HANDOUT FOUR: DIFFERENTIAL CALCULUS ON SURFACES

PETE L. CLARK

1. Analogies between curves in the plane and surfaces in space

We have just finished our study of plane and space curves, or, if you like, of vector-valued functions of a scalar variable. The next unit of the course focuses on the differential calculus of surfaces in three-dimensional space: these include the case of scalar-valued functions of a vector variable, i.e., $z = f(x, y)$. It turns out that that surfaces in $\mathbb{R}^3$ are most analogous to curves in the plane (and not to curves in $\mathbb{R}^3$): we saw an instance of this at the very beginning of the course when we noted that the equation of a line in the plane:

$$ax + by = C$$

and the equation of a plane in $\mathbb{R}^3$:

$$ax + by + cz = C$$

are highly similar, and both are different from the way we give equations for a line in $\mathbb{R}^3$: we must either give *two* equations of intersecting planes or (and this is the method that we have preferred) give three parameteric equations. The reason behind this is that a line – a one-dimensional object – has dimension one less than $\mathbb{R}^2$ and a plane – a two-dimensional object – has dimension one less than $\mathbb{R}^3$, so in each case there is a unique *normal line*, whereas there are infinitely many distinct directions for normal vectors to a line in $\mathbb{R}^3$.

Let us also recall the three different ways we have been dealing with curves in the plane:

1) As the graph of a function $y = f(x)$ (as done in one-variable calculus). E.g., the parabola $y = x^2$.

2) Implicitly defined by an equation in two variables, say $F(x, y) = C$ (this is also done in one-variable calculus, under the heading of **implicit differentiation**), good examples being the family of concentric circles

$$x^2 + y^2 = C$$

and the family of parallel lines

$$x + y = C.$$

Note that in the former case we can view the equation as implicitly defining y as a function of x, which we can make explicit as

$$y = \pm\sqrt{C - x^2},$$

but as the "±" indicates, it is not true "globally" that the circle is of the form $y = f(x)$: indeed near $y = 0$ we cannot write the circle in this way: it fails the vertical line test.

3) Parameterized curves $\mathbf{r}(t) = (x(t), y(t))$. In this way we get a much broader class of curves than just graphs of functions. (Later on the relationship between parameterized curves and implicitly defined curves $F(x,y) = C$ will be clarified.)

Similarly, there are three different ways to represent a surface in three-dimensional space:

1) As the graph of a function of two variables, $z = f(x,y)$. That is, we view the surface as living over the xy-plane (or a nice subset of it): above (or below) every point (x_0, y_0) of the plane there is a point (x_0, y_0, z_0), where $z_0 = f(x_0, y_0)$: the height of the point is given by the function $f(x,y)$.

2) Implicitly defined surfaces, or **level surfaces** of the form $F(x,y,z) = C$.

3) Parameterized surfaces: these are like parameterized curves but instead of tracing the path of one parameter t in space, we have two independent parameters u and v: that is, we have a vector-valued function of a vector variable:

$$\mathbf{R}(u,v) = (x(u,v), y(u,v), z(u,v)).$$

This includes as a very special case the surfaces of the first kind: we can take $x = u$, $y = v$, $z = f(u,v)$ and get that the surface just lies above or below the corresponding point in the uv-plane. But we can get far more complicated surfaces in this way: as a first example, consider the parameterized surface $\mathbf{R}(u,v) = (R\cos u, R\sin u, v)$, which is a right circular cylinder of radius R. This is also a level surface: $F(x,y,z) = x^2 + y^2 = R$. In general, when we have a level surface of the form $F(x,y,z) = G(x,y) = C$ with no z-variable, we get a cylinder on the corresponding level curve in the plane $G(x,y) = C$: the z-variable can be whatever it wants. (A similar statement holds for level surfaces which omit either the x or y variables.)

2. Examples of surfaces; some graphing

In this section we will acquaint ourselves with surfaces by looking at some examples, all in the family of **quadric surfaces**, which are level surfaces $F(x,y,z) = C$ where F is a polynomial of degree at most 2. The method we have for graphing level surfaces is given to us by the biologists: we consider various slices with planes parallel to the coordinate planes: first we take $z = c$ and try to figure out what the corresponding family of curves is (they will be different curves for different values of c but hopefully will have a common shape). Then we try to put these slices together in our mind to form the surface. If we are in doubt (which we usually are), we can consider slices $x = c$ and $y = c$. We should admit that this method certainly presupposes that we can rather easily figure out what the sliced curves $F(x,y,c) = C$ look like, but even graphing plane curves can be very difficult. (If the equations of a surface are even a little more complicated than the examples we consider here, the honest answer to the question of how we are supposed to get an idea of what the surface looks like is that we type the equation of the surface into a mathematical software package like Mathematica or Maple.)

Example: $z = x^2 + y^2$. Taking $z = c$ we get $x^2 + y^2 = c$ the equations of a family of concentric circles. Note that the radius of the slice at c is $\sqrt{c}$, which is *not* a linear function of c. As an aside, there are many, many surfaces whose $z = c$ slices are circles: take the graph of any function $z = f(x)$ and revolve around the z-axis, a **surface of revolution**[1] To figure out which surface of revolution we are getting we slice in another direction: taking $x = c$ we get $z = y^2 + c^2$: this is a family of parabolas in the yz-plane. Similarly taking $y = c$ we get $z = x^2 + c^2$ a family of parabolas in the xz-plane. So it was a parabola that we revolved around the z-axis to get the surface, called an **elliptic paraboloid**. Note that $z \geq 0$: the bottommost point of the paraboloid is $(0, 0, 0)$.

Example: $z^2 = x^2 + y^2$. Taking $z = c$ we get again a family of circles $x^2 + y^2 = c^2$. But this time the radius, c, is *equal* to the z-coordinate. With a little thought, you can probably convince yourself[2] that the circles will fit together to form a cone: fixing a point on a $z = c$ circle and sliding directly upwards on the surface, this point will trace out a straight line. To help us believe it, slice also in another direction $y = c$: $z^2 = x^2 + c^2$ or $z^2 - x^2 = c^2$ is a family of hyperbolas, exactly what we get by taking vertical slices of a right circular cone. Moreover, by taking the slice $y = 0$ we find out what shape we are revolving: at $y = 0$ we get $z^2 = x^2$, or $z = \pm x$: a pair of lines intersecting at the origin in the xz-plane: these do indeed generate a cone upon revolution about the z-axis.

