

Crash Course: Multivariable Calculus

BRYAN LU

September 2019

The A Team employs a relatively frequent use of multivariable calculus in its study of physics. This first lecture is intended to offer the necessary mathematical knowledge about multivariable calculus needed to follow the majority of A Team lectures. New mathematical techniques will be introduced as necessary preceding new topics.

Assume we are strictly working over the real numbers unless otherwise stated.

1 Analyzing Functions of More Than One Variable

The techniques discussed here generalize to functions of arbitrarily many variables, but for ease of visualization we will limit ourselves to functions of two variables.

Before we define the multivariable concept of a derivative, it's first appropriate to discuss what limits look like in multiple variables. To do this, we'll intuitively discuss what the rigorous definition of a limit looks like in one-dimensional BC Calculus, and then build from there.

Recall the following word hash:

Theorem 1 (Limit Definition)

The *limit* of the function f as it approaches a is said to be the constant L if for every $\varepsilon > 0$, there exists some δ such that if $0 < |x - a| < \delta$, $|f(x) - L| < \varepsilon$.

Let's unpack what this is really saying. Suppose we have some function $f(x)$ and we're examining the neighborhood around $x = a$, and the neighborhood around $f(x) = L$. To give these neighborhoods a definite size, say we're only looking a distance of up to δ around $x = a$ and ε around $f(x) = L$. If we can shrink or grow ε at will – specifically, given that we can make ε arbitrarily small, if we can always find points on the function (within some δ away) that encompass those points, the function tends to the limit L as x is said to approach a . Notice that the absolute value is important on $|x - a|$, in the sense that this must approach the same value on both sides of $x = a$ – and obviously, if it doesn't, the limit L cannot take on two different values simultaneously, so it cannot exist.

This perspective can be generalized somewhat to multivariate functions. Suppose we have a function of two variables, $f(x, y)$, and we are examining the neighborhood around $f(x, y) = L$ and $(x, y) = (a, b)$. In this case, our neighborhood in our range space will be ε around $f(x, y) = L$, and the neighborhood in our domain space will be a disk of radius δ around (a, b) . As we shrink ε , we must always have some disk of radius δ whose outputs are contained in ε above and below L . Notice that as we approach the point (a, b) , we must pass through every single possible disk of radius δ corresponding to some ε . So, if while traversing

different two paths to (a, b) we approach different values, the limit cannot exist. In particular, we must have that **all** paths approaching (a, b) must obtain this value, and the fact that infinitely many paths approach this point satisfy this condition is actually insufficient. This is a *much* more stringent restriction on limit existence that makes continuity of a multivariate function (ie. having the limit be the same value as the value of the function) much more difficult to achieve. (For example, consider the function $f(x, y) = \frac{x^2 y}{x^4 + y^2}$ as one approaches $(0, 0)$.) This should be contrasted with our notions of differentiability and partial differentiability as we define them later.

The notion of a partial derivative is defined with a regular limit:

Theorem 2

For a function $f(x, y)$, the *partial derivative* of f with respect to x is given by

$$\frac{\partial f}{\partial x} = f_x(x, y) = \lim_{u \rightarrow x} \frac{f(u, y) - f(x, y)}{u - x}.$$

The idea here is that **ALL** other variables **EXCEPT** the variable that the derivative is being taken with respect to is being held constant. To emphasize here that y is being held constant, one may even see this written as:

$$\left(\frac{\partial f}{\partial x} \right)_y$$

With this structure, all the regular rules about differentiation still apply, since each variable is essentially treated independently of the rest.

We can also have higher-order partial derivatives, denoted similarly to higher-order regular derivatives, as well as mixed-partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} \quad \frac{\partial^n f}{\partial x^n} \quad \frac{\partial^2 f}{\partial x \partial y}$$

This latter notation for mixed partial derivatives can be a little confusing – in particular, $\frac{\partial^2 f}{\partial x \partial y} = f_{yx}$ means taking the partial derivative with respect to y first, then with respect to x , in which case I prefer the subscript notation a little more.

