Translations of

MATHEMATICAL MONOGRAPHS

Volume 166

An Introduction to Algebraic Geometry

Kenji Ueno

Translated by
Katsumi Nomizu

American Mathematical Society
Providence, Rhode Island

代数幾何入門

DAISŪ KIKA NYŪMON
(An introduction to algebraic geometry)
by Kenji Ueno

Originally published in Japanese by Iwanami Shoten, Publishers, Tokyo, 1995
Translated from the Japanese by Katsumi Nomizu

1991 *Mathematics Subject Classification.* Primary 14–01;
Secondary 14B05, 14C40, 14G10, 14H52, 14H55, 14K25.

ABSTRACT. This book offers an invitation to algebraic geometry to students at an early stage and introduces them to the subject with as few prerequisites as possible. A historical and intuitive treatment explains the spirit of algebraic geometry with numerous examples.

Library of Congress Cataloging-in-Publication Data

Ueno, Kenji, 1945–
[Daisū kika nyūmon, English]
An introduction to algebraic geometry / Kenji Ueno ; translated by Katsumi Nomizu.
p. cm. — (Translations of mathematical monographs ; v. 166)
Includes bibliographical references and index.
ISBN 0-8218-0589-4 (alk. paper)
1. Geometry, Algebraic. I. Title. II. Series
QA564.U3713 1997
516.3′5—dc21 97-3030
CIP

---

10 9 8 7 6 5 4 3 2 1 02 01 00 99 98 97

Contents

Preface to the English Edition

Today algebraic geometry plays an important role in several branches of science and technology. The present book is written for non-specialists to explain the main ideas of algebraic geometry. Fortunately, in Japan the original Japanese edition has been widely accepted as an introductory book for algebraic geometry. I hope the present English edition will serve the same role.

My special thanks are due to Professor Nomizu, who not only translated the Japanese edition into English but also suggested improvements of several parts of the text so that the present English edition is more readable than the original Japanese edition. I also express my sincere thanks to Mr. I. Sasaki of Iwanami Shoten, Publishers and my colleague Dr. Y. Shimizu, who found mistakes and misprints in the original Japanese edition.

December 1996

Kenji Ueno

Translator's Note

Professor Alan Landman's sugggestions were highly helpful in getting started in the translation of the book.

Preface

This book offers an invitation to algebraic geometry. Algebraic geometry, as the geometry of figures defined by a number of equations, came into existence when coordinate geometry was introduced by Descartes and Fermat. In the 18th and 19th centuries, mathematicians working on problems in coordinate geometry with the aid of geometric intuition encountered various paradoxes, which necessitated a rigorous development of the theory. In the first half of this century, Zariski, who wanted to put the algebraic geometry of the Italian School on a firm ground, and Weil, who wanted applications to number theory, developed a rigorous foundation of modern algebraic geometry. The theory has made remarkable progress. In particular, the theory of schemes of Grothendieck pushed the algebraic treatment of geometry to its limit, thus making a great contribution.

It was thought for a long time that algebraic geometry was a purely mathematical theory without any applications. Today, however, its relationships to various branches of natural sciences and engineering have been revealed. Weil's algebraic geometry over finite fields has been applied by Goppa to coding theory. In theoretical physics, close relationships between soliton theory and the theory of algebraic curves as well as between string theory and the moduli theory of algebraic curves have been found. Furthermore, elliptic curves now play an important role in testing for prime numbers. They were also essential in Wiles' recent proof of Fermat's Last Theorem. All this seems to indicate that modern mathematics, having achieved a high level of sophistication for theoretical needs, has now reached the level of maturity to make varied applications possible.

Now it is often said that studying algebraic geometry is rather difficult. This may be partly due to the fact that even in an introductory book we run into ideals, sheaves and cohomology, making it difficult to understand what algebraic geometry is all about. Of course, to develop algebraic geometry properly, there are many prerequisites. The need for a variety of preparations comes, on the one hand, from the way algebraic geometry was developed — needing rigorous constructions of the theory to avoid paradoxes arising from naive intuition — and, on the other hand, from the fact that algebraic geometers developed the theory by avidly using whatever tools they needed from many branches of mathematics.

In this book we have tried to develop algebraic geometry with minimal prerequisites. In Chapter 1, we give an introduction to algebraic geometry from a historical point of view. Many of the concepts that appear are refined in Chapter 2. This is because we want the readers to acquire intuitive understanding from the beginning so that they know why certain arguments are necessary and understand the significance of the theory. We go through Chapters 3 and 4 in a hurry compared to Chapter 2. We try to explain by examples the meaning of theorems

and to derive further results rather than to prove theorems. We hope that readers will taste various aspects of algebraic geometry and will be better prepared when they get to work with more advanced books on algebraic geometry. Furthermore, in the Appendix, we explain elementary facts in the theory of commutative rings and fields that are needed in algebraic geometry. We give a lot of detail so that the Appendix can be useful as an introduction and readers will get used to abstract thinking. Throughout the book we often quote previous theorems, lemmas, and examples. To help readers in locating them we provide an index for these items.

A particular feature of this book is that we dare to repeat explanations. The reader will notice that one and the same object comes up in different forms and from different viewpoints. We also give as many concrete examples as possible. In reading a book in mathematics it is necessary to find concrete examples, verify abstract results, and make them your own. Many books avoid repetition as much as possible and expect the readers to read between the lines on their own. The author's experience in teaching at colleges, however, indicates that the number of students who do not understand the importance of reading between the lines on their own is increasing. This phenomenon is probably due to the overemphasis on success in entrance examinations at the expense of "the importance of thinking for oneself."

In this book we have tried to examine many examples from various viewpoints. While doing so we often need concrete computations. We recommend that readers have paper and pencil ready and verify the results by their own computations. As was said in ancient Greece, "there is no royal road to geometry." The royal road in mathematics is to compute on your own until you convince yourself. When you feel that this book gives too much detail and repeats the same thing too often, you will have graduated from it. That is the time when you should begin to work with an authoritative systematic introduction to algebraic geometry (see the references).

This book is based on a draft for the Iwanami Lecture Series in Applied Mathematics. I wrote what I wanted to write about algebraic geometry without worrying about space and without assuming much background. For the Lecture Series, it had to be cut down to about one half. Mr. Hisao Miyauchi of Iwanami Shoten, Publishers, proposed that the original draft be published as a monograph. We thank Mr. Miyauchi, Ms. A. Hamakado, and U. Yoshida of the Editorial Department. Much valuable advice on misprints, errors, and displays was received from the production staff, whose help is greatly appreciated.

December 1994

Kenji Ueno

CHAPTER 1

Invitation to Algebraic Geometry

We might say that algebraic geometry was born at the same time as coordinate geometry. The chief object of algebraic geometry is certainly the study of figures given by equations. But it took a long time before algebraic geometry became a definitive branch of mathematics. The study of figures often depended too much on intuition, which sometimes led to strange paradoxes. To study the theory rigorously without using intuition, one had to wait for the development of mathematical tools. It is said that the algebraic geometry of today is sometimes difficult to follow — that is because we are preoccupied with rigor and forget the underlying geometric intuition.

In this chapter we shall review the history surrounding the birth of algebraic geometry. We wish to show how algebraic geometry gradually emerges as we pass from Euclidean geometry to projective geometry. The theory of algebraic curves will be developed from different perspectives in subsequent chapters. We have omitted some computations; in this chapter we would rather have the readers proceed without necessarily carrying out such computations.

§ 1.1. The birth of geometry

(a) Euclidean geometry. Geometry was developed in various ancient civilizations. It was in ancient Greece that geometry appeared in a complete form that can be regarded as the starting point of today's mathematics. Actually, geometry was developed as a practical science useful for surveying in other old civilizations, but in Greece it was also developed as an object of purely intellectual pursuit. Such developments by the school of Pythagoras had interactions with geometric views on numbers that were related to religious and philosophical ideas. By around 300 B.C. the geometry of Euclid was presented in its complete form in the book *The Elements*. It is not known whether Euclid, supposed to be author of the treatise, actually lived or not. It is certain that many mathematicians participated in its development. The treatise includes the theory of ratios and a geometric treatment of numbers, with geometry itself as the main theme. In geometry, it started from a small number of postulates and logically deduced the properties of lines, triangles, circles, etc., and was regarded as the standard model for a systematic treatise in scholarly writing for a long time to come.

Although the axiomatic system for Euclidean geometry was reorganized by Hilbert and others around the turn of the 20th century, it is amazing that as early as the third and fourth centuries B.C. the method was established to deduce underlying properties of geometric objects. We could say that mathematics itself became an independent science with *The Elements*.

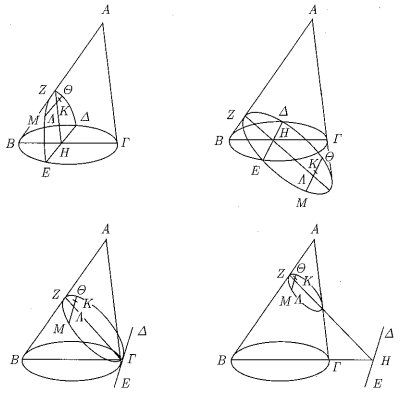


FIGURE 1.1 CONICS OF APOLLONIUS

(b) The theory of conics of Apollonius. The so-called Euclidean geometry treated in *The Elements* was concerned with lines and circles; in other words, it was the geometry of figures that can be drawn with ruler and compass. Although these objects led to interesting geometry, they were not sufficient, as was recognized in ancient Greece. Various other curves were studied, but the methods of describing curves were limited. In particular, in ancient Greece algebra was undeveloped — rather, algebraic problems were solved by using geometry.

Although it has been said that Euclid himself wrote the theory of conics, the *Theory of Conics* written by Apollonius (who was active around 200 B.C.) has been partially preserved. In the book he introduced ellipses, hyperbolas, and parabolas as sections of a cone and studied their properties in detail. (See Figure 1.1.)

§ 1.2. Coordinate geometry

(a) The birth of coordinate geometry. Coordinate geometry was introduced independently by Fermat (1601-65) and by Descartes (1596-1650). Fermat's theory was established before 1629 but was published in 1679 after his death. Descartes published *Geometry* in 1637 as an appendix to *Discourse on the Method* and proposed coordinate geometry. He used cartesian (rectangular) coordinates, but Fermat had a more advanced idea of coordinates (e.g., those that are not rectangular).

The invention of coordinate geometry by Fermat and Descartes was possible because algebra, imported from Arabia, was developed in Europe in the Middle Ages, and algebraic symbols were freely used. Descartes emphasized unified solutions of problems in Euclidean geometry by reducing them to algebraic problems by means of coordinate geometry. Further, the introduction of coordinate geometry formulated geometric objects by algebraic descriptions, thus leading to algebraic

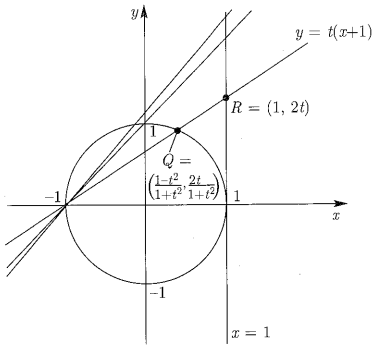


FIGURE 1.2 THE CORRESPONDENCE BETWEEN THE UNIT CIRCLE AND THE LINE

geometry. As a matter of fact, Descartes himself states that conics are nothing but plane quadrics and that quadrics and lines are essentially the same geometric object. Since this point of view of Descartes is important, we shall give some details. For simplicity, let us examine the relationship between the unit circle $C : x^2 + y^2 = 1$ and the line $\ell : x = 1$. (See Figure 1.2.)

For $P = (-1, 0)$ on the unit circle and $Q = (x_0, y_0)$, let $R = (1, y_1)$ be the intersection of the line $\overline{PQ}$ with the line ℓ. We see that there is a one-to-one correspondence between the points other than P of the unit circle C and the points of the line ℓ. Furthermore, as the point Q approaches P from the upper half of the circle, the corresponding point R on ℓ goes upward on ℓ indefinitely. If Q approaches P from the lower half, the point R goes away downward on ℓ.

Let us verify what we said by using equations. Let us represent the line $\overline{PQ}$ by the equation

$$y = t(x + 1), \tag{1.1}$$

with slope t as the parameter. Then the points Q and R are given by

$$Q = \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right), \quad R = (1, 2t). \tag{1.2}$$

This parametrization plays an important role later on. As is immediate from (1.2), for a point $R = (1, s)$ on the line ℓ the intersection Q of the line $\overline{PR}$ and the unit circle C is given by

$$Q = \left(\frac{1 - (s/2)^2}{1 + (s/2)^2}, \frac{s}{1 + (s/2)^2}\right).$$

Conversely, for any point $Q = (x_0, y_0)$ on the unit circle C other than P, the intersection R of the line $\overline{PQ}$ and the line ℓ is given by

$$R = \left(1, \frac{2y_0}{x_0 + 1}\right).$$

It is important to realize that the correspondence between Q and R is given by rational functions of coordinates. That is, the one-to-one correspondence between $C - \{P\}$ and ℓ is expressed in concrete form:

$$\ell \to C - \{P\} \qquad \text{with} \qquad (1, s) \mapsto \left(\frac{1 - (s/2)^2}{1 + (s/2)^2}, \frac{s}{1 + (s/2)^2}\right)$$

and

$$C - \{P\} \to \ell \qquad \text{with} \qquad (x_0, y_0) \mapsto \left(1, \frac{2y_0}{x_0 + 1}\right).$$

How should we think of the excluded point P? As Q approaches P, we may regard the line $\overline{PQ}$ as approaching the tangent $m : x = -1$ of the circle C at P. This corresponds to taking the limit in (1.1) as $t \to \pm\infty$ (by rewriting (1.1) in the form $\frac{1}{t}y = x + 1$). As $t \to \pm\infty$, we have

$$\lim_{t \to \pm\infty} \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right) = (-1, 0), \qquad \text{that is, } Q \to P,$$

and

$$\lim_{t \to \pm\infty} (1, 2t) = (1, \pm\infty);$$

in other words, R goes away upward (for $t \to \infty$) or downward (for $t \to -\infty$). This is obvious because the tangent m and the line ℓ are parallel. It is also possible, however, to imagine a "point at infinity" at the infinite far end of the line ℓ which we can reach by going upward or downward on ℓ "all the way." In this fashion, what we obtain by adjoining this point at infinity to the line ℓ can be identified with the unit circle C. In the next section, we shall show how this is realized in projective geometry. A similar correspondence can be constructed for any circle or ellipse.

What about a hyperbola or parabola? We illustrate the results in Figure 1.3.

We leave the details to the reader. For the hyperbola C, there is a one-to-one correspondence between $C - \{P\}$ and the y-axis with the points $(0, \pm 1)$ deleted. For the parabola C given by $y = x^2 + 1$, there is a one-to-one correspondence between $C - \{P\}$ and the x-axis with the origin deleted. We shall later explain clearly why we need to delete two points or one point from a line from the point of view of projective geometry. Until then, we leave the question for the reader to think about.