Example: $z^2 = x^2 + y^2 - 1$. (This came up in Exercise 12 in Section 9.1) Taking $z = c$ we again get a family of circles $x^2 + y^2 = 1 + c^2$: note that the radius is always at least 1. Taking $y = c$ we get $z^2 - x^2 = (c^2 - 1)$ which is a family of hyperbolas. Taking $y = c = 0$ we get $z^2 - x^2 = -1$ or $x^2 - z^2 = 1$, a hyperbola "opening horizontally" in the xz-plane. Revolving this hyperbola about the z-axis is going to give us a funnel-shaped surface which is narrowest at $z = 0$, a so-called **hyperboloid of one sheet**.

Example: $z^2 = 1 + x^2 + y^2$. Again the slices $z = c$ are circles, this time of radius $\sqrt{c^2 - 1}$: note that all points on this surface have $|z| \geq 1$. Taking $y = c$ we get $z^2 - x^2 = 1 + c^2$, again a family of hyperbolas, but this time always "opening vertically" in the xz-plane. Especially at $y = 0$ we get the vertical hyperbola $z^2 - x^2 = 1$. Revolving this guy around the z-axis gives a surface which is in two pieces, a hyperbolic bowl-shaped thing and its mirror image. This is called a **hyperboloid of two sheets.**

Example: $z^2 = 1 - x^2 - y^2$. This surface has all of its coordinate slices equal to circles: in particular it is obtained by revolving the unit circle in the xz-plane about the z-axis, so it is the unit sphere. And of course it is, since the equation is $1 = x^2 + y^2 + z^2 = ||(x, y, z)||^2$: it is the locus of all points in $\mathbb{R}^3$ which are a unit distance from the origin.

[1]Recall that these come up in one-variable calculus: one computes the volume of the enclosed region.

[2]Since I can, and I am not very good at visualizing three-dimensional objects: trying to picture the intersection of two cylinders makes me want to lie down.

Example: $z = x^2 - y^2$. Note first that, in contrast to the previous examples, this is *not* a surface of revolution: indeed $(1, 0, 1)$ is a point on the surface, but rotating this point by $\pi/2$ about the z-axis gives $(0, 1, 1)$ and this is not a point on the surface. Slicing in the $z = c$ direction we get the family of hyperbolas $x^2 - y^2 = c$. Note that when $c > 0$ these hyperbolas open horizontally; when $c < 0$ they open vertically; at $c = 0$ we get the pair of lines $y = \pm x$. Taking $x = c$ slices we get $z = c^2 - y^2$: these are parabolas opening downward in the yz-plane. Taking $y = c$ slices we get $z = x^2 - c^2$: these are parbolas opening upward in the xz-plane. This surface is called (unenlighteningly enough!) a **hyperbolic paraboloid**, a term that merely reflects its possessing both hyperbolic and parabolic slices. The surface is doing something very interesting at the point $(0, 0, 0)$: to see what, observe that the map of the line $l_x = (t, 0)$ on the surface is the curve $\mathbf{r}_1(t) = (t, 0, t^2)$: this is (as above) a parabola with a **local minimum** at $t = 0$. On the other hand, the map of the line $l_y = (0, t)$ on the surface is $\mathbf{r}_2(t) = (t, 0, -t^2)$, which is a parabola with a **local maximum** at $t = 0$. This type of point – namely a point P_0 with the property that one curve passing through P_0 on the surface has a local minimum at P_0 and another curve has a local maximum at P_0 – is called a **saddle point** of a surface, and indeed "saddle-shaped" is a better description of this surface than hyperbolic paraboloid.

3. Some parameterized surfaces[3]

Example: let $\mathbf{w}_1 = (x_1, y_1, z_1)$ and $\mathbf{w}_2 = (x_2, y_2, z_2)$ be two nonzero vectors in $\mathbb{R}^3$. Consider the parameterized surface

$$\mathbf{R}(u, v) = u\mathbf{w}_1 + \mathbf{v}\mathbf{w}_2 = (ux_1 + vx_2, uy_1 + vy_2, uz_1 + vz_2).$$

This is just giving us every vexctor in $\mathbb{R}^3$ which can be obtained by stretching $\mathbf{w}_1$ and $\mathbf{w}_2$ and adding them together: in the language of linear algebra, we are getting the **span** of $\mathbf{w}_1$ and $\mathbf{w}_2$. This will be a plane in $\mathbb{R}^3$, as long as $\mathbf{w}_1$ and $\mathbf{w}_2$ point in different directions: if they were scalar multiples of each other, the plane degenerates to a line

Here's an example which begins to illustrate the power of parameterized surfaces: let $0 < a < R$ be positive numbers, and consider:

$$\mathbf{R}(u, v) = ((R + a\cos u)\cos v, (R + a\cos u)\sin v, a\sin u).$$

What on earth is this surface? First take $v = 0$: $\mathbf{R}(u, 0) = (R+a\cos u, 0, a\sin u) = (R, 0, 0) + (a\cos u, a\sin u)$. This is a circle of radius a living in the xz-plane and centered at the point $(R, 0, 0)$. On the other hand, if we took $u = 0$, we'd get: $\mathbf{R}(0, v) = ((R + a)\cos v, (R + a)\sin v, 0)$ and this is also a circle, of radius $R + a$ in the xy-plane, centered at the origin. So the surface has these two perpendicular circles lying on it and somehow interpolates between them. Now regard u as *fixed* and put $R_u := R + a\cos u$. Then $\mathbf{R}(u, v) = (R_u \cos v, R_u \sin v, a\sin u)$ describes a circle of radius R_u parallel to the xy-plane and centered at the point $(0, 0, a\sin u)$, i.e. its height is the z-coordinate of some point on the first circle we found. Putting together this information, we may be able to guess that what we are getting is the surface of revolution of the first circle around the z-axis (and having said this, it's

[3]We did not cover much from this section in class.

not to hard to see that the equations have the right symmetry for this to be true): that is, we are getting a **torus** – i.e., the surface of a doughnut! You can check that this surface satisfies the equation

$$F(x,y,z) = (\sqrt{x^2+y^2} - R)^2 + z^2 = a^2;$$

in particular, since the defining equation relates z to $r = \sqrt{x^2+y^2}$, it must be a surface of revolution (and it is).

Parametric equations of a general surface of revolution: Let $\mathbf{r}(t) = (x(t), z(t))$ be a parameterized curve in the plane.[4] Then the paramteric equations of the corresponding surface of revolution are

$$\mathbf{R}(u,v) = (x(u)\cos v, x(u)\sin v, z(u)).$$

If $F(x,z) = C$ is a defining equation for the parameterized plane curve, then $F(\sqrt{x^2+y^2}, z) = C$ expresses this surface as a level surface.

