1. Basic idea and definition

If $f(x_1, \cdots, x_n)$ takes real numbers as values we are interested in how the values of f change when we change one of the variables, say x_i at a fixed location, say $\mathbf{a} = (a_1, \cdots, a_n)$. From first year calculus, this suggests considering

$$\lim_{x_i \to a_i} \frac{f(a_1, \cdots, a_{i-1}, x_i, a_{i+1}, \cdots, a_n) - f(a_1, \cdots, a_n)}{x_i - a_i}$$

Note that this is a one variable limit and we are just computing an ordinary derivative from first year calculus. We denote this value in a couple of ways. Sometimes one is more convenient than the other but they mean the same thing.

$$\frac{\partial f}{\partial x_i} \qquad f_{x_i} \qquad \boxed{f_i \textit{ Terrible!}} \qquad D_i f \qquad D_{x_i} f$$

Just as in first year calculus, *the partial derivative measures the instantaneous rate of change of the function with respect to one of the variables.*

EXAMPLES

(1) $f(x,y) = xy^2$. Find $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ at $(2,3)$. For $\dfrac{\partial f}{\partial x}(2,3)$ consider $f(x,3) = 9x$. Then $\dfrac{\partial f}{\partial x}(2,3) = \dfrac{d\,9x}{d\,x}(2) = 9$. For $\dfrac{\partial f}{\partial y}(2,3)$ consider $f(2,y) = 2y^2$. Then $\dfrac{\partial f}{\partial y}(2,3) = \dfrac{d\,2y^2}{d\,y}(3) = 4y|_{y=3} = 12$.

(2) $h(x,y,z) = x^2 + y^2 + z^2$. Find $\dfrac{\partial h}{\partial x}(2,3,5)$, $\dfrac{\partial h}{\partial y}(2,3,5)$ and $\dfrac{\partial h}{\partial x}(2,3,5)$.

$$\frac{\partial h}{\partial x}(2,3,5) = \frac{d\,(x^2+36)}{d\,x}(2) = 2x|_{x=2} = 4.$$
$$\frac{\partial h}{\partial y}(2,3,5) = \frac{d\,(y^2+20)}{d\,y}(3) = 2y|_{y=3} = 6.$$
$$\frac{\partial h}{\partial z}(2,3,5) = \frac{d\,(z^2+13)}{d\,z}(5) = 2z|_{z=5} = 10.$$

(3) $g(r,t,\theta) = \sin(r\,t\,\theta)$. Find $\dfrac{\partial g}{\partial r}(2,3,\pi)$, $\dfrac{\partial g}{\partial t}(2,3,\pi)$ and $\dfrac{\partial g}{\partial \theta}(2,3,\pi)$.

$$\frac{\partial g}{\partial r}(2,3,\pi) = \frac{d\,(\sin(3\pi r)}{d\,r}(2) = 3\pi\cos(3\pi r)|_{r=2} = 3\pi\cos(6\pi) = 3\pi.$$
$$\frac{\partial g}{\partial t}(2,3,\pi) = \frac{d\,(\sin(2\pi t)}{d\,t}(3) = 2\pi\cos(2\pi t)|_{t=3} = 2\pi\cos(6\pi) = 2\pi.$$
$$\frac{\partial g}{\partial \theta}(2,3,\pi) = \frac{d\,\sin(6\theta)}{d\,\theta}(\pi) = 6\cos(6\theta)|_{\theta=\pi} = 6\cos(6\pi) = 6.$$

For all three of these functions at the point where we did the calculation, the function is increasing as we increase any of the variables.

Another basic interpretations of the derivative in first year calculus is as a slope. That works here as well. For example, look at the graph of $z = f(x, y)$. Then $\dfrac{\partial f}{\partial x}(2, 3)$ is the slope of the following graph at $x = 2$. Intersect the plane $y = 3$ with the 2-dimensional graph $z = f(x, y)$. This gives a 2-dimensional graph $z = f(x, 3) = 9x$. This graph is a straight line with slope $9 = \dfrac{\partial f}{\partial x}(2, 3)$.