It turns out that if the function f is nice enough (ie. continuous and the mixed partials are continuous), the order doesn't actually matter – a result we will refer to by *Clairaut's Theorem* (although other sources will call it Schwarz's Theorem, even though Schwarz already has so many things named after him). To be specific, if the mixed partials of f are continuous, then they in fact are the same, or $f_{xy} = f_{yx}$. The reason this is a non-trivial statement is because it's not obvious that the limits can be done in either order, which is probably best illustrated with an example.

Consider the rather nasty-looking function $f(x, y) = \frac{xy(5x^2 - 2y^2)}{3x^2 + y^2}$, and suppose we would like to find $f_{xy}(0, 0)$ and $f_{yx}(0, 0)$. It would be rather terrible to take derivatives of this function explicitly, so we can instead consider using the limit definition explicitly.

Suppose we are differentiating with respect to x and then to y . In that case, we take

$$f_{xy}(0, 0) = \lim_{v \rightarrow 0} \frac{f_x(0, v) - f_x(0, 0)}{v - 0}$$

These expressions in the numerator need to be evaluated explicitly. Thus, we consider

$$f_x(0, v) = \lim_{u \rightarrow 0} \frac{f(u, v) - f(0, v)}{u} = \lim_{u \rightarrow 0} \frac{\frac{uv(5u^2 - 2v^2)}{3u^2 + v^2} - 0}{u} = -2v$$

so that

$$f_{xy}(0,0) = \lim_{v \rightarrow 0} \frac{-2v}{v} = -2.$$

Similarly, we have that

$$f_{yx}(0,0) = \lim_{u \rightarrow 0} \frac{f_y(u,0) - f_y(0,0)}{u - 0}$$

and we thus have to evaluate $f_y(u,0)$:

$$f_y(u,0) = \lim_{v \rightarrow 0} \frac{f(u,v) - f(u,0)}{v} = \frac{5}{3}u$$

so it turns out that $f_{yx} = \frac{5}{3}$. In this case, we should have known that these expressions weren't going to be equal in advance as the mixed partials can't be continuous at $(0,0)$ if the original function isn't even continuous at $(0,0)$! This is just one of an infinitude of examples where the partials have well-defined values at a point where the function itself is *neither continuous nor differentiable*.

We are obligated to talk differential elements that kind of went ignored in BC alongside our generalized notion of differentiability. These are the dx s that were kind of thrown around that we will pay more attention to. Essentially, recall the local linearity approximation in BC:

$$df = f'(x) dx$$

which came from the tangent line approximation to a function at (x_0, y_0) , that looks like

$$y - y_0 = f'(x_0)(x - x_0).$$

Of course, we don't expect this approximation to be exact - the actual difference in f has an additional error term that we ignore because it usually provides a minimal difference in the second-order.

How do we extend this to multiple variables? With two independent variables, we have a *tangent plane* approximation. Suppose we have a function $f(x, y)$ and we wish to find the tangent plane at (x_0, y_0) if $f(x_0, y_0) = f_0$. We can write any plane passing through this point in the form

$$z - f_0 = A(x - x_0) + B(y - y_0)$$

for some A, B . If we were to hold one variable fixed at a time, and changed the value of the variable, it should be apparent that these coefficients naturally should be the partial derivatives with respect to x and y . In particular, we would have

$$z - f_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

and in order to extend the idea of a differential element to multiple variables, we simply make these differences in x, y, f infinitesimals:

$$df = f_x dx + f_y dy.$$

df here is the differential element of f , written in terms of small differentials of its independent variables, x and y .