Extra credit: Consider the parameterized surface

$$\mathbf{R}(u,v) = (+v\cos(u/2), +v\cos(u/2)\sin u, \sin u/2)$$

defined for arbitrary u and $-1/2 < v < 1/2$. What is this parameterized surface? (You'll know when you get it, it's famous.)

4. Directional derivatives, partial derivatives and the gradient

We want now to bring differential calculus into the game. Suppose we give ourselves a surface of the first type, $z = f(x,y)$. The question of the day is: what should "the derivative" of $f(x,y)$ be, and what is its geometric meaning?

We choose to answer the second question first, using our prior study of derivatives of parameterized curves. Recall that the derivative of a curve in the plane or in space is its velocity vector, which keeps consolidated track of the rate of change of the particle in the x, y and z-directions with respect to time. Suppose that we are at a point $P_0 = (x_0, y_0, z_0) = (x_0, y_0, f(x_0, y_0))$ on the surface. Of course the height of the surface (the z-coordinate) is changing as we walk from P_0 to nearby points, but we cannot hope to measure this change with a single number, since the change in height depends upon the *direction* in which we walk. This can be seen even in the linear function $f(x,y) = x - y$. The surface $z = x - y$ is a plane which is oriented in such a way that if we walk due east, we go uphill at a rate of one unit (i.e., walking to the east by one meter increases our height by one). On the other hand, if we walk due north then our heigh decreases at a rate of one unit. (To check this, observe that the normal vector to this plane is $\mathbf{n} = (1, -1, -1)$.)

Thus it can only make sense to meaure the rate of change of $f(x,y)$ at P_0 in a certain direction. One way to formalize this is to consider curves on the surface: let $\mathbf{r}(t) = (x(t), y(t), 0)$ be any curve in the plane, which we think of as our choice of how to walk along the surface. We suppose that $\mathbf{r}(t)$ passes through the point (x_0, y_0) at time $t = 0$ – i.e., $x(0) = x_0$, $y(0) = y_0$ – and has unit speed, so

[4] It doesn't really matter, but it makes a little more sense to picture this as taking place in the xz-plane.

$\mathbf{r}'(0) = (x'(0), y'(0)) = \mathbf{u}$ is a unit vector. The curve $\mathbf{r}(t)$ gives the projection into the xy-plane of the path actually taken on the surface, which is

$$\mathbf{R}(t) = (x(t), y(t), z(t) = f(x(t), y(t)).$$

This curve has a velocity vector at time $t = 0$ which is of the form

$$\mathbf{R}'(0) = (x'(0), y'(0), z'(0)) = (x'(0), y'(0), 0) + z'(0)\hat{\mathbf{k}} = \mathbf{u} + dz/dt|_{t=0}\hat{\mathbf{k}}.$$

Notice that the x and y-components of the velocity vector are just what we started with – the unit vector $\mathbf{u}$ giving the direction. So the z-component, the dz/dt is giving us the instanteous rate of change of height as we walk in direction $\mathbf{u}$. So it makes sense to define this quantity to be the **directional derivative** of $z = f(x, y)$ in the direction $\mathbf{u}$:[5] we denote it as

$$D_{\mathbf{u}} f(x_0, y_0) = dz/dt_{t=0}.$$

"Example": Take $\mathbf{u} = (1, 0)$, and take the simplest possible curve with velocity vector $\mathbf{u}$ at time 0, namely $\mathbf{r}(t) = (x_0 + t, y_0)$. Then $\mathbf{R}(t) = (x_0 + t, y_0, f(x_0 + t, y_0))$, and the directional derivative in the direction of $\mathbf{u}$ is

$$D_{(1,0)} f(x_0, y_0) = \frac{df(x_0 + t, y_0)}{dt}|_{t=0}.$$

A real example: Suppose $f(x, y) = x^2 y + y^2$. Then we compute $D_{(1,0)} f(x_0, y_0)$ by differentiating $(x_0 + t)^2 y_0 + y_0^2$ with respect to t and setting $t = 0$: we get $2(x_0 + t)y_0 + 0$: setting $t = 0$ we get $2x_0 y_0$.

A time-saving device: Suppose instead of introducing the variable t, we just regarded $f(x, y_0)$ solely as a function of x, differentiated it with respect to x, and then set $x = x_0$. In this case, we'd get the derivative with respect to x of $x^2 y_0 + y_0^2$, which is $2xy_0$: at $x = x_0$ we get $2x_0 y_0$, just as before.

I claim that this "time-saving device" is always legitimate: i.e., that we can always compute $\frac{\partial f}{\partial x}$ at (x_0, y_0) by plugging in $y = y_0$, differentiating with respect to x, and then plugging in $x = x_0$. In other words, the directional derivative in the x-direction can be computed as

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}. \tag{1}$$

This is no problem: the derivative with respect to t of $f(x_0 + t, y_0)$ is

$$\lim_{h \to 0} \frac{f(x_0 + t + h, y_0) - f(x_0 + t, y_0)}{h} \tag{2}$$

but note that Equation 2 is exactly the same limit as in Equation 1 except we evaluate it at the point $(x_0 + t, y_0)$ instead of (x_0, y_0). But now we set $t = 0$ and get exactly the same expression.

To explain what is going on: it is traditional to *define* the partial derivative in the x-direction by Equation 1. However, we have defined it more geometrically as the z-component of the velocity vector of any curve passing through the point $P_0 = (x_0, y_0, f(x_0, y_0))$ and having horizontal velocity given by $(1, 0)$: happily, these turn out to be exactly the same thing.

[5] To be rigorous, we have not shown that this quantity is the same for any choice of path $\mathbf{r}(t)$ with $\mathbf{r}'(0) = \mathbf{u}$. This can – and will – be shown after we have the multi-variable chain rule.

