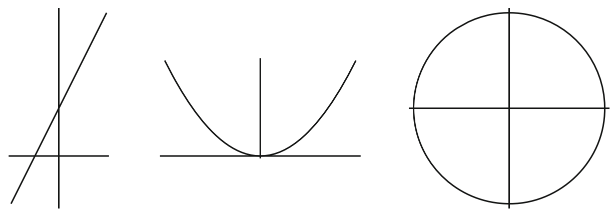
1
Curves in the plane and in space

In this chapter, we discuss two mathematical formulations of the intuitive notion of a curve. The precise relation between them turns out to be quite subtle, so we begin by giving some examples of curves of each type and practical ways of passing between them.

1.1 What is a curve?

If asked to give an example of a curve, you might give a straight line, say $y - 2x = 1$ (even though this is not 'curved'!), or a circle, say $x^2 + y^2 = 1$, or perhaps a parabola, say $y - x^2 = 0$.



Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_1,

All of these curves are described by means of their Cartesian equation

$$f(x, y) = c,$$

where f is a function of x and y and c is a constant. From this point of view, a curve is a set of points, namely

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}. \tag{1.1}$$

These examples are all curves in the plane $\mathbb{R}^2$, but we can also consider curves in $\mathbb{R}^3$ – for example, the x-axis in $\mathbb{R}^3$ is the straight line given by

$$y = 0, \quad z = 0,$$

and more generally a curve in $\mathbb{R}^3$ might be defined by a pair of equations

$$f_1(x, y, z) = c_1, \qquad f_2(x, y, z) = c_2.$$

Curves of this kind are called *level curves*, the idea being that the curve in Eq. 1.1, for example, is the set of points (x, y) in the plane at which the quantity $f(x, y)$ reaches the 'level' c.

But there is another way to think about curves which turns out to be more useful in many situations. For this, a curve is viewed as the path traced out by a moving point. Thus, if $\boldsymbol{\gamma}(t)$ is the position of the point at time t, the curve is described by a function $\boldsymbol{\gamma}$ of a scalar parameter t with vector values (in $\mathbb{R}^2$ for a plane curve, in $\mathbb{R}^3$ for a curve in space). We use this idea to give our first formal definition of a curve in $\mathbb{R}^n$ (we shall be interested only in the cases $n = 2$ or 3, but it is convenient to treat both cases simultaneously).

Definition 1.1.1

A *parametrized curve* in $\mathbb{R}^n$ is a map $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$, for some α, β with $-\infty \leq \alpha < \beta \leq \infty$.

The symbol (α, β) denotes the open interval

$$(\alpha, \beta) = \{t \in \mathbb{R} \mid \alpha < t < \beta\}.$$

A parametrized curve, whose image is contained in a level curve $\mathcal{C}$, is called a *parametrization* of (part of) $\mathcal{C}$. The following examples illustrate how to pass from level curves to parametrized curves and back again in practice.

Example 1.1.2

Let us find a parametrization $\boldsymbol{\gamma}(t)$ of the parabola $y = x^2$. If $\boldsymbol{\gamma}(t) = (\gamma_1(t), \gamma_2(t))$, the components γ_1 and γ_2 of $\boldsymbol{\gamma}$ must satisfy

$$\gamma_2(t) = \gamma_1(t)^2 \tag{1.2}$$

for all values of t in the interval (α, β) where $\boldsymbol{\gamma}$ is defined (yet to be decided), and ideally every point on the parabola should be equal to $(\gamma_1(t), \gamma_2(t))$ for some value of $t \in (\alpha, \beta)$. Of course, there is an obvious solution to Eq. 1.2: take $\gamma_1(t) = t, \gamma_2(t) = t^2$. To get every point on the parabola we must allow t to take every real number value (since the x-coordinate of $\boldsymbol{\gamma}(t)$ is just t, and the x-coordinate of a point on the parabola can be any real number), so we must take (α, β) to be $(-\infty, \infty)$. Thus, the desired parametrization is

$$\boldsymbol{\gamma} : (-\infty, \infty) \to \mathbb{R}^2, \quad \boldsymbol{\gamma}(t) = (t, t^2).$$

But this is not the only parametrization of the parabola. Another choice is $\boldsymbol{\gamma}(t) = (t^3, t^6)$ (with $(\alpha, \beta) = (-\infty, \infty)$). Yet another is $(2t, 4t^2)$, and of course there are (infinitely many) others. So the parametrization of a given level curve is not unique.

Example 1.1.3

Now we try the circle $x^2 + y^2 = 1$. It is tempting to take $x = t$ as in the previous example, so that $y = \sqrt{1 - t^2}$ (we could have taken $y = -\sqrt{1 - t^2}$). So we get the parametrization

$$\boldsymbol{\gamma}(t) = (t, \sqrt{1 - t^2}).$$

But this is only a parametrization of the upper half of the circle because $\sqrt{1 - t^2}$ is always ≥ 0. Similarly, if we had taken $y = -\sqrt{1 - t^2}$, we would only have covered the lower half of the circle.

If we want a parametrization of the whole circle, we must try again. We need functions $\gamma_1(t)$ and $\gamma_2(t)$ such that

$$\gamma_1(t)^2 + \gamma_2(t)^2 = 1 \tag{1.3}$$

for all $t \in (\alpha, \beta)$, and such that *every* point on the circle is equal to $(\gamma_1(t), \gamma_2(t))$ for some $t \in (\alpha, \beta)$. There is an obvious solution to Eq. 1.3: $\gamma_1(t) = \cos t$ and $\gamma_2(t) = \sin t$ (since $\cos^2 t + \sin^2 t = 1$ for all values of t). We can take $(\alpha, \beta) = (-\infty, \infty)$, although this is overkill: any open interval (α, β) whose length is greater than 2π will suffice.

The next example shows how to pass from parametrized curves to level curves.

Example 1.1.4

Take the parametrized curve (called an *astroid*)

$$\boldsymbol{\gamma}(t) = (\cos^3 t, \sin^3 t), \quad t \in \mathbb{R}.$$

Since $\cos^2 t + \sin^2 t = 1$ for all t, the coordinates $x = \cos^3 t$, $y = \sin^3 t$ of the point $\boldsymbol{\gamma}(t)$ satisfy

$$x^{2/3} + y^{2/3} = 1.$$

This level curve coincides with the image of the map $\boldsymbol{\gamma}$. See Exercise 1.1.5 for a picture of the astroid.

In this book, we shall be studying parametrized curves (and later, surfaces) using methods of calculus. Such curves and surfaces will be described almost exclusively in terms of *smooth* functions: a function $f : (\alpha, \beta) \to \mathbb{R}$ is said to be smooth if the derivative $\frac{d^n f}{dt^n}$ exists for all $n \geq 1$ and all $t \in (\alpha, \beta)$. If $f(t)$ and $g(t)$ are smooth functions, it follows from standard results of calculus that the sum $f(t)+g(t)$, product $f(t)g(t)$, quotient $f(t)/g(t)$, and composite $f(g(t))$ are smooth functions, where they are defined.

To differentiate a *vector-valued* function such as $\boldsymbol{\gamma}(t)$ (as in Definition 1.1.1), we differentiate componentwise: if

$$\boldsymbol{\gamma}(t) = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t)),$$

then

$$\frac{d\boldsymbol{\gamma}}{dt} = \left(\frac{d\gamma_1}{dt}, \frac{d\gamma_2}{dt}, \ldots, \frac{d\gamma_n}{dt}\right), \quad \frac{d^2\boldsymbol{\gamma}}{dt^2} = \left(\frac{d^2\gamma_1}{dt^2}, \frac{d^2\gamma_2}{dt^2}, \ldots, \frac{d^2\gamma_n}{dt^2}\right), \quad \text{etc.}$$

To save space, we often denote $d\boldsymbol{\gamma}/dt$ by $\dot{\boldsymbol{\gamma}}(t)$, $d^2\boldsymbol{\gamma}/dt^2$ by $\ddot{\boldsymbol{\gamma}}(t)$, etc. We say that $\boldsymbol{\gamma}$ is *smooth* if the derivatives $d^n\boldsymbol{\gamma}/dt^n$ exist for all $n \geq 1$ and all $t \in (\alpha, \beta)$; this is equivalent to requiring that each of the components $\gamma_1, \gamma_2, \ldots, \gamma_n$ of $\boldsymbol{\gamma}$ is smooth.