As a useful sidenote, if we denote an infinitesimal displacement $d\vec{r} = dx\hat{i} + dy\hat{j}$ as it separates the two different ways we could vary a variable into coordinate components, we can write df as the dot product of another vector ∇f and $d\vec{r}$:

$$df = \nabla f \cdot d\vec{r}$$

where $\nabla f = f_x\hat{i} + f_y\hat{j}$. This vector we called the *gradient* of f and it encompasses how f changes with respect to each of our variables. The upside-down triangle ∇ is called the *del* operator or the *nabla* operator, and we will see it used a lot.

The gradient vector is useful as it allows us to define a directional derivative for the function f , $D_{\hat{u}}f$. To motivate this, consider first when $df = 0$ when moving along some infinitesimal step $d\vec{r}$. This gives

$$\nabla f \cdot d\vec{r} = 0.$$

As we move along curves of constant f , the gradient vector and the steps $d\vec{r}$ are orthogonal (perpendicular). If we generally define the directional derivative

$$D_{\hat{u}}f = \nabla f \cdot \hat{u},$$

where \hat{u} is the direction along which we are taking said derivative, whenever \hat{u} is perpendicular to ∇f , $D_{\hat{u}}f = 0$, which makes sense. Moreover, this definition shows that if we move in the same direction as the gradient, the directional derivative will be maximized, making it so that the function will increase most quickly – or, that the gradient of f points in the direction of greatest increase of f . Notice that if \hat{u} happens to be \hat{i} or \hat{j} , we get f_x and f_y , respectively! In general, then, it's useful to think of the directional derivative of "how fast a function is increasing/decreasing in some direction," which generalizes the partial derivatives along the coordinate axes to changes along general directions. This intuition, however, can lose meaning as we move up in dimensions.

That digression aside, here is a rigorous definition for what it means for a function to be differentiable:

Theorem 3 (Differentiability)

The function $f(x, y)$ is said to be differentiable at (x_0, y_0) if when decomposing $f(x, y)$ such that

$$f(x, y) = f(x_0, y_0) + a(x - x_0) + b(y - y_0) + \xi(x, y)$$

where given any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$ implies $|\xi(x, y)| < \varepsilon \sqrt{(x - x_0)^2 + (y - y_0)^2}$.

If this is true, then $a = f_x(x_0, y_0)$ and $b = f_y(x_0, y_0)$.

Again, that is a lot to unpack. It might be easiest to see this if we suppose a function f is differentiable, in which case, after some rearranging, we have

$$f(x, y) - f(x_0, y_0) = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + \xi(x, y)$$

or, if we let (x, y) approach (x_0, y_0) , we can informally say that

$$\Delta f = f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + \xi(x, y)$$

Notice the similarity to how we defined the differential element for f , but instead having Δ s instead of the proper differential element d . ξ here is the error term that appears similarly in regular calc – when we do a linear approximation, the error term is explicitly everything after the linear term in the Taylor expansion around that point, although it's not good to think that Taylor can be generalized in this sense that easily.

If we wanted to convert this linear approximation perspective into the proper differential statement by taking the limit, we would necessarily want the error function $\xi(x, y)$ that swept up the rest of the mess of the function to vanish as we approach (x_0, y_0) , which is of order greater than linear (as all the linear terms are encompassed in the partials). The condition is essentially saying this – if we want this to vanish in the limit, we want it to vanish faster than linear in ε , or we want $|\xi(x, y)|$ to be bounded by faster than ε .

We may generalize the notion somewhat and relax the constraint that all of the other variables have to be held constant – for example, we could consider $f(x, y) = x^2y^3$, and computing $\frac{\partial f}{\partial x}$ while holding xy^2 constant.

In that case, we consider the differential of f :

$$df = 2xy^3 dx + 3x^2y^2 dy$$

and we consider the differential of xy^2 :

$$d(xy^2) = y^2 dx + 2xy dy$$

If we let xy^2 be constant, we know that

$$y^2 \partial x + 2xy \partial y = 0 \implies \left(\frac{\partial y}{\partial x} \right)_{xy^2} = -\frac{y}{2x}$$

If we divide through by dx , we have

$$\left(\frac{\partial f}{\partial x} \right)_{xy^2} = 2xy^3 + 3x^2y^2 \left(\frac{\partial y}{\partial x} \right)_{xy^2} = 2xy^3 + 3x^2y^2 \cdot -\frac{y}{2x} = \frac{xy^3}{2}$$

which is a surprising result! We will do stuff like this later when we do thermodynamics.