There is an entirely analogous discussion for the directional derivative in the derivation $\mathbf{u} = (0,1)$: we take the curve $\mathbf{r}(t) = (x_0, y_0+t)$, so $\mathbf{R}(t) = (x_0, y_0+t, f(x_0, y_0+t))$, and

$$D_{(0,1)}f(x_0,y_0) = \frac{df(x_0, y_0+t)}{dt}|_{t=0}.$$

If we try this out on our example $f(x,y) = x^2y + y^2$, then this definition tells us to differentiate $x_0^2(y_0+t) + (y_0+t)^2$ with respect to t and then set $t = 0$: we get $x_0^2 + 2(y_0+t)$ and then $x_0^2 + 2y_0$. But again, it would be faster to just plug in $x = x_0$, differentiate with respect to y, and then plug in $y = y_0$: we'd get $d/dy(x_0^2y+y^2) = x_0^2+2y|_{y=y_0} = x_0^2+2y_0$. That is to say, the directional derivative $D_{(0,1)}f(x_0,y_0)$ as we defined it is always equal to

$$\frac{\partial f}{\partial y}(x_0,y_0) = \lim_{\Delta y \to 0} \frac{f(x_0, y_0+\Delta y) - f(x_0,y_0)}{\Delta y}. \tag{3}$$

What if we want the rate of change in a direction that is not $(1,0)$ or $(0,1)$? It turns out that $D_{\mathbf{u}}f(x_0,y_0)$ for any vector $\mathbf{u}$ can be computed easily in terms of $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. Indeed:

$$D_{\mathbf{u}}f(x_0,y_0) = \mathbf{u} \cdot (\frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial x}(x_0,y_0)). \tag{4}$$

We will come back to this formula – including its proof – after we have one more tool, which is the chain rule for functions of multiple variables.

By the way, everything done in this section has a precise analogue for functions of three or more variables (notwithstanding the fact that the geometric arguments would be taking place in $\mathbb{R}^4$ or more: there is no problem here!). In particular, if $f(x,y,z)$ is a function of three variables, then we define not only $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ but also $\frac{\partial f}{\partial z}$ in the expected way: namely by holding all variables constant but one and differentiating with respect to that variable.

Example: If $F(x,y,z) = x^2+y^2+z^2$, then $\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial y} = 2y$, $\frac{\partial f}{\partial z} = 2z$.

5. The Multivariable Chain Rule

Recall that if $y = f(x)$ and $x = x(t)$ is itself a function of t, then there is a very appealing formula telling us how to differentiate the composite function $y = f(x(t))$ with respect to t:

$$\frac{dy}{dt} = \frac{dy}{dx}|_{x(t)}\frac{dx}{dt}.$$

The bit with the bar in the middle is reminding us to evaluate at $x(t)$ after we differentiate with respect to x – our answer must be a function of t.

But consider now the following situation: we have $f(x,y)$ and $x = x(t)$ and $y = y(t)$ are **each** functions of t. Now things are not so clear: it is tempting to write $\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}$ but equally tempting to write $\frac{df}{dt} = \frac{df}{dy}\frac{dy}{dt}$. But they cannot both be right! (And by symmetry, neither one can be right.) The truth is the following, the

multivariable chain rule:

$$\frac{df(x(t),y(t))}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}. \tag{5}$$

Just as the definition of a partial derivative is "nothing new" – it boils down to ordinary derivatives of functions of one variable – the multivariable chain rule can be proven using the one-variable version. But the proof (at least the one I came up with) is rather confusing so is relegated to the end of the notes: I actually advise you not to read it.

Instead, we'll check that the chain rule is at least correct for linear functions.

Example: Suppose $x(t) = x_0 + at$, $y(t) = y_0 + bt$, and $f(x,y) = c_0 + dx + ey$, i.e., all linear functions. On the one hand, $\frac{df}{dx}\frac{dx}{dt} + \frac{df}{dy}\frac{dy}{dt} = da + eb$. On the other hand, $f(x(t),y(t)) = f(x_0 + at, y_0 + bt) = c_0 + d(x_0 + at) + e(y_0 + bt) = c_0 + dx_0 + ey_0 + (da + eb)t$, and clearly the derivative of this expression with respect to t is $da + eb$. Done!

One of the things which makes multivariable calculus fun is that we can sometimes use several variables in a slick way to solve problems in one-variable calculus. Here's an example that I still remember my high school calculus teacher, Dr. Joel Simon, showing me with delight[6].

Example: Consider the function $y = x^x$. What is its derivative? This is impossible to do directly in single-variable calculus: we know that $d/dx(x^a) = ax^{a-1}$ and that $d/dx(a^x) = (\ln a)a^x$, but we don't have a formula when both the base and the exponent are varying! What they tell you to do in such courses is to logarithmically differentiate: consider instead $\ln y = \ln(x^x) = x\ln x$. Now differentiating both sides, we get $y'/y = \ln x + x(1/x) = \ln x + 1$, so $y' = y(\ln x + 1) = (\ln x)(x^x) + x^x = (\ln x)(x^x) + xx^{x-1}$. But Dr. Simon remarks that this formula looks like you couldn't decide whether which of the two above formulas to use to differentiate x^x so you applied both of them and added them together "to be safe." This sounds like nonsense but it is correct in this case because: write $z = x^y$ and $x = t, y = t$. Now $\frac{dz}{dt} = \frac{dz}{dx}\frac{dx}{dt} + \frac{dz}{dy}\frac{dy}{dt}$. But $\frac{dz}{dx}$ means compute the derivative by holding the exponent constant, so using the first formula above, and $\frac{dz}{dy}$ means compute the derivative by holding the base constant, so using the second formula above. Since $x = y = t$ we have $\frac{dz}{dt} = \frac{dy}{dt} = 1$, so the multivarible chain rule really does instruct us to add these two formulas to get the answer!

Extra credit: Use the same technique to find a formula for the derivative of $f(x)^{g(x)}$ and check your answer using logarithmic differentiation.

Back to business: we want to check Equation (4) for the directional derivative. We choose the simplest curve passing through (x_0, y_0) at time $t = 0$ and having

[6]This must have been in late 1993 or early 1994, when dinosaurs roamed the earth and having the Smashing Pumpkins album was considered *avant garde*.

horiztonal velocity vector $\mathbf{u} = (u_x, u_y)$, namely $\mathbf{r}(t) = (x_0 + u_x t, y_0 + u_y t)$, and then

$$D_{\mathbf{u}} f(x_0, y_0) = \frac{d}{dt}(f(x_0 + u_x t, y_0 + u_y t))|_{t=0}.$$

But using the chain rule, this derivative is just

$$\left(\frac{\partial f}{\partial x}(x_0 + u_x t)(x_0 + u_x t)' + \frac{\partial f}{\partial y}(y_0 + u_y t)(y_0 + u_y t)'\right)|_{t=0} =$$

$$\frac{\partial f}{\partial x}(x_0, y_0)u_x + \frac{\partial f}{\partial y}(x_0, y_0)u_y = (\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0)) \cdot \mathbf{u},$$

as we wanted to show.

