CURVES AND SURFACES IN $\mathbb{R}^3$

Let us intuitively define a *curve* as a geometric object which, on a very small scale, looks like a straight line. We need to make this more precise. The simplest example of a curve is the graph of a function of one variable, $y = f(x)$. As we vary the independent variable x over some range, say $a < x < b$, the points (x, y) in the plane $\mathbb{R}^2$ that satisfy $y = f(x)$ lie on a curve.

In single variable calculus, we learn that the derivative of a function at a point gives us the slope of the tangent line to the graph at that point. That is, $f'(c)$ is the slope of the tangent line to the curve at the point $(c, f(c))$. We can write the equation for a straight line passing through a point (x_0, y_0) with slope m as:

$$y - y_0 = m(x - x_0)$$

Applying this in the case $x_0 = c$, $y_0 = f(c)$ and $m = f'(c)$, some rearranging yields:

$$y = f(c) + \frac{df}{dx}(c)\,(x - c)\,. \tag{0.1}$$

This is the *linear approximation* of the function $f(x)$ *near* the point $x = c$. The adjective "linear" means the variable x does not appear with powers higher than one. The term "approximation" means that the function $L(x) = f(c) + \frac{df}{dx}(c)\,(x - c)$ and its first derivative agree with the function $f(x)$ and its derivative *at the point* $x = c$. The linear approximation is a straight line. This is what we mean when we say that a curve "looks like" a straight line on a small scale.

Before continuing our study of curves, let us apply the same ideas to *surfaces*. We will intuitively define a surface to be a geometric object which, on a small scale, looks like a flat plane. The simplest example of a surface is the graph of a function of *two* variables, $z = g(x, y)$. As we vary the independent variables x and y over some region D in the plane $\mathbb{R}^2$, the points (x, y, z) in the space $\mathbb{R}^3$ that satisfy $z = g(x, y)$ lie on a surface.

To find the analogue of linear approximation at a point (a, b) in this case, we need to find a function $L(x, y)$ in which the variables x and y do not appear with powers higher than one, which agrees with $g(x, y)$ and its first derivatives *at the point* (a, b). Note in this case, since the function $g(x, y)$ depends on two variables, we have *two* first derivatives. The conditions therefore are as follows. We need $L(x, y) = A + Bx + Cy$, subject to:

$$L(a, b) = g(a, b) \qquad \frac{\partial L}{\partial x}(a, b) = \frac{\partial g}{\partial x}(a, b) \qquad \frac{\partial L}{\partial y}(a, b) = \frac{\partial g}{\partial y}(a, b)$$

This is enough information to determine the constants A, B, and C, and after rearranging we obtain:

$$z = g(a, b) + \frac{\partial g}{\partial x}(a, b)\,(x - a) + \frac{\partial g}{\partial y}(a, b)\,(y - b) \tag{0.2}$$

as the linear approximation to the function $g(x, y)$ at the point (a, b). It is now easy to see the analogy with equation 0.1. This time, the linear approximation is a plane, and it is the *tangent plane* to the surface at the point $(a, b, g(a, b))$. This is the precise meaning of the statement that a surface "looks like" a flat plane on a small scale.

In general, however, curves and surfaces can be more complicated than simply graphs of functions. For example, a circle of radius R centred at the origin, given by the equation

$$x^2 + y^2 = R^2$$

is something we would like to call a curve: it has a tangent line at every point. But it is not the graph of a function, because if we solve for y as a function of x, we get

$$y = \pm\sqrt{R^2 - x^2}.$$

That is, there are two values of y for each x, (except for $x = \pm R$) which should be clear by considering a sketch of this curve in the plane.

