Multivariate calculus

Last lecture, we looked at the rate of change of a function with one variable, $f(x)$. However, functions can be more complicated than this and may have many variables. We'd still like to study the rate of change of this *multivariate* function, but now we have to take care to specify which variables are going to change. Although this obviously adds extra dimensions of complexity, in practical terms it's actually fairly simple. Suppose we have the function $f(x_1, x_2, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}^m$ (for integers $n, m \geq 1$) and we want to differentiate with respect to x_1 (or your favourite index of choice). Then, we use the concepts and derivative equation that we derived in the last lecture and momentarily treat f as a single-variable function by fixing the other variables:

$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1 + h, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{h} \,. \tag{1}$$

You will notice that the notation has changed slightly, with $\partial f / \partial x_1$ having the curly d, pronounced "del" or "partial" and we call this the **partial derivative**.

Anyway, obviously we can also define the partial derivative with respect to any variable x_i. It is sometimes useful to write a shorthand

$$f_{x_i} = \frac{\partial f}{\partial x_i} \,. \tag{2}$$

Here is an example:

$$f(x, y) = \sin x \cos y \tag{3}$$

and $f_x = \cos x \cos y$ whereas $f_y = -\sin x \sin y$. Here is another example

$$f(x, y) = e^{x^2 y} \tag{4}$$

then $f_x = 2xy \exp(x^2 y)$ and $f_y = x^2 \exp(x^2 y)$.

The chain rule for partial derivatives

Suppose that $z = f(x, y)$ is a differentiable function where $x = g(t)$ and $y = h(t)$ are both differentiable functions of t. Then

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

Nabla

Obviously, $\partial f/\partial x$ gives the rate of change in the x direction and $\partial f/\partial y$ the rate of change in the y direction. It will often be useful to put these together as a vector, this is called the gradient:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \tag{5}$$

The symbol ∇ is often just called "the gradient operator", it is also called "nabla" the name given to it by the mathematician Peter Tait; sometimes it is called "del". The symbol, and the concept of the gradient was introduced by the nineteenth century Irish mathematician William Rowan Hamilton, how also invented the four-dimensional generalization of complex numbers called quoternions. He used the nabla symbol because it is just the Greek letter capital delta, Δ, upside down and was therefore easy for typesetters; the name nabla is from the Greek word for harp. Anyway, whatever its name, here is an example, if $f(x, y) = 2x^3y^2 + y^2$ then

$$\nabla f = (6x^2y^2, 4x^3y + 2y) \tag{6}$$

Often we are interested in how a function $f(x, y)$ changes in some direction that isn't specifically the x or y directions, for this there is the concept of the **derivative along a vector**:

$$\nabla_{\mathbf{w}} f(x, y) = \lim_{h\to 0} \frac{f(x + hw_1, y + hw_2) - f(x, y)}{h} \tag{7}$$

where $\mathbf{w} = (w_1, w_2)$; often a unit vector is used and then we would refer to the *derivative in the direction of* $\mathbf{w}$. Without proving any theorem, you can hopefully see intuitively that the rate f changes along $\mathbf{w}$ is the rate f is changing in the x direction by the amount of $\mathbf{w}$ in the x direction plus the rate f is changing in the y direction by the amount of $\mathbf{w}$ in the y direction. In short, it can be proved that:

$$\nabla_{\mathbf{w}} f(x, y) = w_1 \frac{\partial f}{\partial x} + w_2 \frac{\partial f}{\partial y} \tag{8}$$

or, written using the dot product

$$\nabla_{\mathbf{w}} f(x, y) = \nabla f \cdot \mathbf{w} \tag{9}$$

This leads to a nice interpretation of the gradient. For two vectors **u** and **v** the dot product is given by

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos \theta \tag{10}$$

where θ is the angle between **u** and **v**. Thus, for given lengths, the maximum dot product is when the two vectors point in the same direction. Now since

$$\nabla_{\mathbf{w}} f(x, y) = \nabla f \cdot \mathbf{w} \tag{11}$$

this means the direction along which f changes most is the direction of ∇f, in other words, the gradient points in the direction of highest rate of change. If we are thinking of $f(x, y)$ as giving the height of some landscape over coordinates x and y, this means ∇f "points straight up the hill".

Extrema

Just as in one dimension we define a critical point as $df/dx = 0$ and all extrema are critical points; in higher dimensions a critical point is one where all the partial derivatives are zero. We saw that

$$\nabla_{\mathbf{w}} f = \mathbf{w} \cdot \nabla f \tag{12}$$

and so, given that the rate of change at an extremum must be zero in all directions, all extrema are critical points in higher dimensions too.

Distinguishing between maxima and minima is a small bit more complicated in higher dimesions; it relies on the *Hessian*, the matrix of second-order partial derivatives[1]. Here we will look at two dimensions as an example, though the idea is similar for other numbers of dimensions. In two dimensions the Hessian is

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \tag{13}$$

or, using the obvious notation

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \tag{14}$$

[1] A nice reference for this section is Section 3.7 of the book *Multivariable Calculus* by Rolland Trapp, available in the libaray

Also, note that for all but very weird functions $f_{xy} = f_{yx}$.