While we're here, we might as well show something that was slipped under the rug earlier, namely that the directional derivative does not depend on the **choice** of curve $\mathbf{r}(t)$ so long as $\mathbf{r}(0) = (x_0, y_0)$ and $\mathbf{r}'(0) = \mathbf{u}$. (In the applications, we've been taking straight line parameterizations, but we've claimed the result for any curve.) Indeed let $\mathbf{r}(t)$ be any such curve, and then

$$D_{\mathbf{u}} f(x_0, y_0) = \frac{d}{dt}(f(x(t), y(t))|_{t=0} =$$

$$\frac{\partial f}{\partial x}(x(t), y(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(x(t), y(t))\frac{dy}{dt}|_{t=0} =$$

$$\frac{\partial f}{\partial x}(x_0, y_0)u_x + \frac{\partial f}{\partial y}(x_0, y_0)u_y = (\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0)) \cdot \mathbf{u}.$$

That is, we get the same formula for the directional derivative no matter what curve $\mathbf{r}(t)$ we chose.

Higher order partial derivatives: just as we can differentiate a derivative we can take partial derivatives of partial derivatives. The only problem is that the notation gets rather hairy: e.g. we write $\frac{\partial^2 f}{\partial x \partial y}$ for $\frac{\partial}{\partial x}(\frac{\partial}{\partial y})f$. The operator notation is irrestistible: just as in single-variable calculus we begin to think of $\frac{d}{dt}$ as a free-floating object which may be "applied" to any function $f(t)$, we start thinking the same way about partial derivatives, viewing expressions such as $\frac{\partial^2}{\partial x \partial y}$ as just waiting for a function $f(x, y)$ to be input.

Actually, this reminds me a of a (terrible!) joke. A gathering of functions is crashed by a rogue band of differential operators, who cackle gleefully, "I'll differentiate you down to nothing!" Most of the functions run for their lives, but one differential operator notices that there is a function who is sitting calmly on the couch amid the turmoil. "Are you mad? Aren't you afraid I'll differentiate you down to nothing – or at least, to the point where even your mother wouldn't recognize you?" she asks. The function props his feet up and announces smugly, "No, of course not, I'm e^x. Do your worst, I won't even notice." So the operator sits down, throws an arm around e^x and whispers into his ear, "Glad to meet you, sweetheart. I'm $\frac{\partial}{\partial y}$."[7]

Example: $f(x, y) = e^x \cos y$. Then $\frac{\partial f}{\partial x} = e^x \cos y$ – the joke prepared you for

[7]Question following the joke: if all the functions and all the operators are of just two variables x and y, can you think of a function that really has nothing to fear?

this!. On the other hand $\frac{\partial f}{\partial y} = -e^x \sin y$ (is it getting funnier yet?). We can also compute $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(-e^x \sin y) = -e^x \sin y$ and $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(e^x \cos y) = -e^x \sin y$. Notice that $= \frac{\partial^2 f}{\partial y \partial x}$. This is no accident, but will almost always be the case: for instance, assuming that all the third-order partial derivatives of f exist is more than enough to ensure the equality of the corresponding **mixed partials** of second order, and a similar statement has to hold for higher-order partial derivatives. Also we have $\frac{\partial^2 f}{\partial x^2} = e^x \cos y$ and $\frac{\partial^2 f}{\partial y^2} = -e^x \cos y$. Notice this means that $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$. This differential operator has a name and a symbol:

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

is called the **Laplacian** of f. The equation

$$\Delta f = 0$$

is called **Laplace's Equation**, and a function satisfying it is said to be **harmonic**. Harmonic functions come up all over the place in pure and applied mathematics: see Chapter 19 of your text. By the way, Laplace's equation is an example of a **partial differential equation**, or PDE.[8] Other famous examples are **the wave equation**, **the heat equation** and – very closely related to Laplace's Equation, the **Cauchy-Riemann equations**.

6. Everyone loves Gradient

The vector $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ appearing in our formula for the directional derivative comes up again and again in the subject, so we give it a name: the **gradient** of $f(x, y)$, and a new symbol: if $f(x, y)$ is a differentiable function of two variables, then

$$\nabla(f) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}).$$

This makes equally good sense for functions of three (or more) variables: if $f(x, y, z)$ is differentiable then

$$\nabla(f) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}).$$

The gradient ∇ is sometimes thought of in its own right as an operator: it takes as input a scalar-valued function of several variables and returns a vector-valued function in the same number of variables.

What does $\nabla(f)(x_0, y_0)$ itself measure? And why is it called the "gradient"? This has a very satisfying answer: if $z = f(x, y)$ is the associated surface, suppose we start at $P_0 = (x_0, y_0, f(x_0, y_0))$ and want to walk in such a direction so that we ascend or descend the steepest: for instance, if we pour water onto the surface, the water, acted on by the force of gravity, will naturally run in the direction of steepest descent. It turns out that the gradient precisely measures this: indeed

[8] There are three types of people in the world: people whose eyes light up when you say "PDE," people who cringe and try to change the subject, and (the largest class, alas) people who have no idea what you are talking about. I have to admit that I belong more to the second class than the first.

what we want is to find the direction $\mathbf{u}$ that maximizes the absolute value of the directional derivative $D_{\mathbf{u}}f(x_0, y_0)$, and

$$|D_{\mathbf{u}}f(x_0,y_0)| = |\nabla(f)(x_0,y_0)\cdot\mathbf{u}| = ||\nabla(f)(x_0,y_0)||||\mathbf{u}||\cos\theta = ||\nabla(f)(x_0,y_0)|||\cos\theta|$$

where θ is the angle between $\nabla(f)$ and $\mathbf{u}$. This expression is maximized when $|\cos\theta| = 1$, so when $\theta = 0$ or π, i.e., when $\mathbf{u}$ points either in the direction of $\nabla(f)$ or in precisely the opposite direction.

Example (rapid ascent on a sphere): Suppose you are at the point $P_0 = (x_0, y_0, z_0)$ on an upper hemisphere of radius R. Which direction should you go to get to the top the fastest? To the bottom the fastest?

Solution: The equation of a sphere is $z^2 = 1-x^2-y^2$; to be on the upper hemisphere we take $z = \sqrt{1-x^2-y^2}$. The gradient is

$$\nabla(f) = \left(\frac{-2x}{2\sqrt{1-x^2-y^2}}, \frac{-2y}{2\sqrt{1-x^2-y^2}}\right)$$

In particular, its direction at (x_0, y_0) is given by $(-x_0, -y_0)$. Since $\nabla(f)\cdot(-x_0, -y_0)$ is positive, it is indeed the direction $(-x_0, y_0)$ that gives the steepest ascent, whereas the opposite direction (x_0, y_0) will give the steepest descent. But you should check that if we were on the lower hemisphere – so $z = -\sqrt{1-x^2-y^2}$ – exactly the opposite would be true.