We can "trace" out the circle if we start at the point $(R, 0)$ and travel around it counterclockwise. In this way, every point on the circle is determined uniquely by the angle θ between the positive x-axis and the vector with tip at that point. The coordinates (x, y) are given by

$$x(\theta) = R\cos(\theta) \qquad y(\theta) = R\sin(\theta) \qquad 0 \leq \theta \leq 2\pi. \tag{0.3}$$

In this context the variable θ is called a *parameter*, since it is by considering all allowed values of this parameter that we obtain all the points on the curve.

There are infintely many ways to parametrize any curve. Let us continue with the example of the circle to see this. Another valid parametrization is:

$$x(\theta) = R\cos(2\theta) \qquad y(\theta) = R\sin(2\theta) \qquad 0 \leq \theta \leq \pi.$$

This parametrization also traces around the circle in the counterclockwise direction, but at twice the "speed" since it only takes π units of the parameter to get all the way around. In a similar way we can parametrize the circle to be traversed for any range $0 \leq \theta \leq L$ of the parameter. If we consider now only the top half of the circle, here is yet another parametrization:

$$x(t) = t \qquad y(t) = \sqrt{R^2 - t^2} \qquad -R \leq t \leq R. \tag{0.4}$$

This example illustrates that the symbol for our parameter can be any variable we want (the most common is t), and our range of allowed parameter values does not have to start at zero. You may be wondering which parametrization is the "correct" one. They are all correct. However, in a particular problem one choice might be easier to calculate with than another.

Now we are ready to make our general definition of a *parametrized curve*. A parametrized curve in $\mathbb{R}^3$, is given by

$$\mathbf{r}(t) = (x(t), y(t), z(t)) \qquad a \leq t \leq b \tag{0.5}$$

If the curve lies in the plane $\mathbb{R}^2$ then we only have two functions $x(t)$ and $y(t)$.

A quantity we are often interested in is the *length* of a curve. If we vary the parameter by a small amount Δt, then the curve is approximately a straight line

with length $|\Delta \mathbf{r}| = |\frac{d\mathbf{r}}{dt}|\Delta t$. Adding up all these little pieces and letting $\Delta t \to 0$, we obtain the formula for arc length:

$$\text{Arc length} = \int_a^b |\mathbf{r}'(t)|dt. \tag{0.6}$$

As an illustration of the fact that the choice of parametrization does not matter, let us compute the length of the top half of a circle of radius R using two different parametrizations. Using the parametrization in equation 0.3, we have

$$|\mathbf{r}'(\theta)| = \sqrt{(-R\sin(\theta))^2 + (R\cos(\theta))^2} = R$$

and the arc length is

$$\int_0^\pi R dt = \pi R$$

which is as expected from elementary geometry. (Note the upper limit of integration is π because we are only considering the top half of the circle.) Using the parametrization in equation 0.4, we obtain

$$|\mathbf{r}'(t)| = \sqrt{(1)^2 + \left(\frac{-t}{\sqrt{R^2-t^2}}\right)^2} = \frac{R}{\sqrt{R^2-t^2}}.$$

The arc length integral is

$$R\int_{-R}^{R} \frac{1}{\sqrt{R^2-t^2}} dt = \pi R.$$

The arc length is the same, as we would expect because the length of a curve does not depend on the way we choose to describe it. (To evaluate this integral, make the substitution $t = R\sin(u)$.) This example also illustrates that some parametrizations are easier to calculate with than others.

Let us now turn our attention back to surfaces and discuss the analogous facts. A general surface will not be the graph of a function of two variables. The simplest such example is a sphere of radius R centred at the origin, given by the equation

$$x^2 + y^2 + z^2 = R^2.$$

This is a surface because it has a tangent plane at every point. But it is not the graph of a function, because if we solve for z as a function of x and y, we get

$$z = \pm\sqrt{R^2 - x^2 - y^2}.$$

If we think of the sphere as the Earth, we can describe any point on the surface using two numbers: latitude and longitude. The convention in mathematics is to measure "latitude" θ counterclockwise from the positive x-axis, and "longitude" ϕ from the north pole $(0, 0, R)$. These are called spherical coordinates, and they parametrize the sphere by:

$$x(\theta,\phi) = R\sin(\phi)\cos(\theta) \qquad y(\theta,\phi) = R\sin(\phi)\sin(\theta) \qquad z(\theta,\phi) = R\cos(\phi) \tag{0.7}$$

where $0 \le \theta \le 2\pi$ and $0 \le \phi \le \pi$ are the allowed values of the two parameters. As we vary over all allowed values, we obtain all points on the surface.