(b) Euclidean geometry and affine geometry. By introducing rectangular coordinates and expressing geometric figures by equations, it becomes possible to clearly grasp the essential character of Euclidean geometry. The notion of congruence plays an important role in Euclidean geometry. Intuitively, two figures on the plane are congruent if we can superimpose one figure upon the other. When we think about the operation of superimposing, we realize that it consists of the following three operations:

(1) To translate a figure without rotating; in terms of coordinates, this is expressed by

$$(x, y) \mapsto (x + a, y + b).$$

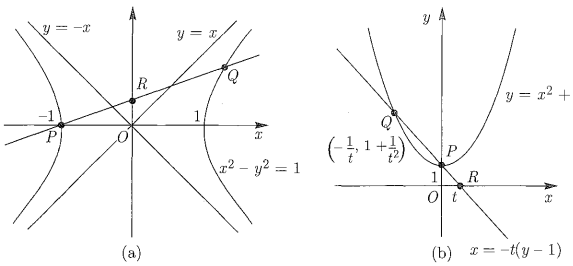


FIGURE 1.3 CORRESPONDENCE BETWEEN THE HYPERBOLA, THE PARABOLA AND THE LINE. (a) Let $P = (-1, 0)$, which is on the hyperbola $C : x^2 - y^2 = 1$. For any other point Q on the hyperbola, let R be the intersection of the line $\overline{PQ}$ with the y-axis. Expressing $\overline{PQ}$ by $y = t(x+1)$, we have $Q = ((1+t^2)/(1-t^2), 2t/(1-t^2))$, $R = (0, t)$, except that Q does not exist for $t = \pm 1$. (b) Let $P = (0, 1)$, which is on the parabola $C : y = x^2 + 1$. The line joining P and another point Q on the parabola meets the x-axis at R. If we express $\overline{PQ}$ by $x = -t(y-1)$, then $Q = (-1/t, 1+t^2)$, $R = (t, 0)$, except that Q does not exist for $t = 0$.

(2) To rotate with a given point as center; for example, a counterclockwise rotation by angle θ around the origin is expressed by

$$(x, y) \mapsto ((\cos\theta)x - (\sin\theta)y, (\sin\theta)x + (\cos\theta)y),$$

in terms of rectangular coordinates.

(3) To reflect about a given line ℓ; for example, the reflection about the y-axis can be represented by

$$(x, y) \mapsto (-x, y).$$

By repeating these types of operations we get a superimposition (or congruence). Using coordinates, the end result can be expressed by

$$f : (x, y) \mapsto (a_{11}x + a_{12}y + b, a_{21}x + a_{22}y + c), \tag{1.3}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

satisfies

$${}^tAA = I_2 \text{ with } I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.4}$$

An operation (1.3) satisfying (1.4) is called a **congruence transformation** or **Euclidean transformation**. The set $O(2)$ of all 2×2 matrices such that ${}^tAA = I_2$ forms a group called the **orthogonal group** of degree 2. The elements of $O(2)$ are called **orthogonal matrices**.

To a congruence transformation f in (1.3) we associate the 3×3 matrix

$$X = \begin{pmatrix} 1 & 0 & 0 \\ b & a_{11} & a_{12} \\ c & a_{21} & a_{22} \end{pmatrix}$$

and write f_X for f. For

$$Y = \begin{pmatrix} 1 & 0 & 0 \\ b' & a'_{11} & a'_{12} \\ c' & a'_{21} & a'_{22} \end{pmatrix}, \quad A' = \begin{pmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{pmatrix},$$

we denote the corresponding congruence transformation by f_Y. The composition $f_Y \circ f_X$ of the two congruence transformations f_X and f_Y, that is, f_X followed by f_Y, turns out to be the congruence transformation f_{YX} corresponding to the matrix YX. We can directly verify that

$$YX = \begin{pmatrix} 1 & 0 & 0 \\ b'' & a''_{11} & a''_{12} \\ c'' & a''_{21} & a''_{22} \end{pmatrix}, \quad A'' = \begin{pmatrix} a''_{11} & a''_{12} \\ a''_{21} & a''_{22} \end{pmatrix} = A'A \in O(2).$$

We can thus conclude that the set $E(2)$ of all congruence transformations (the congruence transformation group or Euclidean motion group) is identified with the set of all matrices

$$\mathcal{E}(2) = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ b & a_{11} & a_{12} \\ c & a_{21} & a_{22} \end{pmatrix} \mid A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in O(2), b, c \in \mathbf{R} \right\};$$

to be more precise, they are isomorphic as groups. To an element

$$X = \begin{pmatrix} 1 & 0 & 0 \\ b & a_{11} & a_{12} \\ c & a_{21} & a_{22} \end{pmatrix}$$

in $\mathcal{E}(2)$ there corresponds the congruence transformation f_X expressed by (1.3).

Now a congruence transformation maps lines to lines, circles to circles, and preserves angle and length. With this in mind we can say that Euclidean geometry is the geometry in which we study the common properties shared by geometric figures that are congruent, that is, transformed to each other by a congruence transformation. We then ask if there is a geometry that has to do with the transformations of the form

$$g : (x, y) \mapsto (a_{11}x + a_{12}y + b, a_{21}x + a_{22}y + c), \tag{1.5}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \det A \neq 0,$$

which is a weaker condition than (1.4). Since A may not be an orthogonal transformation, neither angle nor length is preserved by g. In fact, we can show that area is multiplied by $|\det A|$. (Exercise 1.1). Although g maps lines to lines, it generally maps circles not to circles but to ellipses, unless $A \in O(2)$. It is true that g maps ellipses to ellipses, hyperbolas to hyperbolas, and parabolas to parabolas, as we now verify.

Since the translation part is not essential in our considerations, we may assume that

$$g : (x, y) \mapsto (a_{11}x + a_{12}y, a_{21}x + a_{22}y), \text{ with } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \det A \neq 0.$$

In matrix notation we have

$$g \begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}.$$

First, consider a line m whose equation is

$$m : \alpha x + \beta y + \gamma = 0. \tag{1.6}$$

We can represent it by

$$(\alpha \quad \beta) \begin{pmatrix} x \\ y \end{pmatrix} = -\gamma. \tag{1.6'}$$

For a point (x_0, y_0) on m, let $g(x_0, y_0) = (x'_0, y'_0)$. Then

$$\begin{pmatrix} x'_0 \\ y'_0 \end{pmatrix} = A \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

By using the inverse A^{-1} we have

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = A^{-1} \begin{pmatrix} x \\ y \end{pmatrix}.$$

On the other hand, from the equation (1.6') of the line we obtain

$$-\gamma = (\alpha, \beta) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = (\alpha, \beta) A^{-1} \begin{pmatrix} x'_0 \\ y'_0 \end{pmatrix}.$$

By setting $(\alpha', \beta') = (\alpha, \beta)A^{-1}$, we see that a point on the line m is mapped to a point on the line

$$m' : \alpha' x + \beta' y + \gamma' = 0. \tag{1.7}$$

It also follows that the line m is mapped onto the line m' by g. It is furthermore important that two parallel lines are mapped to parallel lines. This is obvious because a line parallel to (1.6) is given by $\alpha x + \beta y + \delta = 0, \delta \neq \gamma$, and this line is mapped by g to $\alpha' x + \beta' y + \delta' = 0$.

Next, let us consider a quadratic curve

$$Q : q(x, y) = ax^2 + 2bxy + cy^2 + dx + ey + f = 0. \tag{1.8}$$

We use similar arguments. By setting

$$\begin{pmatrix} x \\ y \end{pmatrix} = A^{-1} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} b_{11}x' + b_{12}y' \\ b_{21}x' + b_{22}y' \end{pmatrix}$$

and

$$\begin{aligned} \tilde{q}(X, Y) &= q(b_{11}x' + b_{12}y', b_{21}x' + b_{22}y') \\ &= \tilde{a}x'^2 + 2\tilde{b}x'y' + \tilde{c}y'^2 + \tilde{d}x' + \tilde{e}y' + \tilde{f}, \end{aligned}$$

we define a new quadratic curve

$$\tilde{Q} : \tilde{q}(x, y) = 0.$$

Then we see that g maps Q to $\tilde{Q}$. By a simple computation we get the following relations between the coefficients of the quadratic terms:

$$\begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{b} & \tilde{c} \end{pmatrix} = {}^tA^{-1} \begin{pmatrix} a & b \\ b & c \end{pmatrix} A^{-1}. \tag{1.9}$$

The important cases are as follows.

(i) The symmetric matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ has two positive (or two negative) eigenvalues, and Q is an ellipse;

(ii) The symmetric matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ has one positive and one negative eigenvalue, and Q is a hyperbola;

(iii) The symmetric matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ has one zero eigenvalue and another nonzero eigenvalue, and Q is a parabola.

These properties of the symmetric matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ are invariant under the transformation (1.9). Hence g maps ellipses to ellipses, hyperbolas to hyperbolas, and parabolas to parabolas. We note that, to get the classification as above, we must exclude several cases such as $d = e = f = 0$. Later, we shall show that in algebraic geometry we can simplify the situation so that the quadratic equation (1.8) defines an irreducible quadratic curve or two lines or a double line.

A transformation given by (1.5) is called an **affine transformation**, and the set $A(2)$ of all affine transformations the **affine transformation group**. The group $A(2)$ can be identified with the set $\mathcal{A}(2)$ of all matrices of the form $\begin{pmatrix} 1 & 0 & 0 \\ b & a_{11} & a_{12} \\ c & a_{21} & a_{22} \end{pmatrix}$ with $\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$ and $b, c \in \mathbf{R}$. For an element

$$X = \begin{pmatrix} 1 & 0 & 0 \\ b & a_{11} & a_{12} \\ c & a_{21} & a_{22} \end{pmatrix}$$

in $\mathcal{A}(2)$ we write f_X for the corresponding affine transformation

$$(x, y) \mapsto (a_{11}x + a_{12}y + b, a_{21}x + a_{22}y + c).$$

Then just as in the case of congruence transformations, the composition of transformations corresponds to the product of matrices. The affine transformation group contains the congruence transformation group.

An affine transformation g does not preserve length or angle, unless $g \in \mathcal{E}(2)$; it maps lines to lines, parallel lines to parallel lines, and maps ellipses, hyperbolas, and parabolas to ellipses, hyperbolas, and parabolas, respectively. More generally, a curve of degree n, that is, a curve $f(x, y) = 0$, where $f(x, y)$ is a polynomial of degree n, is mapped to a curve of degree n. On the other hand, we can see that an affine transformation is more general than a congruence transformation from the fact that any given ellipse, hyperbola or parabola can be mapped to the unit circle:

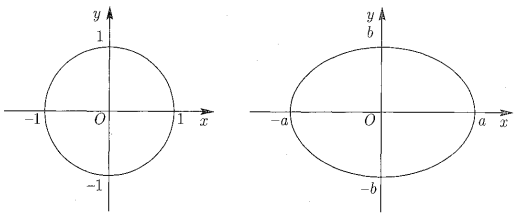


FIGURE 1.4. The affine transformation $(x, y) \mapsto (ax, by)$ takes the unit circle $x^2 + y^2 = 1$ to the ellipse $x^2/a^2 + y^2/b^2 = 1$.

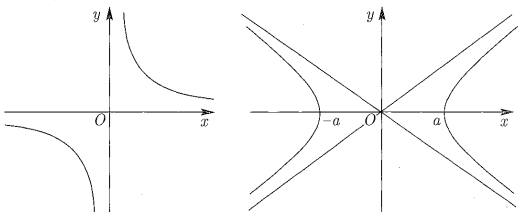


FIGURE 1.5. The transformation $(x, y) \mapsto (a(x+y)/2, b(x-y)/2)$ maps the hyperbola $xy = 1$ to the hyperbola $x^2/a^2 - y^2/b^2 = 1$.

$x^2 + y^2 = 1$, the rectangular hyperbola: $x^2 - y^2 = 1$, or the parabola: $y = x^2$, respectively, by a suitable affine transformation. (See Figures 1.4 and 1.5.)

§ 1.3. Projective geometry

(a) The birth of projective geometry. Projective geometry was introduced by Desargues (1591-1661) and others by necessity in engineering (such as drawings). It was completed as geometry by Monge (1746-1818) and Poncelet (1788-1867).

The fundamental idea for projective geometry is simple and clear. Let us consider two planes H and H' in space. Suppose that there is a light source at a point P outside these planes, and also assume that H is transparent. Then to each point Q on H there corresponds a point Q' on H'. (More precisely, Q' is the point of intersection of the line $\overline{PQ}$ with H'.) In this way, we get a "mapping" $\pi_P : Q \mapsto Q'$ from the plane H to the plane H'. (See Figure 1.6.) A figure α on the plane H is transferred to a figure α'. We easily see that if α is a line, so is α'. If H and H' are parallel, π_P is certainly a mapping, that is, for every point Q, the corresponding point Q' is determined. In this case, π_P takes ellipses, hyperbolas, and parabolas upon ellipses, hyperbolas, and parabolas, respectively.

But in the case where H and H' are not parallel, the situation radically changes (see Figure 1.7). The set of points Q on H such that the line $\overline{PQ}$ is parallel to the plane H' forms a line, say, h. For any point Q on h there is no corresponding point

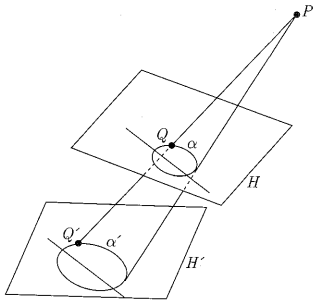


FIGURE 1.6 PROJECTION. From a light source at P outside the planes H and H' we get the image α' on H' for a figure α on H; α' is the image of α by a projective transformation.

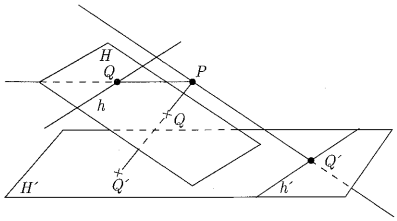


FIGURE 1.7. If the plane spanned by P and a line h on the plane H is parallel to the plane H', any point on h does not have any corresponding point Q'.

Q'; thus π_P is a mapping from $H - h$ to H'. Furthermore, the set of points Q' such that the line $\overline{PQ'}$ is parallel to the plane H forms a line h', and no point on h' can be obtained as projection of a point on H. This means that π_P is a one-to-one mapping from $H - h$ onto $H' - h'$.

This "mapping" π_P has a very interesting property. If an ellipse on H does not intersect the line h, then its image by π_P is also an ellipse. If the ellipse intersects h at two points Q_1 and Q_2, then these two points are not mapped onto the plane H' and the image of the ellipse is a hyperbola on H'. (See Figure 1.8 (a).) If the ellipse is tangent to the line h, then its image by π_P is a parabola (See Figure 1.8 (b).)

Conversely, let us take an ellipse C' on the plane H'. If C' does not intersect the line h', then it is the image by π_P of an ellipse C on H. If C' intersects the line

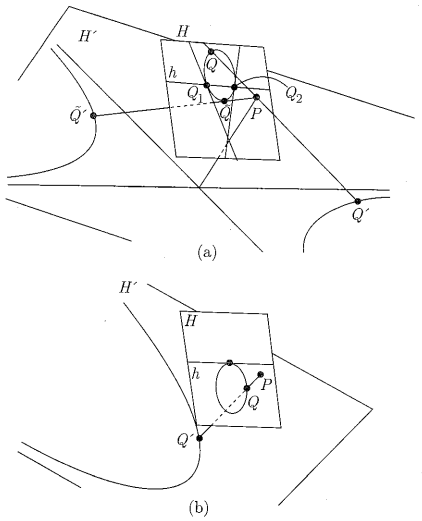


FIGURE 1.8. (a) If an ellipse on H intersects the line h at two points Q_1 and Q_2, its image by π_P is a hyperbola. The images of the tangents to the ellipse at Q_1 and Q_2 are the asymptotes to the hyperbola. (b) If an ellipse on H is tangent to the line h, its image by π_P is a parabola.

h' at two points Q'_1 and Q'_2, then there are no points of H that correspond to Q'_1 and Q'_2, and we see that $C' - \{Q'_1, Q'_2\}$ is the image of a hyperbola on H. Finally, if C' is tangent to h' at a point Q', then $C' - Q'$ is the image of a parabola by π_P.