From now on, all parametrized curves studied in this book will be assumed to be smooth.

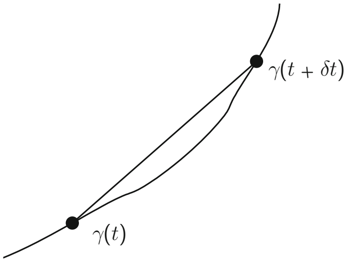
Definition 1.1.5

If $\boldsymbol{\gamma}$ is a parametrized curve, its first derivative $\dot{\boldsymbol{\gamma}}(t)$ is called the *tangent vector* of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$.

To see the reason for this terminology, note that the vector

$$\frac{\boldsymbol{\gamma}(t+\delta t)-\boldsymbol{\gamma}(t)}{\delta t}$$

is parallel to the chord joining the points $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\gamma}(t+\delta t)$ of the image $\mathcal{C}$ of $\boldsymbol{\gamma}$:



As δt tends to zero the length of the chord also tends to zero, but we expect that the *direction* of the chord becomes parallel to that of the tangent to $\mathcal{C}$ at $\boldsymbol{\gamma}(t)$. But the direction of the chord is the same as that of the vector

$$\frac{\boldsymbol{\gamma}(t+\delta t)-\boldsymbol{\gamma}(t)}{\delta t},$$

which tends to $d\boldsymbol{\gamma}/dt$ as δt tends to zero. Of course, this only determines a well-defined direction tangent to the curve if $d\boldsymbol{\gamma}/dt$ is non-zero. If that condition holds, we define the *tangent line* to $\mathcal{C}$ at a point $\mathbf{p}$ of $\mathcal{C}$ to be the straight line passing through $\mathbf{p}$ and parallel to the vector $d\boldsymbol{\gamma}/dt$.

The following result is intuitively clear:

Proposition 1.1.6

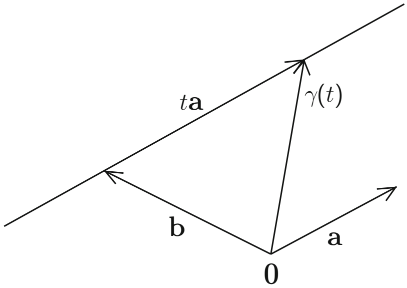
If the tangent vector of a parametrized curve is constant, the image of the curve is (part of) a straight line.

Proof

If $\dot{\boldsymbol{\gamma}}(t)=\mathbf{a}$ for all t, where $\mathbf{a}$ is a constant vector, we have, integrating componentwise,

$$\boldsymbol{\gamma}(t)=\int\frac{d\boldsymbol{\gamma}}{dt}dt=\int\mathbf{a}\,dt=t\,\mathbf{a}+\mathbf{b},$$

where $\mathbf{b}$ is another constant vector. If $\mathbf{a} \neq 0$, this is the parametric equation of the straight line parallel to $\mathbf{a}$ and passing through the point $\mathbf{b}$:



If $\mathbf{a} = 0$, the image of $\boldsymbol{\gamma}$ is a single point (namely, $\mathbf{b}$). □

Before proceeding further with our study of curves, we should point out a potential source of confusion in the discussion of parametrized curves. This is regarding the question what is a 'point' of such a curve? The difficulty can be seen in the following example.

Example 1.1.7

The *limaçon* is the parametrized curve

$$\boldsymbol{\gamma}(t) = ((1 + 2\cos t)\cos t, (1 + 2\cos t)\sin t), \qquad t \in \mathbb{R}$$

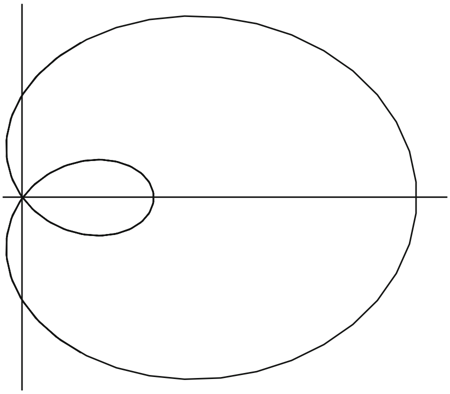
(see the diagram below). Note that $\boldsymbol{\gamma}$ has a self-intersection at the origin in the sense that $\boldsymbol{\gamma}(t) = \mathbf{0}$ for $t = 2\pi/3$ and for $t = 4\pi/3$. The tangent vector is

$$\dot{\boldsymbol{\gamma}}(t) = (-\sin t - 2\sin 2t, \cos t + 2\cos 2t).$$

In particular,

$$\dot{\boldsymbol{\gamma}}(2\pi/3) = (\sqrt{3}/2, -3/2), \quad \dot{\boldsymbol{\gamma}}(4\pi/3) = (-\sqrt{3}/2, -3/2).$$

So what is the tangent vector of this curve at the origin? Although $\dot{\boldsymbol{\gamma}}(t)$ is well-defined for all values of t, it takes different values at $t = 2\pi/3$ and $t = 4\pi/3$, both of which correspond to the point $\mathbf{0}$ on the curve.



This example shows that we must be careful while talking about a 'point' of a parametrized curve $\boldsymbol{\gamma}$: strictly speaking, this should be the same thing as a value of the curve parameter t, and not the corresponding geometric point $\boldsymbol{\gamma}(t) \in \mathbb{R}^n$. Thus, Definition 1.1.5 should more properly read "If $\boldsymbol{\gamma}$ is a parametrized curve, its first derivative $\dot{\boldsymbol{\gamma}}(t)$ is called the *tangent vector* of $\boldsymbol{\gamma}$ at the parameter value t." However, it seems to us that to insist on this distinction takes away from the geometric viewpoint, and we shall sometimes repeat the 'error' committed in the statement of Definition 1.1.5. This should not lead to confusion if the preceding remarks are kept in mind.

EXERCISES

1.1.1 Is $\boldsymbol{\gamma}(t) = (t^2, t^4)$ a parametrization of the parabola $y = x^2$?

1.1.2 Find parametrizations of the following level curves:

(i) $y^2 - x^2 = 1$;

(ii) $\frac{x^2}{4} + \frac{y^2}{9} = 1$.

1.1.3 Find the Cartesian equations of the following parametrized curves:

(i) $\boldsymbol{\gamma}(t) = (\cos^2 t, \sin^2 t)$;

(ii) $\boldsymbol{\gamma}(t) = (e^t, t^2)$.

1.1.4 Calculate the tangent vectors of the curves in Exercise 1.1.3.

1.1.5 Sketch the astroid in Example 1.1.4. Calculate its tangent vector at each point. At which points is the tangent vector zero?

