathbb{R}$: i.e., construct the tangent plane that lies on top of the function at $\mathbf{x}$, shift this plane so it passes thru the origin, then approximate the change in $f$ when you move from $\mathbf{x}$ to $\mathbf{x}+\mathbf{dx}$ by finding the height of this plane above the point $\mathbf{dx}$.

Example 1: Use the example above: $f(x,y) = x^2 + y^2$; $\nabla f(\cdot) = (2x, 2y)$;

Consider $(dx, dy) = (0.1, 0.1)$, set $(x^0, y^0) = (100, 100)$ and approximate $f(x^0 + dx, y^0 + dy) - f(x^0, y^0)$ by $\nabla f(x^0, y^0) * (0.1, 0.1) = (200, 200) * (0.1, 0.1) = 40$. Note that this approximation is virtually the same as the one you get when you use the directional derivative! Now evaluate at $f$ at both $(x^0, y^0) = (100, 100)$ and $(x^0 + dx, y^0 + dy) = (100.1, 100.1)$, and compare: $f(x^0, y^0) = 20,000$ $f(x^0 + dx, y^0 + dy) = 20040.02$ Actual difference is 40.02.

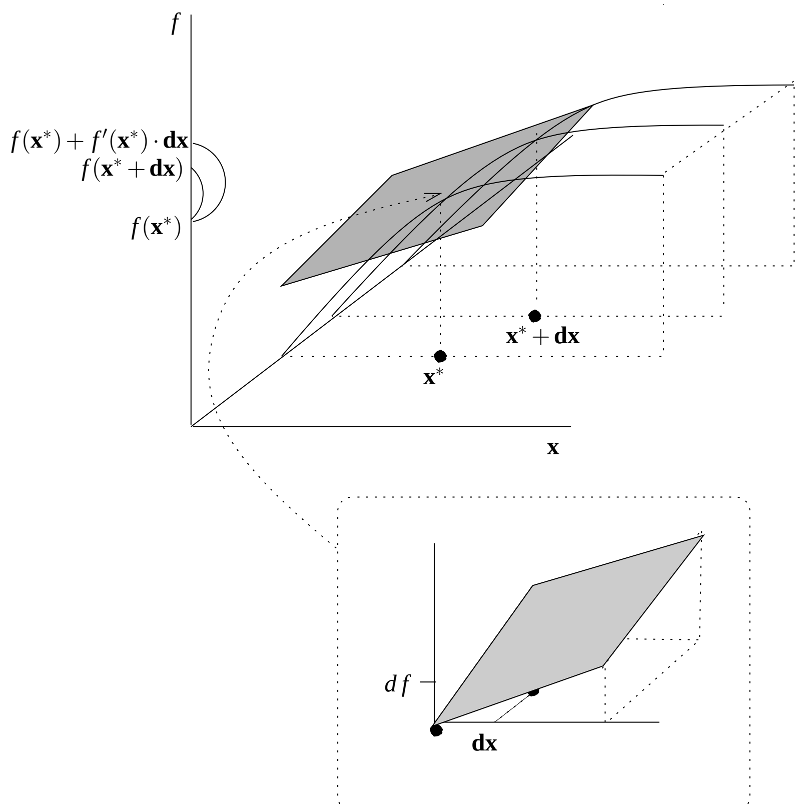


FIGURE 1. Linear Approximations of the change in a nonlinear multivariate function

Example 2: $y = f(\ell, k)$; $\nabla f(\cdot) = (f_\ell(\cdot), f_k(\cdot))$. Fix $(\overline{\ell}, \overline{k})$. Evaluate the difference $f(\overline{\ell}, \overline{k}) - f(\ell, k)$ by the value of the differential, evaluated at $(\overline{\ell}, \overline{k}) - (\ell, k)$:

$$f(\overline{\ell}, \overline{k}) - f(\ell, k) = dy \quad = \quad f_\ell(\overline{\ell}, \overline{k}) d\ell + f_k(\overline{\ell}, \overline{k}) dk$$

Sometimes economists refer to the differential as the *total differential*. It is important to recognize that the total derivative of a function and the total differential are completely different things. The total derivative is related to (but not the same as) a directional derivative (i.e., you take the cross section of the function in a certain direction). The total differential is the name of a linear function.

4.3.5. *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. Having said that, it's not so straightforward to compute it. For computational purposes, it's useful to extrapolate from the following (obvious) 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 $e^i$, 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., $f_i(x) = df^x(e^i)$.
- 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

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

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 $-\mathbf{h}$.