7. Tangent planes and normal lines

Just as a curve has a tangent line at a given point, a smooth surface will have a **tangent plane** at a given point. In this section we discuss how to find this tangent plane, depending upon whether our surface is given 1) as the graph of a function $z = f(x, y)$, 2) as a level surface $F(x, y, z) = C$ or parametrically.

Recall that for a single-variable function $y = f(x)$ the tangent line at $P_0 = (x_0, y_0) = (x_0, f(x_0))$ is the best linear approximation to f at P_0: that is, it is the unique linear function $l(x)$ such that $l(x_0) = y_0 = f(x_0)$ and the derivative of l is (constantly) equal to $f'(x_0)$. Explicitly,

$$l(x) = f(x_0) + \frac{\partial f}{\partial x}(x - x_0).$$

If we stare at this formula a bit and think of $\nabla(f)$ as "the derivative" of $f(x, y)$, then it is conceivable to guess the formula for the tangent plane to $z = f(x, y)$ at (x_0, y_0, z_0), namely

$$P(x, y) = f(x_0, y_0) + \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0).$$

This is indeed correct: let's see how to get it geometrically, rather than by analogy. Geometrically, the tangent plane to (any) surface at a point P_0 should be the plane generated by the velocity vectors of curves $\mathbf{r}(t)$ on the surface passing through P_0 in any direction. But we will know the normal vector to the plane once we know any two vectors in the plane (we will of course take the cross product). But we've been through this already: two tangent vectors to the surface at P_0 are $T_x = (1, 0, \frac{\partial f}{\partial x})$ and $T_y = (0, 1, \frac{\partial f}{\partial y})$. Therefore a normal vector for the tangent plane is

$$\mathbf{n} = T_x \times T_y = -\frac{\partial f}{\partial x}\hat{\mathbf{i}} - \frac{\partial f}{\partial y}\hat{\mathbf{j}} + \hat{\mathbf{k}} = (-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1). \tag{6}$$

Recall that the equation of a plane with normal vector $\mathbf{n}$ passing through (x_0, y_0, z_0) is

$$\mathbf{n} \cdot (x - x_0, y - y_0, z - z_0) = 0.$$

So the equation of our tangent plane is

$$-\frac{\partial f}{\partial x}(x - x0) - \frac{\partial f}{\partial x}(y - y_0) + (z - z_0) = 0,$$

or

$$z = z_0 + \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0),$$

exactly what we had written down above.

Example: Find the tangent plane to $x^2 + y^2 + z^1 = 1$ at $P_0 = (0, 0, 1)$.

Solution: At the moment we must solve for z explicitly (but we'll see soon how to find directly the equation for the tangent plane of a level surface $F(x, y, z) = 0$): our point P_0 is on the upper hemisphere, so $z = f(x, y) = \sqrt{1 - x^2 - y^2}$. Above we computed the gradient of f at any point: at $(x_0, y_0) = (0, 0)$ we have indeed $\nabla(f) = (0, 0)$. Thus the equation of the tangent plane is $z = 1 + 0(x - 0) + 0(y - 0) = 1$, i.e., the tangent plane is horiontal at P_0. This should not be surprising, since P_0 is the top point of the hemisphere – in other words, it is a local (and even global) maximum for the function $f(x, y)$, at which point the tangent plane needs to be horizontal for the same reason that the tangent line must be horizonal at a local maximum of $y = f(x)$: otherwise travelling in the direction $\pm\nabla(f)$ would lead us to a larger value.

Tangent plane to a level surface: suppose we have a level surface $F(x, y, z) = C$ and a point $P_0 = (x_0, y_0, z_0)$ on the surface, and we want to find the equation of the tangent plane. I claim that the normal vector $\mathbf{n} = \nabla(F) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})$.

Before giving the proof, note that this case subsumes the previous calculation: if $z = f(x, y)$, put $F(x, y, z) = f(x, y) - z = 0$. Then $\nabla(F) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1)$, which is -1 times the normal vector of Equation 6 (which is fine: $\mathbf{n}$ and $-\mathbf{n}$ define the same plane).

To prove the claim, we need to see that $\nabla(F)$ has the property of being perpendicular to the velocity vector of every curve $\mathbf{r}(t) = (x(t), y(t), z(t))$ on $F(x, y, z) = C$ such that $\mathbf{r}(0) = P_0$. This is again a job for the chain rule: to say that the curve lies on F is to say

$$F(x(t), y(t), z(t)) = C;$$

differentiating this expression with respect to t using the chain rule and evaluating at $t = 0$, we get exactly that $\nabla F(x(0), y(0), z(0)) \cdot (x'(0), y'(0), z'(0)) = 0$, which is (already!) what we were trying to show.

Example: Find the equation of the tangent plane to the sphere of radius R at any point (x_0, y_0, z_0).

Solution: We can write this sphere as a level surface $F(x,y,z) = x^2+y^2+z^2 = R^2$. By what was just said, a normal vector for the tangent plane at (x_0, y_0, z_0) is just $\nabla(F) = (2x_0, 2y_0, 2z_0) = 2(x_0, y_0, z_0)$ – note this is *easier* than solving for z and differentiating the square root! – so the equation of the plane is

$$x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0) = 0.$$

When $(x_0, y_0, z_0) = (0, 0, 1)$, we recover the equation $z = 1$ that we found before.

Example (contour lines & equipotential surfaces): The preceding discussion is still valid and not without interest if we take it *down* a dimension: that is, we could take a **level curve** $F(x, y) = C$ in the plane, and the same discussion shows that for any point $P_0 = (x_0, y_0)$ on the curve we have $\nabla F(x_0, y_0)$ is perpendicular to the tangent line at P_0; in this case $\nabla F(x_0, y_0)$ is called the **normal line** at P_0. This can equally well be used to give an equation for the tangent line at P_0, namely:

$$\nabla F(x_0, y_0) \cdot (x - x_0, y - y_0) = 0.$$

Moreover, there is a connection to what has just been discussed: if $z = F(x, y)$ is a surface in $\mathbb{R}^3$, then taking $z = C$ gives a family of level curves $F(x, y) = C$. If one plots the level curves for various values of C, one can interpret the result as giving a **contour map** of the surface: these curves are loci of constant height C. So if these curves are widely spaced, it means that the height is changing slowly, whereas if they are bunched together the height is changing rapidly. Moreover, the fact that the gradient is perpendicular to the level curves implies the pleasant geometric fact that a path of steepest ascent/descent on the corresponding surface will intersect every level curve at a right angle.