1.1.6 Consider the ellipse

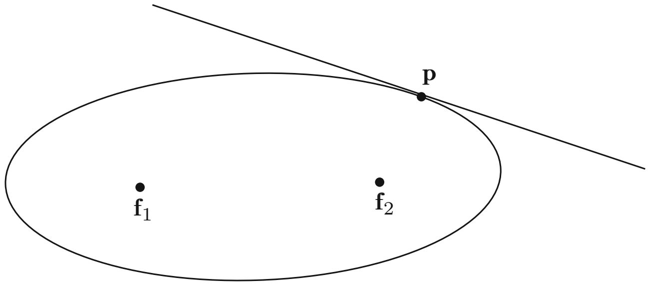
$$\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1,$$

where $p > q > 0$ (see below). The *eccentricity* of the ellipse is $\epsilon = \sqrt{1 - \frac{q^2}{p^2}}$ and the points $(\pm\epsilon p, 0)$ on the x-axis are called the *foci* of the ellipse, which we denote by $\mathbf{f}_1$ and $\mathbf{f}_2$. Verify that $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$ is a parametrization of the ellipse. Prove that

(i) The sum of the distances from $\mathbf{f}_1$ and $\mathbf{f}_2$ to any point $\mathbf{p}$ on the ellipse does not depend on $\mathbf{p}$.

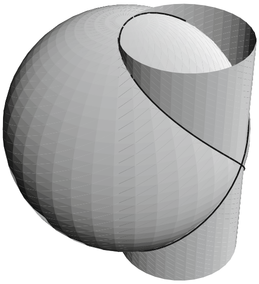
(ii) The product of the distances from $\mathbf{f}_1$ and $\mathbf{f}_2$ to the tangent line at any point $\mathbf{p}$ of the ellipse does not depend on $\mathbf{p}$.

(iii) If $\mathbf{p}$ is any point on the ellipse, the line joining $\mathbf{f}_1$ and $\mathbf{p}$ and that joining $\mathbf{f}_2$ and $\mathbf{p}$ make equal angles with the tangent line to the ellipse at $\mathbf{p}$.



1.1.7 A *cycloid* is the plane curve traced out by a point on the circumference of a circle as it rolls without slipping along a straight line. Show that, if the straight line is the x-axis and the circle has radius $a > 0$, the cycloid can be parametrized as

$$\boldsymbol{\gamma}(t) = a(t - \sin t, 1 - \cos t).$$



1.1.8 Show that $\boldsymbol{\gamma}(t) = (\cos^2 t - \frac{1}{2}, \sin t\cos t, \sin t)$ is a parametrization of the curve of intersection of the circular cylinder of radius $\frac{1}{2}$ and axis the z-axis with the sphere of radius 1 and centre $(-\frac{1}{2}, 0, 0)$. This is called *Viviani's Curve* – see above.

1.1.9 The *normal line* to a curve at a point $\mathbf{p}$ is the straight line passing through $\mathbf{p}$ perpendicular to the tangent line at $\mathbf{p}$. Find the tangent and normal lines to the curve $\boldsymbol{\gamma}(t) = (2\cos t - \cos 2t, 2\sin t - \sin 2t)$ at the point corresponding to $t = \pi/4$.

1.2 Arc-length

We recall that, if $\mathbf{v} = (v_1, \ldots, v_n)$ is a vector in $\mathbb{R}^n$, its *length* is

$$\| \mathbf{v} \| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

If $\mathbf{u}$ is another vector in $\mathbb{R}^n$, $\| \mathbf{u} - \mathbf{v} \|$ is the length of the straight line segment joining the points $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.

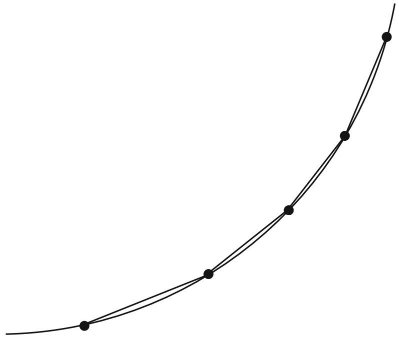
To find a formula for the length of a parametrized curve $\boldsymbol{\gamma}$, note that, if δt is very small, the part of the image $\mathcal{C}$ of $\boldsymbol{\gamma}$ between $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\gamma}(t + \delta t)$ is nearly a straight line, so its length is approximately

$$\| \boldsymbol{\gamma}(t + \delta t) - \boldsymbol{\gamma}(t) \|.$$

Again, since δt is small, $(\boldsymbol{\gamma}(t + \delta t) - \boldsymbol{\gamma}(t))/\delta t$ is nearly equal to $\dot{\boldsymbol{\gamma}}(t)$, so the length is approximately

$$\| \dot{\boldsymbol{\gamma}}(t) \| \, \delta t. \tag{1.4}$$

If we want to calculate the length of a (not necessarily small) part of $\mathcal{C}$, we can divide it into segments, each of which corresponds to a small increment δt in t, calculate the length of each segment using (1.4), and add up the results. Letting δt tend to zero should then give the exact length.



This motivates the following definition:

Definition 1.2.1

The *arc-length* of a curve $\boldsymbol{\gamma}$ starting at the point $\boldsymbol{\gamma}(t_0)$ is the function $s(t)$ given by

$$s(t) = \int_{t_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du.$$

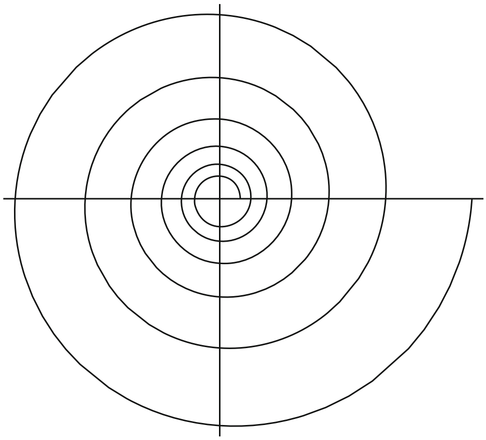
Thus, $s(t_0) = 0$ and $s(t)$ is positive or negative according to whether t is larger or smaller than t_0. If we choose a different starting point $\boldsymbol{\gamma}(\tilde{t}_0)$, the resulting arc-length $\tilde{s}$ differs from s by the constant $\int_{t_0}^{\tilde{t}_0} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du$ because

$$\int_{t_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du = \int_{\tilde{t}_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du + \int_{t_0}^{\tilde{t}_0} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du.$$

Example 1.2.2

For a *logarithmic spiral*

$$\boldsymbol{\gamma}(t) = (e^{kt} \cos t, e^{kt} \sin t),$$



where k is a non-zero constant, we have

$$\dot{\boldsymbol{\gamma}} = (e^{kt}(k \cos t - \sin t), e^{kt}(k \sin t + \cos t)),$$
$$\therefore \quad \parallel \dot{\boldsymbol{\gamma}} \parallel^2 = e^{2kt}(k \cos t - \sin t)^2 + e^{2kt}(k \sin t + \cos t)^2 = (k^2 + 1)e^{2kt}.$$

Hence, the arc-length of $\boldsymbol{\gamma}$ starting at $\boldsymbol{\gamma}(0) = (1, 0)$ (for example) is

$$s = \int_0^t \sqrt{k^2 + 1} e^{ku} \, du = \frac{\sqrt{k^2 + 1}}{k}(e^{kt} - 1).$$

The arc-length is a differentiable function. Indeed, if s is the arc-length of a curve $\boldsymbol{\gamma}$ starting at $\boldsymbol{\gamma}(t_0)$, we have

$$\frac{ds}{dt} = \frac{d}{dt}\int_{t_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du = \parallel \dot{\boldsymbol{\gamma}}(t) \parallel . \tag{1.5}$$

Thinking of $\boldsymbol{\gamma}(t)$ as the position of a moving point at time t, ds/dt is the speed of the point (rate of change of distance along the curve). This suggests the following definition.

Definition 1.2.3

If $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$ is a parametrized curve, its *speed* at the point $\boldsymbol{\gamma}(t)$ is $\parallel \dot{\boldsymbol{\gamma}}(t) \parallel$, and $\boldsymbol{\gamma}$ is said to be a *unit-speed* curve if $\dot{\boldsymbol{\gamma}}(t)$ is a unit vector for all $t \in (\alpha, \beta)$.