Just as in the case of curves, there are infinitely many ways to parametrize any surface. For example, the top half of the sphere (which corresponds to $0 \le \phi \le \frac{\pi}{2}$) can also be parametrized by:

$$x(u,v) = u \qquad y(u,v) = v \qquad z(u,v) = \sqrt{R^2 - u^2 - v^2} \qquad (u,v) \in D \tag{0.8}$$

where D is the disc or radius R in the u-v plane. (The circle together with its interior). This is the allowed set of parameters. The most commonly used symbols for surface parameters are u and v.

Any choice of parametrization of a surface is good, but as in the case of curves for a particular problem one choice might be easier to calculate with than another.

Now we are ready to make our general definition of a *parametrized surface*. A parametrized surface in $\mathbb{R}^3$ is given by

$$\mathbf{X}(u,v) = (x(u,v), y(u,v), z(u,v)) \qquad (u,v) \in D \tag{0.9}$$

where D is a region in the u-v plane called the parameter space.

A quantity we are often interested in is the *area* of a surface. If we vary the parameters by small amounts Δu and Δv, then the corresponding piece of the surface is approximately a flat parallelogram which has area $\left|\frac{\partial \mathbf{X}}{\partial u} \times \frac{\partial \mathbf{X}}{\partial v}\right| \Delta u \Delta v$. Adding up all these little pieces and letting Δu and $\Delta v \to 0$, we obtain the formula for surface area:

$$\text{Surface area} = \iint_D \left|\frac{\partial \mathbf{X}}{\partial u} \times \frac{\partial \mathbf{X}}{\partial v}\right| du dv. \tag{0.10}$$

Let us compute the area of the top half of a sphere of radius R using two different parametrizations. Using the parametrization in equation 0.7, we have (check!)

$$\left|\frac{\partial \mathbf{X}}{\partial \theta} \times \frac{\partial \mathbf{X}}{\partial \phi}\right| = R^2 \sin(\phi)$$

and the area is

$$\int_0^{\frac{\pi}{2}} \int_0^{2\pi} R^2 \sin(\phi) d\theta d\phi = 2\pi R^2$$

which is as expected from elementary geometry. Using the parametrization in equation 0.8, the double integral for the surface area is

$$R \int_{-R}^{R} \int_{-\sqrt{R^2-v^2}}^{\sqrt{R^2-v^2}} \frac{1}{\sqrt{R^2-u^2-v^2}} du dv = 2\pi R^2.$$

As expected, the computed area is the same. (To evaluate this double integral, change to polar coordinates).

Notice that a curve, which we think of as one-dimensional, has one parameter. Also, its tangent "space" is a line, which is one-dimensional. A surface, which we think of as two-dimensional, has two parameters. Its tangents "space" is a plane, which is two-dimensional. In more advanced courses we study *n-dimensional manifolds*, which require n parameters, and their tangent "space" at a point looks like $\mathbb{R}^n$. Of course, we cannot draw pictures of these objects, but otherwise there is essentially no difference in the mathematics.

There is yet another way we can describe curves and surfaces, as level sets of functions. This approach is useful because it provides an easy method to compute a normal vector field to curves in $\mathbb{R}^2$ or surfaces in $\mathbb{R}^3$.