As we have just seen, there are cases where an ellipse, a hyperbola, and a parabola may be transformed from one to another. There are points that have to be excluded, and this situation makes it difficult to give a unified treatment of these cases. We will try to be more imaginative and add points at infinity to planes. That is, for each point Q on the line h in the plane H, let us imagine Q' to be a point at infinity to be added to the plane H'. Hence all the points corresponding to h are added as points at infinity. We need to add one more point to H' as a point Q on h goes away infinitely far. Thus we are adding to H' the set ℓ'_∞ consisting of all points on the line h and one more point at infinity. We call ℓ'_∞ the line at infinity of the plane H'. To the plane H let us add the line at infinity ℓ_∞. By doing this, we consider a point on h' on H' as the image of a point on the line at infinity ℓ_∞ of the plane H. To be more specific, the line ℓ_∞ includes the point at infinity

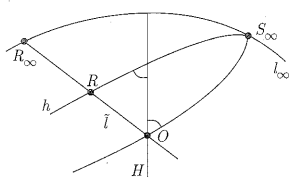


FIGURE 1.9 LINES AT INFINITY OF THE PROJECTIVE PLANE

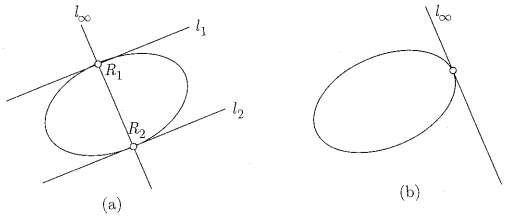


FIGURE 1.10 A HYPERBOLA AND A PARABOLA NEAR A LINE AT INFINITY. (a) A hyperbola meets the line at infinity at two points R_1 and R_2, which correspond to the directions of the asymptotes ℓ_1 and ℓ_2 of the hyperbola. The asymptote ℓ_i is tangent to the hyperbola at R_i. (b) A parabola is tangent to the line at infinity.

corresponding to a point Q' on h' going infinitely far away. In this way, we can see that the mapping π_P is a one-to-one mapping from $H \cup \ell_\infty$ onto $H' \cup \ell'_\infty$.

Let us consider the meaning of $H \cup \ell_\infty$. Take two lines ℓ'_1 and ℓ'_2 that meet at a point Q' on h'. Then they are the images of parallel lines ℓ_1 and ℓ_2 on H. There is no point of H that corresponds to Q', but we have the point Q on the added line ℓ_∞. We may further imagine that every point on ℓ_∞ corresponds to the slope of a line on the plane H. Now take a point O outside the line h. For any line ℓ on H there is a unique line $\tilde{\ell}$ through O that is parallel to ℓ. If $\tilde{\ell}$ is not parallel to h, let R be the intersection with h. The point R_∞ corresponding to the slope of $\tilde{\ell}$ may be considered as the point of intersection of $\tilde{\ell}$ and ℓ_∞. If $\tilde{\ell}$ is parallel to h, $\tilde{\ell}$ does not meet h on H but meets ℓ_∞ at a point S_∞. This is the reason why we added one further point at infinity. All lines m that are parallel to h meet at S_∞. (Figure 1.9)

We call $H \cup \ell_\infty$ a projective plane and often denote it by $\mathbf{P}^2(\mathbf{R})$. In the projective plane, we no longer have parallel lines and any two lines meet at a point. Unlike an ordinary line, a line in the projective plane is closed. That is, since a point at infinity has been added to an ordinary line, by proceeding in the positive or negative direction, we reach the same point at infinity. Strange as it may sound, ellipses, hyperbolas, and parabolas lose their distinction in the projective plane. (Refer to Figure 1.10; we discuss details in the next section.) Furthermore, the

mapping π_P will be a natural mapping from the projective plane $H \cup \ell_\infty$ onto the projective plane $H' \cup \ell'_\infty$ without having to exclude any points. This mapping is a special case of a projective transformation which we shall discuss in the next section.

(b) The projective plane. Projective geometry can be developed starting with a set of axioms, just like Euclidean geometry. Here, however, we shall discuss a coordinate-geometric treatment of projective geometry. Ignoring the historic development of projective geometry, we discuss homogeneous coordinates – the coordinates on which we build projective geometry. We consider only the ratios of coordinates (x_0, x_1, x_2) and write $(x_0 : x_1 : x_2)$ excluding $(x_0, x_1, x_2) = (0, 0, 0)$. (For a precise definition of ratio, see Definition 1.1.)

The relationship to the rectangular coordinates of the ordinary plane can be given by

$$x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}. \tag{1.10}$$

Sometimes, we call (x, y) the inhomogeneous coordinates, which make sense only when $x_0 \neq 0$. Let's take an equation for an ordinary line

$$\alpha x + \beta y + \gamma = 0. \tag{1.11}$$

Using (1.10) we rewrite (1.11) in the form

$$\alpha x_1 + \beta x_2 + \gamma x_0 = 0. \tag{1.12}$$

We can consider the case where $x_0 = 0$, and the solution of (1.12) is given by

$$(0 : \beta : -\alpha).$$

(Note that we consider the ratios only.) This solution represents a point at infinity that does not allow a representation in the form (1.10). Furthermore, the point at infinity is determined by the slope of the line (1.11) and does not depend on γ. That is, two parallel lines

$$\begin{aligned} \alpha x + \beta y + \gamma &= 0 \\ \alpha x + \beta y + \delta &= 0 \end{aligned}$$

can be converted to the form (1.12)

$$\begin{aligned} \alpha x_1 + \beta x_2 + \gamma x_0 &= 0 \\ \alpha x_1 + \beta x_2 + \delta x_0 &= 0, \end{aligned}$$

which pass through the point $(0 : \beta : -\alpha)$, that is, two parallel lines intersect at the point at infinity $(0, \beta : -\alpha)$.

Based on these observations, we shall now define the projective plane $\mathbf{P}^2(\mathbf{R})$. First we make certain what we mean by ratios in general. In the set of all $(n+1)$-tuples $(a_0, a_1, \ldots, a_n) \neq (0, 0, \ldots, 0)$, we introduce an equivalence relation $\sim$ as follows. We say that

$$(a_0, a_1, \ldots, a_n) \sim (b_0, b_1, \ldots, b_n)$$

if there is a real number $\lambda \neq 0$ such that

$$(b_0, b_1, \ldots, b_n) = \lambda(a_0, a_1, \ldots, a_n),$$

that is,

$$b_i = \lambda a_i \text{ for } 0 \le i \le n.$$

The ratio $(a_0 : a_1 : \cdots a_n)$ is, by definition, an equivalence class of $(a_0, a_1, \ldots, a_n) \neq (0, 0, \ldots, 0)$.

DEFINITION 1.1. The set of all ratios $(a_0 : a_1 : a_2)$ is called the projective plane and denoted by $\mathbf{P}^2(\mathbf{R})$. Each $(a_0 : a_1 : a_2)$ is called a point of $\mathbf{P}^2(\mathbf{R})$.

REMARK. Later in §1.4 we extend the notion of ratio $(a_0 : a_1 : \cdots a_n)$ when $a_0, a_1, \cdots : a_n$ are complex numbers by allowing λ to be a nonzero complex number in the definition of the equivalence relation above.

We note that if λ is a nonzero real number, (a_0, a_1, a_2) and $(\lambda a_0, \lambda a_1, \lambda a_2)$ represent the same point, that is,

$$(a_0 : a_1 : a_2) = (\lambda a_0 : \lambda a_1 : \lambda a_2), \quad \lambda \neq 0.$$

In particular, if $a_0 \neq 0$, then we have

$$(a_0 : a_1 : a_2) = (1 : \frac{a_1}{a_0} : \frac{a_2}{a_0}).$$

If we set

$$U_0 = \{(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{R}) | a_0 \neq 0\},$$

then we can define a mapping ϕ_0 from U_0 to the (x, y)-plane H by

$$\phi_0 : (a_0, a_1 : a_2) \in U_0 \longmapsto (\frac{a_1}{a_0}, \frac{a_2}{a_0}) \in H.$$

Conversely, define a mapping ψ_0 from the plane H to U_0 by

$$\psi_0 : (x_0, y_0) \in H \longmapsto (1 : x_0, y_0) \in U_0;$$

then ϕ_0 and ψ_0 are inverse to each other. In this way, we can identify U_0 and the (x, y)-plane H. This explains the meaning of (1.10).

We now set

$$\ell_\infty = \mathbf{P}^2(\mathbf{R}) - U_0 = \{(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{R}) | a_0 = 0\}.$$

In terms of homogeneous coordinates, ℓ_∞ is defined by

$$x_0 = 0$$

and is called the **line at infinity**. As we have seen before, the point $R = (0 : a : b)$ is a point at infinity of the line $bx - ay + c = 0$ in H, that is, the point where all lines parallel to the line $bx - ay + c = 0$ meet.

REMARK. It is also convenient to include the lines $x_1 = 0$ and $x_2 = 0$ as lines at infinity. In fact, given any line ℓ in $\mathbf{P}^2(\mathbf{R})$, we may think of it as a line at infinity, since there is a projective transformation f such that $f(\ell)$ is ℓ_∞.

Just as Euclidean geometry and affine geometry study properties invariant under congruence transformations and affine transformations, respectively, projective geometry studies properties invariant under projective transformations. The projection π_P in §1.3(a) as a mapping of $\mathbf{P}^2(\mathbf{R}) = H \cup \ell_\infty$ to $\mathbf{P}^2(\mathbf{R}) = H' \cup \ell'$ is one special kind of projective transformation, and general projective transformations are obtained by composing projections with center P at different points. In terms

of homogeneous coordinates, a projective transformation can be expressed in the following form:

$$g : (x_0 : x_1 : x_2) \longmapsto \left(\sum_{j=0}^{2} a_{0j} x_j : \sum_{j=0}^{2} a_{1j} x_j : \sum_{j=0}^{2} a_{2j} x_j \right) \tag{1.13}$$

with

$$A = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}, \quad \det A \neq 0.$$

Using matrix forms, we can write

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \longmapsto A \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}. \tag{1.14}$$

We denote by g_A the projective transformation determined by (1.14) with a 3×3 regular matrix A. We note that if ρ is a nonzero real number, then $g_A = g_{\rho A}$, because for homogeneous coordinates it is only the ratios that matter.

In particular, for a matrix

$$B = \begin{pmatrix} 1 & 0 & 0 \\ b_{10} & b_{11} & b_{12} \\ b_{20} & b_{21} & b_{22} \end{pmatrix}, \qquad \det B \neq 0, \tag{1.15}$$

the corresponding projective transformation g_B can be represented by

$$g_B : (x, y) \mapsto (b_{11}x + b_{12}y + b_{10}, b_{21}x + b_{22}y + b_{20}). \tag{1.16}$$

Since $\det B \neq 0$, we get

$$\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \neq 0.$$

(1.16) is an affine transformation. The line at infinity $\ell_\infty : x_0 = 0$ is left invariant under g_B. Conversely, a projective transformation that leaves the line at infinity invariant is of the form (1.14) with

$$A = \begin{pmatrix} a_{00} & 0 & 0 \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix} \text{ with } \det A \neq 0 \text{ and } a_{00} \neq 0.$$

As projective transformations we have $g_A = g_{a_{00}^{-1}A}$, where $a_{00}^{-1}A$ is of the form (1.15). It follows that a projective transformation that leaves the line at infinity invariant, when restricted to the ordinary plane $H = \mathbf{P}^2(\mathbf{R}) - \ell_\infty$, is an affine transformation. Conversely, it is clear from the arguments above that to an affine transformation (1.16) there corresponds a projective transformation that leaves the line at infinity invariant.

In the preceding subsection we explained that there is no distinction between ellipses, hyperbolas and parabolas from the viewpoint of projective geometry. We shall reconfirm this fact by using coordinates. First, let us consider a hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

By virtue of (1.10) the equation in homogeneous coordinates is

$$x_0^2 - \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 0. \tag{1.17}$$

By applying the projective transformation

$$g : \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_0 \\ x_2 \end{pmatrix},$$

the equation (1.17) becomes

$$x_1^2 - \frac{x_0^2}{a^2} + \frac{x_2^2}{b^2} = 0.$$

Using (1.10) again, we get

$$x^2 + \frac{y^2}{b^2} = \frac{1}{a^2},$$

that is, an ellipse. We can also directly check the behavior of the hyperbola around the line at infinity ℓ_∞. Since ℓ_∞ is defined by $x_0 = 0$, we have $x_1 \neq 0$ or $x_2 \neq 0$ at each point on ℓ_∞. So let us consider the region U_1 given by $x_1 \neq 0$. We introduce new affine coordinates (u, v) on U_1 by

$$u = \frac{x_0}{x_1}, \qquad v = \frac{x_2}{x_1}. \tag{1.18}$$

In these coordinates the line at infinity ℓ_∞ is defined by

$$u = 0,$$

that is, (u, v) are the coordinates with the point $(u, v) = (0, 0)$, namely, $(0 : 1 : 0)$, as center. Rewriting (1.17) we get

$$u^2 + \frac{v^2}{b^2} = \frac{1}{a^2},$$

which certainly looks like an ellipse (see Figure 1.11) and has the same representation as in Figure 1.10. What happens if we consider the coordinates

$$w = \frac{x_0}{x_2}, \qquad z = \frac{x_1}{x_2}, \tag{1.19}$$

in the region U_2 given by $x_2 \neq 0$? In this case, the line at infinity ℓ_∞ is defined by

$$w = 0,$$

and the equation (1.17) becomes

$$\frac{z^2}{a^2} - w^2 = \frac{1}{b^2},$$

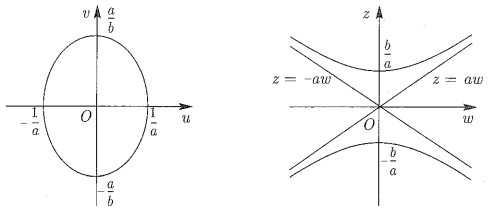


FIGURE 1.11. With (w, z) coordinates, the u-axis is at infinity and thus invisible. $(u, v) = (\pm 1/a, 0)$ correspond to points at infinity on the asymptotes $z = \pm aw$.

which is a hyperbola. This hyperbola and ℓ_∞ meet at two points $(w, z) = (0, \pm a/b)$.

Why does the figure have different appearances depending on whether we use (u, v) or (w, z)? This is clarified by looking at (1.18) and (1.19), the equations that introduce the coordinates (u, v) and (w, z). Between the two coordinates we have the relationships

$$w = \frac{u}{v}, \qquad z = \frac{1}{v}.$$

The u-axis: $v = 0$ is invisible at infinity in the (z, w) coordinates. Conversely, the w-axis $z = 0$ is invisible at infinity in the (u, v) coordinates. Since the w-axis does not meet our curve, we have an ellipse appearing in the (u, v)-plane. The u-axis meets the ellipse, and this intersection goes away in two different directions to infinity in the (w, z)-coordinates, thus making a hyperbola appear.