We shall see many examples of formulas and results relating to curves that take on a much simpler form when the curve is unit-speed. The reason for this simplification is given in the next proposition. Although this admittedly looks uninteresting at first sight, it will be extremely useful for what follows.

We recall that the *dot product* (or scalar product) of vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$ is

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i.$$

If $\mathbf{a}$ and $\mathbf{b}$ are smooth functions of a parameter t, we shall make use of the 'product formula'

$$\frac{d}{dt}(\mathbf{a} \cdot \mathbf{b}) = \frac{d\mathbf{a}}{dt} \cdot \mathbf{b} + \mathbf{a} \cdot \frac{d\mathbf{b}}{dt}.$$

This follows easily from the definition of the dot product and the usual product formula for scalar functions,

$$\frac{d}{dt}(a_i b_i) = \frac{da_i}{dt} b_i + a_i \frac{db_i}{dt}.$$

Proposition 1.2.4

Let $\mathbf{n}(t)$ be a unit vector that is a smooth function of a parameter t. Then, the dot product

$$\dot{\mathbf{n}}(t) \cdot \mathbf{n}(t) = 0$$

for all t, i.e., $\dot{\mathbf{n}}(t)$ is zero or perpendicular to $\mathbf{n}(t)$ for all t.

In particular, if $\boldsymbol{\gamma}$ is a unit-speed curve, then $\ddot{\boldsymbol{\gamma}}$ is zero or perpendicular to $\dot{\boldsymbol{\gamma}}$.

Proof

Using the product formula to differentiate both sides of the equation $\mathbf{n} \cdot \mathbf{n} = 1$ with respect to t gives

$$\dot{\mathbf{n}} \cdot \mathbf{n} + \mathbf{n} \cdot \dot{\mathbf{n}} = 0,$$

so $2\dot{\mathbf{n}} \cdot \mathbf{n} = 0$. The last part follows by taking $\mathbf{n} = \dot{\boldsymbol{\gamma}}$. □

EXERCISES

1.2.1 Calculate the arc-length of the *catenary* $\boldsymbol{\gamma}(t) = (t, \cosh t)$ starting at the point $(0, 1)$. This curve has the shape of a heavy chain suspended at its ends – see Exercise 2.2.4.

1.2.2 Show that the following curves are unit-speed:

(i) $\boldsymbol{\gamma}(t) = \left(\frac{1}{3}(1+t)^{3/2}, \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$.

(ii) $\boldsymbol{\gamma}(t) = \left(\frac{4}{5}\cos t, 1 - \sin t, -\frac{3}{5}\cos t\right)$.

1.2.3 A plane curve is given by

$$\boldsymbol{\gamma}(\theta) = (r\cos\theta, r\sin\theta),$$

where r is a smooth function of θ (so that (r, θ) are the polar coordinates of $\boldsymbol{\gamma}(\theta)$). Under what conditions is $\boldsymbol{\gamma}$ regular? Find all functions $r(\theta)$ for which $\boldsymbol{\gamma}$ is unit-speed. Show that, if $\boldsymbol{\gamma}$ is unit-speed, the image of $\boldsymbol{\gamma}$ is a circle; what is its radius?

1.2.4 This exercise shows that *a straight line is the shortest curve joining two given points*. Let $\mathbf{p}$ and $\mathbf{q}$ be the two points, and let $\boldsymbol{\gamma}$ be a curve passing through both, say $\boldsymbol{\gamma}(a) = \mathbf{p}$, $\boldsymbol{\gamma}(b) = \mathbf{q}$, where $a < b$. Show that, if $\mathbf{u}$ is any unit vector,

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{u} \leq \parallel \dot{\boldsymbol{\gamma}} \parallel$$

and deduce that

$$(\mathbf{q} - \mathbf{p}) \cdot \mathbf{u} \leq \int_a^b \parallel \dot{\boldsymbol{\gamma}} \parallel \, dt.$$

By taking $\mathbf{u} = (\mathbf{q} - \mathbf{p}) / \parallel \mathbf{q} - \mathbf{p} \parallel$, show that the length of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}$ is at least the straight line distance $\parallel \mathbf{q} - \mathbf{p} \parallel$.

1.3 Reparametrization

We saw in Examples 1.1.2 and 1.1.3 that a given level curve can have many parametrizations, and it is important to understand the relation between them.

Definition 1.3.1

A parametrized curve $\tilde{\boldsymbol{\gamma}} : (\tilde{\alpha}, \tilde{\beta}) \to \mathbb{R}^n$ is a *reparametrization* of a parametrized curve $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$ if there is a smooth bijective map $\phi : (\tilde{\alpha}, \tilde{\beta}) \to (\alpha, \beta)$ (the *reparametrization map*) such that the inverse map $\phi^{-1} : (\alpha, \beta) \to (\tilde{\alpha}, \tilde{\beta})$ is also smooth and

$$\tilde{\boldsymbol{\gamma}}(\tilde{t}) = \boldsymbol{\gamma}(\phi(\tilde{t})) \quad \text{for all} \quad \tilde{t} \in (\tilde{\alpha}, \tilde{\beta}). \tag{1.6}$$

Note that, since ϕ has a smooth inverse, $\boldsymbol{\gamma}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$:

$$\tilde{\boldsymbol{\gamma}}(\phi^{-1}(t)) = \boldsymbol{\gamma}(\phi(\phi^{-1}(t))) = \boldsymbol{\gamma}(t) \text{ for all } t \in (\alpha, \beta).$$

Two curves that are reparametrizations of each other have the same image, so they should have the same geometric properties.

Example 1.3.2

In Example 1.1.3, we found that the circle $x^2 + y^2 = 1$ has a parametrization $\boldsymbol{\gamma}(t) = (\cos t, \sin t)$. Another parametrization is

$$\tilde{\boldsymbol{\gamma}}(t) = (\sin t, \cos t)$$

(since $\sin^2 t + \cos^2 t = 1$). To see that $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$, we have to find a reparametrization map ϕ such that

$$(\cos \phi(t), \sin \phi(t)) = (\sin t, \cos t).$$

One solution is $\phi(t) = \pi/2 - t$.

As we remarked in Section 1.2, the analysis of a curve is simplified when it is known to be unit-speed. It is therefore important to know exactly which curves have unit-speed reparametrizations.

Definition 1.3.3

A point $\boldsymbol{\gamma}(t)$ of a parametrized curve $\boldsymbol{\gamma}$ is called a *regular point* if $\dot{\boldsymbol{\gamma}}(t) \neq \mathbf{0}$; otherwise $\boldsymbol{\gamma}(t)$ is a *singular point* of $\boldsymbol{\gamma}$. A curve is *regular* if all of its points are regular.

Before we show the relation between regularity and unit-speed reparametrization, we note two simple properties of regular curves. Although these results are not particularly appealing, they are very important for what is to follow.

Proposition 1.3.4

Any reparametrization of a regular curve is regular.

Proof

Suppose that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ are related as in Definition 1.3.1, let $t = \phi(\tilde{t})$ and $\psi = \phi^{-1}$ so that $\tilde{t} = \psi(t)$. Differentiating both sides of the equation $\phi(\psi(t)) = t$ with respect to t and using the chain rule gives

$$\frac{d\phi}{d\tilde{t}}\frac{d\psi}{dt} = 1.$$

This shows that $d\phi/d\tilde{t}$ is never zero. Since $\tilde{\boldsymbol{\gamma}}(\tilde{t}) = \boldsymbol{\gamma}(\phi(\tilde{t}))$, another application of the chain rule gives

$$\frac{d\tilde{\boldsymbol{\gamma}}}{d\tilde{t}} = \frac{d\boldsymbol{\gamma}}{dt}\frac{d\phi}{d\tilde{t}},$$

which shows that $d\tilde{\boldsymbol{\gamma}}/d\tilde{t}$ is never zero, if $d\boldsymbol{\gamma}/dt$ is never zero. □

Proposition 1.3.5

If $\boldsymbol{\gamma}(t)$ is a regular curve, its arc-length s (see Definition 1.2.1), starting at any point of $\boldsymbol{\gamma}$, is a smooth function of t.

