Chapter 1
The Genesis of Differential Methods

This first chapter is intentionally provocative, and useless! By *useless* (besides being at once provocative) we mean: this first chapter is not formally needed to follow the systematic treatment of the theory of curves and surfaces developed in the subsequent chapters.

So what is this chapter about? Usually, when you open a book on—let us say—the theory of curves in the real plane, you expect to find first "the" precise definition of a plane curve, followed by a careful study of the properties of such a notion. We all have an intuitive idea of what a plane curve is. Everybody knows that the straight line, the circle or the parabola *are* curves, but a single point or the empty set *are not* curves! Nevertheless, all these "figures" can be described by an equation $F(x, y) = 0$, with F a polynomial: for example, $x^2 + y^2 = 0$ is an "equation of the origin" in $\mathbb{R}^2$ while $x^2 + y^2 = -1$ is "an equation" of the empty set. Thus a curve cannot simply be defined via an equation $F(x, y) = 0$, even when F is a "very good" function! For example, consider the picture comprised of 7 hyperbolas, thus 14 branches. Is this *one* curve, or *seven* curves, or *fourteen* curves? After all, it is not so clear what a curve should be!

Starting at once with a precise definition of a curve would give the false impression that this is *the* definition of a curve. Instead it should be stressed that such a definition is *a possible* definition. Discussing the advantages and disadvantages of the various possible definitions, in order to make a sensible choice, is an important aspect of every mathematical approach.

There is also a second aspect that we want to stress. For *Euclid*, a straight line was *What has a length and no width and is well-balanced at each of its points* (see Definition 3.1.1 in [3], *Trilogy I*). The intuition behind such a sentence is clear, but such a "definition" assumes that before beginning to develop geometry, we know what a *length* is. Of course what we want to do concerning a *length* is then to find a formula to compute it, such as $2\pi R$ for a circle of radius R.

With more than two thousand years of further mathematical developments and experience, we now feel quite uneasy about such an approach. How can we establish a formula to compute the *length of a curve* if we did not define first what the *length of a curve* is?

F. Borceux, *A Differential Approach to Geometry*, DOI 10.1007/978-3-319-01736-5_1,

For many centuries—essentially up to the 17th century—mathematicians could hardly handle problems of length for curves other than the straight line and the circle. Differential calculus, with the full power of the theories of derivatives and integrals, opened the door to the study of arbitrary curves. However, in some sense, one was still taking the notion of length (or surface or volume) as something "which exists and that one wants to calculate".

Like many authors today, we adopt in the following chapters a completely different approach: the theory of integration is a well-established part of analysis and we use it to define a length. Analogously the theory of derivatives is a well-established part of analysis and we use it to define a tangent. And so on.

This first chapter is intended to be a "bridge" between the "historical" and the "contemporary" approaches. We present typical arguments developed in the past (and sometimes, still today) to master some geometrical notions (like length, or tangent), but we do that in particular to develop an intuition for the contemporary definitions of these notions. In this introductory chapter, we refer freely to [3] and [4], *Trilogy I* and *II*, when the historical arguments that we have in mind have been developed there.

Various arguments in this chapter can appear quite disconcerting. We often rely on our intuition, without trying to formalize the argument. We freely apply many results borrowed from a first calculus course, taking as a blanket assumption that when we apply a theorem, the necessary assumptions for its validity should be satisfied, even if we have not tried to determine the precise context in which this is the case! This is not a very rigorous attitude, however our point in this chapter is not to *prove* results, but to *guess* what possible "good" definitions should be.

1.1 The Static Approach to Curves

Originally, Greek geometry (see [3], *Trilogy I*) was essentially concerned with the study of two curves: the line and the circle.

> *The line is what has length and no width and is well-balanced around each of its points.*
>
> *The circle is the locus of those points of the plane which are at a fixed distance R from a fixed point O of the plane.*

Passing analogously to three dimensional space, using a circle in a plane and a point not belonging to the plane of the circle, you can then—using lines—construct the cone on this circle with vertex the given point. "Cutting" this cone by another plane then yields new curves that, according to the position of the "cutting plane", you call *ellipse*, *hyperbola* or *parabola*. This is the origin of the theory of curves.

It is common practice to describe a curve by giving its equation with respect to some basis. In this book, we are interested in the study of curves in the real plane $\mathbb{R}^2$. For example a circle of radius R centered at the origin admits the equation (see Chap. 1 in [4], *Trilogy II*)

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

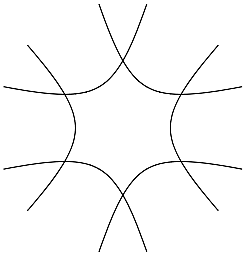

Fig. 1.1

which we can equivalently write as

$$x^2 + y^2 - R^2 = 0.$$

One might be tempted to introduce a general theory of curves by allowing equations of the form

$$F(x, y) = 0$$

where

$$F : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

is an arbitrary function. But it does not take long to realize that:

- choosing $F(x, y) = x^2 + y^2$, we get the equation of a single point: the origin;
- choosing $F(x, y) = x^2 + y^2 + 1$, we even get the equation of the empty set!

In both cases the function $F(x, y)$ is certainly "a very good one": it is even a polynomial, but we do not want a point or the empty set to be considered as a curve.

For more food for thought, look at the picture in Fig. 1.1: should this be considered as *one* curve, or as *six* curves?

In fact, if you look carefully at Fig. 1.1, you will realize that it is comprised of three hyperbolas. The equation of this picture is "simply"

$$\big(x^2 - y^2 - 1\big)\big((x - \sqrt{3}y)^2 - (\sqrt{3}x + y)^2 - 4\big)\big((x + \sqrt{3}y)^2 - (\sqrt{3}x - y)^2 - 4\big) = 0$$

thus again an equation of the form $F(x, y) = 0$ with F a polynomial. But since this is the equation of three hyperbolas, should we consider that the picture represents three curves, not one or six?

If you decide that a hyperbola is *one* curve, then you accept that a curve can have several disjoint branches. Thus you should probably also consider that the picture of Fig. 1.1 represents *one* curve with six branches. Furthermore, you should also consider that a picture comprising 247 straight lines is *one* curve as well. Taking the opposite point of view, the hyperbola is no longer *one* curve, but the union of *two* curves.

If you have not yet given up, the following example may cause you to:

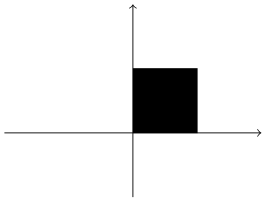

Fig. 1.2

Example 1.1.1 There exist continuous functions $F\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that

$$\left\{(x, y) \in \mathbb{R}^2 \middle| F(x, y) = 0\right\}$$

is a full square.

Proof The following function is one among many examples:

$$F(x, y) = \big(x - |x|\big)^2 + \big(y - |y|\big)^2 + \big((1 - x) - |1 - x|\big)^2 + \big((1 - y) - |1 - y|\big)^2.$$

The condition $F(x, y) = 0$ is indeed equivalent to

$$x \geq 0, \qquad y \geq 0, \qquad 1 - x \geq 0, \qquad 1 - y \geq 0.$$

The corresponding "curve" is the full square of Fig. 1.2. Certainly, you do not want this to be called *a curve*! □

Should we thus give up our attempt to define a curve via a rather general equation of the form $F(x, y) = 0$? For the time being we shall abandon this idea, but we will come back to this problem later, with adequate differential tools.

Nevertheless, let us conclude this section with a comment. Every equation of the form $F(x, y) = 0$ determines a subset of $\mathbb{R}^2$

$$\left\{(x, y) \middle| F(x, y) = 0\right\} \subseteq \mathbb{R}^2$$

and we would like to find conditions on F so that this subset is worthy of being called *a curve*. If we achieve this program, a curve will thus be *a subset of* $\mathbb{R}^2$. Being a subset is a *static* notion: no sense of *movement* is involved here. The full meaning of this comment will be expanded upon in the following Sect. 1.2.

1.2 The Dynamic Approach to Curves

The idea of "separating the variables" of an equation is due to the Swiss mathematician *Leonhard Euler* (1707–1783) (see Chap. 1 in [4], *Trilogy II*). In the case of the circle

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

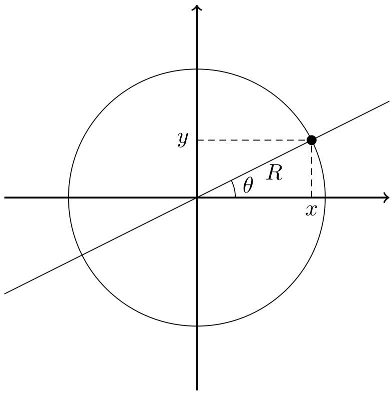

Fig. 1.3

this idea consists, for example, of describing the circle via the classical formulas

$$\begin{cases} x = R\cos\theta \\ y = R\sin\theta \end{cases}$$

where θ is the angle between the x-axis and the radius (see Fig. 1.3).

We thus obtain a *dynamical* description of the circle: when θ runs from $-\infty$ to $+\infty$, we repeatedly travel around the circle.

We are thus tempted to define a plane curve "dynamically" as a function

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad t \mapsto f(t).$$

In this spirit, a curve becomes a "deformation of the real line in the plane". Our intuition of a curve is that such a deformation should at least be continuous. Indeed we cannot imagine calling a "curve" a function such as

$$f(t) = \begin{cases} (t, 1) & \text{if } t \text{ is rational} \\ (t, 0) & \text{if } t \text{ is irrational.} \end{cases}$$

Let us observe that if we want to view a curve as a *continuous deformation of the real line*, then by continuity, every curve will have a single "branch". We discussed the case of the hyperbola in Sect. 1.1: the hyperbola is *not* a continuous deformation of the real line, but each of its two branches is. Thus we slowly begin to realize that choices have to be made and that probably, no optimal choice exists.

In a first "dynamic" approach, let us therefore view a curve as a continuous function

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad t \mapsto f(t)$$

as in Fig. 1.4.

The curve is thus thought of as the *trajectory of a point*, the trajectory expressed in terms of a parameter t which runs along the real line. This parameter t could be

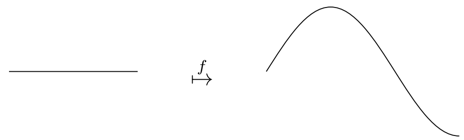

Fig. 1.4

regarded as the "time" calculated (positively or negatively) from a given origin of time: at the instant t, the point has reached the position $f(t)$ in the plane. Alternatively t could also be thought of as the "distance traveled on the curve" from a fixed origin on this curve: after having already traveled a distance t, the point has reached the position $f(t)$. And so on. When you prepare an itinerary for your holiday, you will probably say something like

After 247 km I shall be in Paris.

But when you comment on your travels afterwards, you will probably say

After 2 hours and 36 minutes I was in Paris.

In both cases you are commenting on the same itinerary, using different parameters.

Of course since various functions in terms of various parameters can describe the same curve, each of these functions should better be called a *parametric representation of a curve*.

However, we still have not avoided the "undesirable examples" encountered in the previous section. Simply choose for f the constant function on a point $(a, b) \in \mathbb{R}^2$ (not a particularly convincing "holiday itinerary": you spend your entire holiday at home)! Again we do not want to call this a "representation of a curve". We have a point, not a curve. More surprisingly:

Example 1.2.1 There exist continuous functions

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2$$

whose image covers a full square.

Proof Let us sketch the construction of an example proposed by the Italian mathematician *Peano* in 1890. He defines a sequence

$$f_n : [0, 1] \longrightarrow \mathbb{R}^2, \quad n \in \mathbb{N}$$

of continuous functions, which converges uniformly to a continuous surjective function

$$f : [0, 1] \longrightarrow [0, 1] \times [0, 1].$$

Since moreover

$$f(0) = (0, 0), \qquad f(1) = (1, 1)$$

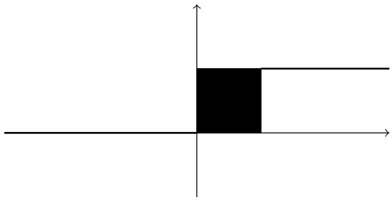

Fig. 1.5

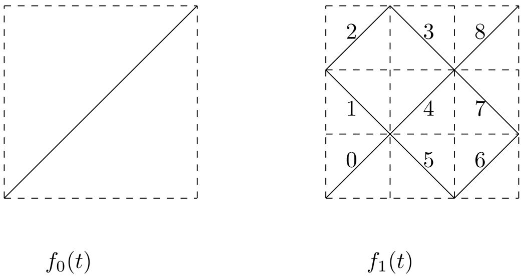

Fig. 1.6

it suffices to extend the definition by

$$\begin{cases} f(t) = (t, 0) & \text{if } t \le 0, \\ f(t) = (t, 1) & \text{if } t \ge 1, \end{cases}$$

to get the expected counterexample as in Fig. 1.5.

The sequence begins with the identity function: $f_0(t) = t$. The graph of the function $f_1(t)$ is then given by the right hand picture in Fig. 1.6. Simply follow the path according to the numbering of the sub-squares 0 to 8.

To obtain $f_2(t)$, replace each diagonal of a small square in the graph of f_1 by an analogous zigzag of nine smaller segments, each starting and ending at the same points as the small diagonal. Repeat the process to pass from f_2 to f_3, and so on. Each function $f_n(t)$ is continuous and the sequence $(f_n)_{n\in\mathbb{N}}$ converges uniformly, since at each level, the further variations are at most the length of the diagonal of the smallest square already constructed. It is then a standard result in analysis that the limit function $f(t)$ is still continuous.

To prove that f is surjective, express t in base 9. The construction shows at once that, writing $a, b, c, d, \ldots$ for the successive digits of the expansion of t in base 9,

$$t = 0.abcde\ldots$$

then

- $f_1(t)$ is in the square numbered a;
- $f_2(t)$ is in the sub-square of the previous square numbered b;

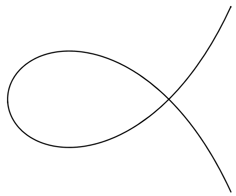

Fig. 1.7

- $f_3(t)$ is in the sub-sub-square numbered c;
- and so on.

