STUDENT MATHEMATICAL LIBRARY
Volume 15

Plane Algebraic Curves

Gerd Fischer

Translated by
Leslie Kay

American Mathematical Society

Contents

Preface to the English Edition

I am very pleased that the AMS has decided to publish an English version of the German text. This was a good opportunity to add a new section (3.9) on the recently discovered Chebyshev curves and to improve the appendix on the implicit function theorem. My thanks go to the AMS and Vieweg for this joint project, to Leslie Kay for her excellent translation, including many clarifications of details, and to my students in Düsseldorf (especially Nadine Engeler, Thorsten Haarhoff, and Thorsten Warmt) for their help in preparing the new sections.

Münich, January 2001 Gerd Fischer

Preface to the German Edition

How many zeros does a polynomial in one variable have? This question is answered definitively by the "fundamental theorem of algebra." But if we go to two variables, the zero sets become infinite in general. These sets can be viewed as geometric objects—more precisely, as plane algebraic curves. So two paths intersect here, one from algebra and one from geometry, and it is hardly surprising that properties of such curves have been pondered for many centuries.

Adding yet another book to the countless books on this topic demands justification, or at least an explanation of some special point of view. I won't conceal the external stimulus: Several years ago, I was encouraged to write something about algebraic curves. My immediate response was that there were already a lot of books—perhaps too many books—about them. But I couldn't resist the temptation to keep giving lectures on the subject and writing up my notes. Let me briefly explain what they eventually turned into.

The text consists of two very different parts. In Chapters 0 to 5, the geometry of curves is explained in as elementary a way as possible: tangents, singularities, inflection points, etc. The most important technical tool is the intersection multiplicity, which is based on the resultant, and the main result is Bézout's theorem on the number of points of intersection of two curves. This part culminates in the Plücker formulas, in Chapter 5. These formulas relate the invariants studied in the preceding chapters.

The Plücker formulas can be given an almost—but not completely—precise proof by elementary techniques. What is missing, in particular, is a deeper understanding of duality and an efficient way to compute the intersection multiplicities that appear. The necessary local and global techniques from analysis are given later, in Chapters 6 to 9. Although the results are relatively easy to state and apply, laying a sound foundation takes some work.

Chapters 6 to 8 therefore contain an introduction to local complex analysis. This is the theory of either convergent power series or holomorphic functions of several variables, depending on one's preferred point of view. Here power series and the algebraic properties of rings of power series are emphasized; this approach goes back to the pioneering work of Rückert [R].

In the last chapter, the local parametrizations are patched together into a Riemann surface. Borrowing from a famous quotation of Felix Klein, one might say that curves are then regarded as freed from their cage—the projective plane—and floating outside a fixed space. The genus formula is ultimately an extension of the elementary Plücker formulas.

The appendices contain some technical tools from algebra and topology that are used repeatedly, as well as supplements to the preceding chapters.

Throughout the text an attempt was made to stay very concrete and, when possible, to give procedures for computing something by using polynomials and power series. The many examples and figures should also help keep things concrete. This aspect of algebraic geometry, long regarded as rather old-fashioned, has regained importance.

As one might expect, almost everything here can be found in a similar form elsewhere. I would especially like to mention Walker [Wa], Burau [Bu], and Brieskorn-Knörrer [B-K]. My goal was as concise a text as possible for an introductory one- or two-semester course. (Following a remark of Horst Knörrer, one could describe this little book as a portable version of the stationary model [B-K].) All that is assumed is some basic background, especially in elementary algebra and complex function theory. A great deal of effort has only strengthened my conviction that there hardly exists a more beautiful approach to algebraic geometry and complex analysis than through algebraic curves. Geometric intuition and "analytic" methods still lie very close together here, and every new technique is completely motivated by clear geometric problems—as in paradise before the many falls from grace.

My thanks go to all who helped bring this book into being: my teacher R. Remmert for his encouragement; my students at Düsseldorf and UC Davis for their suggestions for improvements; Mr. H.-J. Stoppel for his untiring help in countless details and the production of the TEX manuscript; Mr. U. Daub for plotting the first pictures; Mr. C. Töller for the final production of the finished figures; and finally Vieweg, the publishers, who expressed their willingness to publish the book in the German language and at a student-friendly price.