Now imagine the Hessian at a critical point looks like

$$H = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \tag{15}$$

then it would seem that we have a maximum since the x and y derivatives are falling at this point, just like the one-dimensional case. More generally though there are non-zero f_{xy} so what is important is not the diagonal elements but the eigenvalues. If both eigenvalues of the Hessian are negative at a critical point, then it's a maximum, if both are positive it is a minimum. There is another case, where one is positive and one is negative, that implies the critical point is a minumum in one direction and a maximum in another; that can happen in two dimensions and corresponds to a *saddlepoint*, like the point where two hills join and so the land goes up in one set of directions but down in another. Of course, there is always the possibility of a zero eigenvalue, this is like the one-dimensional case where the second derivative is zero, we could still have an extremum, but a very flat one, or it might not be an extremum at all.

This all seems to imply you need the eigenvalues of the Hessian to classify two-dimensional critical points and certainly if you have the eigenvalues and you can do the classification, however, you actually only need the signs of the two eigenvalues and you can usually work that out without actually knowing the eigenvalues. If λ_1 and λ_2 are the eigenvalues of a matrix H then

$$\det H = \lambda_1 \lambda_2 \tag{16}$$

and

$$\mathrm{trH} = \lambda_1 + \lambda_2 \tag{17}$$

There are known as matrix invariants. It means that if the determinant is negative you have a saddlepoint, if it is positive then the two eigenvalues have the same sign, so if the trace is negative you have a maximum, if it is negative, a minimum. If the determinant is zero then there is a zero eigenvalue and the Hessian alone isn't enough to tell you what is happening.

Lets do an example:

$$f(x, y) = x^3 + x^2 y - y^2 - 4y \tag{18}$$

Now $f_x = 3x^2 + 2xy$ and $f_y = x^2 - 2y - 4$. These equations are often a pain, it is hard to come up with good examples were the equations for the critical points are easy to solve. In this example, the first equation looks promising:

$$3x^2 + 2xy = 0 \implies x(3x + 2y) = 0 \tag{19}$$

so the solutions are $x = 0$ or $x = -2y/3$. If $x = 0$ then the second equation gives $y = -2$; if $x = -2y/3$ the second equation is

$$4y^2/9 - 2y - 4 = 0 \implies 2y^2 - 9y - 18 = 0 \tag{20}$$

and this factorizes into $(2y + 3)(y - 6) = 0$ so the solutions are $y = -3/2$ and $y = 6$. Hence the three critical points are $(0, -2)$, $(-4, 6)$ and $(1, -3/2)$. Now lets work out the Hessian:

$$H = \begin{pmatrix} 6x + 2y & 2x \\ 2x & -2 \end{pmatrix} \tag{21}$$

so for the point $(0, -2)$

$$H = \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix} \tag{22}$$

and since this is diagonal matrix we can read off the eigenvalues, they are -4 and -2 and so this is a maximum. For $(-4, 6)$ the Hessian is

$$H = \begin{pmatrix} -12 & -8 \\ -8 & -2 \end{pmatrix} \tag{23}$$

and hence the determinant of the Hessian is -40 which is negative and there is a saddle point. For $(1, -3/2)$ the Hessian is

$$H = \begin{pmatrix} 3 & 2 \\ 2 & -2 \end{pmatrix} \tag{24}$$

so the determinant is -10 and this is another saddle point.

For higher dimensions the story is much the same, but now there are more types of saddle point, for example in three dimensions there are saddle points with two up directions and one down and saddle points with one up direction and two down. These will correspond to two negative and one positive eigenvalues, or one negative and two positive eigenvalues for the Hessian.

Summary

The partial derivatives are the derivative with respect to one variable, to find the partial derivative you just treat the other variable as a constant. The gradient, denoted with a nabla, is the vector of partial derivatives:

$$\nabla f = (f_x, f_y) \tag{25}$$

using the notation

$$f_x = \frac{\partial f}{\partial x} \tag{26}$$

for example. The derivative along a vector $\mathbf{w}$ is

$$\nabla_{\mathbf{w}} f = \mathbf{w} \cdot \nabla f \tag{27}$$

The gradient points in the direction of greatest rate of change.

In higher dimensions there is a critical point when the gradient is zero. To classify the critical point the Hessian is calculated, in two-dimension this is

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} \tag{28}$$

If this has two positive eigenvalues at a critical point then it is a minimum, two negative, a maximum, one of each, a saddle point. If one, or both, of the eigenvalues is zero then it could be any of the three, or it might not be an extremum at all. The sign of the eigenvalues can be figured out from the determinant and trace since the determinant is the product of the eigenvalues and the trace is the sum.