It is obvious that every sequence of square–sub-square–sub-sub-square– $\cdots$ determines a unique point P of the square, and each point P of the square can be determined in this way. Such a sequence is by no means unique since (except for $(0,0)$ and $(1,1)$) each vertex of a small square, at whatever level, belongs to several squares. But nevertheless, choosing one of the possible sequences of square–sub-square–sub-sub-square– $\cdots$ which determines the point P, the list of the numbers 0 to 8 attached to each term of this sequence is then the base 9 expansion of a number $t \in [0,1]$ such that $f(t) = P$. Thus f is surjective. But as we have observed, such a number t is generally not unique, thus f is not injective. □

Again a dead end? Not really! We are now close to a solution. If we think of a parametric representation of a curve

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad t \mapsto f(t)$$

as being the trajectory of a point which "actually" moves in the plane, then when t varies, $f(t)$ should vary as well. Let us then simply impose that f is injective. This immediately eliminates the trivial case $f(t) = (a,b)$, but also Example 1.2.1, as we have seen.

The assumption "*f injective*" is perhaps a little too strong Fig. 1.7 depicts a "curve", even if the "trajectory" passes through the same point twice.

Considering the parametric representation of the circle

$$f(\theta) = (\cos\theta, \sin\theta)$$

as the parameter runs along the real line the corresponding point rotates around the circle infinitely many times. A single loop contains all the required information.

The following definition takes care of these "wishes".

Definition 1.2.2 A *parametric representation* of a plane curve is a continuous function

$$f\colon]a,b[\longrightarrow \mathbb{R}^2, \qquad t \mapsto f(t), \qquad a,b \in \mathbb{R} \cup \{-\infty, +\infty\}$$

which is *locally injective*, that is, injective in a neighborhood of each point.

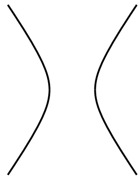

Fig. 1.8

More explicitly, the *local injectivity* means that for every $t_0 \in]a, b[$ one can find $\varepsilon > 0$ such that $]t_0 - \varepsilon, t_0 + \varepsilon[$ is still contained in $]a, b[$ and

$$f :]t_0 - \varepsilon, t_0 + \varepsilon[\longrightarrow \mathbb{R}^2$$

is injective. Allowing a and b to take "infinite values" is a quick way of saying that we allow f to be defined on $\mathbb{R}$ itself, on a half line or on an open interval.

Of course a constant function $f(t) = (a, b)$ is not locally injective. But what about the function f of Example 1.2.1?

Lemma 1.2.3 *The function f of Example 1.2.1 is not locally injective.*

Proof Consider the diagonal of a small square at the level n. This is the injective image under f_n of a small subsegment of [0, 1]. Let us say that this is the subsegment of origin u and length v. Observing the construction of the zigzag in Example 1.2.1, we conclude that all f_i with $i > n$ are such that

$$f_i\left(u + \frac{1}{9}v\right) = f_i\left(u + \frac{5}{9}v\right), \qquad f_i\left(u + \frac{4}{9}v\right) = f_i\left(u + \frac{8}{9}v\right).$$

Thus at the limit we still have

$$f\left(u + \frac{1}{9}v\right) = f\left(u + \frac{5}{9}v\right), \qquad f\left(u + \frac{4}{9}v\right) = f\left(u + \frac{8}{9}v\right).$$

This proves that one can always find points, everywhere in [0, 1], "as close as one wants to each other", which are mapped by f onto the same point. Thus f is not *locally injective*. □

We conclude that the "non-examples" of curves that we gave earlier do not satisfy our Definition 1.2.2 of a curve. Does this mean that Definition 1.2.2 is *the* good one? The *only possible* good one? Certainly not. Nevertheless, the following chapters will give evidence that this is certainly *a possible* good definition.

For example, as already observed, our choice prevents us from considering the hyperbola (Fig. 1.8) as *one* curve, since it has two branches.

To overcome this problem, in the definition of a parametric representation of a curve, we could decide to allow as domain a *union of open intervals*, but probably not any kind of union. It would surely be wise to exclude such unions as

$$\bigcup_{n=1}^{\infty} \left]\frac{1}{2n+1}, \frac{1}{2n}\right[.$$

Reducing one's attention to a *finite* union of open intervals could be a reasonable compromise. However, as already mentioned in Sect. 1.1, do we really want the union of 247 straight lines to be considered as a single curve?

We could also decide to allow *closed* intervals as domains, not only *open* intervals. We would of course not allow these closed intervals to reduce to single points. But then every time we consider a construction using limits or derivatives, at the extremities of a closed interval, we would have to work with "one-sided" limits or derivatives. For example if we define a circle via

$$f\colon [0, 2\pi] \longrightarrow \mathbb{R}^2, \qquad f(\theta) = (\cos\theta, \sin\theta)$$

we have to treat separately the point $f(0) = f(2\pi)$, which by the way, in a circle, should have the same properties as any other point of the circle! As far as possible, we shall avoid entering into these considerations (nevertheless, see Definition 2.14.1).

In Definition 1.2.2 you may also want to impose that f is differentiable, or even of class $\mathcal{C}^\infty$, or some other class. We shall not do this: we will introduce these additional assumptions (or others) when they are needed for some results.

The conclusion of this discussion is thus

Defining a curve is a matter of choice!

But not all choices are sensible. Our choice is Definition 1.2.2.

1.3 Cartesian *Versus* Parametric

In Sect. 1.1 we have tried (without much success up to now) to determine a curve via a *Cartesian equation*

$$F(x, y) = 0$$

while in Sect. 1.2 we have focused our attention on parametric representations

$$f\colon \left]a, b\right[\longrightarrow \mathbb{R}^2, \qquad t \mapsto \big(f_1(t), f_2(t)\big)$$

that is, on a system of *parametric equations*

$$\begin{cases} x = f_1(t) \\ y = f_2(t). \end{cases}$$

Can we switch easily from one approach to the other, and perhaps guess what a good notion of *Cartesian equation of a curve* might be?

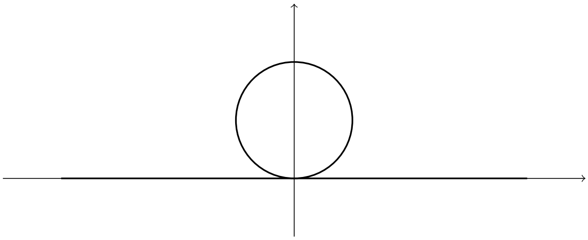

Fig. 1.9

An initial warning must be made. The "static" definition of a curve presents it as "a subset of the plane". The "dynamic" definition of a curve presents is as "a trajectory in the plane". However a curve, regarded as a subset of the plane, can easily be obtained via very different trajectories! Fig. 1.9 presents a curve comprising the x-axis and a circle of radius 1 with center $(0, 1)$. This "curve" admits the equation

$$y\big(x^2 + (y-1)^2 - 1\big) = 0.$$

You can view this as the "smooth" trajectory of a point coming from $(-\infty, 0)$, turning counter-clockwise around the circle, and proceeding next to $(+\infty, 0)$. Having arrived at the origin, you could also very well turn clockwise: as a trajectory, this is completely different! Of course you could also follow both trajectories in the reverse direction, but this is certainly not an essential difference.

Let us thus see how we can pass from a "static" description to a "dynamic" description, and vice-versa. In one direction, the idea is clear. Given the system of parametric equations

$$\begin{cases} x = f_1(t) \\ y = f_2(t) \end{cases}$$

we just need to eliminate the parameter t between the two equations ending up with a Cartesian equation! This is easy to say, but not always that easy to do when f_1 and f_2 are fairly involved functions.

However, analysis is there to help us, at least formally. Let us recall the following important result:

Theorem 1.3.1 (Local Inverse Theorem) *Consider a function $g\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ of class $\mathcal{C}^k$ $(k \geq 1)$. If the matrix*

$$\left(\frac{\partial g_i}{\partial x_j}(a_1, \ldots, a_n)\right)_{1 \leq i, j \leq n}$$

is regular, then the function g is invertible on a neighborhood of the point $(a_1, \ldots, a_n)$ and its inverse is still of class $\mathcal{C}^k$.

Of course when $n = 1$, the condition in Theorem 1.3.1 reduces to $g'(a) \neq 0$. This suggests the following definition:

Definition 1.3.2 A parametric representation of a curve

$$f: \,]a,b[\, \longrightarrow \mathbb{R}^2, \qquad t \mapsto \big(f_1(t), f_2(t)\big)$$

is *regular* when it is of class $\mathcal{C}^1$ and $f'(t) \neq (0,0)$ for each $t \in \,]a,b[$.

We obtain the following result:

Proposition 1.3.3 *Let $f: \,]a,b[\, \longrightarrow \mathbb{R}^2$ be a regular parametric representation of a curve. For every $t_0 \in \,]a,b[$:*

1. *there exists a neighborhood of t_0 on which the curve admits a Cartesian equation $F(x,y)=0$;*
2. *the function $F: \mathbb{R}^2 \longrightarrow \mathbb{R}$ is of class $\mathcal{C}^1$ on this neighborhood;*
3. *at each point of the curve in the given neighborhood, at least one of the partial derivatives of F is non-zero.*

Proof Assume—for example—that $f_1'(t_0) \neq 0$. By Proposition 1.3.1 we can write $t = f_1^{-1}(x)$ on a neighborhood of t_0. This yields

$$y = f_2\big(f_1^{-1}(x)\big)$$

and it suffices to define

$$F(x,y) = f_2\big(f_1^{-1}(x)\big) - y.$$

Notice in particular that

$$\frac{\partial F}{\partial y} = -1 \neq 0.$$ □

Proposition 1.3.3 suggests further to try the following definition:

Definition 1.3.4 (Temporary Definition; see 1.4.2) By a *Cartesian equation* of a plane curve is meant an equation

$$F(x,y) = 0$$

satisfying the following requirements:

- this equation admits solutions;
- $F: \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a function of class $\mathcal{C}^1$;
- at each point (x_0, y_0) such that $F(x_0,y_0)=0$, at least one of the partial derivatives of F is non-zero.

The corresponding *curve* is the set of those points (x,y) such that $F(x,y)=0$.

Let us now consider the opposite problem: how do we pass from a Cartesian equation to a parametric representation? Once more, analysis is there to help us solve our problem. Let us recall a celebrated result:

Theorem 1.3.5 (Implicit Function Theorem) *Consider a function* $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ *of class* $\mathcal{C}^k$ $(k \geq 1)$. *If*

$$F(a_0, \ldots, a_n) = 0, \qquad \frac{\partial F}{\partial x_n}(a_0, \ldots, a_n) \neq 0$$

then there exists

- *a neighborhood* V *of* $(a_0, \ldots, a_{n-1})$ *and*
- *a function* $\varphi \colon V \longrightarrow \mathbb{R}$ *of class* $\mathcal{C}^k$

such that

- $\varphi(a_0, \ldots, a_{n-1}) = a_n$;
- $\forall (x_1, \ldots, x_{n-1}) \in V \;\; F(x_1, \ldots, x_{n-1}, \varphi(x_1, \ldots, x_{n-1})) = 0.$

Moreover, the neighborhood V *can be chosen such that a function* φ *as in the statement is necessarily unique.*

The *implicit function* inferred from F is thus

$$x_n = \varphi(x_1, \ldots, x_{n-1}).$$

Proposition 1.3.6 *Consider a Cartesian equation* $F(x, y) = 0$ *of a plane curve* (*as in Definition* 1.3.4) *and a point* (x_0, y_0) *satisfying this equation. In a neighborhood of* (x_0, y_0), *there exists a regular parametric representation of a curve*

$$f : \,]a, b[\, \longrightarrow \mathbb{R}^2$$

such that each point $(x, y) = (f_1(t), f_2(t))$ *satisfies the equation* $F(x, y) = 0$.

Proof Assume that $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$. With the notation of Theorem 1.3.5 it suffices to define

$$f(x) = \big(x, \varphi(x)\big).$$

The parameter is thus $t = x$ and $f'(x) = (1, \varphi'(x)) \neq (0, 0)$. Considering its first component, we notice that f is injective. □

The slogan suggested by Propositions 1.3.3 and 1.3.6 is thus:

> *In good cases, one can switch* locally *from a system of parametric equations to a Cartesian equation, and vice-versa.*

Locally is certainly *the* point to emphasize here, but it is not the only one. The two processes seem to be "the inverse of each other", but this is definitely a false impression. Let us demonstrate this with some examples.

Consider first the circle of Fig. 1.10 and its parametric equations

$$\begin{cases} x = R\cos\theta \\ y = R\sin\theta. \end{cases}$$

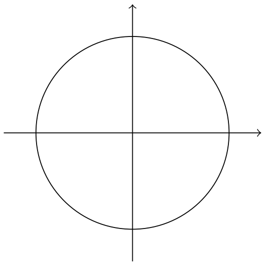

Fig. 1.10

To eliminate θ between the equations you will probably simply square each equation and add the results, to end up with

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

Conversely, you will probably write

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

and observe that each choice of the sign will give you half of the circle (the "upper half" or the "lower half"). You will then obtain

$$f\colon]-R,+R[\to \mathbb{R}^2, \qquad x \mapsto \left(x, \sqrt{R^2 - x^2}\right)$$

as a parametric representation of the upper half of the circle. Working with

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

would give you the "left half" or the "right half". So from the parametric equations, you have obtained the "global" Cartesian equation of the circle, but from that Cartesian equation you have reconstructed—only locally—parametric equations of the circle. Moreover, these are completely different from the original parametric equations!

Let us now try the same with the parametric equations

$$\begin{cases} x = e^t \\ y = e^t \end{cases}$$

which represent the half-diagonal of Fig. 1.11. It suffices to subtract the two equations to eliminate t, and this yields the equation $x = y$ of the full diagonal! Of course one cannot possibly guess, given only the Cartesian equation $x = y$, that it comes from the original parametric equations.