Proof

We have already seen that (whether or not $\boldsymbol{\gamma}$ is regular) s is a differentiable function of t and

$$\frac{ds}{dt} = \parallel \dot{\boldsymbol{\gamma}}(t) \parallel.$$

To simplify the notation, assume from now onwards that $\boldsymbol{\gamma}$ is a plane curve, say

$$\boldsymbol{\gamma}(t) = (u(t), v(t)),$$

where u and v are smooth functions of t, so that

$$\frac{ds}{dt} = \sqrt{\dot{u}^2 + \dot{v}^2}.$$

The crucial point is that the function $f(x) = \sqrt{x}$ is a *smooth* function on the open interval $(0, \infty)$. Indeed, it is easy to prove by induction on $n \geq 1$ that

$$\frac{d^n f}{dx^n} = (-1)^{n-1} \frac{1.3.5.\ldots.(2n-1)}{2^n} x^{-(2n+1)/2}.$$

Since u and v are smooth functions of t, so are $\dot{u}$ and $\dot{v}$ and hence is $\dot{u}^2 + \dot{v}^2$. Since $\boldsymbol{\gamma}$ is regular, $\dot{u}^2 + \dot{v}^2 > 0$ for all values of t, so the composite function

$$\frac{ds}{dt} = f(\dot{u}^2 + \dot{v}^2)$$

is a smooth function of t, and hence s itself is smooth. □

The main result we want is the following.

Proposition 1.3.6

A parametrized curve has a unit-speed reparametrization if and only if it is regular.

Proof

Suppose first that a parametrized curve $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$ has a unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$, with reparametrization map ϕ. Letting $t = \phi(\tilde{t})$, we have $\tilde{\boldsymbol{\gamma}}(\tilde{t}) = \boldsymbol{\gamma}(t)$ and so

$$\begin{aligned} \frac{d\tilde{\boldsymbol{\gamma}}}{d\tilde{t}} &= \frac{d\boldsymbol{\gamma}}{dt}\frac{dt}{d\tilde{t}}, \\ \therefore \quad \left\| \frac{d\tilde{\boldsymbol{\gamma}}}{d\tilde{t}} \right\| &= \left\| \frac{d\boldsymbol{\gamma}}{dt} \right\| \left| \frac{dt}{d\tilde{t}} \right|. \end{aligned}$$

Since $\tilde{\boldsymbol{\gamma}}$ is unit-speed, $\| d\tilde{\boldsymbol{\gamma}}/d\tilde{t} \| = 1$, so $d\boldsymbol{\gamma}/dt$ cannot be zero.

Conversely, suppose that the tangent vector $d\boldsymbol{\gamma}/dt$ is never zero. By Eq. 1.5, $ds/dt > 0$ for all t, where s is the arc-length of $\boldsymbol{\gamma}$ starting at any point of the curve, and by Proposition 1.3.5 s is a smooth function of t. It follows from the inverse function theorem that $s : (\alpha, \beta) \to \mathbb{R}$ is injective, that its image is an open interval $(\tilde{\alpha}, \tilde{\beta})$, and that the inverse map $s^{-1} : (\tilde{\alpha}, \tilde{\beta}) \to (\alpha, \beta)$ is smooth. (Readers unfamiliar with the inverse function theorem should accept these statements for now; the theorem will be discussed informally in Section 1.5 and formally in Section 5.6.) We take $\phi = s^{-1}$ and let $\tilde{\boldsymbol{\gamma}}$ be the corresponding reparametrization of $\boldsymbol{\gamma}$, so that $\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t)$ (see Eq. 1.6). Then,

$$\frac{d\tilde{\boldsymbol{\gamma}}}{ds}\frac{ds}{dt} = \frac{d\boldsymbol{\gamma}}{dt},$$

$$\therefore \quad \left\| \frac{d\tilde{\gamma}}{ds} \right\| \frac{ds}{dt} = \left\| \frac{d\gamma}{dt} \right\| = \frac{ds}{dt} \quad \text{(by Eq. 1.5)},$$
$$\therefore \quad \left\| \frac{d\tilde{\gamma}}{ds} \right\| = 1.$$
□

The proof of Proposition 1.3.6 shows that the arc-length is essentially the only unit-speed parameter on a regular curve:

Corollary 1.3.7

Let γ be a regular curve and let $\tilde{\gamma}$ be a unit-speed reparametrization of γ:

$$\tilde{\gamma}(u(t)) = \gamma(t) \quad \text{for all } t,$$

where u is a smooth function of t. Then, if s is the arc-length of γ (starting at any point), we have

$$u = \pm s + c, \tag{1.7}$$

where c is a constant. Conversely, if u is given by Eq. 1.7 for some value of c and with either sign, then $\tilde{\gamma}$ is a unit-speed reparametrization of γ.

Proof

The calculation in the first part of the proof of Proposition 1.3.6 shows that u gives a unit-speed reparametrization of γ if and only if

$$\frac{du}{dt} = \pm \left\| \frac{d\gamma}{dt} \right\| = \pm \frac{ds}{dt} \quad \text{(by Eq. 1.5)},$$

which is equivalent to $u = \pm s + c$ for some constant c. □

Although every regular curve has a unit-speed reparametrization, this may be very complicated, or even impossible, to write down 'explicitly', as the following examples show.

Example 1.3.8

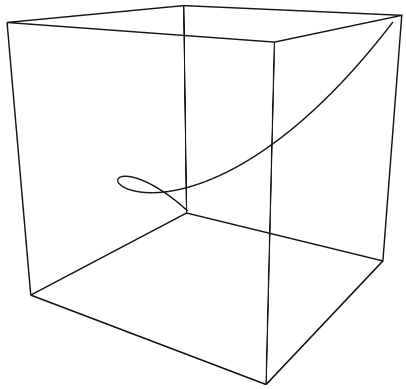
For the logarithmic spiral $\gamma(t) = (e^{kt}\cos t, e^{kt}\sin t)$, we found in Example 1.2.2 that $\| \dot{\gamma} \|^2 = (k^2+1)e^{2kt}$. This is never zero, so γ is regular. The arc-length of γ starting at $(1,0)$ was found to be $s = \sqrt{k^2+1}(e^{kt}-1)/k$. Hence, $t = \frac{1}{k}\ln\left(\frac{ks}{\sqrt{k^2+1}} + 1\right)$, so a unit-speed reparametrization of γ is given by the rather unwieldy formula

$$\tilde{\gamma}(s) = \left(\left(\frac{ks}{\sqrt{k^2+1}} + 1 \right) \cos\left(\frac{1}{k} \ln\left(\frac{ks}{\sqrt{k^2+1}} + 1 \right)\right), \right.$$
$$\left. \left(\frac{ks}{\sqrt{k^2+1}} + 1 \right) \sin\left(\frac{1}{k} \ln\left(\frac{ks}{\sqrt{k^2+1}} + 1 \right)\right) \right).$$

Example 1.3.9

The *twisted cubic* is the space curve given by

$$\boldsymbol{\gamma}(t) = (t, t^2, t^3), \quad t \in \mathbb{R}.$$



We have $\dot{\boldsymbol{\gamma}}(t) = (1, 2t, 3t^2)$ and so

$$\| \dot{\boldsymbol{\gamma}}(t) \| = \sqrt{1 + 4t^2 + 9t^4}.$$

This is never zero, so $\boldsymbol{\gamma}$ is regular. The arc-length starting at $\boldsymbol{\gamma}(0) = \mathbf{0}$ is

$$s = \int_0^t \sqrt{1 + 4u^2 + 9u^4}\, du.$$

This integral cannot be evaluated in terms of familiar functions like logarithms and exponentials, and trigonometric functions. (It is an example of an *elliptic integral.*)

Our final example shows that a given level curve can have both regular and non-regular parametrizations.