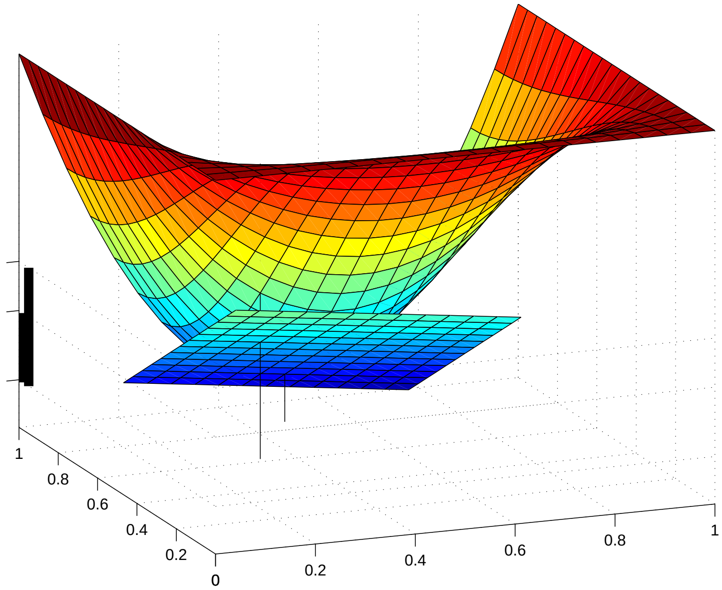


FIGURE 2. Flat board against the graph

4.3.6. *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 $\nabla f(\mathbf{x}_0)$ exists and if for all $\mathbf{h} \in \mathbb{R}^n$,

$$\lim_{|k|\to\infty} \frac{\big(f(\mathbf{x}_0 + \mathbf{h}/k) - f(\mathbf{x}_0)\big) \; - \; \nabla f(\mathbf{x}_0)\cdot\mathbf{h}/k}{||\mathbf{h}||/k} \quad = \quad 0 \tag{2}$$

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). Fig. 2 illustrates how you put a flat board against a graph.

Since the term $\frac{\nabla f(\mathbf{x}_0)\cdot\mathbf{h}/k}{||\mathbf{h}||/k}$ clearly does not depend on $k$, we can rewrite the above definition as:

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

Since the length of the vector $\frac{\mathbf{h}}{||\mathbf{h}||}$ is unity, an equivalent 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."

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 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 $\nabla f(\mathbf{x}_0)$ exists. They rarely make mistakes, but in this case... This illustrates the perils of using word definitions rather than writing out the math (something I, of course, would *never* do).

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 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 $\nabla f(\cdot)$ is a continuous function on $U$. The above theorem establishes that continuous differentiability is a sufficient condition for differentiability. (Given this terminology, it would be embarassing if this were not true!) Continous differentiability 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. 3 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 $f(\cdot)$ (plotted in the right panel) oscillates increasingly wildly as you approach zero, and so is not a continuous function at zero.

The sets of differentiable and continuously differentiable functions are denoted (respectively) by $\mathbb{C}^0$ and $\mathbb{C}^1$.

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

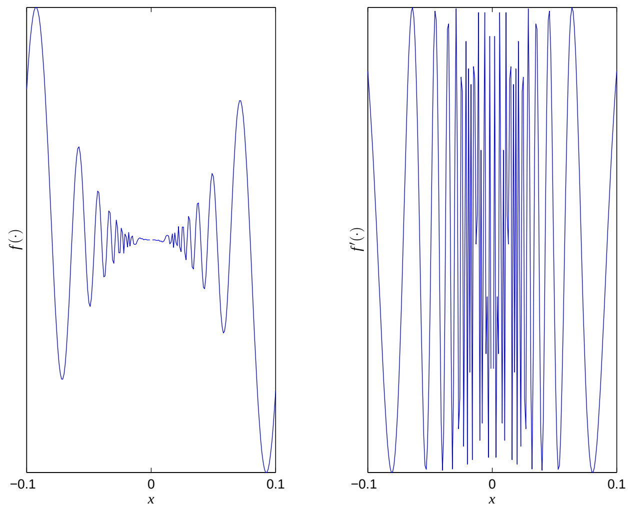


FIGURE 3. A differentiable function that is not continuously differentiable

derivative in the primitive way—i.e., the left hand side of (3)—will yield the same answer as computing it as a linear combination of partials, i.e., the right hand side of (3).

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) The primitive way: (left hand side of (3))

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

(2) The indirect way: (right hand side of (3))

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

4.3.7. *A nondifferentiable function whose partial derivatives exist.* In the homework, you'll consider the function $f(x,y) = \begin{cases} 0 & \text{if } x = y = 0 \\ \frac{xy}{\sqrt{x^2+y^2}} & \text{otherwise} \end{cases}$, which is beautifully behaved, i.e., it is everywhere continuous and directional derivatives exist in every direction, but it is not differentiable.

In this lecture, we'll return to the function we looked at in the graphical overview (Fig 6), which is much less well behaved. Both partial derivatives exist at $(0,0)$ but the function is not continuous at $(0,0)$ and, at $(0,0)$, the only other directional derivatives that exist are those in the directions the positive and negative 45° lines.