Düsseldorf, June 1994 Gerd Fischer

Chapter 0

Introduction

Let an object move through space as time passes. The task of curve theory is to describe this process abstractly and study it in detail. Modern curve theory has many branches, and no attempt will be made here to give an overview of the numerous questions that are treated in this context. Instead we will carefully examine a small, clearly delimited, but very exciting part: the elementary theory of *plane algebraic* curves. The first restriction, *plane*, means that the space in which the motion occurs is only *two-dimensional*; this makes a number of things easier. Before explaining what we mean by an *algebraic* curve, we give a few examples of general plane curves.

The moving object is assumed to be a point. Then its motion in the plane is described by a map

$$\varphi : I \to \mathbb{R}^2, \quad t \mapsto \varphi(t) = (x_1(t), x_2(t)),$$

where $I \subset \mathbb{R}$ denotes an interval. The *parameter* t can be viewed as *time*.

0.1. A *line* can be described by

$$\varphi(t) = v + tw,$$

where v, $w \in \mathbb{R}^2$ are vectors and the direction vector w is not the zero vector. Here we may take I to be $\mathbb{R}$. The same subset $C = \varphi(\mathbb{R}) \subset \mathbb{R}^2$ can be traced in many different ways; that is, there are many different *parametrizations* φ with the same *trace* $\varphi(I)$—just as the railroad, with a fixed network of tracks, can keep setting up new timetables. It will turn out that there is far less freedom of choice in the equations f that *describe* C; that is,

$$C = \{(x_1, x_2) \in \mathbb{R}^2 : f(x_1, x_2) = 0\}.$$

In the case of a line, we always have a linear equation

$$f(x_1, x_2) = a_1x_1 + a_2x_2 + b, \quad \text{with } (a_1, a_2) \neq (0, 0),$$

but every $g = c \cdot f^k$ with $c \in \mathbb{R}^*$ and $k \in \mathbb{N}^*$ obviously describes the same line. In Section 1.6 we will study carefully what other equations there can be.

0.2. The *circle* C with center (z_1, z_2) and radius r has an equation

$$(x_1 - z_1)^2 + (x_2 - z_2)^2 = r^2$$

and a transcendental parametrization

$$\varphi(t) = (z_1 + r\cos t,\ z_2 + r\sin t).$$

There is also a *rational parametrization*, which we construct for the case $(z_1, z_2) = (0, 0)$ and $r = 1$. To do this, we project the circle from the point $p = (0, 1)$ onto the line $x_2 = 0$. It is easy to check that under this projection the point

$$(\varphi_1(t), \varphi_2(t)) = \left(\frac{2t}{t^2+1}, \frac{t^2-1}{t^2+1}\right)$$

is mapped to $(t, 0)$. This results in the parametrization

$$\varphi : \mathbb{R} \to C \setminus \{p\} \subset \mathbb{R}^2, \quad t \mapsto (\varphi_1(t), \varphi_2(t))$$

of the punctured circle; see Figure 0.1. If we adjoin an infinitely distant point, or "point at infinity," ∞ to $\mathbb{R}$, it makes sense to extend φ by setting $\varphi(\infty) = p$. In Chapter 2 we discuss how crucial such points at infinity are.

Rational parametrizations of arbitrary conic sections (ellipses, hyperbolas, parabolas) can be obtained in exactly the same way.