We may sum up by saying that from the projective-geometric point of view ellipses, hyperbolas and parabolas are essentially the same and they take distinct appearances only when they are seen in a finite plane. In projective geometry all lines intersect and all ellipses, hyperbolas and parabolas are indistinguishable, thus simplifying many situations and discussions.

Let us now consider more general curves in the projective plane $\mathbf{P}^2(\mathbf{R})$. We want to consider the set of zeros

$$F(x_0, x_1, x_2) = 0$$

of a polynomial of degree m. Since only the ratios matter for homogeneous coordinates $(x_0 : x_1 : x_2)$, we must have $F(\lambda x_0, \lambda x_1, \lambda x_2) = 0$ for any real number $\lambda \neq 0$. Thus in order to define an algebraic curve

$$C : F(x_0, x_1, x_2) = 0, \tag{1.20}$$

we must assume that $F(x_0, x_1, x_2)$ is homogeneous, that is,

$$F(\lambda x_0, \lambda x_1, \lambda x_2) = \lambda^m F(x_0, x_1, x_2). \tag{1.21}$$

This condition can be rewritten as

$$F(x_0, x_1, x_2) = \sum_{i_0 + i_1 + i_2 = m} a_{i_0 i_1 i_2} x_0^{i_0} x_1^{i_1} x_2^{i_2}, \tag{1.22}$$

where the exponent m for λ on the right-hand side of (1.21) is nothing but the degree of $F(x_0, x_1, x_2)$. A curve C defined by a homogeneous polynomial $F(x_0, x_1, x_2)$ of degree m is called a plane curve of degree m. Ellipses, hyperbolas and parabolas are plane curves of degree 2 in the projective plane.

When a curve is given by

$$E : f(x, y) = 0$$

in the ordinary plane H with a polynomial $f(x, y)$ of degree m, we can, by using (1.10), extend the curve E to a plane curve of degree m in the projective plane

$$\tilde{E} : F(x_0, x_1, x_2) = 0$$

in a natural fashion. To do this, we take the homogeneous polynomial of degree m given by

$$F(x_0, x_1, x_2) = x_0^m f(\frac{x_1}{x_0}, \frac{x_2}{x_0}). \tag{1.23}$$

We have already seen how this is done for curves of degree 2.

Now a plane curve is mapped to a plane curve by a projective transformation. It is easy to show that the degree is invariant. Furthermore, the notion of tangent line is also preserved by any projective transformation. In the projective plane, the equation for a tangent line takes a simple form. Indeed, for a plane curve of degree m

$$C : F(x_0, x_1, x_2) = 0,$$

the tangent line at a point (a_0, a_1, a_2) of C is given by

$$\sum_{i=0}^{2} \frac{\partial F}{\partial x_i}(a_0, a_1, a_2) x_i = 0. \tag{1.24}$$

We shall just show that this equation is obtained by rewriting in homogeneous coordinates the equation of a tangent line in the ordinary Euclidean plane by using (1.10) and (1.23). For simplicity, we assume $a_0 \neq 0$ and set

$$a = \frac{a_1}{a_0}, \qquad b = \frac{a_2}{a_0}$$

and

$$f(x, y) = \frac{1}{x_0^m} F(x_0, x_1, x_2). \tag{1.25}$$

The equation of the tangent line of the curve

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

at (a, b) is given by

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) = 0. \tag{1.26}$$

By using (1.10) and (1.23) we get

$$\begin{aligned}
\frac{\partial F}{\partial x_0} &= m x_0^{m-1} f(x, y) - x_1 x_0^{m-2} \frac{\partial f}{\partial x}(x, y) - x_2 x_0^{m-2} \frac{\partial f}{\partial y}(x, y) \\
\frac{\partial F}{\partial x_1} &= x_0^{m-1} \frac{\partial f}{\partial x}(x, y) \\
\frac{\partial F}{\partial x_2} &= x_0^{m-1} \frac{\partial f}{\partial y}(x, y)
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{\partial F}{\partial x_0}(1, a, b) &= -a \frac{\partial f}{\partial x}(a, b) - b \frac{\partial f}{\partial y}(a, b) \\
\frac{\partial F}{\partial x_1}(1, a, b) &= \frac{\partial f}{\partial x}(a, b) \\
\frac{\partial F}{\partial x_2}(1, a, b) &= \frac{\partial f}{\partial y}(a, b).
\end{aligned}$$

Substituting these relations into (1.26), we obtain

$$\frac{\partial F}{\partial x_0}(1, a, b) + \frac{\partial F}{\partial x_1}(1, a, b) \frac{x_1}{x_0} + \frac{\partial F}{\partial x_2}(1, a, b) \frac{x_2}{x_0} = 0.$$

Multiplying both sides by $x_0 a_0^{m-1}$, we arrive at (1.24). (Note that $\frac{\partial F}{\partial x_i}$ is homogeneous of degree $m - 1$, and we have $\frac{\partial F}{\partial x_i}(a_0, a_1, a_2) = a_0^{m-1} \frac{\partial F}{\partial x_i}(1, a, b)$.)

We remark that a homogeneous polynomial $F(x_0, x_1, x_2)$ of degree m satisfies Euler's identity

$$\sum_{i=0}^{2} x_i \frac{\partial F}{\partial x_i}(x_0, x_1, x_2) = m F(x_0, x_1, x_2). \tag{1.27}$$

(See Exercise 1.3.) In particular, if

$$\frac{\partial F}{\partial x_i}(a_0, a_1, a_2) = 0, \quad i = 0, 1, 2, \tag{1.28}$$

then we see from (1.27) that (a_0, a_1, a_2) lies on the curve C. At such a point, we cannot define the tangent line. We call a point of the plane curve C a singular point if (1.28) holds. We give more detail on singular points in §2.5.

Let us confirm that the asymptotes $y = \pm \frac{b}{a} x$ of the hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

are the tangent lines at the points at infinity. In homogeneous coordinates the hyperbola can be expressed by

$$x_0^2 - \frac{x_1^2}{a^2} + \frac{x_0^2}{b^2} = 0$$

and has points at infinity $(0 : a : b)$ and $(0 : -a : b)$. At $(0 : a : b)$ the equation (1.24) of the tangent line is

$$-\frac{1}{a} x_1 + \frac{1}{b} x_2 = 0,$$

and the tangent line at $(0:-a:b)$ is

$$\frac{1}{a}x_1+\frac{1}{b}x_2=0.$$

From (1.10) these lines are nothing but $y=\frac{b}{a}x$ and $y=-\frac{b}{a}x$. In this way we see that the asymptotes of the hyperbola are the part of the tangent lines at the points at infinity that appears on the ordinary plane. In general, we may say that any asymptote to a plane curve is the tangent line at a point at infinity, although we have to generalize the notion of tangent line when the point at infinity happens to be a singular point. We shall discuss this question in §2.5.

We shall check on the tangent line at the point at infinity of a parabola

$$y=ax^2.$$

In homogeneous coordinates, this parabola is written as

$$x_0x_2-ax_1^2=0,$$

and the point at infinity on the curve is $(0:0:1)$. The tangent line at this point is

$$x_0=0,$$

as we see from (1.24). This is nothing but the line at infinity ℓ_∞. In other words, the parabola has ℓ_∞ as a tangent line. Every point on the parabola other than the point at infinity can be written in the form $(1:\alpha:a\alpha^2)$. When $\alpha\neq 0$, we see that the point

$$(1:\alpha:a\alpha^2)=(\frac{1}{a\alpha^2}:\frac{1}{a\alpha}:1)$$

approaches the point at infinity $(0:0:1)$ as $\alpha\to\pm\infty$. On the other hand, the tangent line at $(1:\alpha:a\alpha^2)$ is

$$a\alpha^2x_0-2a\alpha x_1+x_2=0,$$

which can be rewritten in the form

$$x_0-\frac{2}{\alpha}x_1+\frac{1}{a\alpha^2}x_2=0.$$

This line approaches

$$x_0=0$$

as $\alpha\to\infty$. Thus it is natural to regard the line at infinity ℓ_∞, being the limit of the tangents to the parabola, as the tangent line at the point at infinity.

§ 1.4. Introduction of complex numbers

(a) The introduction of complex numbers. We have seen that in projective geometry the conics are all plane curves of degree 2 without any distinction between ellipses, hyperbolas or parabolas. However, there is one problem we have avoided which makes our considerations rather unsatisfactory. For example, it does not make sense to speak of the plane curve

$$x_0^2+x_1^2+x_2^2=0, \tag{1.29}$$

because the equation has only $(0,0,0)$ as a real solution and thus does not define a figure in the projective plane $\mathbf{P}^2(\mathbf{R})$. It would also be strange to call

$$(x_1-ax_0)^2+(x_2-bx_0)^2=0 \tag{1.30}$$

a plane curve of degree 2, because the solutions of this equation can be written in the form $(x_0,x_1,x_2)=(\alpha,a\alpha,b\alpha)$, which represent only one point $(1:a:b)$ in the projective plane $\mathbf{P}^2(\mathbf{R})$.

Such inconveniences are not confined to projective geometry and appear in the study of Euclidean geometry by means of coordinates. The equation

$$x^2+y^2=1$$

determines a unit circle, but the equation

$$x^2+y^2=0$$

is satisfied only at the origin, which normally is not regarded as a curve. Such inconveniences arise also when we look for the intersections of the line

$$y=ax+b$$

with the circle

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

Substituting the first equation in the second, we get the quadratic equation

$$x^2+(ax+b)^2=1,$$

which may not have any solution within the real number system. It is a practically important question to find the condition on (a,b) under which there are solutions, but general discussions get complicated because a given line may or may not intersect a given circle.

The only way we can avoid such inconveniences is not to work with real numbers only but to use complex numbers. According to the fundamental theorem of algebra, a polynomial with complex coefficients

$$a_0x^n+a_1x^{n-1}+\cdots+a_n=0$$

has a solution within complex numbers. By allowing complex numbers the line and the unit circle always intersect. The equation

$$x^2+y^2=0$$

also has solutions $(x,y)=(\alpha,\pm i\alpha)$, where α is an arbitrary complex number. In fact,

$$x^2+y^2=(x+iy)(x-iy)=0,$$

and the equation $x^2+y^2=0$ represents two lines $x\pm iy=0$. The equations (1.29) and (1.30) have infinitely many complex solutions and hence define some geometric figures. As for (1.29), the change of variables

$$y_0=ix_0,\quad y_1=x_1,\quad y_2=x_2$$

changes it to

$$-y_0^2+y_1^2+y_2^2=0.$$

Using inhomogeneous coordinates $X = y_1/y_0, Y = y_2/y_0$, we obtain

$$X^2 + Y^2 = 1,$$

namely, the unit circle. In this way, the plane curve of degree 2 defined by (1.29) and the unit circle are essentially the same.

Mathematicians, however, hesitated for a long time to introduce complex numbers because intuition does not allow "numbers" whose squares are negative. This feeling is well expressed in the name of "imaginary number". If the real world is described by real numbers, then the world described by complex numbers will include "imaginary" parts, or so they felt. But in projective geometry itself it was slowly recognized that the reality is not entirely revealed unless imaginary parts are also included.

Now if we decide to study figures by extending real numbers to complex numbers, it becomes necessary to throw away naive intuition. If we allow complex numbers as coordinates (x, y) for the plane H, then writing

$$x = s + it, \quad y = u + iv, \text{ where } s, t, u, v \in \mathbf{R},$$

we identify the pair of complex numbers (x, y) with the quadruple (s, t, u, v) of real numbers, and thus H becomes a 4-dimensional space. Also for the line

$$y = ax + b, \tag{1.31}$$

y is determined for any complex number x. Thus the degree of freedom is the same as that of complex numbers, and (1.31) determines a 2-dimensional figure. On the other hand, we still want to call the figure determined by (1.31) a line. Therefore we say that the plane H has complex dimension 2 and that the figure ℓ determined by (1.31) is a line of complex dimension 1. This is a natural definition, because the freedom of complex numbers is 2 for H and 1 for ℓ. The dimension in the ordinary sense will now be called the real dimension.

The projective plane can also be extended to the complex projective plane (denoted $\mathbf{P}^2(\mathbf{C})$) by using complex numbers. The projective plane $\mathbf{P}^2(\mathbf{R})$ is then called the real projective plane. Now the complex projective plane $\mathbf{P}^2(\mathbf{C})$ is defined as follows. We take the set of all triples of complex numbers $(a_0, a_1, a_2) \neq (0, 0, 0)$ and consider only the ratios, just as in the case of $\mathbf{P}^2(\mathbf{R})$. The homogeneous coordinates $(x_0 : x_1 : x_2)$ also work in the same way, except that each x_i is a complex number, and likewise for inhomogeneous coordinates (x, y), where $x = x_1/x_0$ and $y = x_1/x_0$. Leaving the discussion of plane curves in the complex projective plane to the next section, we shall deal with projective lines. For simplicity, let us consider the line L defined by the equation

$$x_2 = 0.$$

Using inhomogeneous coordinates we have

$$y = 0,$$

which represents the x-axis in the plane H. In the complex projective plane, we have furthermore the point at infinity $(0 : 1 : 0)$. Hence we may write

$$L = \mathbf{C} \cup \{\infty\},$$

where ∞ denotes the point $(0 : 1 : 0)$. This projective line L is indeed the same as the Riemann sphere that is well-known in complex function theory; this may be seen as follows. The line L defined by $x_2 = 0$ is the set of all points $(a_0 : a_1 : 0)$. If $a_1 \neq 0$, then we have

$$(a_0 : a_1 : 0) = (1 : \frac{a_1}{a_0} : 0),$$

and

$$\phi : L - \{(0 : 1 : 0)\} \longrightarrow \mathbf{C} \text{ with } (a_0 : a_1 : 0) \longmapsto \frac{a_1}{a_0}$$

is a bijection, with which we identify $L - \{(0 : 1 : 0)\}$ and $\mathbf{C}$. On the other hand, for $z \neq 0 \in \mathbf{C}$ we have

$$\phi^{-1}(z) = (1 : z : 0) = (\frac{1}{z} : 1 : 0), \tag{1.32}$$

and this point approaches $(0 : 1 : 0)$ as $|z| \to +\infty$. Similarly,

$$\psi : L - \{(1 : 0 : 0)\} \longrightarrow \mathbf{C} \text{ with } (a_0 : a_1 : 0) \longmapsto \frac{a_0}{a_1}$$

is a bijection and

$$\psi^{-1}(w) = (w : 1 : 0).$$

For $w \neq 0$ we have

$$(w : 1 : 0) = (1 : \frac{1}{w} : 0). \tag{1.32$'$}$$

Since

$$L = (L - \{(1 : 0 : 0)\}) \cup (L - \{(0 : 1 : 0)\}),$$

by virtue of (1.32) and (1.32$'$), we may obtain L by pasting two copies of $\mathbf{C}$ by using the relation

$$z = \frac{1}{w},$$

that is, L can be regarded as the same as the Riemann sphere. (For the details, see §2.1.) Since a general projective line can be mapped onto the line L by a projective transformation, it can be identified with the Riemann sphere.