Now go *up* a dimension again: consider a family of level surfaces $F(x, y, z) = C$ for varying C. These surfaces are loci of constant values of the function of three variables $F(x, y, z)$, and there are several interesting physical examples of this: we will see later that F can be viewed as a **potential function**, whose level surfaces are called **equipotential surfaces**. The gradient $\nabla(F)$ can be interpreted as a **field of forces** (e.g. gravitational, electrical, magnetic...). Any two points on $F(x, y, z) = C$ are said to have the same amount of potential energy, and (the same fact again!) the force is always perpendicular to the equipotential surface. Here is an important example: $F(x, y, z) = -1/2(x^2 + y^2 + z^2)^{-1/2}$. Then

$$\nabla(F) = \left(\frac{-x}{||\mathbf{r}||^3}, \frac{-y}{||\mathbf{r}||^3}, \frac{-z}{||\mathbf{r}||^3}\right) = \frac{-\mathbf{r}}{||\mathbf{r}||^3},$$

where $\mathbf{r} = (x, y, z)$. This is (up to a constant) precisely the gravitational force felt by a particle at a distance $\mathbf{r}$ from the origin under Newton's inverse square law. The fact that the equipotential surfaces $F = C$ are spheres means i) that all particles the same distance away from the central mass have the same amount of potential energy, and ii) the final interpretation of $\nabla F \perp (F(x, y, z) = C)$ is that it requires **no work** to move a particle about in a given equipotential surface.[9]

[9] Don't worry if this made no sense. It's just a preview of a future topic in the course: conservative vector fields.

Tangent plane to a parameterized surface: let

$$\mathbf{R}(u,v) = (x(u,v), y(u,v), z(u,v))$$

be a parameterized surface in $\mathbb{R}^3$. We want to find the equations for the tangent plane at the point $\mathbf{R}(u_0,v_0)$. As above, we need only find two tangent vectors pointing in different directions at this point. Ironically, for this we use partial derivatives: if we hold v constant, $\mathbf{R}(u,v_0)$ traces a parameterized curve on the surface, and its velocity vector at $u=0$ is just

$$\mathbf{T}_u := (\frac{dx}{du}(u_0,v_0), \frac{dy}{du}(u_0,v_0), \frac{dz}{du}(u_0,v_0)).$$

Similarly, we have

$$\mathbf{T}_v := (\frac{dx}{dv}(u_0,v_0), \frac{dy}{dv}(u_0,v_0), \frac{dz}{dv}(u_0,v_0)),$$

and the normal vector we want is just

$$\mathbf{n} = \mathbf{T}_u \times \mathbf{T}_v.$$

Example: We find the tangent plane to the torus

$$\mathbf{R}(u,v) = ((R + a\cos u)\cos v, (R + a\cos u)\sin v, a\sin u)$$

at $u=v=0$. At $(u,v)=(0,0)$, the values of the various partial derivatives are as follows: $\frac{dx}{du}=0$, $\frac{dx}{dv}=0$, $\frac{dy}{du}=0$, $\frac{dy}{dv}=R+a$, $\frac{dz}{du}=a$, $\frac{dz}{dv}=0$. Thus $\mathbf{T}_u=(0,0,a)$ and $\mathbf{T}_v=(0,R+a,0)$. We find therefore that $\mathbf{n}$ is a multiple of $\hat{\mathbf{i}}$, i.e., the tangent plane is $x=R+a$. This is consistent with the fact that $\mathbf{R}(0,0)$ is the rightmost point on the torus.

8. Extra: Gradients and Implicitly Defined Functions

In this section we want to use the gradient to examine in what sense it's really true that an equation like $F(x,y)=C$ or $F(x,y,z)$ implicitly defines a function.

The point is that just because we write down some expression relating x and y (or x, y and z), what assures us that the set of all ponts in the plane (or in space) satisfying this expression is actually a nice curve or a surface? Obviously we must show some care: e.g.

$$F(x,y) = x^2 + y^2 = C$$

does not define a curve when $C<0$ – the equation has no solution – or when $C=0$ – we just get the single point $(0,0)$. But suppose we have a point $P_0=(x_0,y_0)$ satisfying $F(x_0,y_0)=C$: are we then entitled to claim that there's a nice curve going through that point which implicitly defines y as a function of x?

Not quite: even for $x^2+y^2=1$, at the point $(1,0)$ this does not implicitly define y as a function of x, since the circle is multiple-valued there. On the other hand, it does define x as a function of y.

The amazing truth is that the gradient $\nabla f(x_0,y_0)$ keeps track of all this information automatically: recall we have already said that the gradient – as long as it is nonzero – gives the direction for the normal line to the curve at (x_0,y_0). Indeed, much more is true: the gradient $\nabla f(x_0,y_0) \neq (0,0)$ is necessary and sufficient for the locus to look like a curve near (x_0, y_0). More precisely: if $\nabla f(x_0, y_0) \neq (0, 0)$, then either $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$ or $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$ (or both). If $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$, this means that the tangent line (which recall, is perpendicular to the gradient) is *not* horizontal, and a nonhorizontal line is the linear approximation to some function x of y, at least near the point. Similarly, if $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$, then the tangent line is *not* vertical, so we can implcitly define y as a function of x. Thus, as long as $\nabla f(x_0, y_0) \neq (0, 0)$, we can *either* view the equation as defining x implicitly as a function of y or as defining y implictly as a function of x. Indeed, looking at the case of the unit circle $F(x, y) = x^2 + y^2 = 1$, we have $\nabla F = (2x, 2y)$, and this is only the zero vector at the origin, which is *not* a point on the unit circle. We get precisely that we can view the circle as a function of x except when $y = 0$ (correct) and as a function of y except when $x = 0$ (also correct). If we looked at the equation $F(x, y) = x^2 + y^2 = 0$, the gradient is the same $(2x, 2y)$ but now this *does* vanish at the unique point $(0, 0)$, which is also correct because there is no nice function here, just a single point!