The slope interpretation in more than two variables is difficult to visualize. Given three variables the graph lies in four space, we intersect it with the two space given by fixing two of the variables and in that two space we will see a curve. The slope of that curve at the given point is given by the partial derivative.

2. The gradient

Since we often want to know all the partials it is convenient to write them out as a vector. This vector will be extremely important for us! The definition is simple. Suppose $f(x_1, \cdots, x_n)$ is a function of n variables. Then *the gradient of f at the point $\boldsymbol{a}$* is

$$\nabla f(\mathbf{a}) = \left\langle \frac{\partial f}{\partial x_1}(\mathbf{a}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right\rangle$$

Hence from the examples above

- $\nabla f(2, 3) = \langle 9, 12 \rangle$
- $\nabla h(2, 3, 5) = \langle 4, 6, 10 \rangle$
- $\nabla g(2, 3, \pi) = \langle 3\pi, 2\pi, 6 \rangle$

For now this is all there is to it. You calculate it by calculating the partial derivatives. In a few lectures we will explore further and see that the gradient is telling us many things and comes up in the solution of many important problems.

3. Partial derivatives as functions

Just as in first year calculus, once we have the definition of a partial derivative at a point, we can consider what happens as we change the point. This is the definition of a function.

EXAMPLES

(1) $f(x, y) = xy^2$. Find $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ as functions of (x, y).
   - $\dfrac{\partial f}{\partial x} = \dfrac{\partial\, xy^2}{\partial x} = y^2$
   - $\dfrac{\partial f}{\partial y} = \dfrac{\partial\, xy^2}{\partial y} = 2xy$
   - $\nabla f = \langle y^2, 2xy \rangle$

(2)
$$\frac{\partial h}{\partial x} = \frac{d\,(x^2+y^2+z^2)}{d\,x} = 2x$$
$$\frac{\partial h}{\partial y} = \frac{d\,(x^2+y^2+z^2)}{d\,y} = 2y$$
$$\frac{\partial h}{\partial z} = \frac{d\,(x^2+y^2+z^2)}{d\,z} = 2z$$

- $\nabla h = 2\,\langle x, y, z\rangle$

(3) $g(r,t,\theta) = \sin(r\,t\,\theta)$.
$$\frac{\partial g}{\partial r} = \frac{d\,\sin(r\,t\,\theta)}{d\,r} = t\,\theta\cos(r\,t\,\theta)$$
$$\frac{\partial g}{\partial t} = \frac{d\,\sin(r\,t\,\theta)}{d\,t} = r\,\theta\cos(r\,t\,\theta).$$
$$\frac{\partial g}{\partial \theta} = \frac{d\,\sin(r\,t\,\theta)}{d\,\theta} = r\,t\cos(r\,t\,\theta).$$

- $\nabla g = \cos(r\,t\,\theta)\,\langle t\,\theta, r\,\theta, r\,t\rangle$

4. Higher partial derivatives

Since a partial derivative of a function is itself a function, one can take the partial derivatives of a partial derivative. In one variable the derivative of the derivative is called the second derivative and is very important since one of its interpretations is acceleration and another is convexity.

As in first year calculus, the only technique you will have to compute a higher partial derivative is to keep iterating the partial derivative. For $f(x,y) = xy^2$,

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial\, y^2}{\partial\, x} = 0$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial\, y^2}{\partial\, y} = 2y$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial\, 2xy}{\partial\, x} = 2y$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial\, 2xy}{\partial\, y} = 2x$$

As long as the problem is posed with parentheses it is clear which partial should be taken when. There is shorter notation which allows you to drop the parentheses. In fact there are two versions. Using the ∂ symbol,

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y\,\partial x}$$

and we read the bottom from right to left. Hence

$$\frac{\partial^4 h}{\partial x\,\partial y\,\partial x\,\partial z} = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}\left(\frac{\partial h}{\partial z}\right)\right)\right)$$

The exponent on the ∂ in the numerator is equal to the number of ∂'s in the denominator. If two or more adjacent ∂'s have the same variable, we just write one of them an stick an exponent on it to indicate how many times it is to be repeated. Hence

$$\frac{\partial^4 h}{\partial x\,\partial x\,\partial x\,\partial z} = \frac{\partial^4 h}{\partial^3 x\,\partial z}$$

an now the exponent on the ∂ in the numerator is equal to the sum of the exponents in the denominator where ∂ with no exponent is as usual ∂^1.