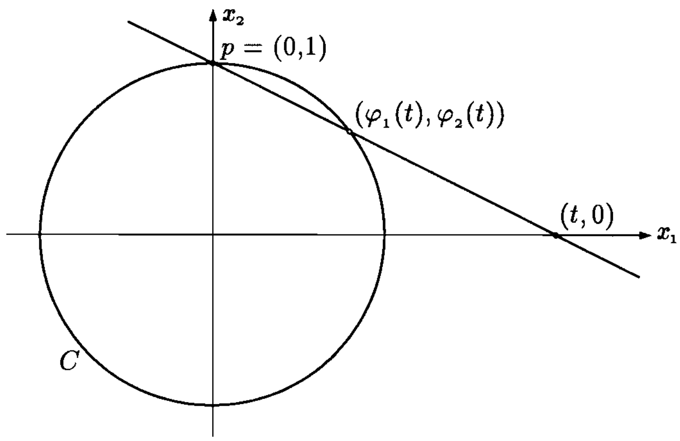

Figure 0.1. Rational parametrization of the circle

0.3. The *cuspidal cubic* (or *Neil's parabola*) $C \subset \mathbb{R}^2$ is given by the parametrization

$$\varphi(t) = (t^2, t^3)$$

and has the equation

$$x_1^3 - x_2^2 = 0.$$

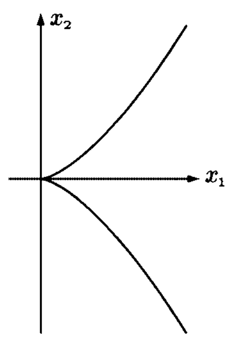

Figure 0.2. Cuspidal cubic

This is a polynomial of degree three, so the curve is called a *cubic*. The tangent vector is given by

$$\dot{\varphi}(t) = (2t, 3t^2), \quad \text{so} \quad \dot{\varphi}(0) = (0, 0).$$

At time $t = 0$ the velocity with which C is traced reverses direction, and its magnitude is zero. It can be shown that $\dot{\psi}(0) = (0, 0)$ for any differentiable parametrization

$$\psi : \mathbb{R} \to C \subset \mathbb{R}^2 \quad \text{with} \quad \psi(0) = (0, 0).$$

For sufficiently differentiable ψ_i, this follows easily from $\psi_1^3 = \psi_2^2$. It takes more work if ψ can be differentiated only once. This phenomenon can occur only at a *singular* point; the cusp of the cuspidal cubic is the simplest and most important example of a *singularity*.

0.4. Newton's *nodal cubic* is given by

$$C = \{(x_1, x_2) \in \mathbb{R}^2 : x_2^2 = x_1^2(x_1 + 1)\}.$$

To obtain a picture of the curve, it is useful to determine the points of intersection of C with the lines $x_1 = \lambda$. For $\lambda < -1$ there are none, for $\lambda = -1$ and $\lambda = 0$ there is one, and for all other λ there are two, with the square roots of $\lambda^3 + \lambda^2$ as abscissas.

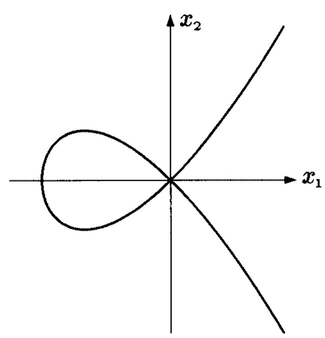

Figure 0.3. Nodal cubic

A rational parametrization

$$\varphi : \mathbb{R} \to C, \quad t \mapsto (t^2 - 1, t - t^3)$$

can be obtained by projecting the curve from the origin to the line $x_1 = -1$. Under this projection $\varphi(1) = \varphi(-1) = (0, 0)$. The origin is an *ordinary double point*; around it the curve has two *branches*, which correspond to the distinct values ± 1 of the parameter t.