2 Multiple Integrals

Before talking about the physically important subject of vector fields, we do have to talk about multiple integrals – specifically, double and triple integrals. In this case, we are integrating over regions of space or on a plane or a surface, not just over the real line. Generally, this is done by parameterizing the space we are integrating over, usually with two or more integrals, and then integrate first with respect to one variable and then with respect to another. Usually, we will assume we can interchange the order of integration at will as long as we change the limits appropriately – this is *Fubini's Theorem* which is a nontrivial result which states when the limits present in the Riemann sums can be interchanged, like Clairaut.

3 On Vector Fields and Their Behavior

We define a *vector field* as a function that assigns a vector to every point in space. These are commonly seen in electric/magnetic fields, or when modeling wind/fluid flow or currents. The gradient itself is also a vector field, associated with a scalar field (better known as a multivariate function).

Often, we would like to describe vector fields and how they behave locally (specifically, how they rotate around a point or diverge from a point), and we motivate this with several different types of integral operations we may perform with vector fields.

3.1 Line Integrals and Curl

Line integrals are most lucrative for us when we are integrating a vector field \vec{F} over a curve C , parameterized by $\vec{r}(t)$. You may have seen this when computing the work done to an object along a path. We essentially take infinitesimal steps along C , taking the dot product of the displacement vector with the vector field at each step.

If it turns out that \vec{F} is a gradient of some function $f(x, y, z)$, ie. $\vec{F} = \nabla f$, we have the *Fundamental Theorem of Line Integrals*:

$$\int_C \vec{F} \cdot d\vec{r} = \int_C \nabla f \cdot d\vec{r} = f \Big|_{end}^{start}$$

which is very similar to the ordinary Fundamental Theorem of Calculus, in which the integral is equal to evaluation of the "anti-derivative" at the boundary of the curve. This implies that this particular line integral is path-independent, which is rather nice. We have a special name for such vector fields \vec{F} if they are a gradient – we call them *conservative*, like conservative forces. With this theorem, you can convince yourself that if we integrate a conservative vector field \vec{F} around any loop in space, we have

$$\oint_C \vec{F} \cdot d\vec{r} = 0.$$

What if we are doing loop integrals of general vector fields? It turns out that we can turn a loop integral around a planar loop into double integrals via *Green's Theorem*, which states that if the vector field $\vec{F} = P\hat{i} + Q\hat{j}$, then we have that

$$\oint_C \vec{F} \cdot d\vec{r} = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA.$$

This double integral may be easier to evaluate in certain cases if integrating the left hand side is difficult, and vice versa.

We can check that this makes sense by looking at the right hand side for conservative vector fields. If we know that \vec{F} is a gradient, then Clairaut's Theorem is satisfied, and so the two terms being subtracted from each other have to be the same.

The $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ term on the right-hand side of Green's Theorem can be thought of as the "circulation" of the vector field \vec{F} around the loop, per unit area, that is being integrated over the entire loop. This is, however, oriented with respect to the xy -plane, so suppose we define this as the "curl" of \vec{F} , projected along the \hat{k} direction, as this is the corresponding normal to this plane. If we extend this to the \hat{i} and \hat{j} directions, we would have that if $\vec{F} = P\hat{i} + Q\hat{j} + R\hat{k}$, then

$$\text{curl } \vec{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \hat{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \hat{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \hat{k}$$

Notice how this looks a lot like a "cross product"-esque form! In fact, if we would like to abuse some notation a little, and define a vector operator $\nabla = \hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y} + \hat{k} \frac{\partial}{\partial z}$, then we can think of this as

$$\text{curl } \vec{F} = \nabla \times \vec{F}$$

Of course, this is not the proper way to actually think about this, but as its form suggests a cross product, the above $\nabla \times$ notation can be helpful to remember how to compute it. I will, however, opt for this notation because I don't want to type out "curl" everywhere I want to use it.