(b) Complex plane curves. The theory of plane curves becomes clear in the complex projective plane $\mathbf{P}^2(\mathbf{C})$. The zero set of a homogeneous polynomial $F(x_0, x_1, x_2)$ of degree m with complex coefficients, namely, the figure C in $\mathbf{P}^2(\mathbf{C})$ defined by

$$F(x_0, x_1, x_2) = 0,$$

is called a **complex plane curve of degree** m, or simply a **plane curve of degree** m. We often denote C by $V(F)$ to indicate that C is defined by F. If $F(x_0, x_1, x_2)$ is irreducible, we say that $C = V(F)$ is an **irreducible plane curve**. If $F(x_0, x_1, x_2)$ is reducible, C is called a **reducible plane curve**. In the latter case, if F is factored

$$F(x_0, x_1, x_2) = G(x_0, x_1, x_2)H(x_0, x_1, x_2),$$

then we can assume that G and H are homogeneous polynomials. We have obviously

$$V(F) = V(G) \cup V(H).$$

If

$$F(x_0, x_1, x_2) = G(x_0, x_1, x_2)^m,$$

then

$$V(F) = V(G).$$

Therefore, we mainly discuss the case where $F(x_0, x_1, x_2)$ is irreducible.

A projective transformation of the complex projective plane $\mathbf{P}^2(\mathbf{C})$ can be expressed by

$$g : (x_0 : x_1 : x_2) \longmapsto (\sum_{j=0}^{2} a_{0j}x_j : \sum_{j=0}^{2} a_{1j}x_j : \sum_{j=0}^{2} a_{2j}x_j), \tag{1.33}$$

$$A = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}, \qquad \det A \neq 0.$$

The only difference from (1.13) is that A is a complex matrix. Just like (1.14), it is convenient to write

$$g : \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \mapsto A \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \tag{1.34}$$

and denote by g_A the projective transformation determined by a regular matrix A. If ρ is a nonzero complex number, we have

$$g_A = g_{\rho A}.$$

Now let us study the image $g_A(V(F))$ of a plane curve $V(F)$ of degree m by the projective transformation g_A. Given

$$F(a_0, a_1, a_2) = 0,$$

we set

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = A \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}.$$

Then $(b_0 : b_1 : b_2) \in g_A(V(F))$. Now set

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = A^{-1} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} b_{00} & b_{01} & b_{02} \\ b_{10} & b_{11} & b_{12} \\ b_{20} & b_{21} & b_{22} \end{pmatrix}$$

and

$$G(x_0, x_1, x_2) = F(y_0, y_1, y_2) = F(\sum_{j=0}^{2} b_{0j}x_j, \sum_{j=0}^{2} b_{1j}x_j, \sum_{j=0}^{2} b_{2j}x_j). \tag{1.35}$$

Then

$$G(b_0, b_1, b_2) = F(a_0, a_1, a_2)$$

and hence

$$g_A(V(F)) = V(G). \tag{1.36}$$

It is clear from (1.35) that if $F(x_0, x_1, x_2)$ is homogeneous of degree m, then so is $G(x_0, x_1, x_2)$. If $F(x_0, x_1, x_2)$ is irreducible (resp. reducible), then $G(x_0, x_1, x_2)$ is irreducible (resp. reducible). The purpose of projective geometry is to study properties invariant under projective transformations. The degree of a plane curve

is invariant under projective transformations. A plane curve of degree 1 is usually called a line. We sometimes call it a **projective line** to emphasize that it is in the projective plane. Given a line

$$\ell : a_0x_0 + a_1x_1 + a_2x_2 = 0, \text{ where } (a_0, a_1, a_2) \neq (0, 0, 0),$$

we can find $a_{10}, a_{11}, a_{12}, a_{20}, a_{21}, a_{22}$ such that the matrix

$$A = \begin{pmatrix} a_0 & a_1 & a_2 \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}$$

has nonzero determinant, and, using (1.35) and (1.36), we find that $g_A(\ell)$ is defined by

$$x_0 = 0,$$

that is, it is the line at infinity ℓ_∞. We have thus shown that any line can be mapped to ℓ_∞ by a projective transformation. It follows that any two lines ℓ_1 and ℓ_2 can be transformed to each other by a projective transformation.

Next we consider a plane curve of degree 2

$$\mathbf{Q} : \sum_{i,j=0}^{2} q_{ij}x_ix_j = 0. \tag{1.37}$$

In matrix form, we have

$$(x_0, x_1, x_2)Q \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = 0, \quad Q = (q_{ij}).$$

We may take Q to be a symmetric matrix. From (1.35) we see that the image $g_A(\mathbf{Q})$ by the projective transformation g_A is the plane curve of degree 2 defined by

$$(x_0, x_1, x_2)Q' \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = 0, \text{ where } Q' = {}^tA^{-1}QA^{-1}. \tag{1.38}$$

By an appropriate choice of A we can make Q' into one of the following matrices:

$$Q_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.39}$$

The corresponding plane curves of degree 2 are respectively

$$\begin{aligned} x_0^2 &= 0 \\ -x_0^2 + x_1^2 &= 0 \\ -x_0^2 + x_1^2 + x_2^2 &= 0. \end{aligned} \tag{1.40}$$

The curve of degree 2 defined by the first equation is the line at infinity ℓ_∞ counted twice. The second equation defines two lines

$$x_0 \pm x_1 = 0.$$

The third equation defines an irreducible curve of degree 2, which is expressed in inhomogeneous coordinates by

$$x^2 + y^2 = 1,$$

which is like the unit circle in the complex domain. Once again, ellipses, hyperbolas and parabolas considered in the complex domain are irreducible plane curves of degree 2 that can be transformed to a unit circle by a projective transformation. The second and third equations in (1.40) can be transformed to

$$x_0^2+x_1^2 = 0$$
$$x_0^2+x_1^2+x_2^2 = 0.$$

The third equation in (1.40) can be mapped by a projective transformation upon

$$-x_0^2 + x_1x_2 = 0.$$

In terms of inhomogeneous coordinates, this is expressed as the rectangular hyperbola

$$xy = 1.$$

We have just seen that in the complex projective plane any two irreducible curves of degree 2 can be transformed into each other by a projective transformation. To study projective properties of an irreducible conic, we can choose a unit circle or a rectangular hyperbola to represent a figure suitably, depending on the given problem.

Here we shall review once more the correspondence (Figure 1.2) between the unit circle and the line that we studied in §1.2(a). We consider the problem in the complex projective plane $\mathbf{P}^2(\mathbf{C})$. The unit circle C can be represented by

$$-x_0^2 + x_1^2 + x_2^2 = 0,$$

and the line ℓ

$$-x_0 + x_1 = 0$$

is tangent to C at the point $(1 : 1 : 0)$. An arbitrary line through the point $P = (1 : -1 : 0)$ is given by

$$u_1(x_0 + x_1) - u_0x_2 = 0. \tag{1.41}$$

The lines given by (1.41) for the same ratio $(u_0 : u_1)$ are the same line, which we denote by $\ell_{(u_0:u_1)}$. To find the intersection R of the line $\ell_{(u_0:u_1)}$ and ℓ, we solve the system of equations

$$\begin{cases} u_1(x_0 + x_1) - u_0x_2 = 0 \\ -x_0 + x_1 = 0 \end{cases}$$

and find

$$R = (u_0 : u_0 : 2u_1). \tag{1.42}$$

On the other hand, let Q be the intersection other than P of the line $\ell_{(u_0:u_1)}$ and the unit circle C. By solving the system of equations

$$\begin{cases} -x_0^2 + x_1^2 + x_2^2 = 0 \\ u_1(x_0 + x_1) - u_0x_2 = 0 \end{cases}$$

we find

$$Q = (u_0^2 + u_1^2 : u_0^2 - u_1^2 : 2u_0u_1). \tag{1.43}$$

If we set $t = u_1/u_0$ and use inhomogeneous coordinates, we get the same result as (1.2). Furthermore, for $u_0 = 0$, the line

$$\ell_{(0:u_1)} : x_0 + x_1 = 0$$

is tangent to C at the point $P = (1 : -1 : 0)$. Then we get

$$R = (0 : 0 : 1), \qquad Q = (1 : -1 : 0) = P,$$

and we have a natural interpretation of the observation we made in §1.2 (a) as $t \to \pm\infty$. In this way we see that the point Q on C and the point R on ℓ correspond one-to-one by (1.42) and (1.43). To P there corresponds the point at infinity $(0 : 0 : -1)$ on ℓ. Furthermore, the unit circle C, the line ℓ and the set of ratios $(u_0 : u_1)$ of all pairs of complex numbers $(u_0, u_1) \neq (0, 0)$ correspond in a one-to-one manner by associating the point $(u_0^2 + u_1^2 : u_0^2 - u_1^2 : 2u_0u_1)$, the point $(u_0, u_0, 2u_1)$ on ℓ, and the ratio $(u_0 : u_1)$. We now denote by $\mathbf{P}^1(\mathbf{C})$ the set of ratios of all pairs of complex numbers $(u_0, u_1) \neq (0, 0)$, and call it the **complex projective line**. (The reader may wish to refer to the Remark following Definition 1.1.) The point of $\mathbf{P}^1(\mathbf{C})$ corresponding to the ratio $(u_0 : u_1)$ is also denoted by $(u_0 : u_1)$. Let us also note that $\mathbf{P}^1(\mathbf{C})$ corresponds to the set of all slopes of the lines (1.41). Before we study the relationship between the complex projective line $\mathbf{P}^1(\mathbf{C})$ and the unit circle C in detail, let us make a few observations. When $u_0 \neq 0$, we have $(u_0, u_1) = (1 : u_1/u_0)$ and hence the mappings

$$\phi : \mathbf{P}^1(\mathbf{C}) - \{(0 : 1)\} \longrightarrow \mathbf{C}$$
$$\text{with } (u_0 : u_1) \longmapsto \frac{u_1}{u_0}$$

and

$$\psi : \mathbf{C} \longrightarrow \mathbf{P}^1(\mathbf{C}) - \{(0 : 1)\}$$
$$\text{with } z \longmapsto (1 : z)$$

are inverse to each other. We identify $\mathbf{P}^1(\mathbf{C}) - \{(0 : 1)\}$ and the complex plane $\mathbf{C}$ by these mappings. When furthermore $z \neq 0$, we have

$$(1 : z) = (\frac{1}{z} : 1)$$

and hence

$$\lim_{|z|\to\infty} (1 : z) = \lim_{|z|\to\infty} (\frac{1}{z} : 1) = (0 : 1),$$

which we may regard as the point at infinity ∞ of the Riemann sphere in the complex function theory of one variable. As a result, we have

$$\mathbf{P}^1(\mathbf{C}) = \mathbf{C} \cup \{(0 : 1)\} = \mathbf{C} \cup \{\infty\},$$

and the complex projective line $\mathbf{P}^1(\mathbf{C})$ is the Riemann sphere.

With these preparations we study the correspondence between $\mathbf{P}^1(\mathbf{C})$, the line ℓ, and the unit circle C by means of (1.42) and (1.43). First, the mapping

$$\mathbf{P}^1(\mathbf{C}) \longrightarrow \ell$$
$$\text{with } (u_0 : u_1) \longmapsto (u_0 : u_0 : 2u_1)$$

gives a one-to-one correspondence between $\mathbf{P}^1(\mathbf{C})$ and ℓ. Thus an arbitrary line can be identified with $\mathbf{P}^1(\mathbf{C})$ and hence with the Riemann sphere by a projective transformation.

Next, using (1.43) we consider the mapping

$$\tilde{\phi} : \mathbf{P}^1(\mathbf{C}) \longrightarrow C \tag{1.44}$$
$$\text{with } (u_0 : u_1) \longmapsto (u_0^2 + u_1^2 : u_0^2 - u_1^2 : 2u_0u_1).$$

This is well-defined because

$$(u_0 : u_1) = (\lambda u_0 : \lambda u_1), \quad \lambda \in \mathbf{C} - \{0\},$$

implies

$$\begin{aligned}((\lambda u_0)^2+(\lambda u_1)^2 : (\lambda u_0)^2 - (\lambda u_1)^2 : 2\lambda^2 u_0u_1) \\ = (\lambda^2(u_0^2 + u_1^2) : \lambda^2(u_0^2 - u_1^2) : 2\lambda^2 u_0u_1) \\ = (u_0^2 + u_1^2 : u_0^2 - u_1^2 : 2u_0u_1).\end{aligned}$$

That this mapping is a bijection is clear from the geometric correspondence between the point Q on C and the point R on ℓ, but we can derive it from (1.44) only. First, suppose

$$\tilde{\phi}(u_0 : u_1) = \tilde{\phi}(u_0' : u_1'), \text{ where } (u_0, u_1), (u_0' : u_1') \in \mathbf{P}^1(\mathbf{C}),$$

that is,

$$(u_0^2 + u_1^2 : u_0^2 - u_1^2 : 2u_0u_1) = ({u_0'}^2 + {u_1'}^2 : {u_0'}^2 - {u_1'}^2 : 2u_0'u_1').$$

If $u_0u_1 \neq 0$, then we have

$$\frac{u_0^2 + u_1^2}{u_0u_1} = \frac{{u_0'}^2 + {u_1'}^2}{u_0'u_1'}$$
$$\frac{u_0^2 - u_1^2}{u_0u_1} = \frac{{u_0'}^2 - {u_1'}^2}{u_0'u_1'},$$

which implies

$$\frac{u_1}{u_0} = \frac{u_0'}{u_1'}$$

and thus

$$(u_0 : u_1) = (u_0' : u_1').$$

If $u_0 = 0$, then

$$\tilde{\phi}(0 : u_1) = (1 : -1 : 0),$$

and hence $u_0'u_1' = 0$ and

$${u_0'}^2 + {u_1'}^2 = -({u_0'}^2 - {u_1'}^2).$$

It follows that $u_0' = 0$, that is, $(u_0' : u_1') = (0 : 1)$. Also in the case where $(u_0 : u_1) = (1 : 0)$ a similar argument leads to $(u_0' : u_1') = (1 : 0)$. We have seen that $\tilde{\phi}$ is a one-to-one mapping.

