

Parametric curves in the plane

1. The idea of parametric equations. Up to now, we've been used to describing curves in the xy -plane by specifying a single equation that relates x and y , such as $y = x^2$ to define a parabola or $x^2 + y^2 = 2$ to define the circle of radius $\sqrt{2}$ centered at the origin. This way of describing a curve is sometimes called *extrinsic*, because it specifies the curve by giving a way to test whether a given point lies on the curve. For instance, the point $(2, 3)$ is not on the parabola $y = x^2$ because $3 \neq 2^2$, but the point $(3, 9)$ is because $9 = 3^2$.

In basic calculus, we sometimes made a distinction between “implicitly” and “explicitly” defined curves. The $y = x^2$ definition of the parabola is “explicit” because it clearly defines y as a function of x : to find the (only) point on the parabola that has $x = 5$, you just substitute $x = 5$ into x^2 to learn that $y = 25$, so the point $(5, 25)$ is on the parabola, so that finding points on the parabola involves substituting values in for x but not having to solve any equations (I realize that this is belaboring the point). On the other hand, in $x^2 + y^2 = 2$, if you're told that $x = 2/3$, then you still have to solve the equation $4/9 + y^2 = 2$ for y to find the (two) points on the circle, namely $(2/3, \sqrt{14}/3)$ and $(2/3, -\sqrt{14}/3)$, which have $x = 2/3$. So the equation $x^2 + y^2 = 2$ *implicitly* defines y as a function of x , once we've decided which of the positive or negative square roots to use.

The point here is that in these extrinsic ways of describing a curve, we start with the 2-dimensional plane (where the values of x and y are unrestricted), and place one equation constraint on x and y , which reduces the dimension of points under consideration by one – so the resulting curves such as the parabola $y = x^2$ and the circle $x^2 + y^2 = 1$ are one-dimensional objects.

An alternative, *intrinsic* way to define a curve in the plane is via parametric equations. This way of defining a curve involves using an additional variable, called the *parameter*, and x and y are given as functions of this new variable. The letter t is often used for the parameter and in many applications t depends on time, and so writing $x = f(t)$, $y = g(t)$ for some functions f and g tells, for each value of the time t , where some object is in the xy -plane.

There are several basic examples of parametric equations that are good to know:

- (1) *Graphs of functions.* Using parametric equations is a true generalization of the $y = f(x)$ explicit extrinsic way to define a curve. To see this, consider the parabola $y = x^2$ again. To write this parametrically, we could write $x = t$, $y = t^2$, and it's obvious that for any function $f(x)$ the curve $y = f(x)$ can be expressed parametrically as $x = t$, $y = f(t)$.

There are a lot of other ways to express the parabola, though – in fact we can set x equal to any expression in t , and then y should be the square of that expression, and we'll get at least part of the parabola. For instance, the graph of $x = e^t$, $y = e^{2t}$ will give the right half of the parabola, whereas $x = \cos t$, $y = \cos^2 t$ will give the part of the parabola for x between -1 and 1 .

- (2) *Straight lines.* It's always important to know how to write the equation of a line. Recall that to specify a line, we need to specify one point on the line and the direction

of the line. In extrinsic form, this is usually done using the slope of the line, so the line through the point (a, b) with slope m is $y = b + m(x - a)$ (how can you make a direct translation of this into parametric form?).

When we're thinking parametrically, it's better to specify the direction of a line by giving a vector that points in the direction the line is supposed to go. Our textbook writes vectors in the plane as $\langle p, q \rangle$ for the vector that points p units to the right and q units up (you may have written this as $p\mathbf{i} + q\mathbf{j}$ in physics class). The slope of a line that goes in the direction of $\langle p, q \rangle$ is q/p , and the parametric form of the line that goes through the point (a, b) in this direction is $x = a + pt, y = b + qt$. Not so bad.

One thing made easier by this way of thinking is writing the line through the two points (a_1, b_1) and (a_2, b_2) . The vector $\langle a_2 - a_1, b_2 - b_1 \rangle$ points from the first point to the second, so the line is given parametrically by $x = a_1 + (a_2 - a_1)t, y = b_1 + (b_2 - b_1)t$.

As we shall see, one advantage of this parametric form of the equation of a line is that when we look at lines in higher dimensional spaces, the parametric form generalizes pretty easily, where the explicit extrinsic form does not.

- (3) *Circles.* The (implicit, extrinsic) equation of the unit circle centered at the origin is $x^2 + y^2 = 1$. To put this in parametric form, we need two functions of t (one for x and one for y) whose squares add up to 1. So we could let $x = \cos t$ and $y = \sin t$, for instance. If we think of the values $(1, 0)$ of x and y for $t = 0$ as where the curve "starts", and then see which points on the circle we get as t increases, we can see that the parametric equations $x = \cos t, y = \sin t$ describes the circle starting at $(1, 0)$ and traversed counterclockwise at a steady speed so that it takes 2π time units to go once around the circle.