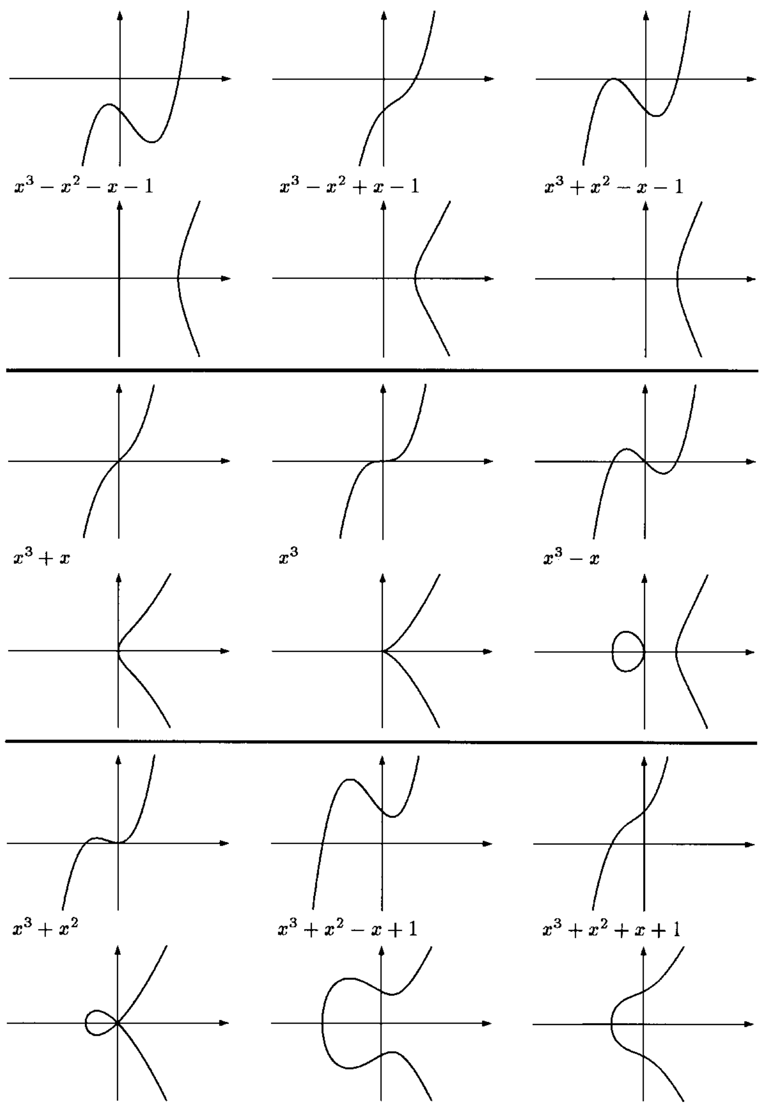

Figure 0.4. Newton's *diverging parabolas*: The curves $y = g(x)$ and $y^2 = g(x)$

In Newton's classification of cubic curves [Ne], as it was published in 1710, both the nodal cubic and the cuspidal cubic belong to the family of "diverging parabolas." These are defined in general by an equation of the form $x_2^2 = g(x_1)$, where g is a cubic polynomial. Some examples of curves $x_2 = g(x_1)$ and $x_2^2 = g(x_1)$ can be seen in Figure 0.4. There $x = x_1$ and $y = x_2$.

0.5. The *folium of Descartes* (named after R. Descartes) looks similar to the nodal cubic but, according to Newton, belongs to the family of "defective hyperbolas." The usual equation is

$$x_1^3 + x_2^3 - 3x_1x_2 = 0.$$

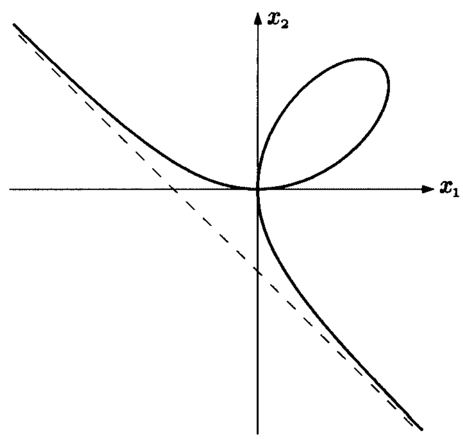

Figure 0.5. Folium of Descartes

The essential difference between this and the nodal cubic is the existence of an asymptote, which has the equation

$$x_1 + x_2 + 1 = 0.$$

If we rotate so that the axis of symmetry becomes $x_2 = 0$, then shift the asymptote to $x_1 = 0$, the folium of Descartes has an equation of the form

$$x_1x_2^2 = g(x_1),$$

where g is a cubic polynomial. According to Newton, this is characteristic of the defective hyperbolas. Newton's list of cubics contains

72 "species." Later it was completed and, by switching to a coarser equivalence relation (complex projective instead of real affine—see Chapter 2), considerably simplified. Once this switch is made, the equation of any smooth cubic can be brought into Hesse normal form, which contains one complex parameter (see [B-K]).

The passage from quadrics to cubics already indicates that as the degrees of the equations increase, the classification problem becomes more and more difficult, and soon becomes hopeless. From degree 4 on, the list of examples can only be sporadic.