Next we shall show that, given a point $(a_0 : a_1 : a_2) \in C$, there is a point $(u_0 : u_1) \in \mathbf{P}^1(\mathbf{C})$ such that $\tilde{\phi}((u_0, u_1)) = (a_0 : a_1 : a_2)$. Since $(a_0 : a_1 : a_2) \in \mathbf{C}$, we have

$$-a_0^2 + a_1^2 + a_2^2 = 0. \tag{1.45}$$

If $a_2 = 0$, then $a_1^2 = a_0^2$ and $(a_0 : a_1 : a_2) = (1 : \pm 1 : 0)$. It follows that

$$\tilde{\phi}((1 : 0)) = (1 : 1 : 0)$$
$$\tilde{\phi}((0 : 1)) = (1 : -1 : 0).$$

Now suppose $a_2 \neq 0$. Then we have

$$\frac{a_1 - a_0}{a_2} \cdot \frac{a_1 + a_0}{a_2} = 1. \tag{1.46}$$

From this relation we obtain

$$a_1 \pm a_0 \neq 0.$$

Using the mapping (1.44) we consider the point $(1 : (a_0 - a_1)/a_2) \in \mathbf{P}^1(\mathbf{C})$. (Note that if $\tilde{\phi}((u_0 : u_1)) = (a_0 : a_1 : a_2)$, then $u_1/u_0 = (a_0 - a_1)/a_2$.) By (1.45), we get

$$\begin{aligned}\tilde{\phi}((1 : \frac{a_0 - a_1}{a_2})) &= (\frac{a_2^2 + (a_0 - a_1)^2}{a_2^2} : \frac{a_2^2 - (a_0 - a_1)^2}{a_2^2} : \frac{2(a_0 - a_1)}{a_2}) \\ &= (a_2^2 + (a_0 - a_1)^2 : a_2^2 - (a_0 - a_1)^2 : 2a_2(a_0 - a_1)) \\ &= (2(a_0^2 - a_0a_1) : 2(a_0a_1 - a_1^2) : 2a_2(a_0 - a_1)) \\ &= (a_0 : a_1 : a_2),\end{aligned}$$

and see that $(a_0 : a_1 : a_2)$ is in the image of $\tilde{\phi}$. We have thus seen that we can identify $\mathbf{P}^1(\mathbf{C})$ and the curve C by $\tilde{\phi}$. From the argument above the inverse of $\tilde{\phi}$ is given by

$$\psi_+ : C \longrightarrow \mathbf{P}^1(\mathbf{C}) \tag{1.47}$$
$$\text{with } (a_0 : a_1 : a_2) \longmapsto (a_2 : a_0 - a_1)$$

Here, however, ψ_+ is not defined at $(1 : 1 : 0)$, where $a_2 = a_0 - a_1 = 0$, If $a_2 \neq 0$, then (1.46) gives

$$\begin{aligned}(a_2 : a_0 - a_1) &= (1 : \frac{a_0 - a_1}{a_2}) \\ &= (1 : \frac{a_2}{a_0 + a_1}) = (a_0 + a_1 : a_2).\end{aligned}$$

Now if we define

$$\psi_- : C \longrightarrow \mathbf{P}^1(\mathbf{C}) \tag{1.48}$$
$$\text{by } (a_0 : a_1 : a_2) \longmapsto (a_0 + a_1 : a_2),$$

then this mapping is not defined at $(1 : -1 : 0)$. From the discussions above, we see that ψ_+ and ψ_- coincide on $C - \{(1 : 1 : 0), (1 : -1 : 0)\}$. Thus ψ_+ and ψ_- together define

$$\tilde{\psi} : C \longrightarrow \mathbf{P}^1(\mathbf{C}).$$

It is easy to verify that $\tilde{\phi}$ and $\tilde{\psi}$ are inverse to each other. They are both expressible by using homogeneous polynomials in the coordinates. Such maps are called **algebraic morphisms**. Algebraic geometry is the study of geometric properties invariant under bijections that are algebraic morphisms. Thus from the viewpoint of algebraic geometry, lines (nothing but plane curves of degree 1) and irreducible plane curves of degree 2 are regarded as isomorphic to the projective line $\mathbf{P}^1(\mathbf{C})$. This is a precise formulation of what Descartes stated in his *Geometry*.

§1.5. The birth of algebraic geometry

(a) Plane curves and intersection theory. Historically, constructing the intersection theory of plane curves was a driving force for the birth of algebraic geometry. The starting point is to count the number of intersections of a plane curve of degree m

$$C : F(x_0, x_1, x_2) = 0 \tag{1.49}$$

and a plane curve of degree n

$$D : G(x_0, x_1, x_2) = 0$$

in the complex projective plane $\mathbf{P}^2(\mathbf{C})$. Naively, we have only to find a solution of the system

$$\begin{cases} F(x_0, x_1, x_2) = 0 \\ G(x_0, x_1, x_2) = 0 \end{cases}$$

by solving a new equation obtained by eliminating one of the variables in the system. For this purpose, the theory of elimination was developed, occasionally with somewhat bizarre findings. The effort of resolving such difficulties contributed to the development of algebraic geometry.

Before dealing with the general intersection theory, let us treat the simplest case, namely, we shall find the intersection of the curve C of degree $m \geq 2$ given by (1.49) and a line L. If we take two distinct points $P = (a_0 : a_1 : a_2)$ and $Q = (b_0 : b_1 : b_2)$ on L, then the line L can be expressed in parametric form

$$(\lambda a_0 + \mu b_0 : \lambda a_1 + \mu b_1 : \lambda a_2 + \mu b_2), \quad (\lambda : \mu) \in \mathbf{P}^1(\mathbf{C}). \tag{1.50}$$

We can derive this fact by observing that the line through P and Q can be expressed by

$$\begin{vmatrix} a_0 & b_0 & x_0 \\ a_1 & b_1 & x_1 \\ a_2 & b_2 & x_2 \end{vmatrix} = 0. \tag{1.51}$$

(Through P and Q there is a unique line L , and the line determined by (1.51) goes through P and Q. Hence L is given by (1.51).) Then, for any point $(c_0 : c_1 : c_2)$ on L we must have

$$\begin{vmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0,$$

which implies that we can write

$$(c_0, c_1, c_2) = \lambda(a_0, a_1, a_2) + \mu(b_0, b_1, b_2).$$

Now to find the intersections of the curve C and the line L, we substitute (1.50) in (1.49) and solve

$$F(\lambda a_0 + \mu b_0, \lambda a_1 + \mu b_1, \lambda a_2 + \mu b_2) = 0 \tag{1.52}$$

for the ratio $(\lambda : \mu)$. Since F is homogeneous of degree m, (1.52) is homogeneous of degree m relative to (λ, μ). This means that unless the left-hand side vanishes identically, the equation (1.52) has m solutions counting multiplicities. If $(\lambda_0 : \mu_0)$ appears with multiplicity k, the point $R = (\lambda_0 a_0 + \mu_0 b_0 : \lambda_0 a_1 + \mu b_1 : \lambda_0 a_2 + \mu_0 b_2)$ is counted k times. We say in this case that the curve C and the line L are **tangent with multiplicity** k at R.

In the case where the left-hand side of (1.52) vanishes identically, the curve C contains the line L. This means that the curve C is reducible, namely,

$$F(x_0, x_1, x_2) = G(x_0, x_1, x_2)H(x_0, x_1, x_2),$$

where

$$G(x_0, x_1, x_2) = \begin{vmatrix} a_0 & b_0 & x_0 \\ a_1 & b_1 & x_1 \\ a_2 & b_2 & x_2 \end{vmatrix}.$$

In the following we assume that C does not contain L. Thus C and L intersect at m points including multiplicities. Let us examine (1.52) in detail. Since F is a homogeneous polynomial, we get its Taylor expansion at $(\lambda a_0, \lambda a_1, \lambda a_2)$ in the following form:

$$\begin{aligned} &F(\lambda a_0 + \mu b_0, \lambda a_1 + \mu b_1, \lambda a_2 + \mu b_2) \\ &\quad = \lambda^m F(a_0, a_1, a_2) + \lambda^{m-1}\mu \Delta_b^{(1)} F(a) \\ &\qquad + \lambda^{m-2}\mu^2 \Delta_b^{(2)} F(a) + \cdots + \frac{\lambda \mu^{m-1}}{(m-1)!} \Delta_b^{(m-1)} F(a) \\ &\qquad + \frac{\mu^m}{m!} \Delta_b^{(m)} F(a), \end{aligned} \tag{1.53}$$

where for $z = (z_0, z_1, z_2)$

$$\Delta_z^{(1)} F = z_0 \frac{\partial}{\partial x_0} + z_1 \frac{\partial}{\partial x_1} + z_2 \frac{\partial}{\partial x_2}$$

and $\Delta_z^{(i)} F(y)$ means letting $\Delta_z^{(1)}$ operate on $F(x_0, x_1, x_2)\, i$ times and evaluating the result at $(x_0, x_1, x_2) = (y_0, y_1, y_2)$. (Exercise 1.4.) If we apply Taylor's formula at $(\mu b_0, \mu b_1, \mu b_2)$, then instead of (1.53) we get

$$\begin{aligned} &F(\lambda a_0 + \mu b_0, \lambda a_1 + \mu b_1, \lambda a_2 + \mu b_2) \\ &\quad = \mu^m F(b_0, b_1, b_2) + \mu^{m-1}\lambda \Delta_a^{(1)} F(b) \\ &\qquad + \mu^{m-2}\lambda^2 \Delta_a^{(2)} F(b) + \cdots + \frac{\mu \lambda^{m-1}}{(m-1)!} \Delta_a^{(m-1)} F(b) \\ &\qquad + \frac{\lambda^m}{m!} \Delta_a^{(m)} F(b). \end{aligned} \tag{1.54}$$

Since C does not contain L, we may take two points $P = (a_0 : a_1 : a_2)$ and $Q = (b_0 : b_1 : b_2)$ on L that do not lie on C, that is,

$$F(a_0, a_1, a_2) \neq 0, \quad F(b_0, b_1, b_2) \neq 0.$$

Thus both (1.53) and (1.54) are homogeneous of degree m relative to (λ, μ) and admit m zeros (λ_j, μ_j), $1 \leq j \leq m$, allowing multiplicities. We have thus verified that L and C have m intersections counting multiplicities.

Next, we consider the case where L and C intersect at $(a_0 : a_1 : a_2)$. In this case, we have from (1.53)

$$\lambda^{m-1}\mu \Delta_b^{(1)} F(a) + \lambda^{m-2}\mu^2 \Delta_b^{(2)} F(a) + \cdots + \frac{\mu^m}{m!} \Delta_b^{(m)} F(a) = 0. \tag{1.55}$$

Now if

$$\Delta_b^{(1)}F(a) = 0, \tag{1.56}$$

then L and C meet at $(a_0 : a_1 : a_2)$ with multiplicity at least 2, and thus L is tangent to C at $(a_0 : a_1 : a_2)$. In this argument, the point $(b_0 : b_1 : b_2)$ is free to move on the line L. Therefore from (1.56) we see that the equation of the tangent line is

$$\frac{\partial F}{\partial x_0}(a)x_0 + \frac{\partial F}{\partial x_1}(a)x_1\frac{\partial F}{\partial x_2}(a)x_2 = 0. \tag{1.57}$$

This is nothing but (1.24). For the tangent to make sense we must have

$$(\frac{\partial F}{\partial x_0}, \frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}) \neq (0,0,0).$$

If

$$(\frac{\partial F}{\partial x_0}(a), \frac{\partial F}{\partial x_1}(a), \frac{\partial F}{\partial x_2}(a)) = 0,$$

we say that $P = (a_0 : a_1 : a_2)$ is a **singular point** or **multiple point** of the curve C. In this case, the equation (1.57) does not make sense, but we may consider the situation as follows. The intersections of the line $L = \overline{PQ}$ with C can be found by solving (1.55). At a singular point $P = (a_0 : a_1 : a_2)$ we have $\Delta_b^{(1)}F(a) = 0$, and $(\lambda : \mu) = (1 : 0)$ is a multiple solution of (1.55). Thus for P, we have

$$\Delta_b^{(1)}F(a) = 0, \Delta_b^{(2)}F(a) = 0, \cdots, \Delta_b^{(n-1)}F(a) = 0,$$

but if

$$\Delta_b^{(n)}F(a) \neq 0$$

for some point $Q = (b_0 : b_1 : b_2)$, then we say that $P = (a_0 : a_1 : a_2)$ is an **n-fold point** or a **multiple point of order** n; n is called the **multiplicity** of the singular point P. Now then, if P is an n-fold point, then for a point $Q = (b_0 : b_1 : b_2)$ in general, $(\lambda : \mu) = (1 : 0)$ is a solution of (1.55) with multiplicity n . That is, the line $L = \overline{PQ}$ intersects C with multiplicity at least n at P. But for a special point Q, the line $L = \overline{PQ}$ intersects C at P with multiplicity $n+1$ at least. We may regard such a line L as tangent to C at the singular point P. In this case, the curve of degree n given by

$$\Delta_x^{(n)}F(a) = 0,$$

namely,

$$\sum_{i_0+i_1+i_2=n} \frac{n!}{i_0!i_1!i_2!}\frac{\partial^n F}{\partial x_0^{i_0}x_1^{i_1}x_2^{i_2}}(a_0,a_1,a_2)x_0^{i_0}x_1^{i_1}x_2^{i_2} = 0 \tag{1.58}$$

is called the **tangent cone** of C at P. This is a natural extension of (1.57). We examine several examples.

EXAMPLE 1.1. Let us consider a cubic curve

$$F = x_0x_1^2 + x_1^3 - x_2^2x_0 = 0.$$

For $P = (a_0 : a_1 : a_2)$ we have

$$\Delta_x^{(1)}F(a) = (a_1^2 - a_2^2)x_0 + a_1(2a_0 + 3a_1)x_1 - 2a_0a_2x_2$$
$$\Delta_x^{(2)}F(a) = 4a_1x_0x_1 + 2(a_0 + 3a_1)x_1^2 - 2a_0x_2^2 - 4a_2x_2x_0.$$

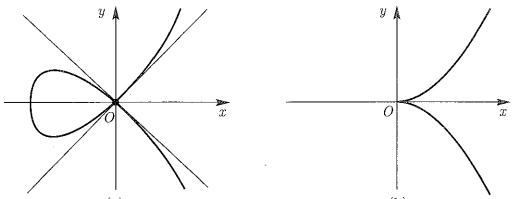


FIGURE 1.12 ORDINARY DOUBLE POINT AND ORDINARY CUSP. (a) If we set $x = x_1/x_0, y = x_2/x_0$, then $V(F)$ in the (x,y)-plane is represented by $y^2 = x^2(x+1)$. The two lines $y = \pm x$ form the tangent cone at the origin. (b) If we set $x = x_1/x_0, y = x_2/x_0$, then $V(G)$ in the (x,y)-plane is represented by $y^2 = x^3$. The origin is a singular point, and the x-axis counted doubly is the tangent cone at the origin.

Thus $P = (1 : 0 : 0)$ is a double point of the cubic curve $C = V(F)$, and the tangent cone at P is given by

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

This represents two lines, $x_1 - x_2 = 0$ and $x_1 + x_2 = 0$. This double point$P$ is called an **ordinary double point**. See Figure 1.12 (a).

EXAMPLE 1.2. Consider the cubic curve

$$G = x_1^3 - x_2^2x_0 = 0.$$

For a point $P = (a_0 : a_1 : a_2)$ we have

$$\Delta_x^{(1)}G(a) = -a_2^2x_0 + 3a_1^2x_1 - 2a_2a_0x_2$$
$$\Delta_x^{(2)}G(a) = 6a_1x_1^2 - 4a_2x_2x_0 - 2a_0x_2^2.$$

The point $P = (1 : 0 : 0)$ is a double point on the cubic curve $V(G)$, and the tangent cone at P is given by

$$x_2^2 = 0.$$

This means that we take the line $x_2 = 0$ doubly. The double point P is called an **ordinary cusp** (or simply a cusp). See Figure 1.12. (b).

It is known that a double point is either an ordinary double point (for which the tangent cone consists of two lines) or a cusp (for which the tangent cone is a line counted doubly). In the examples above, the tangent cones are described by lines. This is always valid, that is, the tangent cone at a k-fold point of a plane curve consists of at most k lines, in fact, exactly k lines counting multiplicities.

Now let us find the number of intersections of a plane curve of degree m in $\mathbf{P}^2(\mathbf{C})$

$$C : F(x_0, x_1, x_2) = 0$$

and a plane curve of degree n

$$D : G(x_0, x_1, x_2) = 0,$$

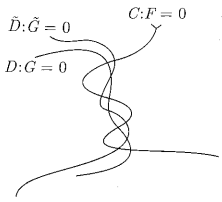


FIGURE 1.13. The curve $\tilde{D}$ determined by a variation $\tilde{G}$ of G may be thought of as having the same number of intersections with the curve C as D does.

counting multiplicities. The following theorem is fundamental.