Another slogan should thus be

Be careful ...

But we should perhaps also add *Be sorry!*, as the next section will explain.

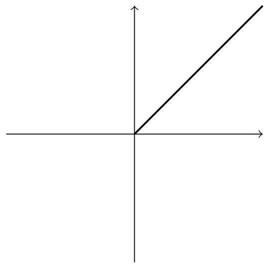

Fig. 1.11

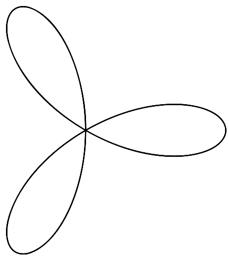

Fig. 1.12

1.4 Singularities and Multiplicities

We have already observed in Sect. 1.1 that not all real polynomials $F(X, Y)$ yield the equation $F(x, y) = 0$ of a curve in $\mathbb{R}^2$. But we know many examples where $F(x, y) = 0$ does describe something which is worthy of being called a "curve". For example

$$x^3 - 3xy^2 = \left(x^2 + y^2\right)^2$$

which yields the "curve" of Fig. 1.12.

This equation can thus be written in the form $F(x, y) = 0$ with

$$F(x, y) = x^3 - 3xy^2 - \left(x^2 + y^2\right)^2.$$

Since each term is of degree at least 2 (in fact, of degree 3 or 4),

$$\frac{\partial F}{\partial x}(0, 0) = 0, \qquad \frac{\partial F}{\partial y}(0, 0) = 0.$$

So unfortunately, this "curve" is not a curve in the sense of our Definition 1.3.4.

Not recapturing the "decent algebraic curves" in our theory is rather unsatisfactory. The present Section, deeply inspired by the considerations of Chap. 7 in [4], *Trilogy II*, will now discuss this "difficulty" further. We observe first that:

Proposition 1.4.1 *Let $f(X,Y)\in\mathbb{R}[X,Y]$ be a non-constant polynomial without any multiple factors. There are at most finitely many points (X,Y) such that $f(X,Y)=0$ and both derivatives of f vanish at (X,Y).*

Proof Let us consider the family $(a_i,b_i)_{i\in I}$ of those points such that

$$f(a_i,b_i)=0,\qquad \frac{\partial f}{\partial X}(a_i,b_i)=0,\qquad \frac{\partial f}{\partial Y}(a_i,b_i)=0.$$

We must prove that there are only finitely many of them.

Let us write $F(X,Y,Z)$ for the homogeneous polynomial associated with F (see [4], *Trilogy II*, Sect. C.2). Thus $f(X,Y)=F(X,Y,1)$ and the factors of f and F correspond to each other via the "homogenizing" process. Both polynomials f and F have the same degree: let us say, n. By Euler's formula (see C.1.5 in [4], *Trilogy II*)

$$nF=X\frac{\partial F}{\partial X}+Y\frac{\partial F}{\partial Y}+Z\frac{\partial F}{\partial Z}.$$

Applying this formula at the points $(a_i,b_i,1)$ yields

$$\frac{\partial F}{\partial Z}(a_i,b_i,1)=0.$$

Thus the points $(a_i,b_i,1)$ are *multiple points* (see Definition 7.4.5, [4], *Trilogy II*) of the complex projective curve $F(X,Y,Z)=0$.

By assumption, $f(X,Y)$ and thus $F(X,Y,Z)$ do not have any multiple factor as real polynomials. If we can prove that analogously $F(X,Y,Z)$ does not have any multiple factors in $\mathbb{C}[X,Y,Z]$, then the number of multiple points of $F(X,Y,Z)$ is bounded by $n(n-1)$ (see Sect. 7.9 in [4], *Trilogy II*). Thus there are at most $n(n-1)$ points (a_i,b_i) as above.

Let us recall that splitting all coefficients into their real and their imaginary parts, every complex polynomial $\alpha(X,Y,Z)$ can be written as

$$\alpha(X,Y,Z)=\beta(X,Y,Z)+i\gamma(X,Y,Z)$$

where α and β are polynomials with real coefficients. This shows at once that given a non-constant real polynomial $\delta(X,Y,Z)$, if $\alpha\delta$ is a real polynomial, then $\gamma(X,Y,Z)=0$. In other words, if a non-constant real polynomial $\delta(X,Y,Z)$ divides another real polynomial in $\mathbb{C}[X,Y,Z]$, it divides it in $\mathbb{R}[X,Y,Z]$.

Replacing the coefficients of α by their conjugates then yields

$$\overline{\alpha}(X,Y,Z)=\beta(X,Y,Z)-i\gamma(X,Y,Z).$$

It follows at once that, just as for complex numbers

$$\alpha(X,Y,Z)\overline{\alpha}(X,Y,Z)=\beta(X,Y,Z)^2+\gamma(X,Y,Z)^2$$

that is, a polynomial with real coefficients.

Write now

$$F(X,Y,Z) = G_1(X,Y,Z) \cdots G_m(X,Y,Z)$$

with the $G_k(X,Y,Z)$ irreducible. We must prove that each factor G_k is simple.

Since F has real coefficients, passing to the conjugates yields

$$F(X,Y,Z) = \overline{G_1}(X,Y,Z) \cdots \overline{G_m}(X,Y,Z).$$

Of course, the $\overline{G_k}$ are still irreducible, because conjugation is a homomorphism of fields.

If some G_k has real coefficients, it divides $F(X,Y,Z)$ in $\mathbb{R}[X,Y,Z]$ as we have just seen. Therefore by assumption, it is a simple factor.

Otherwise by uniqueness of the decomposition into irreducible factors, there exists an index $j \neq k$ such that $G_j = \overline{G_k}$. Then $G_k G_j = G_k \overline{G_k}$ is a polynomial with real coefficients which divides $F(X,Y,Z)$. By assumption, it is a simple factor. Dividing by this polynomial and repeating the argument allows as to conclude that all non-real G_k are simple factors as well. □

Of course replacing a multiple factor of $f(x,y)$ by the same factor with degree 1 does not modify the set of points (x,y) such that $f(x,y) = 0$. Thus the assumption in Proposition 1.4.1 is not really a restriction, as far as the study of curves is concerned.

All this suggests modifying Definition 1.3.4 in the following way:

Definition 1.4.2 By a *Cartesian equation* of a plane curve is meant an equation

$$F(x,y) = 0$$

where:

- $F\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a function of class $\mathcal{C}^1$;
- there exist solutions (x,y) where at least one partial derivative of F does not vanish;
- there are at most finitely many solutions (x,y) where both derivatives of F vanish.

The corresponding *curve* is the set of those points (x,y) such that $F(x,y) = 0$.

Of course now, Proposition 1.3.6 holds only for those points where at least one of the partial derivatives is not zero.

Still inspired by the considerations of Chap. 7 in [4], *Trilogy II*, it is also sensible to define:

Definition 1.4.3 Let $F(x,y) = 0$ be a Cartesian equation of a plane curve. The points (x,y) of the curve where both partial derivatives of F vanish are called the *multiple points* of the curve.

In the curve of Fig. 1.12 there is thus one single multiple point, namely, $(0,0)$.

However, if we define *multiple points* when the curve is given by a Cartesian equation, we are immediately faced with the challenge of defining a corresponding notion for a parametric representation. Since at a multiple point Proposition 1.3.6 does not hold, we would be tempted, in the case of a parametric representation, to consider those points where the "converse" Proposition 1.3.3 does not hold:

Definition 1.4.4 Given a parametric representation of class $\mathcal{C}^1$ of a curve, a point of parameter t is *singular* when $f'(t) = (0,0)$.

It is important to stress two facts concerning this notion:

- "being a singular point" is a property of the parametric representation which does not necessarily exhibit a "singularity" of the corresponding subset of $\mathbb{R}^2$;
- being a "singular point" for a parametric representation is by no means equivalent to being a "multiple point" for a corresponding Cartesian equation.

For example

$$f: \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad t \mapsto (t^3, 0)$$

is a parametric representation of class $\mathcal{C}^\infty$ of the x-axis, which on the other hand admits the Cartesian equation $y = 0$. Observe that

$$f'(t) = (3t^2, 0), \qquad f'(0) = (0,0)$$

thus the origin is a *singular point* of the parametric representation f. But the origin is by no means a *multiple point* of the x-axis, that is, the algebraic curve $y = 0$.

If we consider the more usual parametric representation of the x-axis

$$g: \mathbb{R} \longrightarrow \mathbb{R}, \qquad t \mapsto (t, 0)$$

then

$$g'(t) = (1,0) \neq (0,0)$$

and there is no singular point at all.

Next consider the curve of Fig. 1.12 and its multiple point at the origin. It is routine to verify that

$$\begin{cases} x = \cos\theta \cdot \cos 3\theta \\ y = \sin\theta \cdot \cos 3\theta \end{cases}$$

is a system of parametric equations of the same "curve". Since $\sin\theta$ and $\cos\theta$ do not vanish together, the origin $(0,0)$ is reached when $\cos 3\theta = 0$, that is (up to 2π) for

$$\theta = \frac{\pi}{6}, \qquad \theta = \frac{\pi}{2}, \qquad \theta = \frac{5\pi}{6}.$$

A straightforward computation shows that, writing f for the parametric representation, $f'(\theta) \neq (0,0)$ at these three points (in fact, the parametric representation f

is regular!) Thus these three values of the parameter θ are *not singular*, while the corresponding point is *multiple*.

The conclusion is clear: parametric representations are not appropriate for the study of multiple points in the sense of the theory of algebraic curves! Intuitively, if you travel "regularly" along a curve, there is nothing special about passing through a point you have passed through earlier.

You might claim that passing through the same point several times *is* something special. When working with a parametric representation f, we might then try to say that a point $P \in \mathbb{R}^2$ is *multiple* if $P = f(t)$ for several values of the parameter t. But then all points of the circle

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad \theta \mapsto (\cos\theta, \sin\theta)$$

are multiple, and even of *multiplicity* ∞! I am sure this is not what you had in mind!

You should certainly now be convinced that defining *a curve* is definitely a matter of choice. If you strengthen the conditions in order to avoid some pathologies, then you eliminate some examples that you would like to keep, and conversely. Moreover, working with parametric equations or with a Cartesian equation lead rather naturally to non-equivalent choices of definitions.

In this book, we shall adopt Definitions 1.2.2 and 1.4.2, and we shall stop our endless search for possible improvements of these definitions.

1.5 Chasing the Tangents

The Greek geometers defined tangents in the following way:

Definition 1.5.1 A *tangent* to a circle at one of its points P is a line whose intersection with the circle is reduced to the point P.

They proved (see Proposition 3.3.2, [3], *Trilogy I*):

Proposition 1.5.2 *Given a point P of a circle, there exists a unique tangent at P to the circle, namely, the perpendicular to the radius at P (see Fig. 1.13).*

Very trivially, such a definition does not work at all for arbitrary curves. Just have a look at Fig. 1.14: a tangent can cut the curve at a second point, and a line which cuts the curve at exactly one point has no reason to be a tangent.

Consider the trivial case of a straight line: the tangent to a straight line should be the line itself, which certainly takes us very far from a "unique" point of intersection, globally or locally. Also keep in mind that a tangent can "cut" the curve at the point of tangency, as in Fig. 1.15: thus "touching without cutting" is an inadequate definition. Finally do not forget the case of multiple points, as in Fig. 1.12. At a multiple point, there could be several tangents, not just a single one.

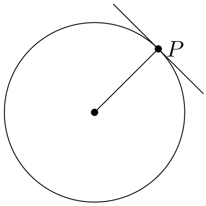

Fig. 1.13

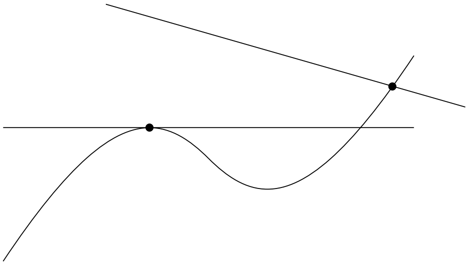

Fig. 1.14

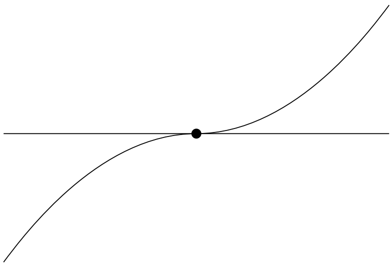

Fig. 1.15

Fortunately, the fact of not having a good definition of a tangent did not prevent mathematicians from calculating tangents!

In the 1630's, *Fermat* and *Descartes* proposed methods to calculate the tangent to a curve given by a polynomial equation $F(x, y) = 0$ (see Sect. 1.9 of [4], *Trilogy II*). The idea was that

A tangent is a line having a double point of intersection with the curve.

The notion of "double point of intersection" was in those days (1630–1640) more heuristic than precisely defined, but today it has been formalized in rigorous contemporary algebraic terms (see Definition 7.4.5 in [4], *Trilogy II*). In this book, we shall instead turn our attention to some attempts which prefigure contemporary differential methods.

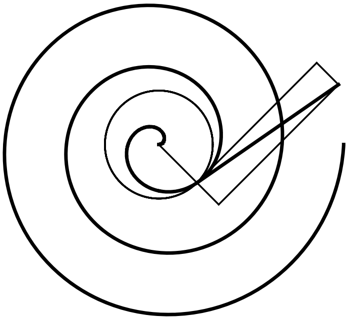

Fig. 1.16

Greek geometers could thus calculate the tangent to a circle and of course the tangent to a line *Archimedes* (287–212 AC) had already made the conceptual leap of regarding a curve as the trajectory of a moving point, and in this context he treated the tangent as follows:

Definition 1.5.3 The *tangent* to a curve is the line in the direction of the instantaneous movement of a point traveling on that curve.