How do the parametric equations $x = \sin t, y = \cos t$ describe the circle? How about $x = \cos 3t, y = \sin 3t$?

It's not too hard to see that the parametric equations $x = a + r \cos t, y = b + r \sin t$ describe the circle of radius r centered at the point (a, b) . This will come in handy when we do some animations in Maple.

2. Calculus with parametric equations. All the fundamental calculus concepts such as limits, derivatives and integrals can be adapted to the parametric setting. There are no new techniques to learn, but some interpretation of the concepts is needed.

Limits. There's nothing new as regards limits – it's just that since there are two functions $x(t)$ and $y(t)$, you have to do two limits to figure out which point $(x(t), y(t))$ approaches as t goes to some value. And from limits comes the idea of continuity, and again there's nothing new.

Derivatives. Again, since there are two functions, you have to take two derivatives. If we recall what we used derivatives for before, namely for giving the slope of a graph at a point, you can see easily that, since

$$dx = \frac{dx}{dt} dt \quad \text{and} \quad dy = \frac{dy}{dt} dt$$

and since the slope of a curve at a point is given by the derivative dy/dx of y with respect to x , we have

$$\text{slope} = \frac{dy}{dx} = \frac{dy/dt}{dx/dt}$$

in parametric form.

You can use the slope to find the equation of tangent lines to parametric graphs, but it's more natural (and generalizable to higher dimensions) to use the parametric form of lines described above to get equations of tangent lines. So, if for a certain value t_0 of t , it is the case that $x(t_0) = a$, $y(t_0) = b$, $x'(t_0) = c$ and $y'(t_0) = d$, then the parametric equations of the line tangent to the curve at (a, b) are

$$x = a + ct, \quad y = b + dt,$$

since the tangent line should go through the point (a, b) and go in the direction of the vector $\langle c, d \rangle$.

One thing to recall is that we used to find the highest and lowest points on curves by setting $dy/dx = 0$ and solving for x . We set $dy/dx = 0$ because we were looking for points where the tangent line is horizontal. In parametric form, this will happen when $dy/dt = 0$ – you can solve this for t and then substitute the values obtained back into both $x(t)$ and $y(t)$ to get candidates for the highest and lowest points on parametric curves.

But parametric curves can also have vertical tangents – these happen when $dx/dt = 0$.

Integrals. We used integrals to find areas, arclengths etc. We'll do these two tasks for parametric curves. When we had a curve described by expressing y as a function of x , we calculated the area between a part of the curve and the x -axis by doing an integral of the form $\int y dx$. Likewise, to calculate the area between a part of a curve and the y axis we would express x as a function of y and do an integral of the form $\int x dy$.

If x and y are both functions of a parameter t , then we can express dx as $x'(t) dt$ and dy as $y'(t) dt$. This gives us (at least) two ways of doing areas related to parametrically-described curves. One thing we have to be careful of is the sign of dx or dy – in other words, we have to check whether x and/or y is increasing or decreasing as the parameter t increases. Here's an example:

The circle of radius 2 centered at the origin is given parametrically by $x = 2 \cos t$, $y = 2 \sin t$. Note that the circle is traversed counter-clockwise as t increases. So to get the area between the top half of the circle and the x -axis, we note that the top half of the circle corresponds to parameter values $0 \leq t \leq \pi$ but the semi-circle is traversed from right to left. So we can calculate the area using the integral

$$-\int_0^\pi y dx = -\int_0^\pi 2 \sin t (-2 \sin t dt) = \int_0^\pi 4 \sin^2 t dt = 2\pi.$$

To get the area of between the x axis and the *bottom* half of the circle, we use parameter values $\pi \leq t \leq 2\pi$. The bottom half of the circle goes from left to right, but since it's below the x axis we still want an integral of the form

$$\int_\pi^{2\pi} (-y) dx = \int_\pi^{2\pi} (-2 \sin t)(-2 \sin t dt) = \int_\pi^{2\pi} 4 \sin^2 t dt = 2\pi.$$

It's not an accident that the two integrals we had to do were the same, except for different values of the parameter. In general, to find the area of a *closed* curve that is traversed counter-clockwise, we can integrate $-y dx$ "once around the curve", in other words, over a parameter

interval that takes us exactly once around the curve. If the curve were traversed clockwise, we would integrate $+y dx$.