0.6. The path traced by the valve on a bicycle tire is an example of a *cycloid*. It can be parametrized by

$$x_1 = t - \sin t,\ x_2 = 1 - \cos t.$$

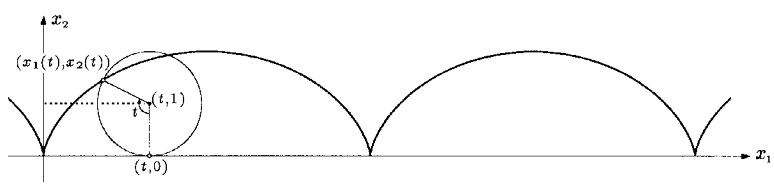

Figure 0.6. Cycloid

Since it meets the line $x_2 = 0$ in infinitely many points, it cannot be described by a polynomial (see Section 1.7). If a circle of radius r is permitted to roll along the inside of a circle of radius $R > r$, the path traced by a point on the inner circle is called a *hypocycloid*. It is closed when r/R is rational. If, say, $R = 1$ and $r = 1/3$, then the center of the small circle has coordinates $z = \frac{2}{3}(\cos t, \sin t)$ and the moving point is

$$p = (x_1, x_2) = z + \frac{1}{3}(\cos 2t, -\sin 2t) = \frac{1}{3}(2\cos t + \cos 2t, 2\sin t - \sin 2t).$$

Applying a few trigonometric identities gives

$$3(x_1^2 + x_2^2)^2 + 8x_1(3x_2^2 - x_1^2) + 6(x_1^2 + x_2^2) = 1$$

as the equation of the *hypocycloid of three cusps*. This polynomial has degree four, so it is called a *quartic*.

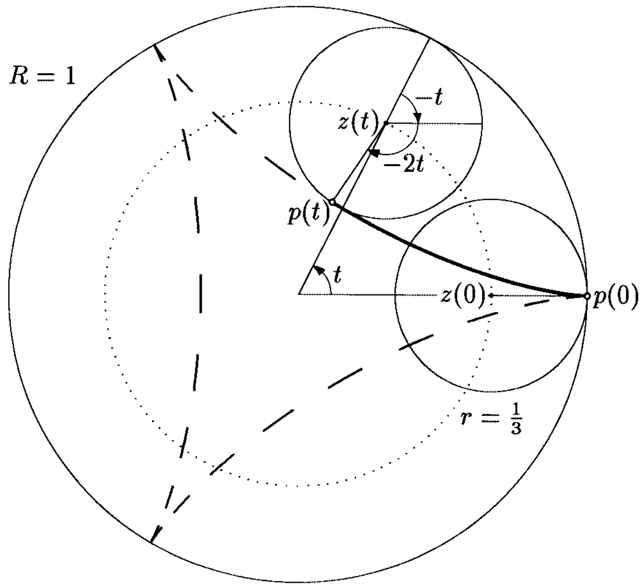

Figure 0.7. Constructing the hypocycloid of three cusps

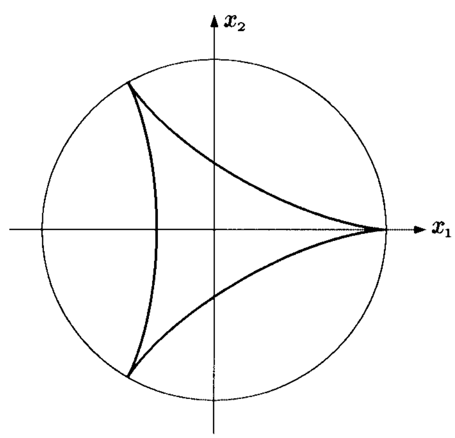

Figure 0.8. Hypocycloid of three cusps

A *rational parametrization* can be obtained by setting $\tau = \tan(t/2)$. Then

$$(\cos t, \sin t) = \left(\frac{1-\tau^2}{1+\tau^2}, \frac{2\tau}{1+\tau^2}\right)$$

(see Section 0.2). Hence

$$p = (x_1, x_2) = \frac{1}{3(1+\tau^2)^2}(3 - 6\tau^2 - \tau^4, 8\tau^3).$$

When $t = \pi$, we have $p = (-1/3, 0)$; the corresponding parameter is $\tan(t/2) = \infty$.