Archimedes then computed the tangent to certain curves by "decomposing" the movement into a combination of linear and circular movements, and assuming that the direction of the tangent can be decomposed analogously. Let us follow his argument on a precise example (see Sect. 4.5 in [3], *Trilogy I*, for more comments on this curve).

Definition 1.5.4 The *spiral of Archimedes* (see Fig. 1.16) is the trajectory of a point in a plane, which moves at constant speed along a line, while the line turns at constant speed around one of its points, called the *center* of the spiral (see Fig. 1.16).

Example 1.5.5 Archimedes' construction of the tangent to his spiral.

Proof The global movement has two components: one resulting from the uniform linear movement of the point on the line, one resulting from the uniform circular movement of the line.

The component resulting from the uniform linear movement of the point on the line is expressed by a segment oriented along this line. Its length is the distance traveled on the line during a unit of time: let us say (to keep the picture on a page), during the time necessary for the line to make a half turn. This length is then half the distance between two turns of the spiral.

To obtain the component of the movement resulting from the uniform circular movement:

- consider the circle centered at the center of the spiral and which passes through the point at which you want to compute the tangent;

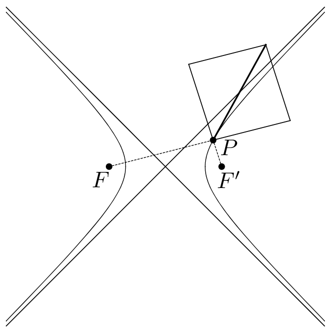

Fig. 1.17

- the circular component of the movement is a segment in the direction of the tangent to this circle;
- this component has a length equal to the distance traveled on this circle during a unit of time, that is, half the length of the circle.

Adding these two components via the parallelogram rule, *Archimedes* gets the direction of the tangent. □

This is a beautiful theoretical result, but of course, quite a disturbing one for a Greek geometer! Indeed Greek geometers could not, by ruler and compass constructions, draw a segment whose length is equal to the length of the circle, and we have known since the 19th century that this is in fact impossible: "circle squaring", that is, constructing π, is impossible by ruler and compass constructions (see Corollary B.3.3 in [3], *Trilogy I*).

In 1636, the French mathematician *Roberval* systematized *Archimedes*' idea to compute what he called *the touching line*. He applied this method to a wide variety of curves: various spirals, the conchoids, the cycloid, and so on (see Chap. 3 for a description of these curves). But let us focus here on his treatment of the tangent to a conic.

Proposition 1.5.6

1. *The tangent at a point P to a hyperbola with foci F, F' is a bisector of the two lines FP, $F'P$.*
2. *The tangent at a point P to an ellipse with foci F, F' is a bisector of the two lines FP, $F'P$.*
3. *The tangent at a point P to a parabola with focus F and directrix f is a bisector of the line FP and the perpendicular to f through P.*

Proof Consider first the case of the hyperbola (Fig. 1.17). As proved in Proposition 1.12.1 [4], *Trilogy II*, the hyperbola is the locus of those points P such that the difference of the distances

$$\left|d(P,F)-d(P,F')\right|$$

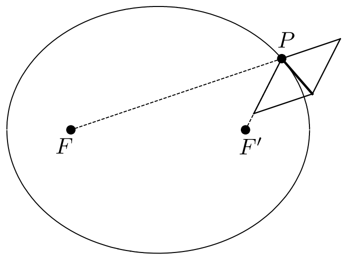

Fig. 1.18

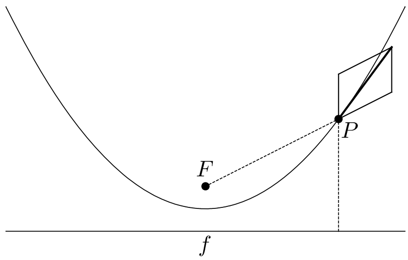

Fig. 1.19

to the two foci remains constant. When you move along a branch of the hyperbola—let us say—away from the origin, both distances increase. But since the difference between the two distances remains the same, both distances increase at the same rate. Roberval decomposes the movement into two instantaneous movements: one along the line FP, one along the line $F'P$. Since these two movements have equal amplitudes, the corresponding "parallelogram of movements" is a diamond, and the length of a side has no influence on the direction of the diagonal. Therefore, the tangent is simply the bisector of the two lines joining the point P and the two foci F and F'.

An analogous argument holds for the ellipse (see Fig. 1.18): this time, by Proposition 1.11.1 in [4], *Trilogy II*,

$$d(F, P) + d(F, P)$$

is constant. Thus one distance increases in the same way as the other one decreases. This again yields a diamond as "parallelogram of movements".

Finally for the parabola (see Fig. 1.19) with focus F and directrix f, when you move away from the origin, the two distances $d(F, P)$ and $d(f, P)$ increase at the same rate. Therefore the "parallelogram of movements" is a diamond with one side perpendicular to f and the other one in the direction FP. □

Think what you want of such arguments, they were nevertheless efficient in a period when differential calculus did not exist!

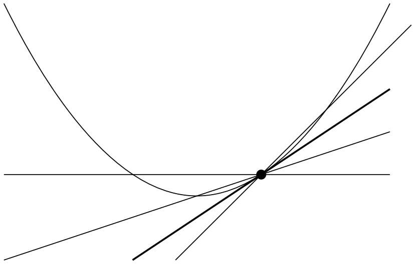

Fig. 1.20

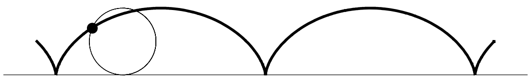

Fig. 1.21

1.6 Tangent: The Differential Approach

Although it is nice to see how tangents were computed historically, today everybody "with a basic mathematical culture" knows that:

> *The* tangent *to an arbitrary curve at one of its points P is the limit of the secant through P and another point Q of the curve, as Q converges to P (see Fig.* 1.20).

Indeed, this "dynamic" definition of the tangent, taking full advantage of the notion of limit, recaptures precisely our intuition of what a tangent should be (see Fig. 1.20).

Of course this is no longer the case at a multiple point (see Fig. 1.12): there we should consider separately the various "branches" of the curve, whatever that means! Perhaps we should decide if at a vertex of a square, there are two tangents, or no tangent at all. In the case—for example—of the cycloid: the trajectory of a point of a circle which rolls on a line (see Fig. 1.21), we should decide whether or not there is a tangent at each *cusp point*.

What might be a possible tangent at the origin for the curve with parametric representation

$$f: \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad t \mapsto \begin{cases} \left(t, t^k \sin \frac{1}{t}\right) & \text{if } t \neq 0 \\ (0,0) & \text{if } t = 0, \end{cases} \qquad k \in \mathbb{N}?$$

It is no longer clear which curves have a tangent and which do not. We still need a precise definition. The trouble with the "definition" above is that we can define the limit of a family of points in $\mathbb{R}$, and the limit of a family of vectors in $\mathbb{R}^2$, but how are we to precisely define the limit of a family of lines?

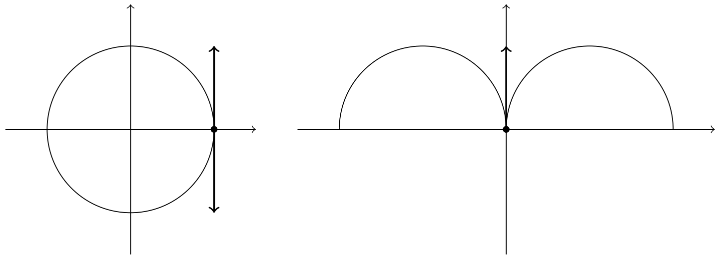

Fig. 1.22

Let us take a parametric representation of the curve

$$f\colon\,]a,b[\longrightarrow \mathbb{R}^2$$

and consider the point $P = f(t_0)$. When t converges to t_0, the point $Q = f(t)$ converges to $P = f(t_0)$. The secant through P and Q is the line

- passing through $P = f(t_0)$;
- of direction $\overrightarrow{PQ} = f(t) - f(t_0)$.

The tangent will thus be the line

- passing through $P = f(t_0)$;
- of direction $\lim_{t\to t_0}(f(t) - f(t_0))$.

Unfortunately this does not make any sense because by continuity of f, the limit is simply $f(t_0) - f(t_0)$, that is the zero vector.

But the difficulty is easy to overcome. Two vectors in the same direction define the same secant, thus let us simply work with vectors of length 1, so that the limit should remain of length 1. The tangent should thus be the line

- passing through $P = f(t_0)$;
- of direction

$$\lim_{t\to t_0} \frac{f(t) - f(t_0)}{\|f(t) - f(t_0)\|}.$$

When this limit exists, of course. Unfortunately, this limit most often does not exist.

Look at the following two curves, represented in Fig. 1.22.

- the circle with parametric representation

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad \theta \mapsto (\cos\theta, \sin\theta);$$

- the curve constituted of two half-circles, whose parametric representation is given by

$$g\colon\,]-1,+1[\longrightarrow \mathbb{R}^2, \qquad t \mapsto \bigl(t, \sqrt{1-x^2}\bigr).$$

In both cases, consider the point with parameter 0. What about the tangent, at these points, in the sense of the "definition" above?

- In the case of the circle, the expected limit does not exist! Indeed, we observe immediately that

$$\lim_{\substack{t\to 0\\ t<0}} \frac{f(t)-f(t_0)}{\|f(t)-f(t_0)\|} = (-1,0), \qquad \lim_{\substack{t\to 0\\ t>0}} \frac{f(t)-f(t_0)}{\|f(t)-f(t_0)\|} = (1,0).$$

  Both results are different, thus the limit does not exist.
- In the case of the two half-circles, the same kind of computation shows at once that

$$\lim_{t\to 0} \frac{g(t)-g(t_0)}{\|g(t)-g(t_0)\|} = (1,0)$$

  and the limit exists.

What does this mean? Although defining a tangent as "a limit of secants" is a good idea, when you try to make precise what "a limit of secants means", you easily run into severe problems. For example, with the attempt above, the circle does not have a tangent while the curve comprising of two half circles does! This first attempt to define a "limit of secants", because of the "counterexample" of the circle, is certainly unacceptable.

Note that in the case of the circle, the limits for $t < t_0$ and $t > t_0$ are opposite vectors, thus define the same direction, thus the same line. So one could modify our definition of the tangent by saying that both limits

$$\lim_{\substack{t\to t_0\\ t<t_0}} \frac{f(t)-f(t_0)}{\|f(t)-f(t_0)\|}, \qquad \lim_{\substack{t\to t_0\\ t>t_0}} \frac{f(t)-f(t_0)}{\|f(t)-f(t_0)\|}$$

on the right and on the left should exist, be non-zero, and be proportional vectors. However, one cannot expect to be able to prove elegant results and make computations with such a convoluted definition of the tangent.

The sensible thing to do is indeed to replace the vector $f(t) - f(t_0)$ by a vector proportional to it, but not a vector of length 1. Consider instead

$$\lim_{t\to t_0} \frac{f(t)-f(t_0)}{t-t_0}$$

which, when it exists, is simply $f'(t_0)$. Of course for such an approach to be efficient, not only must the derivative exist, but it must be non-zero! Therefore we make the following definition:

Definition 1.6.1 Consider a regular parametric representation of a curve

$$f\colon\,]a,b[\longrightarrow \mathbb{R}^2.$$

- The *tangent* to this curve at the point with parameter t_0 is the line containing $f(t_0)$ and of direction $f'(t_0)$.
- The *normal* to this curve at the point with parameter t_0 is the perpendicular to the tangent at this point.

Now of course, since the same curve can be described by various parametric representations, one should verify that Definition 1.6.1 of the tangent does not depend on the choice of the representation. We shall treat this question in Sect. 2.4. We recall that our point here is not to prove theorems, but to "guess" good definitions! Analogously one should see what happens to this definition when the curve is given by a Cartesian equation, but again this will be done in Sect. 2.4.

To conclude this section, let us insist once more on the fact that *defining the tangent is a matter of choice*.

Of course with Definition 1.6.1, our parametric representation f of the circle now yields a tangent at each point, because it is regular.

In our first attempt, the curve comprising two half circles also had a tangent at each point, but the parametric representation g of this curve is not differentiable at $t = 0$. So using Definition 1.6.1, the curve represented by g does not have a tangent at the origin.

Therefore one might want to further modify the definition of a tangent, to get the best of the two attempts. For example, by requiring only the proportionality of the left derivative and the right derivative, not the existence of the derivative. Then the curve represented by g would also have a tangent at the origin.

If we consider the parametric representation

$$h: \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad t \mapsto (t^3, 0)$$

of the x-axis, we observe that $h'(0) = (0, 0)$, thus in view of Definition 1.6.1, the x-axis—when represented by h—does not have a tangent at the origin, which is less than satisfactory. Again one might want to revise the definition of a tangent to avoid such a situation.

However, we shall not enter into these considerations: we adopt once and for all Definition 1.6.1. We are now well aware that in doing so, we exclude examples where a more involved definition would have produced a "sensible tangent".

1.7 Rectification of a Curve

As far as the length of a curve is concerned, the greatest achievement of Greek geometry was (see Theorem 3.1.4 in [3], *Trilogy I*):

Theorem 1.7.1 *The ratio between the length of a circle and the length of its diameter is a constant, independent of the size of the circle. This constant is written* π.

Proof This result was proved by the so-called *exhaustion method* due to *Eudoxus* (around 380 AC): this method was the direct ancestor of the notion of *limit*. Greek geometers first proved a corresponding result for regular polygons inscribed in a circle and then "by a limit process", inferred the result for the circle. □

The importance of this result is often hidden by the systematic use of the well-know formulas $2\pi R$ and πR^2 for the length and the area of a circle of radius R. *These formulas hold because the number π involved is independent of the size of the circle!* Today many of us consider that these formulas answer the question fully. For Greek geometers, they were only a beginning: what is the precise value of this quantity π? The famous problem of *squaring the circle* consisted equivalently of finding a construction of a segment of length π. However, all attempts in this direction seemed to be hopeless.