THEOREM 1.1 (BÉZOUT'S THEOREM). *If a plane curve C of degree m and a plane curve D of degree n have no common component (that is, if the defining polynomials F and G have no common factor), then the number of intersections of C and D is mn, counting multiplicities.*

A rigorous proof of this theorem requires mathematical preparations of various kinds. The fact that proving intuitively obvious propositions requires much work has been a cause for increased bias toward algebraic geometry. In the past, we sometimes drew a wrong conclusion by counting the number of intersections while forgetting the fact that C and D have a common irreducible component. Reflections on such experiences have led to the development of algebraic geometry. It is often a practically difficult problem to decide whether C and D have a common irreducible component.

Suppose two plane curves C and D intersect at a point P. When we vary the coefficients of

$$G(x_0,x_1,x_2)=\sum_{i_0+i_1+i_2=n} b_{i_0i_1i_2}x_0^{i_0}x_1^{i_1}x_2^{i_2}$$

slightly, the locations of intersections may vary a little but the number of intersections will not change (see Figure 1.13). Therefore by varying the coefficients of G little by little we can deform G to the product of n distinct linear factors. By this variation D will change to n distinct lines. Since the number of intersections of a line and a plane curve of degree m is m, it follows that the number of intersections of C and D is equal to mn. Now this argument sounds correct but has a serious defect – it overlooks the case where C and D are reducible and contain a common plane curve. As is clear in the extreme case where C contains D, the number of intersections of C and D may be infinite and cannot be defined in a naive intuitive sense. Although the argument above might be correctable, we would have to exclude the exceptional cases that we mentioned.

How can we compute the **intersection multiplicity** $I_P(C,D)$ of C and D at P? This number is often called the **local intersection number** of C and D at P. We have to deal with the particular case where P is a singular point of C or D.

Using Bézout's theorem, let us study the intersections of the plane curves in Examples 1.1 and 1.2:

$$C: x_0x_1^2+x_1^3-x_2^2x_0=0 \tag{1.59}$$

$$D: x_1^3-x_2^2x_0=0. \tag{1.60}$$

The curves C and D are irreducible. From the equations, the intersection is on

$$x_0x_1^2=0.$$

If $x_0=0$, then (1.59) gives $x_1=0$ and the intersection is $(0:0:1)$. If $x_1=0$ and $x_0\neq 0$, (1.59) gives $x_2=0$ and the intersection is $(1:0:0)$. The point $P=(1:0:0)$ is a double point of C and D. At $Q=(0:0:1)$ the curves C and D are nonsingular; the tangent at Q is $x_0=0$ for both C and D, which are tangent at Q. By Bézout's theorem we have

$$I_P(C,D)+I_Q(C,D)=9. \tag{1.61}$$

In order to find $I_P(C,D)$ we transform (1.59) and (1.60) into

$$x^2+x^3-y^2=0$$
$$x^3-y^2=0$$

in terms of the inhomogeneous coordinates $x=x_1/x_0$, $y=x_2/x_0$. The point P corresponds to the origin $(0,0)$. By varying the coefficients of D slightly, we count the number of intersections with D in a neighborhood of the origin. Consider instead of $x^3-y^2=0$ the equation

$$x^3-y^2=-\epsilon, \text{ where } \epsilon>0 \text{ is small.}$$

We solve the system of equations

$$\begin{cases} x^2+x^3-y^2=0 \\ x^3-y^2=-\epsilon. \end{cases}$$

We get $x^2=\epsilon$ and hence $x=\pm\sqrt{\epsilon}$. Substituting this into the first equation, we get

$$y^2=\epsilon\pm\sqrt{\epsilon}.$$

Thus the solutions of the system provide 4 points

$$(\sqrt{\epsilon},\pm\sqrt{\epsilon+\epsilon\sqrt{\epsilon}}),\qquad (-\sqrt{\epsilon},\pm\sqrt{\epsilon-\epsilon\sqrt{\epsilon}}),$$

which all converge to the origin as $\epsilon\to 0$. We may thus obtain

$$I_P(C,D)=4.$$

From (1.61) we conclude that $I_Q(C,D)=5$. We shall verify this directly. We take inhomogeneous coordinates $u=x_0/x_2, v=x_1/x_2$ so that the point $Q=(0:0:1)$ corresponds to the origin. Then (1.59) and (1.60) can be rewritten

$$uv^2+v^3-u=0$$
$$v^3-u=0.$$

Using the same idea as above, we solve the system

$$\begin{cases} uv^2 + v^3 - u = 0 \\ v^3 - u = -\epsilon. \end{cases}$$

We get $uv^2 = \epsilon$. If $\epsilon \neq 0$, we have $uv \neq 0$. Substituting $u = \frac{\epsilon}{v^2}$ into the first equation, we obtain

$$v^5 + \epsilon v^2 - \epsilon = 0.$$

Let $\omega_i, 1 \leq i \leq 5$, be the roots of this equation. The system has 5 solutions $(\epsilon\omega_i^{-2}, \omega_i), 1 \leq i \leq 5$, confirming that $I_Q(C, D) = 5$. In the discussions above we varied D; we could have varied C or both. We get the same result if we use a different variation. In fact, we can take an infinitesimal variation and make the theory clearer. We shall come back to a rigorous theory of intersections in §2.3.

As we saw above, the problem of finding the number of intersections is intuitively clear, but to develop a rigorous theory is not as easy as it appears. Reflection on the intuitive treatment through the development of the theory has stimulated the vigorous development of algebraic geometry and gave rise to a rigorous algebraic treatment of the subject.

Now consider two distinct plane curves both of degree n

$$C : F(x_0, x_1, x_2) = 0$$
$$D : G(x_0, x_1, x_2) = 0.$$

If $(\lambda, \mu) \neq (0, 0)$, then

$$\lambda F(x_0, x_1, x_2) + \mu G(x_0, x_1, x_2) = 0$$

determines a plane curve $C_{(\lambda:\mu)}$; note that it is determined by the ratio $(\lambda : \mu)$ only. When $(\lambda : \mu)$ varies in $\mathbf{P}^1(\mathbf{C})$ the family of plane curves $\{C_{(\lambda:\mu)}\}_{(\lambda:\mu)\in\mathbf{P}^1(\mathbf{C})}$ is called a **pencil** of plane curves. When C and D have no common component, they intersect at n^2 points including multiplicities. These n^2 points are called the **base points** of the pencil, because every curve belonging to the pencil goes through these points.

By the way, a plane curve of degree n is determined by the ratio of the coefficients in its defining equation

$$\sum_{i_0+i_1+i_2=n} a_{i_0 i_1 i_2} x_0^{i_0} x_1^{i_1} x_2^{i_2} = 0.$$

Since the number of monomials of degree n in 3 variables is equal to

$$\binom{n+2}{2} = (n+2)(n+3)/2,$$

we may say that the plane curves of degree n have

$$(n+2)(n+1)/2 - 1 = n(n+3)/2$$

parameters. For a curve of degree n to pass through a point $P = (w_0 : w_1 : w_2)$, it is necessary that

$$\sum_{i_0+i_1+i_2=n} a_{i_0 i_1 i_2} w_0^{i_0} w_1^{i_1} w_2^{i_2} = 0.$$

This is a linear polynomial in $\{a_{i_0 i_1 i_2}\}$. The set of solutions is a subspace of codimension 1. By passing to the ratios, we see that the set of all plane curves of degree n passing through P has

$$n(n+3)/2 - 1$$

parameters. Continuing this argument, we might think that the family of plane curves of degree n passing through a given set of m points will have

$$n(n+3)/2 - m$$

parameters. That is, we are tempted to think that for each additional point to pass through, the resulting family has one less parameter. However, this idea is not correct as we can show by using the pencil of plane curves determined by C and D in our discussions above. They intersect at n^2 points, which every curve in the pencil $\{C_{(\lambda:\mu)}\}_{(\lambda:\mu)\in\mathbf{P}^1(\mathbf{C})}$ goes through. The pencil has at least one parameter. On the other hand, we have the obvious inequality

$$\frac{1}{2}n(n+3) - n^2 \leq 0 \text{ for } n \geq 3;$$

that is, something is wrong with the idea mentioned above. This fact was pointed out by Cramer in the early stage of the theory of algebraic curves (in the mid-18th century) and remained a paradox that plagued mathematicians for a long time. At present, we can explain this paradox as follows. For the plane curve of degree n defined by

$$\sum_{i_0+i_1+i_2=n} a_{i_0 i_1 i_2} x_0^{i_0} x_1^{i_1} x_2^{i_2} = 0$$

to pass through m points $(b_0^j : b_1^j : b_2^j)$, it is necessary that we have

$$\sum_{i_0+i_1+i_2=n} b_0^{(j)^{i_0}} b_1^{(j)^{i_1}} b_2^{(j)^{i_2}} a_{i_0 i_1 i_2} = 0, \qquad 1 \leq j \leq m. \tag{1.62}$$

Now consider this as a system of linear equations in $\{a_{i_0 i_1 i_2}\}$. If its rank (that is, the number of independent equations) is k, then the number of parameters is $n(n+3)/2 - k$. In general, we have $k \leq m$. For the pencil of plane curves $\{C_{(\lambda:\mu)}\}_{(\lambda:\mu)\in\mathbf{P}^1(\mathbf{C})}$, we have $k < n^2$, namely, the equations (1.62) are not all linearly independent. This explains the paradox. In algebraic geometry, the question whether a given system of equations is independent is often important. In the example above, finding the rank of the system (1.62) for m given points is reduced to the computation of minors. In many problems in algebraic geometry, we cannot compute the rank even in principle. We are then led to situations contrary to our intuition. This is one of the reasons why algebraic geometry gives the impression of being a difficult subject.

(b) Dual curves and Plücker's formula. It was the theory of Plücker (1801-68) that brought essential progress from the primitive coordinate-geometric treatment of plane curves. In this subsection, we give a cursory view of Plücker's theory.

For the equation of a line in $\mathbf{P}^2(\mathbf{C})$

$$a_0 x_0 + a_1 x_1 + a_2 x_2 = 0,$$

we can determine a point $(a_0 : a_1 : a_2)$ in $\mathbf{P}^2(\mathbf{C})$ from the ratio of the coefficients. Conversely, a point $(b_0 : b_1 : b_2)$ in $\mathbf{P}^2(\mathbf{C})$ determines a line

$$b_0x_0 + b_1x_1 + b_2x_2 = 0.$$

This type of correspondence between the lines and the points is the basis for the duality principle in projective geometry. The set of all lines in $\mathbf{P}^2(\mathbf{C})$ may be considered as a projective plane, which we call the **dual projective plane** of the original projective plane and denote it by $\mathbf{P}^2(\mathbf{C})^*$. That is, a point $(b_0 : b_1 : b_2)$ of $\mathbf{P}^2(\mathbf{C})^*$ represents a line

$$b_0x_0 + b_1x_1 + b_2x_2 = 0.$$

Now what object in $\mathbf{P}^2(\mathbf{C})$ does a line $\ell : c_0y_0+c_1y_1+c_2y_2 = 0$ in $\mathbf{P}^2(\mathbf{C})^*$ represent? Since the line ℓ is determined by joining two distinct points $P = (\alpha_0 : \alpha_1 : \alpha_2)$ and $Q = (\beta_0 : \beta_1 : \beta_2)$, the equation of ℓ can be expressed by

$$\begin{vmatrix} x_0 & \alpha_0 & \beta_0 \\ x_1 & \alpha_1 & \beta_1 \\ x_2 & \alpha_2 & \beta_2 \end{vmatrix} = 0,$$

that is, we have

$$c_0 : c_1 : c_2 = \begin{vmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{vmatrix} : \begin{vmatrix} \alpha_2 & \beta_2 \\ \alpha_0 & \beta_0 \end{vmatrix} : \begin{vmatrix} \alpha_0 & \beta_0 \\ \alpha_1 & \beta_1 \end{vmatrix}.$$

On the other hand, the lines in $\mathbf{P}^2(\mathbf{C})$ corresponding to the points P and Q in $\mathbf{P}^2(\mathbf{C})^*$ are

$$\begin{aligned} \alpha_0x_0 + \alpha_1x_1 + \alpha_2x_2 &= 0 \\ \beta_0x_0 + \beta_1x_1 + \beta_2x_2 &= 0, \end{aligned}$$

whose intersection is given by

$$\left(\begin{vmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{vmatrix} : \begin{vmatrix} \alpha_2 & \alpha_0 \\ \beta_2 & \beta_0 \end{vmatrix} : \begin{vmatrix} \alpha_0 & \alpha_1 \\ \beta_0 & \beta_1 \end{vmatrix} \right) = (c_0 : c_1 : c_2).$$

This means that to a line in $\mathbf{P}^2(\mathbf{C})^*$ there corresponds a point in $\mathbf{P}^2(\mathbf{C})$ determined by the ratios of the coefficients.

Let us now try to generalize the correspondence between the points and the lines. First, consider a plane curve of degree n

$$C : f(x_0, x_1, x_2) = 0$$

without singular points. For each point P of C, we take the tangent line to C and think of it as a point of the dual projective plane. In this way we have a mapping

$$\delta : (a_0 : a_1 : a_2) \in C \longmapsto \left(\frac{\partial F}{\partial x_0}(a) : \frac{\partial F}{\partial x_1}(a) : \frac{\partial F}{\partial x_2}(a) \right) \in \mathbf{P}^2(\mathbf{C})^*. \tag{1.63}$$

The image of δ turns out to be a plane curve in $\mathbf{P}^2(\mathbf{C})^*$, which we call the **dual curve** of C and denote by C^*. The degree ℓ of C^* is called the **class** of C. The curve C is said to have degree n and class ℓ.

EXAMPLE 1.3. Consider a plane curve of degree $n = 2, 3$

$$C : \quad x_0^n + x_1^n + x_2^n = 0.$$

The dual curve C^* is given by

$$\delta : (a_0 : a_1 : a_2) \in C \longmapsto (na_0^{n-1} : na_1^{n-1} : na_2^{n-1}) \in \mathbf{P}^2(\mathbf{C})^*,$$

and is hence obtained by rationalizing

$$(y_0^{1/(n-1)})^n + (y_1^{1/(n-1)})^n + (y_2^{1/(n-1)})^n = 0.$$

For $n = 2$ we get back the conic

$$y_0^2 + y_1^2 + y_2^2 = 0.$$

Thus C is a plane curve with degree 2 and class 2. For $n = 3$, the dual curve is

$$(y_0^3 + y_1^3 - y_2^3)^2 - 4y_0^3y_1^3 = 0.$$

Thus C is a plane curve of degree 3 and class 6.

In the discussion above we assumed that $C : F(x_0, x_1, x_2) = 0$ has no singular points. This is because at a singular point $P = (a_0 : a_1 : a_2)$ we have

$$\frac{\partial F}{\partial x_0}(a) = \frac{\partial F}{\partial x_1}(a) = \frac{\partial F}{\partial x_2}(a) = 0,$$

and the map (1.63) cannot be defined. However, it turns out that if C_{reg} denotes the set of all nonsingular points of C, then the image $\delta(C_{\text{reg}})$ by δ is the plane curve C^* with a finite number of points removed. In this case, we also call C^* the dual curve of C. This makes sense because the mapping

$$\delta : (a_0 : a_1 : a_2) \in C \longmapsto \left(\frac{\partial F}{\partial x_0}(a) : \frac{\partial F}{\partial x_1}(a) : \frac{\partial F}{\partial x_2}(a) \right) \in \mathbf{P}^2(\mathbf{C})^*$$

is defined by using the ratios of the homogeneous polynomials in a_0, a_1, and a_2. Such a mapping is called a **rational mapping**. Although δ is not defined for all points, we can treat it as if it were a mapping; this is because our objects of study are confined to figures and mappings defined by polynomials.