Example 1.3.10

For the parametrization $\boldsymbol{\gamma}(t) = (t, t^2)$ of the parabola $y = x^2$, $\dot{\boldsymbol{\gamma}}(t) = (1, 2t)$ is obviously never zero, so $\boldsymbol{\gamma}$ is regular. But $\tilde{\boldsymbol{\gamma}}(t) = (t^3, t^6)$ is also a parametrization of the same parabola. This time, $\dot{\tilde{\boldsymbol{\gamma}}} = (3t^2, 6t^5)$, and this is zero when $t = 0$, so $\tilde{\boldsymbol{\gamma}}$ is *not* regular.

EXERCISES

1.3.1 Which of the following curves are regular?

(i) $\gamma(t) = (\cos^2 t, \sin^2 t)$ for $t \in \mathbb{R}$.

(ii) The same curve as in (i), but with $0 < t < \pi/2$.

(iii) $\gamma(t) = (t, \cosh t)$ for $t \in \mathbb{R}$.

Find unit-speed reparametrizations of the regular curve(s).

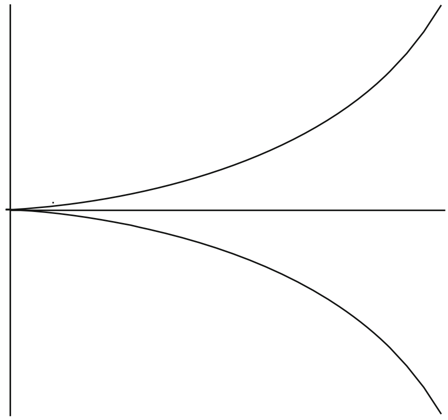
1.3.2 The *cissoid of Diocles* (see below) is the curve whose equation in terms of polar coordinates (r, θ) is

$$r = \sin\theta \tan\theta, \quad -\pi/2 < \theta < \pi/2.$$

Write down a parametrization of the cissoid using θ as a parameter and show that

$$\gamma(t) = \left(t^2, \frac{t^3}{\sqrt{1-t^2}}\right), \quad -1 < t < 1$$

is a reparametrization of it.



1.3.3 The simplest type of singular point of a curve $\boldsymbol{\gamma}$ is an *ordinary cusp*: a point $\mathbf{p}$ of $\boldsymbol{\gamma}$, corresponding to a parameter value t_0, say, is an ordinary cusp if $\dot{\boldsymbol{\gamma}}(t_0) = \mathbf{0}$ and the vectors $\ddot{\boldsymbol{\gamma}}(t_0)$ and $\dddot{\boldsymbol{\gamma}}(t_0)$ are linearly independent (in particular, these vectors must both be non-zero). Show that:

(i) The curve $\boldsymbol{\gamma}(t) = (t^m, t^n)$, where m and n are positive integers, has an ordinary cusp at the origin if and only if $(m, n) = (2, 3)$ or $(3, 2)$.

(ii) The cissoid in Exercise 1.3.2 has an ordinary cusp at the origin.

(iii) If $\boldsymbol{\gamma}$ has an ordinary cusp at a point $\mathbf{p}$, so does any reparametrization of $\boldsymbol{\gamma}$.

1.3.4 Show that:

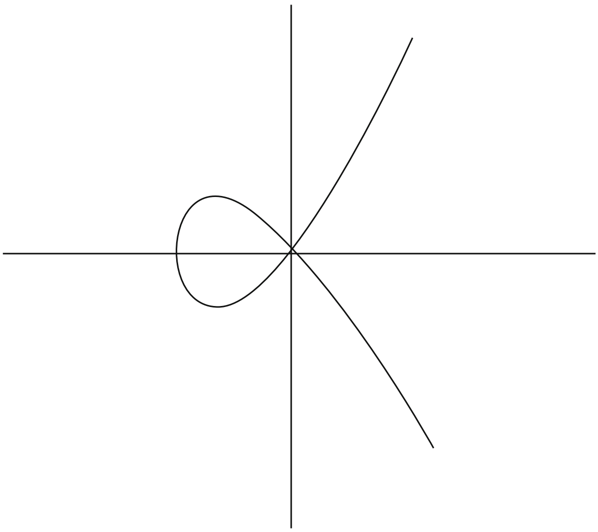
(i) If $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of a curve $\boldsymbol{\gamma}$, then $\boldsymbol{\gamma}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$.

(ii) If $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$, and $\hat{\boldsymbol{\gamma}}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$, then $\hat{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$.

1.4 Closed curves

It is obvious that some curves 'close up', like a circle or an ellipse, while some do not, like a straight line or a parabola. If a point moves, say at constant speed, around a curve that closes up, it will return to its starting point after some time interval, and will then trace out the same curve all over again. On the other hand, if a point moves at constant speed along a straight line or a parabola, it will never return to its starting point. But there are some intermediate cases like

$$\boldsymbol{\gamma}(t) = (t^2 - 1, t^3 - t);$$



a point moving at constant speed along this curve may return to its starting point if the starting point is the origin, but will not do so otherwise. So a careful definition of what it means for a curve to 'close up' is needed.

Definition 1.4.1

Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a smooth curve and let $T \in \mathbb{R}$. We say that γ is *T-periodic* if

$$\gamma(t+T) = \gamma(t) \qquad \text{for all } t \in \mathbb{R}.$$

If γ is not constant and is T-periodic for some $T \neq 0$, then γ is said to be *closed*.

Thus, if γ is T-periodic, a point moving around γ returns to its starting point after time T, whatever the starting point is. Of course, every curve is 0-periodic.

Remark

If γ is T-periodic, it is clear that γ is determined by its restriction to any interval of length $|T|$. Conversely, closed curves are often given to us as curves defined on a closed interval, say $\gamma : [a, b] \to \mathbb{R}^n$. If γ and all its derivatives take the same value at a and b,[1] there is a unique way to extend γ to a $(b-a)$-periodic (smooth) curve $\gamma : \mathbb{R} \to \mathbb{R}^n$. Thus, the discussion below can be applied to curves defined on closed intervals.

It is clear that if a curve γ is T-periodic then it is $(-T)$-periodic because

$$\gamma(t-T) = \gamma((t-T)+T) = \gamma(t).$$

It follows that if γ is T-periodic for some $T \neq 0$, then it is T-periodic for some $T > 0$.

Definition 1.4.2

The *period* of a closed curve γ is the smallest positive number T such that γ is T-periodic.

It is actually not quite obvious that this number T exists (remember that not every set of positive real numbers has a smallest element). A proof that it does exist can be found in the exercises.

Example 1.4.3

The ellipse $\gamma(t) = (p\cos t, q\sin t)$ (Exercise 1.1.6) is a closed curve with period 2π because both of its components are (by well-known properties of trigonometric functions).

[1] The derivatives at the endpoints a and b must be defined in the one-sided sense.