The level set of a function of $g(x,y)$ of two variables is a curve in $\mathbb{R}^2$. We have already seen an example of this: the circle $x^2 + y^2 = R^2$ is the level set $\{g(x,y) = R^2\}$ of the function $g(x,y) = x^2 + y^2$. Consider a point (x_0, y_0) on the circle. The *vector* with components (x_0, y_0) points in the radial direction, and is clearly normal to the circle at that point. It is also proportional to the gradient $\nabla g = (2x, 2y)$ evaluated at the point (x_0, y_0). This is true in general. To see why, supposed we have a curve γ which is the level set of a function $\{g(x,y) = C\}$. We

can parametrize γ by $\mathbf{r}(t) = (x(t), y(t))$. Then the *composed function* $g(\mathbf{r}(t)) = g(x(t), y(t))$ is constant as a function of t since g is constant on γ. Taking the derivative with respect to t and using the chain rule, we obtain:

$$\frac{d}{dt} g(\mathbf{r}(t)) = \nabla g(x(t), y(t)) \cdot \mathbf{r}'(t) = 0.$$

But the velocity vector $\mathbf{r}'(t)$ is tangent to the curve, so the gradient $\nabla g(x, y)$ must be normal to the curve at the point (x, y).

We can do the same thing for surfaces. A surface M in $\mathbb{R}^3$ can be given as the level set of a function $h(x, y, z)$ of *three* variables. For example, the level sets of $h(x, y, z) = x^2+y^2+z^2$ are spheres centred at the origin. The exact same calculation as above, with a parametrized curve $\mathbf{r}(t) = (x(t), y(t), z(t))$ on the surface shows that the gradient $\nabla h(x, y, z)$ must be perpendicular to the surface M at the point (x, y, z), because $\mathbf{r}'(t)$ is tangent to the curve on M and hence tangent to M.

Let us finally consider these facts in the special cases when a curve γ in $\mathbb{R}^2$ is the graph of a function $f(x)$ and when a surface M in $\mathbb{R}^3$ is the graph of a function $g(x, y)$. We can easily parametrize γ as $\mathbf{r}(t) = (t, f(t))$, which gives us the tangent vector $\mathbf{r}'(t) = (1, f'(t))$ and from there the normal vector $\mathbf{n} = (-f'(t), 1)$. We can also view γ as being the level set $\{F(x, y) = 0\}$ for the function $F(x, y) = y-f(x)$. With this approach, a normal vector $\vec{n}$ to the curve is given by $\nabla F(x, y) = (-f'(x), 1)$, which is the same normal obtained using the parametrization. (Recall $x = t$ in our parametrization.)

We can parametrize our surface M as $\mathbf{X}(u, v) = (u, v, g(u, v))$, which gives us two tangent vectors $\mathbf{X}_u = (1, 0, \frac{\partial g}{\partial u})$ and $\mathbf{X}_v = (0, 1, \frac{\partial g}{\partial v})$. This gives us the normal vector $\mathbf{n} = \mathbf{X}_u \times \mathbf{X}_v = (-\frac{\partial g}{\partial u}, -\frac{\partial g}{\partial v}, 1)$. We can also view M as the level set $\{G(x, y, z) = 0\}$ for the function $G(x, y, z) = z - g(x, y)$. With this approach, a normal vector $\mathbf{n}$ is given by $\nabla G(x, y, z) = (-\frac{\partial g}{\partial x}, -\frac{\partial g}{\partial y}, 1)$, which is the same as the normal obtained from the parametrization since $x = u$ and $y = v$ in this case.

This point is one that is very often confusing to students in this course. The gradient vector gives a normal to a curve or a surface only if you are taking the gradient of the function whose level set is your curve or surface. For example, if you have a surface which is the graph of $g(x, y)$, the gradient ∇g is a vector in $\mathbb{R}^2$, which cannot possibly be normal to a surface in $\mathbb{R}^3$. The correct normal is the gradient of $G(x, y, z) = z - g(x, y)$, a function of three variables, whose gradient is a vector in $\mathbb{R}^3$.