Two thousand years later, in 1637, the French mathematician and philosopher *Descartes* wrote

> *The relations between straight lines and curves are not known and, I think, cannot be discovered by the human mind; for that reason, no conclusion at all based on such relations can be accepted as rigorous and exact.*

So, even if you say that the length of a circle is $2\pi R$, you still do not know the length of the circle since you do not know the precise value of π! You are unable to construct (with ruler and compass) a segment having the same length as the circle, and if you cannot do this for the circle, how could you possibly hope to do it for more complicated curves?

One year later Descartes studied the movement of a body falling on the Earth, while the Earth is itself was considered as a body in rotation. For that he introduced the so-called *logarithmic spiral* (see Fig. 1.23).

Definition 1.7.2 The *logarithmic spiral* is the trajectory of a point moving on a line, at a speed proportional to the distance already travelled on this line, while the line itself turns at constant speed around one of its points.

In an irony of history, the logarithmic spiral was the first curve to be *rectified*, that is, a precise construction was given to produce a segment whose length is equal to the length of a given arc of the curve. This result is due to the Italian mathematician *Torricelli* (1608–1647), a student of *Galilee Galileo*. At the same time, *Torricelli* rectified various other curves, such as the *cycloid* (see Definition 1.9.1 and Proposition 1.9.4).

Proposition 1.7.3 *The length of an arc of a logarithmic spiral, from its origin O to a given point P, is equal to the length of the segment joining P and the intersection Q of the tangent at P and the perpendicular at O to the radius OP.*

Proof Let R be the length of the radius OP. By definition of the spiral, the component of the movement at P along the radius is kR, for a fixed constant k. The

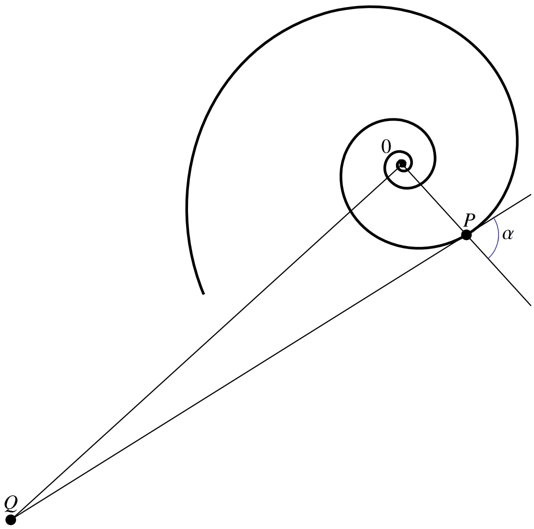

Fig. 1.23

circular component of the movement at P is oriented along the tangent to the circle with center O and radius R, that is, perpendicular to the radius OP; it has a length equal to the length of the circle, that is $2\pi R$. The parallelogram of movements is thus a rectangle with sides kR and $2\pi R$; its diagonal is thus oriented as that of a rectangle with sides k and 2π. Therefore the angle between the tangent and the radius OP is independent of R and is thus a constant α.

Now "unroll" the spiral along its tangent at P, starting from P. As we have just seen, the various successive radii OP' will keep forming an angle α with the tangent. Therefore the movement of the point O during this "unrolling" process is perpendicular to the direction of these radii. So the point O moves on the perpendicular to the radius OP at the point O. When O finally reaches the intersection point Q with the tangent at P, the spiral is entirely unrolled. □

Most probably, you are not fully convinced by these "dynamical" arguments. However, we must bear in mind that differential calculus did not exist in Toricelli's time. In modern terms, writing θ for the angle of rotation of the line, the logarithmic spiral is defined by the differential equation

$$\frac{dR(\theta)}{d\theta} = kR(\theta).$$

This yields as possible solution

$$R(\theta) = ae^{k\theta}, \quad a \in \mathbb{R}.$$

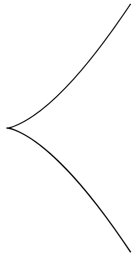

Fig. 1.24

A parametric representation of the logarithmic spiral is then

$$f(\theta) = \left(ae^{k\theta}\cos\theta, ae^{k\theta}\sin\theta\right).$$

One can now obtain the result by brute computation rather than imagination!

This result was doubly amazing for the mathematicians of the time:

- first, as already mentioned, the result finds a precisely defined segment whose length is equal to that of a piece of a curve;
- second, this segment has a finite length, while the piece of the curve winds infinitely many times before reaching the origin.

However, this is not really a counterexample to *Descartes*' statement. Indeed the curve itself was considered as "badly defined": the curve was described in a dynamic way, but its equation could not be written. Of course in those days, the exponential function could by no means be considered as a function and, even less, as a "well defined function".

Nevertheless this first attempt raised the hope of being able to rectify some curves and, perhaps, all curves. The British mathematician *Neil* (1659), the Dutch mathematician *van Heuraet* (1659) and the French mathematician *Fermat* (1660) were able to "rectify" the semi-cubic parabola, that is, the "well-defined" curve with equation

$$y^2 = x^3$$

(see Fig. 1.24).

The method of *Neil* and *van Heuraet* consisted of approaching the curve by a polygonal line inscribed to the curve and letting the distance between two consecutive points tend to zero. The method of *Fermat* consisted instead of approaching the curve by a polygonal line tangent to the curve (see Fig. 1.25). The most amazing point is the fact that all of them succeeded in computing the limit of the lengths of these polygonal lines before the birth of differential calculus in 1676. *Newton* (1642) and *Leibniz* (1646) were at the time already teenagers who were starting to get interested in these questions!

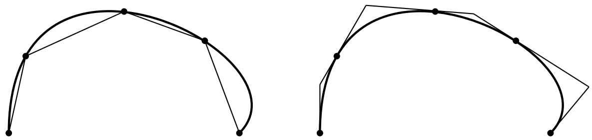

Fig. 1.25

1.8 Length *Versus* Curve Integral

The tricky computations of *Neil*, *van Heuraet* and *Fermat* (see Sect. 1.7) to compute the length of an arc of a cubic parabola are of course based on the following "definition":

> *Given a curve, we approximate it by a polygonal line as in Fig.* 1.25. *The length of the curve is the limit of the lengths of all possible polygonal lines as the length of all segments tends to zero.*

Once more, the intuition behind this "definition" is clear, but the terms contained in it should now be given a precise mathematical meaning. To achieve this, let us first work with this "definition" as such, without asking too many questions about its precise meaning and about the assumptions needed to develop the following proof.

Proposition 1.8.1 *Consider a parametric representation*

$$f\colon\,]a,b[\, \longrightarrow \mathbb{R}^2, \qquad t \mapsto f(t)$$

of class $\mathcal{C}^1$ *of a plane curve and two points* $c < d$ *in* $]a,b[$. *The length of the arc of the curve between the points of parameters* c *and* d *is equal to*

$$\int_c^d \|f'(t)\| dt.$$

Proof We thus call the *length* of the arc of the curve the limit of the lengths of the polygonal lines inscribed to the curve, as the length of each side tends to zero. For a natural number $n \neq 0$, put

$$\Delta_n(t) = \frac{d-c}{n}$$

and consider the polygonal line determined by the values

$$t_i = c + i\,\Delta_n(t), \quad 0 \leq i \leq n$$

of the parameter. The length of this polygonal line is equal to

$$\sum_{i=0}^{n} \| f(t_{i+1}) - f(t_i) \|.$$

The function f is continuous, thus it is uniformly continuous on the compact interval $[c,d]$. Therefore when n tends to ∞, that is as $\Delta_n(t)$ tends to 0, each side of the polygonal line has a length which also tends to 0.

But for "good" functions f, the Taylor expansion of f tells us that

$$f(t_{i+1}) = f(t_i) + \Delta_n(t) f'(t_i) + \mathcal{O}_1\big(\Delta_n(t)\big)$$

where $\mathcal{O}_1$ has the property

$$\lim_{x\to 0} \frac{\mathcal{O}_1(x)}{x} = 0.$$

This suggests to re-write the length of the polygonal line as

$$\sum_{i=0}^{n} \left\| f'(t_i) + \frac{\mathcal{O}_1(\Delta_n(t))}{\Delta_n(t)} \right\| \cdot \Delta_n(t).$$

When the number n of sides of the polygonal line tends to ∞, $\Delta_n(t)$ tends to 0 and this sum should thus have as limit

$$\int_c^d \|f'(t)\| dt.$$ □

This "proof" is not very rigorous, and nor should we expect it to be, after all it concerns a definition whose terms have bot been given a precise meaning. Nevertheless, the formula in Proposition 1.8.1 should remind us of well-known result in analysis:

Theorem 1.8.2 *Consider an injective function of class* $\mathcal{C}^1$

$$f\colon [c,d] \longrightarrow \mathbb{R}^n$$

and another continuous function

$$g\colon [c,d] \longrightarrow \mathbb{R}.$$

Then the curve integral of g *along* f *exists and is equal to*

$$\int_c^d g(t) \cdot \|f'(t)\| \, dt.$$

This immediately suggests the following definition

Definition 1.8.3 Let $f\colon]a,b[\longrightarrow \mathbb{R}^2$ be an injective parametric representation of class $\mathcal{C}^1$ of a curve. Given $a < c < d < b$, the *length* of the arc of the curve between the points with parameters c and d is by definition the *curve integral* of the constant function 1 along $f\colon [c,d] \longrightarrow \mathbb{R}^2$.

Since a parametric representation is always locally injective, the restriction of injectivity can easily be overcome: it suffices to compute a length "by pieces".

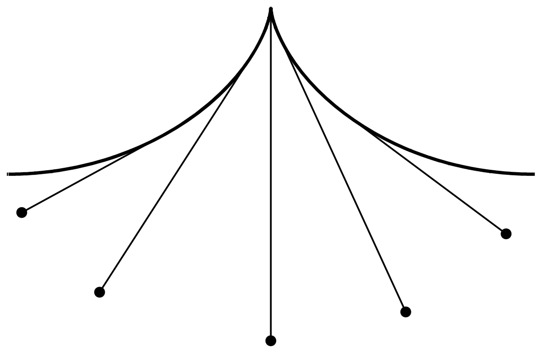

Fig. 1.26

1.9 Clocks, Cycloids and Envelopes

Still in the years 1650–1560, thus before the invention of differential calculus, the Dutch physicist *Huygens* was very much concerned with constructing the best clocks of his time. The most immediate way to construct a clock is based on the *pendulum principle*: attach a weight at the extremity of a chord and let it swing! Our physics courses tell us that the frequency of such a *pendulum* is "more or less" independent of the amplitude of the movement, at least in the case of oscillations of "small amplitude". We make these various qualifications, "more or less", "small amplitudes", and so on, but is it not possible to construct an *isochronal* pendulum: a pendulum which always swings at the same frequency, whatever the amplitude of the oscillations?

The frequency of a pendulum increases when the "chord" of the pendulum is shorter. When the amplitude of the oscillation increases, the frequency of the pendulum decreases. Thus to obtain a pendulum whose frequency is independent of the amplitude of the oscillation, it would "suffice" to have a chord of variable length: a length which diminishes as the pendulum moves away from its position of equilibrium, and gets longer again as the pendulum moves back towards its bottom position. How can one realise such a pendulum?

Huygens' idea was to attach the chord between two templates, so that the chord "rolls" on these templates (see Fig. 1.26) while the pendulum is swinging. The chord has its full length in vertical position and this length becomes shorter and shorter as the pendulum moves away from this equilibrium position. But what form should you give to the templates in order to get a pendulum whose frequency is independent of the amplitude of the oscillations?

Huygens discovers that the solution to the problem is obtained via the *cycloid*: a curve already mentioned in Sect. 1.6 (see Fig. 1.27).

Definition 1.9.1 The *cycloid* is the trajectory of a fixed point of a circle, as this circle rolls on a line.

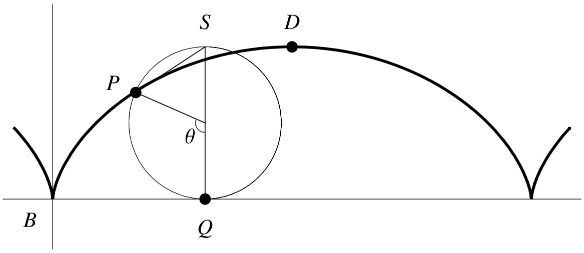

Fig. 1.27

The Italian mathematician *Torricelli* (1608–1647) and the French mathematician *Roberval* (1602–1675) had already studied problems of tangency, length and area for various curves, including the cycloid. *Huygens* knew these results and pursued the study of the cycloid further. Let us establish the necessary results with the contemporary techniques of Sects. 1.6 and 1.8.

Proposition 1.9.2 *Choosing the radius of the rolling circle as unit of length, a parametric representation of the cycloid is given by*

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad f(\theta) = (\theta - \sin\theta, 1 - \cos\theta).$$

Proof In Fig. 1.27, when the circle has turned by an angle θ, the length of the arc PQ is equal to θ. But by definition of the cycloid, this length is also that of the segment BQ. Observing further that the center of the circle moves on the line with equation $y = 1$, we obtain the announced formula. □

The following result was first discovered by *Roberval*, using his technique of *composition of movements*.

Proposition 1.9.3 *In Fig. 1.27, the tangent to the cycloid at a point P is the line joining P and the point S diametrically opposite to Q on the circle.*

Proof In view of Definition 1.6.1 and Proposition 1.9.2, the direction of the tangent at P is $(1 - \cos\theta, \sin\theta)$. The equation of this tangent is thus

$$\sin\theta\big(x - (\theta - \sin\theta)\big) - (1 - \cos\theta)\big(y - (1 - \cos\theta)\big) = 0.$$

The point Q has coordinates $(\theta, 0)$, thus S has coordinates $(\theta, 2)$. It is immediate that the coordinates of S satisfy the equation of the tangent. □

Here we present (with contemporary proof) *Toricelli's* result concerning the rectification of the cycloid.