You should convince yourself, using the example of the circle again, that we could also calculate the area enclosed by a closed loop traversed counterclockwise by integrating $x dy$. But as a bonus, since we get the area either by integrating $x dy$ or $-y dx$, we could also add them together and divide by two and still get the area. And in a surprising number of situations, this simplifies things a lot. For example, for our circle, we have

$$x = 2 \cos t \quad y = 2 \sin t$$

so

$$dx = -2 \sin t dt \quad dy = 2 \cos t dt$$

and so we can calculate the area of the circle by integrating $\frac{1}{2}(x dy - y dx)$ from 0 to 2π (once around, counterclockwise) as follows:

$$\begin{aligned} \text{Area} &= \frac{1}{2} \int_0^{2\pi} (x dy - y dx) = \frac{1}{2} \int_0^{2\pi} (2 \cos t(2 \cos t dt) - 2 \sin t(-2 \sin t dt)) \\ &= \frac{1}{2} \int_0^{2\pi} (4 \cos^2 t + 4 \sin^2 t) dt = \frac{1}{2} \int_0^{2\pi} 4 dt = 4\pi \end{aligned}$$

as we should have expected – but the integral in this case was certainly easier than integrating $\sin^2 t$ or $\cos^2 t$.

Here's a more interesting example. The curve given extrinsically by the equation $x^3 + y^3 = 3xy$ is called the *folium of Descartes*. Here's a graph of it:

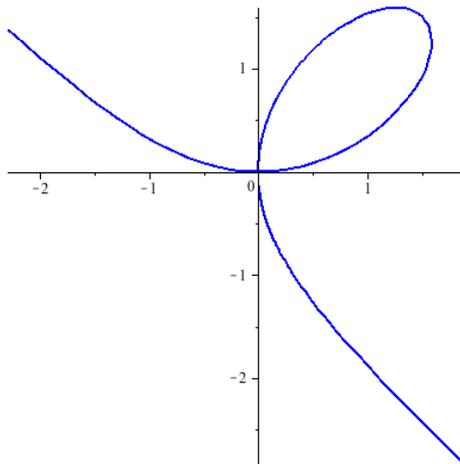


FIGURE 1. Folium of Descartes

Because of the x^3 and y^3 terms, the equation of the folium is not easy to solve for either x or y . But there is a useful set of parametric equations for this curve. To find it, we choose a “geometrically meaningful” parameter – in this case, for each point (x, y) on the curve, we’ll let t be the slope of the line through the origin and (x, y) . In other words, $t = y/x$. So we now have two equations that relate the three variables x , y , and t :

$$x^3 + y^3 = 3xy \quad \text{and} \quad t = \frac{y}{x}.$$

And whenever we have two equations, we can expect (or at least hope) to be able to solve for two of the variables in terms of the other(s). So in this case, we'll solve for x and y as functions of t . Substitute $y = tx$ into the first equation to get $x^3 + t^3x^3 = 3tx^2$, or $x^2((1+t^3)x - 3t) = 0$, which we can solve for x to get

$$x = \frac{3t}{1+t^3}$$

(ignoring the trivial solution $x = 0$). And since $y = tx$, we have that the parametric equations for the folium of Descartes are

$$x = \frac{3t}{1+t^3} \quad y = \frac{3t^2}{1+t^3}.$$

The folium has a loop that seems to start and end at the origin – let's find its area. Now there's only one value of the parameter, namely $t = 0$, that gives $x = 0$ and $y = 0$, but a moment's thought (or L'Hôpital's rule) reveals that

$$\lim_{t \rightarrow \infty} \frac{3t}{1+t^3} = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} \frac{3t^2}{1+t^3} = 0$$

(see how parametric limits work the same way as the ones we already know?), so to find the area of the loop we have to integrate one of our area quantities for t going from 0 to infinity. You should convince yourself that as t goes from 0 to infinity, that the loop is traversed counterclockwise.

Now we calculate:

$$x \, dy = \frac{3t}{1+t^3} \frac{(1+t^3)(6t) - (3t^2)(3t^2)}{(1+t^3)^2} dt = \frac{9t^2(2-t^3)}{(1+t^3)^3} dt$$

and

$$y \, dx = \frac{3t^2}{1+t^3} \frac{(1+t^3)(3) - (3t)(3t^2)}{(1+t^3)^2} dt = \frac{9t^2(1-2t^3)}{(1+t^3)^3} dt.$$

Neither of these looks like much fun to integrate. But note:

$$\frac{1}{2}(x \, dy - y \, dx) = \frac{1}{2} \frac{9t^2}{(1+t^3)^3} (2-t^3-1+2t^3) dt = \frac{9t^2(1+t^3)}{2(1+t^3)^3} dt = \frac{9t^2}{2(1+t^3)^2} dt$$

and we can integrate this by u -substitution ($u = 1+t^3$), as follows:

$$\text{Area} = \int_0^\infty \frac{9t^2}{2(1+t^3)^2} dt = \int_1^\infty \frac{3}{2u^2} du = -\frac{3}{2} \frac{1}{u} \Big|_1^\infty = \frac{3}{2}.$$

Now, recall that we calculated the length of a curve given by $y = f(x)$ by integrating

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx.$$

But this expression seems to treat the two variables x and y differently. Moreover, we derived this expression for the element of arclength by beginning with the fact that on a small bit of the graph we have that x changes by dx and y changes by dy , so that the length of the small bit of the graph should be $\sqrt{(dx)^2 + (dy)^2}$. We then wrote $dy = y' dx$ and factored out $(dx)^2$.