3.2 Surface Integrals and Divergence

Surface integrals are also something we may perform by integrating a vector field \vec{G} across a surface S . Over the entire surface, we usually would like to determine the flux of the vector field \vec{G} through the surface, or how much of \vec{G} is going through it. This is done by taking a dot product of \vec{G} with a normal vector to the surface at every point on the surface, and then integrating over both parameters. Essentially, if we think about the flux as "how much of the field is going out," this makes sense because we are projecting the vector field onto the "outward direction," and we are adding up all of the contributions of the vector field that go out of the surface, multiplied by the area of each small section of the surface.

How do we find this normal vector for every point on the surface? A surface can be fully parameterized with two independent coordinates, say, u and v , in the form $\vec{r}(u, v)$. We essentially take the differential of this parameterization with respect to u and v , $d\vec{r}_u$ and $d\vec{r}_v$ respectively, which will be two vectors tangent to this

surface. Their cross product will be normal to the surface, which will be our surface element $d\vec{S} = d\vec{r}_u \times d\vec{r}_v$. We then essentially perform a dot product and do the double integral with appropriate bounds on u and v , as usual.

This is generally difficult, but it turns out if \vec{G} is a curl, ie. $\vec{G} = \nabla \times \vec{F}$, its surfaces are independent of the surface, but only constrained by the loop that bounds the surface. In particular, this is described by *Stokes' Theorem*, which states that

$$\oint_C \vec{F} \cdot d\vec{r} = \iint_S (\nabla \times \vec{F}) \cdot d\vec{S} = \iint_S \vec{G} \cdot d\vec{S}$$

which is dependent on the line integral of \vec{F} around the loop enclosing the surface S , and nothing else. This is what we mean by having surface integrals being independent of the surface. In particular, if we integrate over a closed surfaces, and look at the flux through it, by splitting it into two oppositely-oriented surfaces, we can see that if \vec{G} is a vector field such that $\vec{G} = \nabla \times \vec{F}$ for some \vec{F} , then

$$\oiint_S \vec{G} \cdot d\vec{S} = 0.$$

Again, we have a special name for such vector fields \vec{G} – they are called *solenoidal*, like magnetic fields in a solenoid.

Like before, what if we want to do closed-surface integrals of a general vector field? It turns out that we can, just like in Green's Theorem, turn this surface integral into a volume integral across the interior of the surface. If one lets $\vec{G} = P\hat{i} + Q\hat{j} + R\hat{k}$, then we have that

$$\oiint_S \vec{G} \cdot d\vec{S} = \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dV = \iiint_V (\nabla \cdot \vec{G}) dV$$

This quantity is called the *divergence* of \vec{G} , and it is denoted $\nabla \cdot \vec{G}$ or $\text{div } \vec{G}$. Again, I will opt for the former. This quantity measures the degree to which a quantity is expanding out of a given volume, and we can see this intuitively if we make the volume and surface that we are integrating over infinitesimally small. Just as its symbol denotes, we can think of this as taking a "dot product" of sorts with the ∇ vector operator with \vec{G} , although that technically isn't proper.

With that, we have most of the necessary tools we need to deal with vector and scalar fields. Before we conclude, however, want to define the *Laplacian*, which is a sort of "second derivative" operator which is the divergence of the gradient, $\nabla \cdot \nabla$, or otherwise denoted ∇^2 . This essentially means to sum the second derivative of the desired function with respect to all the variables. It's also abused for vector fields as well, in which case the Laplacian is meant to act on each component individually. This operator will appear often going forward, and its meaning in every individual scenario will depend on the context.