Proposition 1.9.4 *In Fig. 1.27, write D for the middle point of the full arch of the cycloid. The length of the arc of the cycloid between P and D is equal to twice the* *length of the tangent segment* PS (*see Proposition* 1.9.3). *As a consequence, the length of a full arch of the cycloid is equal to eight times the radius of the rolling circle.*

Proof Going back to the proof of Proposition 1.9.3, we have

$$P = (\theta - \sin\theta, 1 - \cos\theta), \qquad S = (\theta, 2).$$

Therefore

$$\|\overrightarrow{PS}\| = \sqrt{2 + 2\cos\theta} = 2\cos\frac{\theta}{2}$$

since

$$1 + \cos\theta = 2\cos^2\frac{\theta}{2}.$$

On the other hand the length of the arc PD of the cycloid is given by (see Proposition 1.8.1)

$$\begin{aligned}\int_\theta^\pi \|f'\| &= \int_\theta^\pi \sqrt{2 - 2\cos\theta}\,d\theta \\ &= 2\int_\theta^\pi \sin\frac{\theta}{2}\,d\theta \\ &= -4\left(\cos\frac{\pi}{2} - \cos\frac{\theta}{2}\right) \\ &= 4\cos\frac{\theta}{2}\end{aligned}$$

where this time, we have used the formula

$$1 - \cos\theta = 2\sin^2\frac{\theta}{2}.$$

The length of the arc PD of the cycloid is thus indeed twice the length of the segment PS.

The length of a full arch is therefore four times the length of the tangent vector PS, when $P = B$. Except that at $P = B$, the argument above does not apply! Indeed for $\theta = 0$, the tangent vector $f'(\theta)$ becomes simply $(0, 0)$: the cusp points of the cycloid are singular points (see Definition 1.4.4). But taking the limit of the lengths of the arcs PD as P converges to B yields

$$\lim_{\theta\to 0} 4\cos\frac{\theta}{2} = 4$$

as expected. □

What *Huygens* proved about the cycloid is the following theorem.

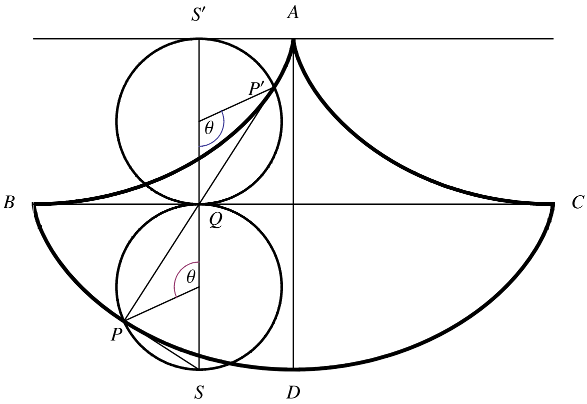

Fig. 1.28

Theorem 1.9.5 *Put a cycloid upside-down in a gravitational field. Attach at a cusp point of this cycloid a pendulum whose length is equal to half the length of an arch of the cycloid. The trajectory of this pendulum is another cycloid of the same size and the frequency of the pendulum is independent of the amplitude of the oscillation.*

Proof In Fig. 1.28, consider the lower cycloid, obtained when the lower circle of radius 1 rolls on the middle horizontal line. Analogously consider the upper cycloid, obtained when the upper circle with the same radius 1 rolls on the upper horizontal line. Write P, P' for the fixed points on these circles whose trajectories are the cycloids. Write further Q for the contact point of these two circles with the middle horizontal line and S, S' for the points of the circles diametrically opposite to Q.

By Proposition 1.9.3, PS is the tangent to the lower cycloid at P while $P'Q$ is the tangent to the upper cycloid at P'. Since both circles have already turned by the same angle θ after leaving B, the points P and P' are symmetric to each other with respect to Q. In other words, P, Q, P' are on the same line and both segments PQ, $P'Q$ have the same length. But by Proposition 1.9.4, the length of the segment $P'Q$ is half the length of the arc BP' of the cycloid. Thus the arc BP' of the cycloid has the same length as the segment PP'.

Thus the length of the arc AP' of the cycloid, augmented by the length of the segment $P'P$, yields the same result as the length of the full arc AB of the cycloid. In other words, if a pendulum is attached at A, with a length of cord equal to the arc AB of the upper cycloid, when this pendulum swings, its trajectory is exactly the lower cycloid.

Now call τ the angle between the tangent at P and the horizontal line. Writing g for the gravitational force, the acceleration of the pendulum along its trajectory is thus equal to $-g\sin\tau$. But, still by Proposition 1.9.4

$$\sin\tau = \frac{\|PS\|}{2} = \frac{\frac{1}{2}\text{arc}\,DP}{2}.$$

Write s for the length of the arc DP, viewed as the position of the pendulum on the cycloid. The acceleration of the pendulum along the cycloid is thus characterized by the differential equation

$$\ddot{s}(t) = -\frac{g}{4}s(t)$$

where t is the time. Integrating this differential equation yields

$$s(t) = \alpha \cos\left(\frac{1}{2}\sqrt{g}\,t\right) + \beta \sin\left(\frac{1}{2}\sqrt{g}\,t\right).$$

Let us assume that at the instant $t = 0$ the pendulum reaches its highest position $s = s_0$. The symmetry of the problem forces $s(t) = s(-t)$, which implies $\beta = 0$. Putting $t = 0$ yields $\alpha = s_0$, so that the equation of the movement is

$$s(t) = s_0 \cos\left(\frac{1}{2}\sqrt{g}\,t\right).$$

The point D, that is $s = 0$, is thus reached at the time t_0 such that

$$0 = s_0 \cos\left(\frac{1}{2}\sqrt{g}\,t_0\right)$$

that is

$$\frac{1}{2}\sqrt{g}\,t_0 = \frac{\pi}{2}.$$

The time necessary for the pendulum to reach its bottom position D is thus $\pi\sqrt{\frac{1}{g}}$: this time is indeed independent of s_0, the amplitude of the oscillation.

Not surprisingly, given that he was essentially trying to solve a differential equation before the invention of differential calculus, *Huygens*' argument for this last point was fairly convoluted. Amazingly, he nevertheless managed to solve the problem. □

This study of the cycloid underlines another important geometrical notion: the *envelope* of a family of curves.

Proposition 1.9.6 *In the situation depicted in Fig.* 1.28, *the upper cycloid is tangent to all the normals to the lower cycloid: therefore the upper cycloid is called the envelope of these normals (see Fig.* 1.29).

Proof The angle QPS is inscribed in a half circle, thus it is a right angle. Since PS is tangent to the lower cycloid, PQP' is normal to this cycloid, but it is also the tangent to the upper cycloid. □

The idea of considering the *envelope* of the normals to a given curve is much older that the work of *Huygens*. The first result in this direction is probably the case

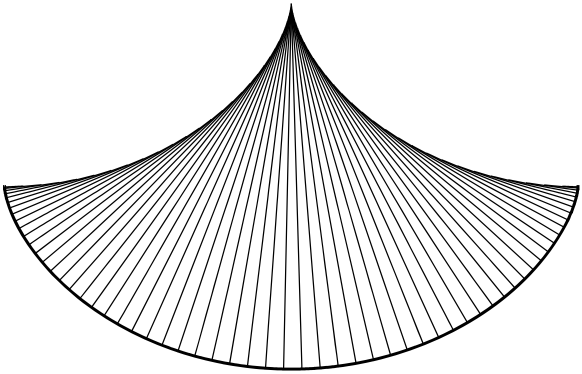

Fig. 1.29

of the normals to a parabola, studied once again by *Apollonius*, around 220 BC! The first case of an envelope of a family of arbitrary curves (not just straight lines) is probably due to *Torricelli* around 1642. We shall come back to these problems in Sect. 2.6.

1.10 Radius of Curvature and Evolute

The study of the cycloidal pendulum, as developed by *Huygens* (see Sect. 1.9) gives rise to a very interesting observation. An ordinary pendulum moves along a circle whose radius is the length of the chord of this pendulum. But the cycloidal pendulum has a chord of variable length, originating from a variable point: in Fig. 1.28, the chord is $P'P$ with origin P'. The instantaneous movement of the cycloidal pendulum at the point P thus coincides with the instantaneous movement of an ordinary pendulum whose chord would be $P'P$. Therefore the circle of radius $P'P$ is the best circular approximation of the cycloid at its point P. *Huygens* calls P' the *center of curvature* of the cycloid at P and $P'P$, the corresponding *radius of curvature*. The upper cycloid is thus the locus of the centers of curvature of the lower cycloid: what one calls the *evolute* of the lower cycloid.

Again it was *Huygens* who first succeeded in handling the problem of the *radius of curvature* in a quite general setting. Let us follow his argument, taking for granted that when we refer to a tangent, an intersection or a limit, it does exist!

> *Consider a fixed point P on a given curve and a variable point Q on the same curve.*

- The *center of curvature* at P is the limit of the intersection of the normal at P and the normal at Q, when Q converges to P.
- The distance between P and the corresponding center of curvature is called the *radius of curvature* at P.

- The locus of all the centers of curvature is called the *evolute* of the given curve.

As already mentioned, we take for granted that this definition makes sense, which is of course false, even for very good curves! For example if the curve is a straight line, the two normals at P and Q are parallel and you cannot even start the process! Our purpose is therefore once more to guess what a "good" contemporary definition should be. Let us translate *Huygens*' argument in contemporary terms.

Proposition 1.10.1 *Consider a plane curve with parametric representation f. In "good cases", the radius of curvature is given by the formula*

$$\rho = \frac{\|f'\|^3}{|f_2' f_1'' - f_1' f_2''|}.$$

Proof Let us consider the fixed point $P = f(t_0)$ and the variable point $Q = f(t)$. The normal vector at Q is thus $n(t) = (f_2'(t), -f_1'(t))$ and analogously at P. The intersection of the two normals is thus such that

$$f(t_0) + \alpha_t n(t_0) = f(t) + \beta_t n(t)$$

for two scalars α_t and β_t that we have now to determine.

This equality yields the system of equations

$$f_1(t_0) + \alpha_t f_2'(t_0) = f_1(t) + \beta_t f_2'(t)$$
$$f_2(t_0) - \alpha_t f_1'(t_0) = f_2(t) - \beta_t f_1'(t)$$

so that, by Cramer's rule for solving such a system and the well-known properties of determinants

$$\alpha_t = \frac{\det\begin{pmatrix} f_1(t) - f_1(t_0) & -f_2'(t) \\ f_2(t) - f_2(t_0) & f_1'(t) \end{pmatrix}}{\det\begin{pmatrix} f_2'(t_0) & -f_2'(t) \\ -f_1'(t_0) & f_1'(t) \end{pmatrix}} = \frac{\det\begin{pmatrix} \frac{f_1(t) - f_1(t_0)}{t - t_0} & -f_2'(t) \\ \frac{f_2(t) - f_2(t_0)}{t - t_0} & f_1'(t) \end{pmatrix}}{\det\begin{pmatrix} f_2'(t_0) & \frac{f_2'(t_0) - f_2'(t)}{t - t_0} \\ -f_1'(t_0) & \frac{f_1'(t) - f_1'(t_0)}{t - t_0} \end{pmatrix}}.$$

When t converges to t_0, we obtain

$$\alpha = \lim_{t \to t_0} \alpha_t = \frac{\det\begin{pmatrix} f_1'(t_0) & -f_2'(t_0) \\ f_2'(t_0) & f_1'(t_0) \end{pmatrix}}{\det\begin{pmatrix} f_2'(t_0) & -f_2''(t_0) \\ -f_1'(t_0) & f_1''(t_0) \end{pmatrix}} = \frac{\|f'(t_0)\|^2}{f_2'(t_0) f_1''(t_0) - f_1'(t_0) f_2''(t_0)}.$$

The center of curvature at P is then the point

$$f(t_0) + \alpha\, n(t_0).$$

The radius of curvature is simply

$$\begin{aligned}\|\alpha\, n(t_0)\| &= \left| \frac{\|f'(t_0)\|^2}{f_2'(t_0)\,f_1''(t_0) - f_1'(t_0)\,f_2''(t_0)} \sqrt{\big(f_2'(t_0)\big)^2 + \big(f_1'(t_0)\big)^2} \right| \\ &= \frac{\|f'(t_0)\|^3}{|f_2'(t_0)\,f_1''(t_0) - f_1'(t_0)\,f_2''(t_0)|}\end{aligned}$$

which is indeed the announced formula. □

1.11 Curvature and Normality

The treatment of the *radius of curvature* as in Sect. 1.10 is certainly very intuitive, but raises many questions. It refers again to a notion of "limit computed on a family of lines" and as we have seen in Sect. 1.6, such a notion of limit is not always as simple as one might imagine. So we now want to translate the ideas of Sect. 1.10 in "decent differential terms".

The idea is the following. What measures the *curvature* of a curve is the speed at which the tangent to the curve changes direction as you travel along this curve. But the measure of the variation of a quantity is something well-known today: this is the derivative of the quantity. The direction of the tangent to a regular curve represented by $f(t)$ is given by the tangent vector $f'(t)$ (see Definition 1.6.1). One could thus be tempted to define the *curvature* as the derivative of this tangent vector, that is as the vector $f''(t)$.

However, $f'(t)$ is a vector in $\mathbb{R}^2$: it thus has a direction and a length. When this vector varies, it can vary both in direction and in length. What we are interested in, is only the *variation in direction*. Of course if it turns out that the vector $f'(t)$ has a constant length, for all values of the parameter t, then the derivative $f''(t)$ measures exactly the variation in direction of the tangent vector $f'(t)$, and we end up with an elegant way of defining the curvature. But is such a situation possible?

Proposition 1.11.1 *Choose as parameter for describing a curve, the length s of the arc of the curve from an arbitrary point on the curve chosen as origin. If the corresponding parametric representation $f(s)$ is regular, the tangent vector $f'(s)$ is of length 1, for every value of the parameter s. Such a parametric representation is called normal.*

Proof Applying Proposition 1.8.1 to such a special parametric representation, we get

$$s = \int_0^s \|f'\|.$$

Deriving both sides with respect to s yields $1 = \|f'\|$. □

We therefore make the following definition:

Definition 1.11.2 Let $f(s)$ be a normal representation of class $\mathcal{C}^2$ of a curve (see Proposition 1.11.1). The *curvature* at the point with parameter s is by definition the quantity $\|f''(s)\|$.