The second version uses the f_{x_i} notation.

$$f_{xy} = (f_x)_y = \frac{\partial^2 f}{\partial y\,\partial x}$$

where now we read from left to right and there is no shortening of the notation for repeated partials, $\dfrac{\partial^4 h}{\partial^3 x\,\partial z} = h_{zxxx}$.

These two notations could be annoying were it not for Clairaut's Theorem. This says that provided two higher partials differentiate the same number of times with respect to each variable and if both answers are continuous in a neighborhood of a point, then the two higher partials are equal at that point.

So it was no accident that $\dfrac{\partial^2 f}{\partial x\,\partial y} = \dfrac{\partial^2 f}{\partial y\,\partial x}$ in the example above.

5. Reprise

Let $x_i = a_i + h$ and then notice

$$\lim_{x_i \to a_i} \frac{f(a_1, \cdots, a_{i-1}, x_i, a_{i+1}, \cdots, a_n) - f(a_1, \cdots, a_n)}{x_i - a_i} = \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{x}=\mathbf{a}} = \lim_{h \to 0} \frac{f(\mathbf{a} + h \cdot \mathbf{e}_i) - f(\mathbf{a})}{h}$$

where $\mathbf{e}_i = \langle 0, 0, \cdots, 0, 1, 0 \cdots, 0\rangle$ where the 1 is in the i^{th} position.

If we rewriting the formula as above,it suggests reinterpret the partial derivative as the *instantaneous rate of change of f in the direction of $\boldsymbol{e}_i$*. Given any direction $\mathbf{u}$, we can think about the *instantaneous rate of change of f in the direction of $\boldsymbol{u}$* as

$$\lim_{h \to 0} \frac{f(\mathbf{a} + h \cdot \mathbf{u}) - f(\mathbf{a})}{h}$$

In general, why restrict to unit vectors? Why not think about

$$\lim_{h \to 0} \frac{f(\mathbf{a} + h \cdot \mathbf{v}) - f(\mathbf{a})}{h}$$

for any vector $\mathbf{v}$?

As $h \to 0$, $h \cdot \mathbf{v} \to \mathbf{0}$ which suggests that we might want to understand

$$f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a})$$

as $\mathbf{v} \to \mathbf{0}$.

A function f is *differentiable at a point* $\boldsymbol{a}$ provided

$$\lim_{\mathbf{v} \to \mathbf{0}} \frac{\left| \big(f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a}) \big) - \nabla f(\mathbf{a}) \bullet \mathbf{v} \right|}{|\mathbf{v}|} = 0$$

A vector function $\mathbf{F}(\mathbf{x}) = \langle F_1(\mathbf{x}), \cdots, F_m(\mathbf{x}) \rangle$ is defined to be *differentiable at a point* $\boldsymbol{a}$ if and only if each function F_i is differentiable in the above sense.

In general, it is difficult to tell if a function is differentiable. Fortunately there is a theorem which handles the cases we care about.

Theorem: *If ∇f exists and is continuous in a neighborhood of $\boldsymbol{a}$ then f is differentiable in a neighborhood of $\boldsymbol{a}$.*

This is good since many theorems coming up depend on knowing if a function is differentiable.

Corollary: *If f is differentiable in a neighborhood of $\boldsymbol{a}$ then*

$$\lim_{h \to 0} \frac{f(\boldsymbol{a} + h \cdot \boldsymbol{v}) - f(\boldsymbol{a})}{h} = \nabla f(\boldsymbol{a}) \bullet \boldsymbol{v}$$

A really big corollary is the Chain Rule which is discussed in the next handout.