Example (d) of Section 5.1 gives a more elegant way of using the relationship between the circle and the hypocycloid.

If the ratio of the radii is irrational, then the cusps of the hypocycloid are dense in the outer circle. This is an immediate consequence of the following theorem.

Kronecker's Theorem. *Let $\alpha \in \mathbb{R}$ be irrational, and let $\xi \in \mathbb{R}$ be arbitrary. Then for every $\varepsilon > 0$ there exist integers n and p such that*

$$|n\alpha - \xi - p| < \varepsilon.$$

In short: the multiples $n \cdot \alpha$ are dense mod 1 (see [Cha], VIII).

0.7. Felix Klein constructed an interesting family of quartics as follows: Start with two ellipses C_1, C_2, with equations

$$f_1 = x_1^2 + \tfrac{1}{4}x_2^2 - 1 = 0,$$
$$f_2 = \tfrac{1}{4}x_1^2 + x_2^2 - 1 = 0.$$

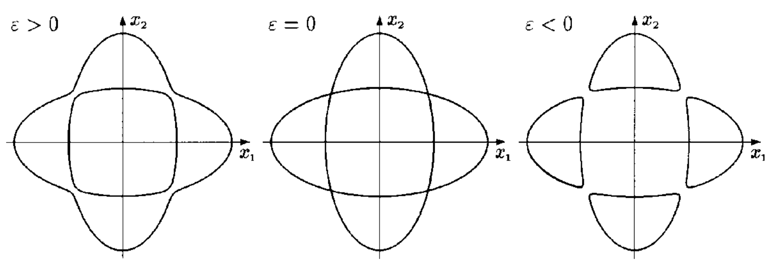

Figure 0.9. Three members of Felix Klein's family of quartics

The equation $f_1 \cdot f_2 = 0$ describes the curve $C_0 = C_1 \cup C_2$. For small real ε, let C_ε be the curve described by $f_1 \cdot f_2 = \varepsilon$. If we consider the signs of the functions f_1, f_2, and $f_1 \cdot f_2$, we get an idea of how C_ε

looks: for $\varepsilon < 0$, the curve consists of four kidney-shaped pieces; for $\varepsilon > 0$, it splits into two belts.

The kidney-shaped quartic is remarkable because (in contrast to quadrics and cubics) it has *bitangents*, which have two points of tangency with the curve. A careful count gives 28 of them.

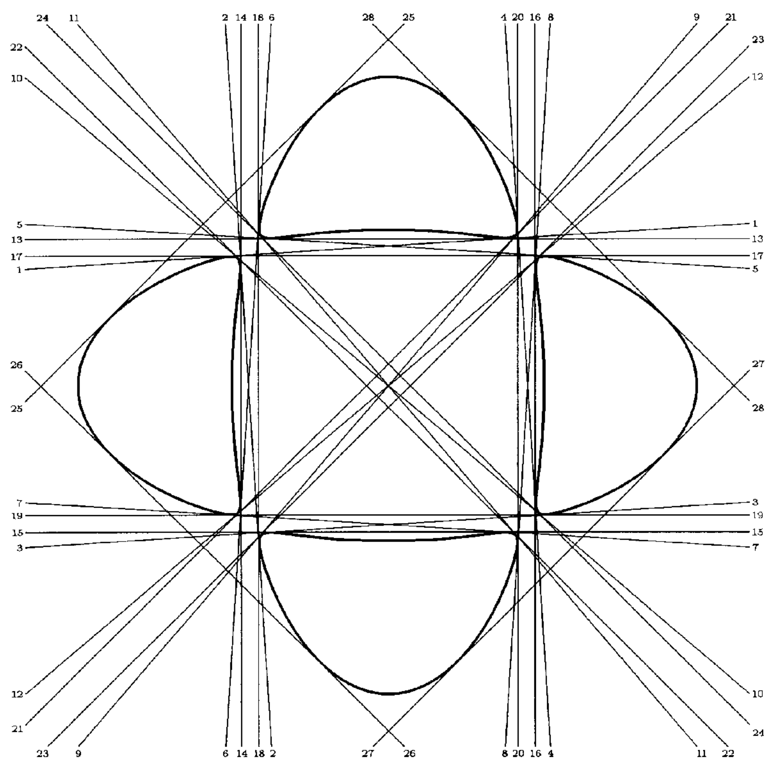

Figure 0.10. The 28 bitangents to the kidney-shaped quartic

For $\varepsilon > 0$, on the other hand, C_ε has only four real bitangents.