It is fairly immediate to observe that this definition does not depend on the normal representation chosen: we shall see this in more detail in Sect. 2.9.

To conclude this section, we should now exhibit the link between the notion of *curvature* in Definition 1.11.1 and the more intuitive idea of *radius of curvature* obtained via the intersection of normals. For that it suffices to remember that the derivative of a scalar product can be computed via a formula analogous to that of the derivative of an ordinary product.

Lemma 1.11.3 *Consider two functions of class* $\mathcal{C}^1$

$$f, g\colon \mathbb{R} \to \mathbb{R}^2$$

and the corresponding function

$$(f|g)\colon \mathbb{R} \longrightarrow \mathbb{R}, \qquad t \mapsto \big(f(t)\big|g(t)\big).$$

Under these conditions

$$(f|g)' = \big(f'|g\big) + \big(f|g'\big).$$

Proof Indeed

$$\begin{aligned}(f|g)' &= (f_1g_1 + f_2g_2)' \\ &= \big(f_1'g_1 + f_1g_1'\big) + \big(f_2'g_2 + f_2g_2'\big) \\ &= \big(f_1'g_1 + f_2'g_2\big) + \big(f_1g_1' + f_2g_2'\big) \\ &= \big(f'\big|g\big) + \big(f\big|g'\big)\end{aligned}$$

by the ordinary formula for the derivative of a product. □

We then have:

Proposition 1.11.4 *Let* $f(s)$ *be a normal representation of class* $\mathcal{C}^2$ *of a curve. The radius of curvature at a given point is the inverse of the curvature at this point, provided of course that this curvature is not zero.*

Proof By Proposition 1.11.1, a normal representation f has a tangent vector f' of constant length 1. Differentiating the equality $(f'|f') = 1$, we obtain $2(f'|f'') = 0$. Thus f'' is orthogonal to f' and therefore, the vector $v = (-f_2'', f_1'')$ perpendicular to f'' is parallel to f'. It follows that

$$\big(f'\big|v\big) = \big\|f'\big\| \cdot \|v\| \cdot \cos k\pi = \pm\|v\| = \pm\sqrt{\big(f_2''\big)^2 + \big(f_1''\big)^2} = \pm\big\|f''\big\|.$$

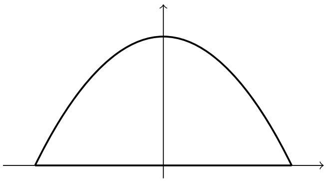

Fig. 1.30

But

$$\left(f'\middle|v\right) = -f_1' f_2'' + f_2' f_1''.$$

Introducing these values into the formula of Proposition 1.10.1 gives

$$\rho = \frac{1}{\|f''\|}$$

as announced. □

1.12 Curve Squaring

"Circle squaring", that is, constructing by ruler and compass a square having the same area as a circle, is a problem which puzzled mathematicians for more than two millenniums. This problem has largely been discussed in Sect. 2.4 of [3], *Trilogy I*; its impossibility was finally proved in 1882 as a corollary of a famous result of *Lindemann*: the number π is transcendental, that is, it cannot be obtained as a solution of an equation with rational coefficients (see Sect. B.3 of [3], *Trilogy I*).

Of course, today, "squaring" a portion of the plane is no longer seen as a "ruler and compass" problem, but as a question of integral calculus. Therefore "curve squaring" is generally not considered as part of curve theory and is instead treated in an analysis course: we thus direct the reader towards an analysis book for a systematic treatment of these questions. Notice that making clear which curves can be "squared" is already a challenging problem.

Nevertheless, due to the historical importance of these questions, it is sensible to present here a short section on this curve squaring problem, focusing on some historically important examples. Our first example was treated by *Archimedes* (see Sect. 4.4 in [3], *Trilogy I*).

Example 1.12.1 The area of the portion of the plane delimited by the x-axis and the parabola of equation $y = 1 - x^2$ is equal to $\frac{4}{3}$ (see Fig. 1.30).

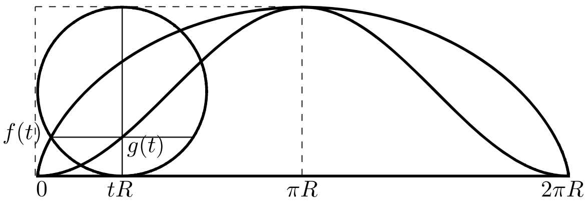

Fig. 1.31

Proof We know that this area is given by the integral

$$\int_{-1}^{+1} 1-x^2\,dx = \left[x-\frac{x^3}{3}\right]_{-1}^{+1} = \left(1-\frac{1}{3}\right)-\left(-1+\frac{1}{3}\right)=\frac{4}{3}. \qquad \square$$

Don't forget that the parabola is a section of a circular cone by a plane. Greek geometers were able to "square" the parabola, but not the circle which for them was thus a priori a "more elementary" curve. A rather intriguing situation! Today we know the easy explanation for this phenomenon: the parabola admits an equation of the form $y = p(x)$ with $p(x)$ a polynomial: the integration yields another polynomial $\int p(x)\,dx$, that is a formula in terms of the four arithmetical operations, and the four arithmetical operations (as well as the square root) can be performed by ruler and compass constructions. On the other hand, viewing the upper half of a circle of radius R as the graph of the function $y = \sqrt{R^2 - x^2}$, the area of the circle is given by

$$2\int_{-R}^{+R} \sqrt{R^2-x^2}\,dx$$

and such an integral involves inverse trigonometric functions!

Again it was during the 17th century that several mathematicians were able to compute—using precursors of differential methods—areas delimited by various curves. A celebrated achievement of this type, before the invention of differential calculus by *Newton* and *Leibniz*, is the squaring of the cycloid by *Roberval*, in 1634. *Roberval's* subtle computation of the integral, before the discovery of that notion, is described below.

Proposition 1.12.2 *The area of an arch of a cycloid, with generating circle of radius* R, *is equal to* $3\pi R^2$.

Proof We shall follow the argument in modern terms (see Fig. 1.31). We know already that a parametric representation of the cycloid is given by (see Proposition 1.9.2)

$$f(t) = R(t - \sin t, 1 - \cos t).$$

An arch of the cycloid is obtained when t varies from 0 to 2π.

Let us write $g(t)$ for the orthogonal projection of the point $f(t)$ on the instantaneous vertical radius of the rolling circle. Thus

$$g(t) = R(t, 1 - \cos t).$$

The function g represents the so-called *Roberval curve*. It is immediate that this curve g, for t varying from 0 to π, is symmetric with respect to its middle point $g(\frac{\pi}{2})$: that is

$$\frac{g(\frac{\pi}{2}+t) - g(\frac{\pi}{2}-t)}{2} = g\left(\frac{\pi}{2}\right).$$

Therefore the area under this curve g, between 0 and π, is equal to half the area of the corresponding rectangle; that is, $\frac{1}{2}(\pi R)(2R) = \pi R^2$. It follows that the area under the *Roberval curve*, between the points with parameters $t = 0$ and $t = 2\pi$, is equal to $2\pi R^2$.

It remains to compute the area between the *Roberval curve* and the cycloid. *Roberval* simply observes that the segment joining $f(t)$ and $g(t)$ is equal to half the corresponding chord of the circle. "Pushing" all the segments $f(t)g(t)$ to the left, in order to align their right extremities $g(t)$ vertically, for t varying from 0 to π, *Roberval* concludes that these segments fill the left half of the generating circle. This half circle has area $\frac{1}{2}R^2$. Thus as t varies from 0 to 2π, the area between the *Roberval curve* and the cycloid is equal to R^2.

Putting all of this together, an arch of the cycloid has an area equal to $3\pi R^2$. □

This proof is very interesting for two reasons. First—choosing the radius R as unit length—the *Roberval curve* is simply the curve

$$y = 1 - \cos x.$$

Up to a translation, the *Roberval curve* is thus the graph of a cosine (or sine) function. It seems that this is the first time that the graph of the sine function appears in the mathematical treatment of a problem. The way the corresponding integral is computed is particularly imaginative.

Second, the "pushing" argument of *Roberval* may appear rather strange to us. Let us nevertheless observe that writing the equations of the first halves of the cycloid and the *Roberval* curve in the form

$$x = c(y), \quad x = r(y), \quad 0 \le x \le R\pi,\ 0 \le y \le 2R$$

this "pushing" argument corresponds precisely to the modern formula

$$\int (c - r) = \int c - \int r.$$

To conclude this short section, let us recall a celebrated result of integral calculus: the so-called *Green–Riemann* formula.

Theorem 1.12.3 (Green–Riemann) *Let $K \subseteq \mathbb{R}^2$ be a compact subset whose boundary is constituted of finitely many curves of class $\mathcal{C}^1$. Moreover assume that the boundary of K is oriented in such a way that K is always on the left of its boundary. Given a differential form $P(x, y)dx + Q(x, y)dy$ of class $\mathcal{C}^1$ defined on an open subset containing K, one has*

$$\int_{\partial K} \big(P(x, y)dx + Q(x, y)dy\big) = \int_K \left(\frac{\partial Q}{\partial x}(x, y) - \frac{\partial P}{\partial y}(x, y)\right) dx\, dy.$$

Corollary 1.12.4 *The area delimited by a closed plane curve $\mathcal{C}$, the boundary of a compact subset $K \subseteq \mathbb{R}^2$ as in the* Green–Riemann *theorem, is equal to either of the following equal quantities*

$$\left|\int_{\mathcal{C}} x\, dy\right|, \qquad \left|\int_{\mathcal{C}} y\, dx\right|.$$

Proof The area is simply the integral

$$\int\int_K dx\, dy$$

of the constant function 1 on K. Putting $P = 0$ and $Q = x$ in the *Green–Riemann* formula yields the first formula of the statement; putting $P = y$ and $Q = 0$ yields the second formula. □

Example 1.12.5 The area delimited by an ellipse of half axis a and b is equal to πab.

Proof A parametric representation of the ellipse $\mathcal{E}$ is given by

$$f(\theta) = (a \cos\theta, b \sin\theta).$$

By Corollary 1.12.4, the corresponding area is thus

$$\begin{aligned}
\left|\int_{\mathcal{E}} a\, \cos\theta\, d(b\, \sin\theta)\right| &= \left|\int_0^{2\pi} a\, \cos\theta\, b \cos\theta\, d\theta\right| \\
&= \left|ab \int_0^{2\pi} \cos^2\theta\, d\theta\right| \\
&= \left|ab\left[\frac{\theta}{2} + \frac{\sin 2\theta}{4}\right]_0^{2\pi}\right| \\
&= ab\pi.
\end{aligned}$$

When $a = R = b$, we recapture the usual formula πR^2 for the area of a circle of radius R. □

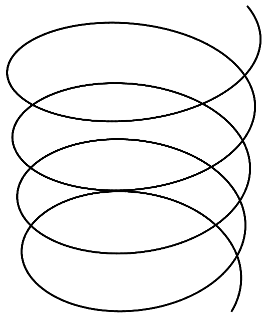

Fig. 1.32

1.13 Skew Curves

Let us now switch to the case of *skew curves*, or *space curves*, that is: curves in the three dimensional space $\mathbb{R}^3$.

The systematic study of skew curves was initiated in 1731 by the French mathematician *Clairaut*. His idea is to present a skew curve as the intersection of two surfaces, just as a line can be presented as the intersection of two planes. A skew curve is thus described by a system of two equations

$$\begin{cases} F(x,y,z)=0 \\ G(x,y,z)=0. \end{cases}$$

The tangent line to the skew curve at a given point is then obtained as the intersection of the tangent planes to the surfaces

$$F(x,y,z)=0, \qquad G(x,y,z)=0$$

at this same point. As you might suspect, the technicalities inherent to such an approach are quite heavy.

Let us for example focus on the question of the curvature. In the plane, a curve with constant curvature is a circle. However, in three dimensional space, there are many more curves with constant curvature. The easiest example is that of the *circular helix* (see Fig. 1.32): the curve having the shape of a screw. Of course this curve must be "identical" at all points: this is why you can screw a bolt through a nut! In particular, the curvature must be the same at all points. There are many more examples of curves with constant curvature, as we shall see later.

The study of the curvature of a skew curve was initiated by the French mathematician *Monge* in 1771, thus still before the introduction of parametric equations. In three dimensional space, the *normal* to the curve now becomes a *normal plane* to the curve: the plane perpendicular to the tangent at a given point. Therefore *Monge* makes the following definitions:

- *The* axis of curvature *at a point P of the skew curve is the limit of the intersection of the normal plane at P and the normal plane at a point Q of the curve, as Q converges to P.*
- *The* radius of curvature *at P is the distance between P and the axis of curvature at P.*

The fact of having an *axis* of curvature instead of a *center* of curvature explains in particular why curves with the same curvature can have very different shapes. The orientation of the axis of curvature is generally not constant and its variations in direction affect in an essential way the shape of the curve.

It remains an excellent exercise of technical virtuosity to compute the axis of curvature, starting from a system of Cartesian equations as in the time of *Clairaut* and *Monge*.

The ideas of *Monge* were clarified and developed in 1805 by his student *Lancret*, who introduced what we call today the *osculating plane* and the *torsion*, in order to study the variations in direction of the *axis of curvature*. The *osculating* plane is in a sense the *tangent plane* to the curve: the plane in which the curve tends to fit locally. Provided the following can be made precise, the idea is this:

> *The* osculating plane *at a fixed point P of a skew curve is the limit of the planes through P, Q, R, when Q and R are two other points of the curve converging to P.*

Lancret observed that the axis of curvature is perpendicular to the osculating plane.