If $\boldsymbol{\gamma}$ is a regular closed curve, a unit-speed reparametrization of $\boldsymbol{\gamma}$ is always closed. To see this, note that since every point in the image of a closed curve $\boldsymbol{\gamma}$ of period T is traced out as the parameter t of $\boldsymbol{\gamma}$ varies through any interval of length T, for example, $0 \leq t \leq T$, it is reasonable to define the *length of* $\boldsymbol{\gamma}$ to be

$$\ell(\boldsymbol{\gamma}) = \int_0^T \parallel \dot{\boldsymbol{\gamma}}(t) \parallel dt.$$

By the proof of Proposition 1.3.6, using the arc-length

$$s = \int_0^t \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du$$

of $\boldsymbol{\gamma}$ as the parameter gives a unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ (so that $\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t)$). Note that

$$s(t+T) = \int_0^{t+T} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du = \int_0^T \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du + \int_T^{t+T} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du = \ell(\boldsymbol{\gamma}) + s(t),$$

since, putting $v = u - T$ and using $\boldsymbol{\gamma}(u - T) = \boldsymbol{\gamma}(u)$ (and hence by differentiation $\dot{\boldsymbol{\gamma}}(u - T) = \dot{\boldsymbol{\gamma}}(u)$), we get

$$\int_T^{t+T} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du = \int_0^t \parallel \dot{\boldsymbol{\gamma}}(v) \parallel \, dv = s(t).$$

Hence,

$$\tilde{\boldsymbol{\gamma}}(s(t)) = \tilde{\boldsymbol{\gamma}}(s(t')) \Longleftrightarrow \boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t') \Longleftrightarrow t' - t = kT \Longleftrightarrow s(t') - s(t) = k\ell(\boldsymbol{\gamma}),$$

where k is an integer. This shows that $\tilde{\boldsymbol{\gamma}}$ is a closed curve with period $\ell(\boldsymbol{\gamma})$. Note that, since $\tilde{\boldsymbol{\gamma}}$ is unit-speed, this is also the length of $\tilde{\boldsymbol{\gamma}}$. In short, *we can always assume that a closed curve is unit-speed and that its period is equal to its length.*

Returning to the curve illustrated at the beginning of this section, it is clearly not closed; nevertheless, if a point starts at the origin and moves at constant speed around the loop in the region $x < 0$ it will return to its starting point. This suggests the following definition.

Definition 1.4.4

A curve $\boldsymbol{\gamma}$ is said to have a *self-intersection* at a point $\mathbf{p}$ of the curve if there exist parameter values $a \neq b$ such that

(i) $\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b) = \mathbf{p}$, and

(ii) if $\boldsymbol{\gamma}$ is closed with period T, then $a - b$ is not an integer multiple of T.

Example 1.4.5

The limaçon in Example 1.1.7 is a closed curve with period 2π. It is clear from the picture that it has exactly one self-intersection, at the origin. (This can also be verified analytically – cf. Exercise 1.4.1 and its solution.)

EXERCISES

1.4.1 Show that the *Cayley sextic*

$$\boldsymbol{\gamma}(t) = (\cos^3 t \cos 3t, \cos^3 t \sin 3t), \qquad t \in \mathbb{R},$$

is a closed curve which has exactly one self-intersection. What is its period? (The name of this curve derives from the fact that its Cartesian equation involves a polynomial of degree 6.)

1.4.2 Give an example to show that a reparametrization of a closed curve need not be closed.

1.4.3 Show that if a curve $\boldsymbol{\gamma}$ is T_1-periodic and T_2-periodic, then it is $(k_1T_1 + k_2T_2)$-periodic for any integers k_1, k_2.

1.4.4 Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a curve and suppose that T_0 is the smallest positive number such that $\boldsymbol{\gamma}$ is T_0-periodic. Prove that $\boldsymbol{\gamma}$ is T-periodic if and only if $T = kT_0$ for some integer k.

1.4.5 Suppose that a *non-constant* function $\gamma : \mathbb{R} \to \mathbb{R}$ is T-periodic for some $T \neq 0$. This exercise shows that there is a *smallest* positive T_0 such that γ is T_0-periodic. The proof uses a little real analysis. Suppose for a contradiction that there is no such T_0.

(i) Show that there is a sequence $T_1, T_2, T_3, \dots$ such that $T_1 > T_2 > T_3 > \cdots > 0$ and that γ is T_r-periodic for all $r \geq 1$.

(ii) Show that the sequence $\{T_r\}$ in (i) can be chosen so that $T_r \to 0$ as $r \to \infty$.

(iii) Show that the existence of a sequence $\{T_r\}$ as in (i) such that $T_r \to 0$ as $r \to \infty$ implies that γ is constant.

1.4.6 Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a non-constant curve that is T-periodic for some $T > 0$. Show that $\boldsymbol{\gamma}$ is closed.

1.5 Level curves versus parametrized curves

We shall now try to clarify the relation between the two types of curves we have considered in previous sections.

Level curves in the generality we have defined them are not always the kind of objects we would want to call curves. For example, the level 'curve' $x^2 + y^2 = 0$ is a single point. The correct conditions to impose on a function $f(x, y)$ in order that $f(x, y) = c$, where c is a constant, will be an acceptable level curve in the plane are contained in the following theorem, which shows that such level curves can be parametrized. Note that we might as well assume that $c = 0$ (since we can replace f by $f - c$).

Theorem 1.5.1

Let $f(x, y)$ be a smooth function of two variables (which means that all the partial derivatives of f, of all orders, exist and are continuous functions). Assume that, at every point of the level curve

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\},$$

$\partial f/\partial x$ and $\partial f/\partial y$ are not both zero. If $\mathbf{p}$ is a point of $\mathcal{C}$, with coordinates (x_0, y_0), say, there is a regular parametrized curve $\boldsymbol{\gamma}(t)$, defined on an open interval containing 0, such that $\boldsymbol{\gamma}$ passes through $\mathbf{p}$ when $t = 0$ and $\boldsymbol{\gamma}(t)$ is contained in $\mathcal{C}$ for all t.

The proof of this theorem makes use of the inverse function theorem (one version of which has already been used in the proof of Proposition 1.3.6). For the moment, we shall only try to convince the reader of the truth of this theorem. The proof will be given later (Exercise 5.6.2) after the inverse function theorem has been formally introduced and used in our discussion of surfaces.

To understand the significance of the conditions on f in Theorem 1.5.1, suppose that $(x_0 + \Delta x, y_0 + \Delta y)$ is a point of $\mathcal{C}$ near $\mathbf{p}$, so that

$$f(x_0 + \Delta x, y_0 + \Delta y) = 0.$$

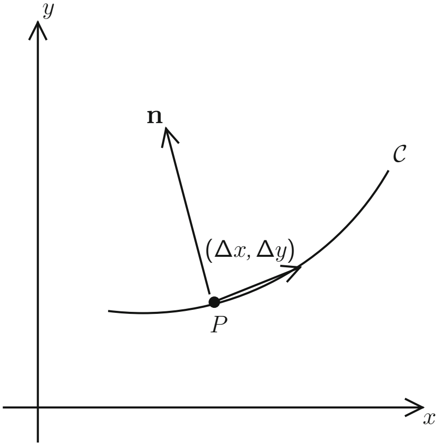
By the two-variable form of Taylor's theorem,

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

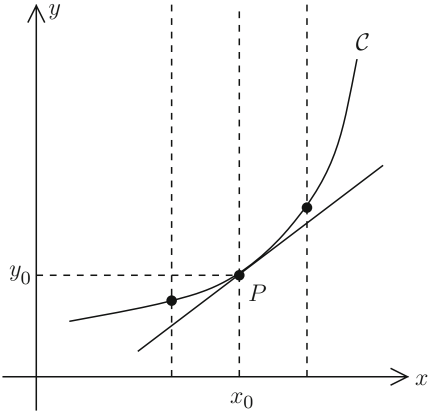
neglecting products of the small quantities Δx and Δy (the partial derivatives are evaluated at (x_0, y_0)). Hence,

$$\Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} = 0. \tag{1.8}$$

Since Δx and Δy are small, the vector $(\Delta x, \Delta y)$ is nearly tangent to $\mathcal{C}$ at $\mathbf{p}$, so Eq. 1.8 says that *the vector* $\mathbf{n} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ *is perpendicular to* $\mathcal{C}$ *at* $\mathbf{p}$.