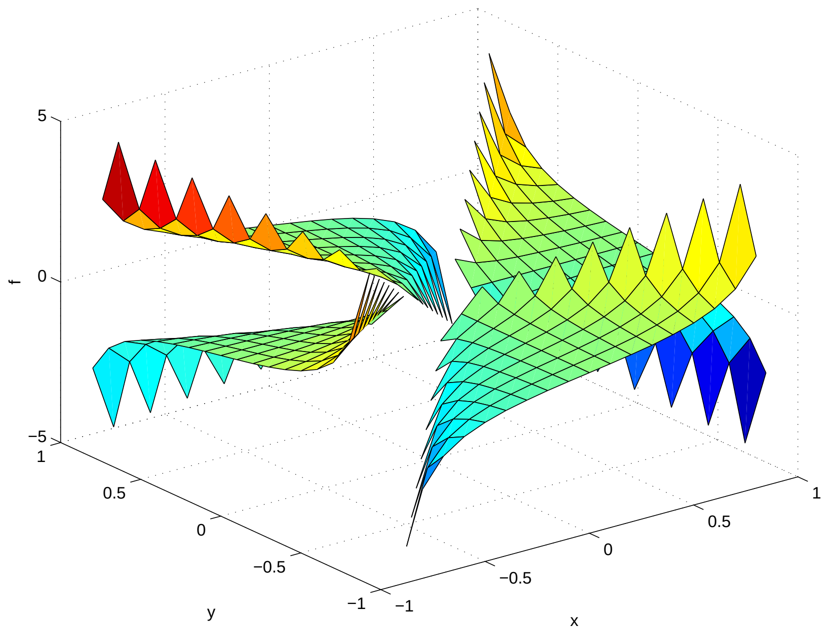


FIGURE 4. Graph of $f$: partials convey no information about other directional derivatives

Consider the function $f(x,y) = \begin{cases} \frac{xy}{x^2-y^2} & \text{if} |x| \neq |y| \\ 0 & \text{otherwise} \end{cases}$ and evaluate this function at $(x,y) = (0,0)$. The partials w.r.t. $x$ and $y$ both exist and $\nabla f(\mathbf{0}) = \mathbf{0}$. 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. 4 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 $(0,0)$. Moreover, though it is hard to see this from the graph, the function is not even continuous at $(0,0)$, which means that directional derivatives certainly can't be defined in all directions. Indeed, consider the direction defined by the condition that $y = \alpha x$, $\alpha \in (0,1)$. In this case, evaluating the function, we have

$$f(x,y) = \begin{cases} \frac{\alpha x^2}{(1-\alpha)x^2} = \frac{\alpha}{1-\alpha} & \text{if} |x| \neq |y| \\ 0 & \text{otherwise} \end{cases}$$

Clearly, along this direction, $f(\cdot,\cdot)$ is constant *except* at zero, where it blips down to zero. Since a necessary condition for a directional derivative to exist is that the function is continuous along the direction, directional derivatives fail to exist for all of the directions defined by $\alpha \in (0,1)$.

4.3.8. *Total Derivative.* This concept appears to have been invented by economists but never appears in a math book. With good reason. It's not really a derivative at all; rather it's nothing more than a straightfoward application of the chain rule. The *only* reason why it is taking up space in these notes is because economists have been using it for years, so you need to be familiar with it. It would be better to erase it from your brain (after this lecture), or if you want to leave it in your brain, then at least give it another name, which does not involve the word 'derivative'.

In any event, the total derivative should not to be confused with the total 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 "$\partial$") and is defined by

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

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? In the good old days, pre-enlightenment, I used to tell people that the total derivative was the *directional derivative* in the direction $(dp, q'(p)dp)$. I was **wrong**. The following reasoning explains why what I said was wrong:

- the *total derivative* of $\pi(\bar{p}, q(\bar{p}))$ tells you how much $\pi$ 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 $\pi$ 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 $\pi$ 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 competitive industry with supply function $q(p) = \sqrt{p}$ and zero costs. (Economics is a little wierd, but this is a math course after all...) Industry profits are $\pi(p,q) = pq(p)$. Set $p = 1$ and note that $q(p) = 1$ while $q'(p) = 0.5/\sqrt{p} = 0.5$. Note also that $\nabla\pi(1,1) = (q,p) = (1,1)$. All four of the routes below tell you how much $\pi$ 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) The derivative of $\pi$, viewed as a function of *p only*:

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

(2) The total derivative of $\pi$, viewed as a function of $p$ and $q(p)$:

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

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

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

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

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

Now if you add to $(1,1)$ a vector pointing in the direction $\mathbf{h}$ of length $\sqrt{1.25}$, you increase $\pi$ by

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