Following the comments of the previous sections, we need not re-emphasize the fact that the definitions of *Monge* and *Lancret*, however intuitive, raise endless difficulties! Again, this is not the point here: in the "good cases" these definitions should recapture what we have in mind. What remains to be done is to work out these unpolished ideas to end up with a rigorous alternative presentation in decent differential terms!

Let us recall, as already mentioned in Sect. 1.2, that *Euler* introduced in 1775 his idea of *separating the variables*. This allows us to define a skew curve via three parametric equations

$$\begin{cases} x = f_1(t) \\ y = f_2(t) \\ z = f_3(t) \end{cases}$$

that is, finally, via a parametric representation

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}^3, \qquad t \mapsto f(t) = \big(f_1(t), f_2(t), f_3(t)\big).$$

This approach, together with the full strength differential calculus introduced a century earlier by *Newton* and *Leibniz*, allows us to transpose to skew curves most of the considerations developed in the previous sections. We do this immediately.

- A *parametric representation* of a skew curve is a locally injective continuous function

$$f\colon\,]a,b[\longrightarrow \mathbb{R}^3.$$

- The parametric representation is *regular* when it is of class $\mathcal{C}^1$ and $f'(t) \neq 0$ at each point.
- In the regular case, the *tangent* to the curve at the point with parameter t is the line through $f(t)$ and of direction $f'(t)$.
- The *normal plane* to the curve at a point is the plane perpendicular to the tangent at this point.
- When f is injective of class $\mathcal{C}^1$, the *length* of the arc of the curve between the points with parameters $c < d$ is the integral of the constant function 1 along this arc; it is also equal to $\int_c^d \|f'\|$.
- The parametric representation f is *normal* when the parameter is the length traveled on the curve from an arbitrary origin.
- Given a normal representation of class $\mathcal{C}^2$, $\|f'\| = 1$ and f' is orthogonal to f''.
- Given a normal representation of class $\mathcal{C}^2$, the *curvature* is the quantity $\|f''\|$.

Let us follow *Lancret's idea* and investigate first the case of the *osculating plane*.

Proposition 1.13.1 *Let $f(t)$ be a parametric representation of a skew curve. "Under suitable assumptions", the* osculating plane *at a point $f(t)$ is the plane through $f(t)$ whose direction is determined by the vectors $f'(t)$ and $f''(t)$.*

Proof Of course for this statement to make sense, f should be at least of class $\mathcal{C}^2$, with $f'(t)$ and $f''(t)$ linearly independent, in order to determine a plane. But our point here is not to exhibit all the "suitable" assumptions.

We thus fix a point $P = f(t_0)$ and consider two variable points $Q = f(t_1)$, $R = f(t_2)$ on the curve. We are interested in the plane

- containing the point $f(t_0)$;
- whose direction contains the vectors $f(t_1) - f(t_0)$ and $f(t_2) - f(t_0)$.

We now have to let t_1 and t_2 converge to t_0. With the considerations of Sect. 1.6 on the tangent in mind, we might be tempted to divide the two vectors $f(t_i) - f(t_0)$ by $t_i - t_0$ and let t_i converge to t. But of course this cannot possibly work since in both cases, the limit would be the same vector $f'(t_0)$. *One* vector no longer determines a plane! So let us handle separately the points Q and R. We consider first that the direction of the plane is equivalently given by

$$\frac{f(t_1) - f(t_0)}{t_1 - t_0}, \qquad f(t_2) - f(t_0)$$

and we let t_1 tend to t_0. This yields a plane whose direction contains the vectors

$$f'(t_0), \qquad f(t_2) - f(t_0).$$

Using a Taylor expansion we write

$$f(t_2) = f(t_0) + (t_2 - t_0) f'(t_0) + \frac{1}{2}(t_2 - t_0)^2 f''(t_0) + \mathcal{O}(t_2 - t_0)$$

where

$$\lim_{x \to 0} \frac{\mathcal{O}(x)}{x^2} = 0.$$

This allows us to further characterize the direction of the plane by the vectors

$$f'(t_0), \qquad (t_2 - t_0) f'(t_0) + \frac{1}{2}(t_2 t_0)^2 f''(t_0) + \mathcal{O}(t_2 - t_0).$$

Working on linear combinations of these two vectors, the direction is also determined by the vectors

$$f'(t_0), \qquad f''(t_0) + 2\,\frac{\mathcal{O}(t_2 - t_0)}{t_2 - t_0}.$$

Letting t_2 tend to t_1, we get

$$f'(t_0), \qquad f''(t_0).$$

□

Next, let us make the link with the ideas of *Monge*.

Proposition 1.13.2 *"Under suitable assumptions", the axis of curvature is perpendicular to the osculating plane and the radius of curvature is the inverse of the curvature.*

Proof Let us work with a parametric representation

$$f\colon\,]a,b] \longrightarrow \mathbb{R}^3, \qquad t \mapsto f(t) = \big(f_1(t), f_2(t), f_3(t)\big)$$

and let us assume that Proposition 1.13.1 applies: in particular $f'(t)$ and $f''(t)$ are linearly independent at each point. We study the curvature at $f(t_0)$ and there is no loss of generality in choosing a rectangular system of coordinates with origin $f(t_0)$ and such that $f'(t_0)$ is oriented along the third axis. Thus

$$f(t_0) = (0,0,0), \qquad f'(t_0) = \big(0, 0, f_3'(t_0)\big).$$

The normal plane at $f(t_0)$ is thus the plane with equation $z = 0$.

The normal plane at $f(t)$ admits the equation

$$f_1'(t)\big(x - f_1(t)\big) + f_2'(t)\big(y - f_2(t)\big) + f_3'(t)\big(z - f_3(t)\big) = 0.$$

Its intersection with the plane $z = 0$ is thus the line with equation

$$f_1'(t)\big(x - f_1(t)\big) + f_2'(t)\big(y - f_2(t)\big) = f_3'(t) f_3(t)$$

in the (x, y)-plane. Keeping in mind the very special form of the coordinates of $f(t_0)$ and $f'(t_0)$, and dividing by $t - t_0$, this equation can equivalently be re-written as

$$\frac{f_1'(t) - f_1'(t_0)}{t - t_0}\big(x - f_1(t)\big) + \frac{f_2'(t) - f_2'(t_0)}{t - t_0}\big(y - f_2(t)\big) = f_3'(t)\frac{f_3(t) - f_3(t_0)}{t - t_0}.$$

Passing to the limit when t converges to t_0 yields

$$f_1''(t_0)x + f_2''(t_0)y = \big(f_3'(t_0)\big)^2.$$

Assume now that f is a normal representation. Then $\|f'\| = 1$, thus $f_3'(t_0) = 1$. Moreover f'' is perpendicular to f':

$$f_1'f_1'' + f_2'f_2'' + f_3'f_3'' = 0$$

thus, $f_3''(t_0) = 0$. The system of equations of the axis of curvature can thus be written as

$$\begin{cases} f_1''(t_0)x + f_2''(t_0)y + f_3''(t_0)z = 1 \\ z = 0. \end{cases}$$

The first plane is orthogonal to $f''(t_0)$ and the second one is orthogonal to $f'(t_0)$, thus their intersection—the axis of curvature—is perpendicular to these two vectors, which by Proposition 1.13.1 span the osculating plane.

The radius of curvature is the distance, in the (x, y)-plane, between the origin and the line with equation

$$f_1''(t_0)x + f_2''(t_0)y = 1.$$

This distance is simply

$$\sqrt{\big(f_1''(t_0)\big)^2 + \big(f_2''(t_0)\big)^2} = \sqrt{\big(f_1''(t_0)\big)^2\big(f_2''(t_0)\big)^2\big(f_3''(t_0)\big)^2} = \|f''(t_0)\|. \qquad \square$$

Let us conclude with the definition of the *torsion* of the curve which measures *the variation of the osculating plane*, that is, the *variation of the axis of curvature*. The symbol $\times$ indicates the cross product (see Sect. 1.7 in of [4], *Trilogy II*).

Definition 1.13.3 The *torsion* of a skew curve is the variation of its osculating plane. More precisely, given a normal representation $f(s)$ of the curve, it is the quantity $\|n'(s)\|$ where

$$n(s) = \frac{f'(s) \times f''(s)}{\|f'(s) \times f''(s)\|}$$

is the vector of length 1 perpendicular to the osculating plane at $f(s)$.

Subsequent work of the French mathematicians *Cauchy*, *Frenet*, *Serret* and *Darboux* (among others) then established the modern bases of the theory of skew curves, studied in our Chap. 4.

1.14 Problems

1.14.1 Show that the "non-curve" of Example 1.1.1 can also be obtained via a function $F(x, y)$ of class $\mathcal{C}^1$.

1.14.2 Show that the curve with equation

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

has the shape of the curve in Fig. 1.7. Find a corresponding parametric representation.

1.14.3 It is possible to modify Example 1.2.1 so that the continuous functions f_n are all injective, and still converge uniformly to a function which is surjective from the "unit interval" to the "unit square", but the limit function is nevertheless not locally injective. Can you imagine such an example?

1.14.4 When $n = 1$, the *Local inverse theorem* (see 1.3.1) reduces to the simple fact that a function $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ of class $\mathcal{C}^1$ whose derivative is non-zero at a point is *monotone*, and thus bijective, in a neighborhood of this point.

1.14.5 An irreducible polynomial $F(X, Y)$ yields a curve $F(x, y) = 0$ in the sense of Definition 1.4.2 as soon as this "curve" is not empty.

1.14.6 Determine the tangent to a cycloid using the method of "composition of movements".

1.14.7 Prove Proposition 1.7.3 using the differential definitions of a tangent and a length.

1.14.8 Calculate the length of an arc of the cubic parabola, the curve with equation $y^2 = x^3$.

1.14.9 In Propositions 1.8.1 and 1.13.1, what about the case where f is of class $\mathcal{C}^\infty$ but the Taylor expansion of f does not converge to f on a neighborhood of t_0?

1.14.10 Prove *Apollonius*' result attesting that every normal to the parabola $y^2 = 2px$ is tangent to the semi-cubic parabola $27py^2 = 8(x - p)^3$.

1.14.11 Using the differential techniques, prove that the evolute of a cycloid is another cycloid of the same size.

1.14.12 Is there a normal parametric representation of the circle of radius R?

1.14.13 Prove that when a curve admits a regular parametric representation, all its normal representations are regular.

1.14.14 Calculate the area of an arch of a cycloid using the *Green–Riemann* formula.

1.14.15 Find a system of two Cartesian equations describing the circular helix.

1.14.16 Calculate the axis of curvature of the circular helix.

1.14.17 Prove that the circular helix has constant curvature and constant torsion.

1.14.18 Consider a *logarithmic helix* with parametric representation

$$f(\theta) = \big(e^\theta \cos\theta, e^\theta \sin\theta, \alpha(\theta)\big).$$

Determine the differential equation that α must satisfy in order for this spiral to have a constant curvature.

1.14.19 Given two functions of class $\mathcal{C}^1$

$$f, g \colon]a, b[\longrightarrow \mathbb{R}^3$$

and their cross product

$$f \times g \colon]a, b[\longrightarrow \mathbb{R}^3$$

(see Sect. 1.7, Vol. 2), prove that $f \times g$ is still of class $\mathcal{C}^1$ and

$$(f \times g)' = \big(f' \times g\big) + \big(f \times g'\big).$$

1.15 Exercises

1.15.1 Find a polynomial equation $F(x, y) = 0$ whose set of solutions comprises n points (a_i, b_i), $i = 1, \ldots, n$.

1.15.2 Find an equation $F(x, y) = 0$, with F a continuous function, whose set of solutions is the full circle of radius 1 centered at the origin.

1.15.3 Find an open subset U of the real line and an injective function of class $\mathcal{C}^\infty$,

$$f \colon U \longrightarrow \mathbb{R}^2$$

whose image is the hyperbola with equation

$$x^2 - y^2 = 1.$$

1.15.4 Give a parametric representation of the curve in Fig. 1.9.

1.15.5 Consider the curve represented by

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad \theta \mapsto (\cos k\theta \cos\theta, \cos k\theta \sin\theta), \quad k \in \mathbb{R}.$$

Draw a picture of this curve for $k = 0$, $k = 1$, $k = 2$ and $k = 3$. Find the corresponding Cartesian equations. Are there multiple points? Can you guess what the shape of the curve becomes in the case of an irrational parameter k?

1.15.6 Find the two tangents to the ellipse of equation $x^2 + 2y^2 = 1$, passing through the point $(3, 3)$. Determine the position of the two foci and observe the bisector property mentioned in Proposition 1.5.6.2.

1.15.7 Find the two tangents to the hyperbola with equation $x^2 - 2y^2 = 3$, passing through the point $(1, 2)$. Determine the position of the two foci and observe the bisector property mentioned in Proposition 1.5.6.1.

1.15.8 Find the two tangents to the parabola with equation $y = 3x^2 + 2x - 1$, passing through the point $(-2, -3)$. Determine the position of the focus and that of the directrix and observe the bisector property mentioned in Proposition 1.5.6.3.

1.15.9 In $\mathbb{R}^2$, consider an ellipse, a hyperbola or a parabola with equation $p(x, y) = 0$, with p a polynomial of degree 2. Consider further the two subsets

$$\mathcal{Q}_- = \{(x, y) \big| p(x, y) < 0\}, \qquad \mathcal{Q}_+ = \{(x, y) \big| p(x, y) > 0\}.$$

Show that through a point in one of these subsets, there are always two distinct tangents (or asymptotes) to the conic, while through a point in the other subset, there is no tangent at all.

1.15.10 Consider the cycloidal pendulum as in Fig. 1.26. Find a parametric equation of the trajectory of the pendulum when the two arches of the cycloid are substituted by quarters of circles.

1.15.11 In a coordinate system of your choice, find a parametric representation of the skew curve obtained as the intersection of a sphere and a cone "in arbitrary positions".

http://www.springer.com/978-3-319-01735-8

A Differential Approach to Geometry
Geometric Trilogy III
Borceux, F.
2014, XVI, 452 p. 159 illus., Hardcover
ISBN: 978-3-319-01735-8