To be sure, we did not prove this bold claim about the gradient not being zero implying that $F(x, y) = C$ implicitly defines a function. This is a real theorem, the **Implicit Function Theorem**. But it is definitely good to keep this behavior in mind. Indeed, we call a point (x_0, y_0) such that $F(x_0, y_0) = C$ a **smooth** (or **regular** or **nonsingular**) point if $\nabla f(x_0, y_0) \neq (0, 0)$ and otherwise we say that (x_0, y_0) is **singular**. Even if $F(x, y)$ is itself a perfectly nice function (like $x^2 + y^2$) there will in general be some singular points, and knowing about these singular points is often the single most important feature to understanding what's going on. Let's look at some more examples:

Example: $F(x, y) = y^2 - x^2(1 - x) = 0$. The gradient is $\nabla F = (-2x + 3x^2, 2y)$. Both coordinates vanish when $2y = 0$ (so $y = 0$ and when $-2x + 3x^2 = 0$, so $x = 0$ or $x = 2/3$. But the point $(2/3, 0)$ does not lie on the curve, so the only singular point is the origin $(0, 0)$. If you graph this curve there is indeed something fishy going on here: the curve circles back and intersects itself at $(0, 0)$, so there is *not* a well-defined tangent line at this point. This type of singularity is called a **node** and is the most basic kind of singularity a plane curve can have.

Example: $f(x, y) = y^2 - x^3 = 0$. Here $\nabla f = (-3x^2, 2y)$ and the only point where both components vanishes is the origin $(0, 0)$. There is again something funny going on at this point, a **cusp**: as we round the cusp the velocity vector switches from pointing to the left to pointing to the right.

Singularities in families of level curves: Say we have a surface $z = F(x, y)$, and consider the associated family of level curves $F(x, y) = C$. Typically, for most values of C the level curves will have no singularities, but all of a sudden for an isolated value of C the curve acquires a singularity: e.g. $F(x, y) = y^2 - x^3 = C$: you can check that for any nonzero value of C this is a smooth cubic curve, but when $C = 0$ the curve "degenerates" to having a cusp. Another way to say this is that the gradient $\nabla F(x, y)$ is the same no matter what C is, and the pair of equations $\frac{\partial f}{\partial x}(x_0, y_0) = 0$, $\frac{\partial f}{\partial y}(x_0, y_0) = 0$ will typically have only finitely many solutions. These are singular points for the corresponding level curves, with $C = F(x_0, y_0)$.

The point that I want to leave you with is: it would certainly be nice if *none* of the level curves were singular. But life doesn't work that way: almost any family of level curves $F(x,y) = C$ will have some singular curves in it. And far from ignoring the singular curves, we should pay the utmost attention to them, because near all non-singular points the surface is changing smoothly, but by passing through a singular point the surface can change abruptly, or qualitatively!

For example, consider again the hyperbolic paraboloid $z = x^2 - y^2$. The level curves $z = c$ are for positive c hyperbolas opening horizontally in the xy-plane and for negative c hyperbolas opening vertically in the xy-plane. How does a hyperbola change from being horizontal to being vertical? Only by passing through the special value $c = 0$, at which point we have the **degenerate hyperbola** $x^2 = y^2$ or $x = \pm y$, which has a singular point (a node) at the origin. If we are trying to understand the behavior of this surface, the (saddle) point $(0,0,0)$, such that $(x_0, y_0) = (0,0)$ lies on the singular level curve $c = 0$ is the most important point on the whole surface!

There is an entirely analogous discussion for level surfaces $F(x,y,z) = C$ in $\mathbb{R}^3$: they really do implicitly define a surface, with a well-defined tangent plane near a point (x_0, y_0, z_0) if and only if $\nabla F(x_0, y_0, z_) \neq (0,0,0)$. At a singular point, something is going wrong with the surface: two branches of the surface may meet each other, velocity vectors may switch direction abruptly, and so on.

There is an extremely rich theory of singularities of curves and surfaces, including a notion of **stable** and **unstable** singularities. Roughly speaking, a singularity is unstable if when we perturb the curve (or surface) just a little bit in any direction, we get a different type of singularity. For instance, if you poke a curve with a cusp a tiny bit, then (if you poke it in one way) the curve will become nonsingular and if you poke it in another direction then the branches will cross each other and you'll get a node. But if you close your eyes and poke it a little bit, then it is extremely unlikely that it will stay a cusp, so this is an unstable singularity. But if you poke a nodal curve, there is a good chance that it will stay a node. In fact any singularity of a plane curve, when poked, will probably either go away or become a node, so the nodes are the "stable singularities" for plane curves. Another example of this is that for surfaces, it is very unlikely that you will have at least three branches all meeting at a point, a **triple point**.

9. Appendix: Proof of the Chain Rule

We give the proof of Equation 5. The left hand side of (5) is

$$\lim_{h \to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h}.$$

We do a little algebraic manipulation on this expresion: it is

$$1/h(f(x(t+h), y(t+h)) - f(x(t), y(t))) =$$

$$1/h(f(x(t+h), y(t+h)) - f(x(t+h), y(t)) + f(x(t+h), y(t)) - f(x(t), y(t))) =$$

(7)
$$\left[\frac{f(x(t+h),y(t+h)) - f(x(t+h),y(t))}{h}\right] + \left[\frac{f(x(t+h),y(t)) - f(x(t),y(t))}{h})\right].$$

Let $F_1(u) := f(u, y(t))$. Then $F_1 \circ x(t+h) = f(x(t+h), y(t))$ and taking the limit at $h \to 0$ in the second term of Equation (7) the same as computing $d/dt(F_1 \circ x(t))$, which by the ordinary chain rule is $F_1'(x(t))x'(t) = \frac{\partial f}{\partial x}(x(t))x'(t)$.

To evaluate the first term we need an extra trick: set $s := t + h$. The first expression can be rewritten as

$$\lim_{h\to 0} \frac{f(x(s),y(s)) - f(x(s),y(s-h))}{h},$$

which, since $-h \to 0$ as $h \to 0$, is the same as

$$\lim_{h\to 0} -\frac{f(x(s),y(s)) - f(x(s),y(s+h))}{h} =$$
$$\lim_{h\to 0} \frac{f(x(s),y(s+h)) - f(x(s),y(s))}{h}.$$

Now we can proceed almost as above: define $F_2(u) := f(x(t), u)$. Then the limit as $h \to 0$ of the first term in (7) is $d/dtF_2(y(s))$, which by the ordinary chain rule is

$$d/ds(F_2(y(s))) = \frac{\partial f}{\partial y}(y(s))dy/ds.$$

But at $h = 0$, $s = t$, so we get $/dfdy(y(t))dy/dt$. Thus the sum of the two terms is

$$\frac{\partial f}{\partial x}(x(t))dx/dt + \frac{\partial f}{\partial y}(y(t))dy/dt.$$