0.8. There is a good reason why almost all the curves introduced so far have had polynomial equations. You can already see from the rationality condition for hypocycloids how rare this is for curves parametrized in an elementary way.

Things can become quite pathological when the curve admits only a continuous parametrization. One example is the *Peano curve*, a continuous surjective map

$$\varphi : I \to I \times I, \quad \text{where} \quad I = [0, 1].$$

φ is constructed as the uniform limit of piecewise linear maps. In 1890 Hilbert, in Bremen, illustrated it to the Association of German Natural Scientists and Physicians as follows:

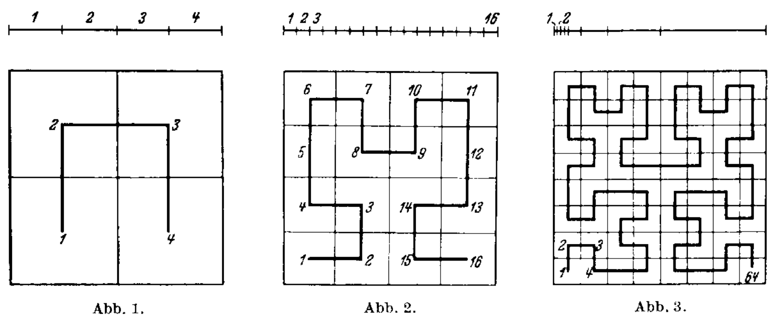

Figure 0.11. Peano curve

The trace of the curve in this case is the whole square, so just looking at the trace gives absolutely no idea how the "curve" was formed.

In constructing the *snowflake curve* (see Figure 0.12), we can think of a geographer who wants to draw the coastline of Brittany with greater and greater accuracy.

We start with an equilateral triangle and, at each stage, attach a triangle with sides of length $a/3$ to each existing side of length a.

The length of the curve increases at each step by the factor $4/3$. The uniform limit of this sequence is a continuous curve that is nowhere rectifiable; we can no longer write an equation for its trace.

The last examples should show above all that whoever is interested in particular regularity properties cannot avoid restricting the class of curves considered. The existence of a polynomial equation is a very rigid condition. But in this case we can expect more precise statements about (for instance) possible singularities, inflection

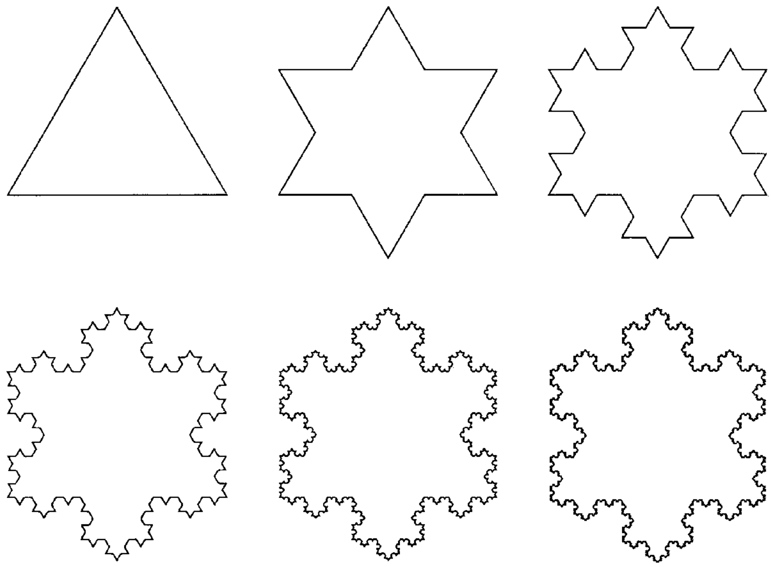

Figure 0.12. Snowflake curve

points, bitangents, and relations among their numbers. Finding these precise statements is the goal of the following chapters.

Exercise. Investigate the symmetry group of the kidney-shaped quartic of Section 0.7 and its action on the 28 bitangents.