In parametric form, the two variables x and y are treated the same. And since $dx = x'(t) dt$ and $dy = y'(t) dt$, we have

$$ds = \sqrt{(x'(t))^2 + (y'(t))^2} dt.$$

This is what we integrate to get arclength of a curve given in parametric form.

For example for our circle of radius 2, we have

$$x = 2 \cos t \quad y = 2 \sin t$$

and so

$$ds = \sqrt{4 \sin^2 t + 4 \cos^2 t} dt = 2 dt,$$

and we compute the length (circumference) of the circle easily:

$$\text{Length} = \int_0^{2\pi} 2 dt = 4\pi.$$

3. Pythagorean triples. Here's an interesting example of using a non-standard parametrization of the circle to derive an interesting fact. You remember from high-school trig and physics that there are some right-angle triangles which have all three sides integers, such as (3,4,5) or (5,12,13). These triples of numbers (a, b, c) satisfy the famous Pythagorean equation $a^2 + b^2 = c^2$. When you have a solution for which a , b and c are all whole numbers, you call it a *Pythagorean triple*. You may have wondered just how many really distinct Pythagorean triples there are ("really distinct" means that the triangles corresponding to them are not similar – so the triples (3,4,5) and (6,8,10) are not really distinct, but (3,4,5) and (5,12,13) are).

Pythagorean triples correspond with points on the circle $x^2 + y^2 = 1$ whose x and y coordinates are both rational numbers – we just let $x = a/c$ and $y = b/c$. And there is an interesting parametrization of the unit circle which allows us to find lots of (in fact, infinitely many) points with both coordinates rational.

The parameter t in this parametrization is "geometrically meaningful" in a way that is reminiscent of the folium of Descartes. But instead of being the slope of the line connecting the origin to (x, y) on the circle, t will be the slope of the line connecting the point $(1, 0)$ to (x, y) , so we have

$$t = \frac{y}{x-1}, \text{ or } y = t(x-1).$$

We substitute this into $x^2 + y^2 = 1$ and get

$$x^2 + t^2(x-1)^2 = 1,$$

which we can solve for x as follows: Rewrite the preceding equation as a quadratic equation in x :

$$(t^2 + 1)x^2 - 2t^2x + (t^2 - 1) = 0$$

and then the quadratic formula gives

$$x = \frac{2t^2 \pm \sqrt{4t^4 - 4(t^2 + 1)(t^2 - 1)}}{2(t^2 + 1)} = \frac{2t^2 \pm 2}{2(t^2 + 1)} = \frac{t^2 - 1}{t^2 + 1}$$

(this is the only interesting solution, since using the + sign just gives $x = 1$). Thus

$$x = \frac{t^2 - 1}{t^2 + 1}.$$

Since $y = t(x - 1)$, we can calculate

$$y = t \left(\frac{t^2 - 1}{t^2 + 1} - 1 \right) = -t \left(\frac{t^2 - 1}{t^2 + 1} - \frac{t^2 + 1}{t^2 + 1} \right) = t \frac{-2}{t^2 + 1}.$$

So the following are parametric equations for the unit circle:

$$x = \frac{t^2 - 1}{t^2 + 1} \quad y = \frac{-2t}{t^2 + 1}.$$

This is called a *rational parametrization* of the circle because both x and y are quotients of polynomials with integer coefficients. And because we have a rational parametrization, we can generate lots of points on the circle with both x and y rational numbers, simply by substituting an integer or rational number for t .

For instance, if $t=10$, then we get $x = 99/101$ and $y = -20/101$. And it's true that $20^2 + 99^2 = 101^2$ (both sides equal 10,201). And, from our parametrization, we're led pretty directly to believe that for any integer t , the triple $(t^2 - 1, 2t, t^2 + 1)$ will be Pythagorean, which is pretty easy to verify. Let's list the first few:

t	triple
2	(3, 4, 5)
3	(8, 6, 10)
4	(15, 8, 17)
5	(24, 10, 26)
6	(35, 12, 37)
7	(48, 14, 50)
8	(63, 16, 65)
9	(80, 18, 82)
10	(99, 20, 101)

There are some interesting patterns in this table, for even versus odd values of t . But you can see that we'll generate infinitely many really distinct Pythagorean triples this way. Do you think there are any others that are distinct from all the triples on this list? How would you either find one of them or else prove that none exist?