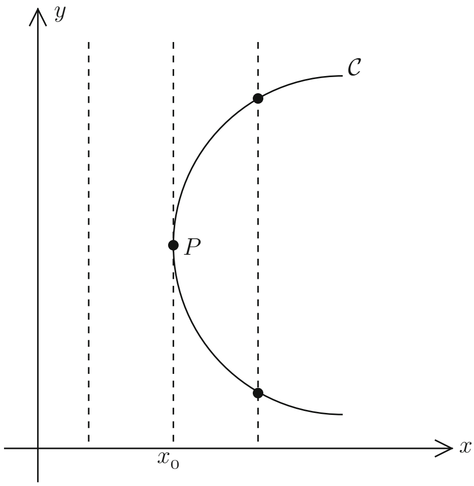
The hypothesis in Theorem 1.5.1 tells us that the vector $\mathbf{n}$ is non-zero at every point of $\mathcal{C}$. Suppose, for example, that $\frac{\partial f}{\partial y} \neq 0$ at $\mathbf{p}$. Then, $\mathbf{n}$ is not parallel to the x-axis at $\mathbf{p}$, so the tangent to $\mathcal{C}$ at $\mathbf{p}$ is not parallel to the y-axis.



This implies that vertical lines $x =$ constant near $x = x_0$ all intersect $\mathcal{C}$ in a unique point (x, y) near $\mathbf{p}$. In other words, *the equation*

$$f(x, y) = 0 \tag{1.9}$$

has a unique solution y near y_0 for every x near x_0. Note that this may fail to be the case if the tangent to $\mathcal{C}$ at $\mathbf{p}$ is parallel to the y-axis (i.e., if $\partial f/\partial y = 0$):

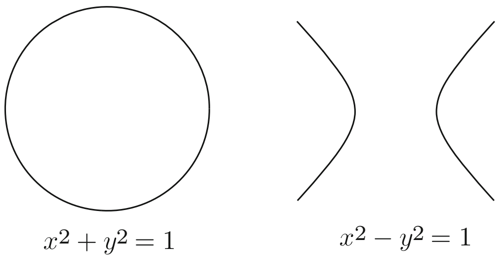


In this example, lines x = constant just to the left of $x = x_0$ do not meet $\mathcal{C}$ near $\mathbf{p}$, while those just to the right of $x = x_0$ meet $\mathcal{C}$ in more than one point near $\mathbf{p}$.

The italicized statement about f in the last paragraph means that there is a function $g(x)$, defined for x near x_0, such that $y = g(x)$ is the unique solution of Eq. 1.9 near y_0. We can now define a parametrization $\boldsymbol{\gamma}$ of the part of $\mathcal{C}$ near $\mathbf{p}$ by

$$\boldsymbol{\gamma}(t) = (t, g(t)).$$

If we accept that g is smooth (which follows from the inverse function theorem), then $\boldsymbol{\gamma}$ is certainly regular since $\dot{\boldsymbol{\gamma}} = (1, \dot{g})$ is obviously never zero. This 'proves' Theorem 1.5.1.



It is actually possible to prove slightly more than we have stated in Theorem 1.5.1. Suppose that $f(x, y)$ satisfies the conditions in the theorem, and assume in addition that the level curve $\mathcal{C}$ given by $f(x, y) = 0$ is *connected*.

For readers unfamiliar with point set topology, this means roughly that $\mathcal{C}$ is in 'one piece'. For example, the circle $x^2 + y^2 = 1$ is connected, but the hyperbola $x^2 - y^2 = 1$ is not (see above). With these assumptions on f, there is a regular parametrized curve $\boldsymbol{\gamma}$ whose image is *the whole* of $\mathcal{C}$. Moreover, if $\mathcal{C}$ is not closed $\boldsymbol{\gamma}$ can be taken to be injective; if $\mathcal{C}$ is closed, then $\boldsymbol{\gamma}$ maps some closed interval $[\alpha, \beta]$ *onto* $\mathcal{C}$, $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\gamma}(\beta)$ and $\boldsymbol{\gamma}$ is injective on the open interval (α, β).

A similar argument can be used to pass from parametrized curves to level curves:

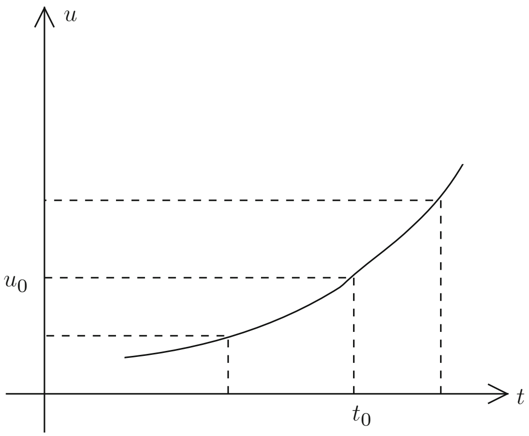
Theorem 1.5.2

Let $\boldsymbol{\gamma}$ be a regular parametrized plane curve, and let $\boldsymbol{\gamma}(t_0) = (x_0, y_0)$ be a point in the image of $\boldsymbol{\gamma}$. Then, there is a smooth real-valued function $f(x, y)$, defined for x and y in open intervals containing x_0 and y_0, respectively, and satisfying the conditions in Theorem 1.5.1, such that $\boldsymbol{\gamma}(t)$ is contained in the level curve $f(x, y) = 0$ for all values of t in some open interval containing t_0.

The proof of Theorem 1.5.2 is similar to that of Theorem 1.5.1. Let

$$\boldsymbol{\gamma}(t) = (u(t), v(t)),$$

where u and v are smooth functions. Since $\boldsymbol{\gamma}$ is regular, at least one of $\dot{u}(t_0)$ and $\dot{v}(t_0)$ is non-zero, say $\dot{u}(t_0)$. This means that the graph of u as a function of t is not parallel to the t-axis at t_0:



As in the proof of Theorem 1.5.1, this implies that any line parallel to the t-axis close to $u = x_0$ intersects the graph of u at a unique point $u(t)$ with t close to t_0. This gives a function $h(x)$, defined for x in an open interval containing

x_0, such that $t = h(x)$ is the unique solution of $u(t) = x$ if x is near x_0 and t is near t_0. The inverse function theorem tells us that h is smooth. The function

$$f(x, y) = y - v(h(x))$$

has the properties we want.

It is not in general possible to find a *single* function $f(x, y)$ satisfying the conditions in Theorem 1.5.1 such that the image of $\boldsymbol{\gamma}$ is contained in the level curve $f(x, y) = 0$, for $\boldsymbol{\gamma}$ may have self-intersections like the limaçon in Example 1.1.7. It follows from the inverse function theorem that no single function f satisfying the conditions in Theorem 1.5.1 can be found that describes a curve near such a self-intersection.

EXERCISES

1.5.1 Show that the curve $\mathcal{C}$ with Cartesian equation

$$y^2 = x(1 - x^2)$$

is not connected. For what range of values of t is

$$\boldsymbol{\gamma}(t) = (t, \sqrt{t - t^3})$$

a parametrization of $\mathcal{C}$? What is the image of this parametrization?

1.5.2 State an analogue of Theorem 1.5.1 for level curves in $\mathbb{R}^3$ given by $f(x, y, z) = g(x, y, z) = 0$.

1.5.3 State and prove an analogue of Theorem 1.5.2 for curves in $\mathbb{R}^3$ (or even $\mathbb{R}^n$). (This is easy.)

In the remainder of this book, we shall speak simply of 'curves', unless there is serious danger of confusion as to which type (level or parametrized) is intended.