EXAMPLE 1.4. We want to find the dual curve of the plane curve of degree 3 with an ordinary double point

$$C : F = x_0x_1^2 + x_1^3 - x_2^2x_0 = 0.$$

The rational mapping is

$$\delta : (a_0, a_1 : a_2) \in C \longmapsto (a_1^2 - a_2^2 : 2a_0a_1 + 3a_1^2 : -2a_2a_0) \in \mathbf{P}^2(\mathbf{C})^*,$$

which is not defined at the point $(1 : 0 : 0)$ only.

We try to eliminate a_0, a_1, a_2 from the equations

$$y_0 = a_1^2 - a_2^2, \quad y_1 = 2a_0a_1 + 3a_1^2, \quad y_2 = -2a_2a_0$$

$$a_0a_1^2 + a_1^3 - a_2^2a_0 = 0.$$

For this purpose, we set

$$X = \frac{y_0}{y_2}, \quad Y = \frac{y_1}{y_2}$$

$$x = \frac{a_1}{a_0}, \quad y = \frac{a_2}{a_0},$$

and find that

$$\begin{aligned} y^2 &= x^2(x+1) \\ X &= -\frac{1}{2}\left(\frac{x^2}{y} - y\right) = -\frac{1}{2y}(x^2 - y^2) \\ Y &= -\left(\frac{x}{y} + \frac{3}{2}\cdot\frac{x^2}{y}\right) = -\frac{x}{2y}(2+3x). \end{aligned}$$

From the first equation we get

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

Setting $w = y/x$, we have

$$\begin{aligned} x &= w^2 - 1 \\ y &= w(w^2 - 1). \end{aligned}$$

Hence

$$\begin{aligned} 2X &= \frac{(w^2-1)^2}{w} \\ 2Y &= -\frac{3w^2 - 1}{w}. \end{aligned}$$

To make the computation easier to follow, we set

$$u = 2X, \quad v = 2Y, \quad z = w^2;$$

we have

$$\begin{aligned} u^2 &= \frac{(z-1)^4}{z} \\ uv &= \frac{-(z-1)^2(3z-1)}{z} \\ v^2 &= \frac{(3z-1)^2}{z}. \end{aligned}$$

It follows that

$$uv^3 - 36uv + 8v^2 = -\frac{(3z-1)^3(z+1)}{z} = -27z^3 + 18z - 8 + \frac{1}{z}.$$

On the other hand, we have

$$27u^2 - 36uv + 8v^2 = 27z^3 - 18z + 24 - \frac{1}{z}.$$

We thus have

$$uv^3 - v^4 + 27u^2 - 36uv + 8v^2 - 16 = 0.$$

Using X, Y we get

$$4XY^3 - 4Y^4 + 27X^2 - 36XY + 8Y^2 - 4 = 0.$$

Rewriting this equation in homogeneous coordinates, we get

$$C^* : \quad 4y_0y_1^3 - 4y_1^4 + y_2^2(27y_0^2 - 36y_0y_1 + 8y_1^2 - 4y_2^2) = 0,$$

which is of degree 4. Thus the curve C has degree 3 and class 4.

EXAMPLE 1.5. Take a plane curve

$$C : G = x_1^3 - x_2^2x_0 = 0$$

with an ordinary cusp. The dual curve C^* is obtained from the rational map

$$\delta : (a_0 : a_1 : a_2) \in C \longmapsto (-a_2^2 : 3a_1^2 : -2a_2a_0) \in \mathbf{P}^2(\mathbf{C})^*$$

by eliminating a_0, a_1, a_2 from

$$y_0 = -a_2^2, \quad y_1 = 3a_1^2, \quad y_2 = -2a_2a_0.$$

By a simple computation we get

$$C^* : y_0y_2^2 + \frac{4}{27}y_1^3 = 0.$$

Thus C has degree 3 and class 3.

The degree n of a plane curve C is also the number of intersections of C with a line. The class of C is the degree of the dual curve C^* and hence equal to the number of intersections of C^* with a line in $\mathbf{P}^2(\mathbf{C})$. But we know that a line ℓ^* in $\mathbf{P}^2(\mathbf{C})$ corresponds to a point P in $\mathbf{P}^2(\mathbf{C})$, and the intersection Q^* of ℓ^* and C^* corresponds to the tangent line of C in $\mathbf{P}^2(\mathbf{C})$. It follows that the class of C is equal to the number of tangent lines from a general P to the curve C. If we set $P = (b_0 : b_1 : b_2)$, then the tangent line to C at $Q = (a_0 : a_1 : a_2)$ on C goes through P if

$$\Delta_b^{(1)}F(a) = 0,$$

where $F(x) = 0$ is the defining equation of C. Suppose $Q = (a_0 : a_1 : a_2)$ is a solution of the system of equations

$$\begin{cases} F(x) = 0 \\ \Delta_b^{(1)}F(x) = 0. \end{cases}$$

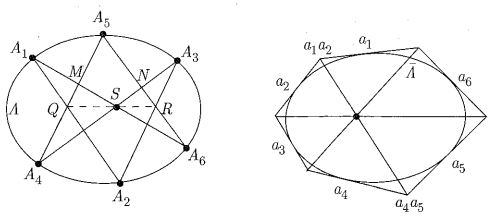
If Q is not a singular point of C, then the tangent line to C at Q is given by

$$\Delta_x^{(1)}F(a) = 0.$$

Thus we see that the number of tangent lines through P does not exceed the number of solutions of the system. The equation

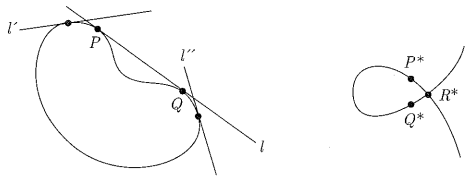
$$\Delta_b^{(1)}F(x) = 0$$

is a plane curve of degree $m-1$, where m is the degree of $F(x)$. The curve is called the **polar curve** of C. Since the point P is general, the curve C and the polar curve of C have no common components, and the number of intersections of C and its polar curve is $m(m-1)$. Therefore the class of C is at most $m(m-1)$. In particular, if C has no singular point, its class is $m(m-1)$. As we stated before, the tangent line at a singular point R should be regarded as a line appearing as a component of the tangent cone. If we take a point P outside C in general position, the line joining P and R is not tangent to C. On the other hand, we can show that the curve C and its polar curve intersect at a singular point with multiplicity greater than or equal to 2. In this way, when the curve C has a singular point, its class is less than $m(m-1)$. More precisely, we have the following result.



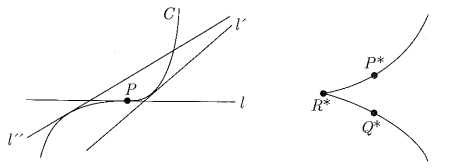
(a) (b)

FIGURE 1.14. (a) Pascal's theorem, (b) Brianchon's theorem.



(a) (b)

FIGURE 1.15. A bitangent corresponds to an ordinary double point of the dual curve. P^*: the point on $\mathbf{P}^2(\mathbf{C})^*$ corresponding to ℓ', Q^*: the point $\mathbf{P}^2(\mathbf{C})^*$ corresponding to ℓ'', R^*: the point in $\mathbf{P}^2(\mathbf{C})^*$ corresponding to ℓ.



(a) (b)

FIGURE 1.16. An inflection point corresponds to a cusp of the dual curve. (a) At an inflection point P, the tangent ℓ intersects C triply. (b) P^*, Q^*, R^* are the points in $\mathbf{P}^2(\mathbf{C})^*$ corresponding to ℓ', ℓ'', ℓ, respectively.

THEOREM 1.2 (PLÜCKER'S FORMULA).. *Assume that a plane curve C of degree m has s ordinary double points and t ordinary cusps as its singular points. Then the class of C is equal to $m(m-1)-2s-3t$.*

The computational results in Examples 1.3, 1.4, and 1.5 confirm Theorem 1.2 for these curves. Now we may ask: what is the dual curve $(C^*)^*$ of the dual curve C^*? We have the following remarkable result.

THEOREM 1.3. *The dual curve $(C^*)^*$ of the dual curve C^* coincides with C.*

This theorem completes the duality principle. The principle says that if a proposition concerning points, lines, and plane curves in $\mathbf{P}^2(\mathbf{C})$ is true, so is the dual proposition concerning lines, points, and dual plane curves in the dual projective plane $(\mathbf{P}^2(\mathbf{C})^2)^*$. For example, Pascal's theorem and Brianchon's theorem are dual to each other. (See Figure 1.14.)

So far we have studied plane curves from a fairly naive viewpoint. Our last observation is about singular points of the dual curve. When a plane curve C has no singular point, the dual curve C^* may still have singular points. For example, if the tangent line ℓ to C at a point P is also tangent to C at another point Q, we call ℓ a **bitangent**. In this case, we have $\delta(P) = \delta(Q)$ and a singular point of C^* appears. Since the tangent lines to C near P and near Q are distinct from each other, the singular point is in general an ordinary double point (see Figure 1.15). When the tangent ℓ at P to C has multiplicity 3, that is, $I_P(\ell, C) = 3$, we call P an **inflection point** of C. (When $I_P(\ell, C) \geq 4$, P is an inflection point of higher order.) In this case, the corresponding point R^* on the dual curve is a cusp (see Figure 1.16). This may not be easy to see intuitively; however, one may guess the result from the fact that the slope of the tangent near R^* to C^* corresponds to the way the tangent line to C changes near P.

(c) The development of algebraic geometry. We have given a classical treatment of plane curves. Though some of the things are intuitively clear, the reader may have felt that the discussions in detail are complicated and that there may be some logical difficulties. To give a rigorous development of the arguments given so far, we need algebraic tools such as the theory of commutative rings. Anyway, the arguments so far are elementary-geometric. It was the theory of Abelian integrals by Riemann (1826-66) that radically changed the viewpoint in algebraic geometry and built the foundation for its new development.

Around that time the integration of algebraic functions was one of the central problems in mathematics. The simplest example is an elliptic integral

$$\int \frac{dx}{\sqrt{x^3 - g_2x - g_3}}.$$

More generally, under the restriction

$$f(x, y) = 0, \tag{1.64}$$

that is, when y is regarded as a multi-valued function of x, consider the integral

$$\int \frac{g(x,y)dx}{h(x,y)} \tag{1.65}$$

for polynomials g and h. It was Abel (1802-29) who first noticed that the integral has a deep relation with the geometric properties of the curve $f(x, y) = 0$. Riemann went beyond Abel in recognizing that (1.65) is an integral over the figure defined by (1.64), namely, the Riemann surface. He understood the Riemann surface as a covering surface of the projective line $\mathbf{P}^1(\mathbf{C})$ (that is, consider x as a point of $\mathbf{P}^1(\mathbf{C})$

and get a finite number of solutions y of $f(x,y)=0$, thus determining a covering surface of $\mathbf{P}^1(\mathbf{C})$). By developing complex function theory over a Riemann surface Riemann clarified integration theory. Riemann's theory of Abelian integrals was far ahead of its time in dealing with the fundamental properties of Riemann surfaces and in introducing theta functions. His viewpoint was to consider plane curves as one-dimensional complex manifolds, namely, Riemann surfaces, and it led to very important results in algebraic geometry as well. One of his basic discoveries was that a Riemann surface is uniquely determined by the set of all rational functions on it (called the field of rational functions). This was the birth of **birational geometry**. We shall give detailed discussions of this viewpoint. According to Riemann, two different equations define the same Riemann surface if the fields of rational functions determined by the equations are isomorphic. This might be considered as a generalization of the statement in §1.4 (b) that a projective line and a conic can be identified. Following Riemann's viewpoint, the Riemann surfaces defined by

$$\begin{aligned} y^2 - x^3 &= 0 \\ y^2 - x^2(x+1) &= 0 \\ y - \alpha x &= 0 \end{aligned}$$

are the same. As we have already seen, these plane curves have delicate differences. The curve defined by the first equation has a cusp, the second has an ordinary double point, and the third is a projective line. If we resolve the singular points of the first and the second, they become isomorphic to the projective line (see §2.5). In fact, it is known that the Riemann surface determined by $f(x,y)=0$ is obtained by resolving the singular points of the plane curve defined by $f(x,y)=0$. This is an important result. The significant contribution by Riemann was that the defining equations for plane curves define the same geometric object if their fields of rational functions are the same, and that these geometric objects have a rich structure, which he studied by the techniques of complex function theory. We shall give a simple description of his techniques in Chapter 4.

Algebraic geometry has made much progress since Riemann's theory of Abelian integrals. In particular, Max Noether (1844-1921) and Clebsch (1833-72) gave an algebraic and geometric treatment of Riemann's theory. Dedekind (1831-1916) and Weber (1842-1913) completed the framework for the algebraic theory of algebraic functions, which emphasized an analogy with number theory while the geometric viewpoint receded. Later, Weil (1906-) reorganized algebraic geometry once more in order to solve Riemann's conjecture on congruence zeta functions. We discuss congruence zeta functions in the case of algebraic curves in §3.4.

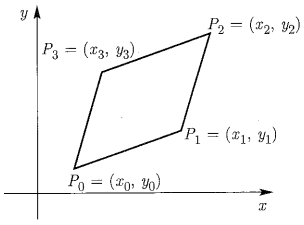
On another front, the object of study in algebraic geometry extended from curves to algebraic surfaces in the period from the late 19th century to the early 20th century, except that intersection theory remained on shaky ground with too much reliance on intuition, as explained in §1.5 (a). It was in the 1940s that algebraic geometry was placed on mathematically strong foundations, thanks to Weil and Zariski (1899-1986). Zariski aimed at placing the intuitive theory of algebraic surfaces of the Italian School on rigorous foundations, and Weil wanted to construct an algebraic geometry that would be sufficient to prove Riemann's conjecture on congruence zeta functions for algebraic curves. Independently in the 1940s, they succeeded in constructing algebraic geometry not only over the complex field $\mathbf{C}$ but also over fields of characteristic p. This was the birth of algebraic geometry in its true sense. Serre (1926-) introduced a still new viewpoint by using the sheaf theory originally developed for the theory of functions of several complex variables. Grothendieck (1928-) used Serre's theory to construct algebraic geometry over an arbitrary commutative ring in order to solve Riemann's (Weil's) conjecture for congruence zeta functions on arbitrary algebraic varieties. Weil's conjecture was finally solved by Deligne (1944-). Grothendieck's theory of schemes (see the box at the end of this chapter) is the most natural approach to algebraic geometry conveniently organized for applications, but it requires very advanced mathematical tools. The theories of Weil and Zariski, Serre's theory, and Grothendieck's theory have different outlooks but the same purpose, namely, to study the geometry of figures defined algebraically. To convey its essence will not require big theories. We want to put the emphasis on the geometric viewpoint and develop an introduction to algebraic geometry.

Problems

1.1. (i) When the coordinates of the four vertices P_0, P_1, P_2, P_3 are as indicated in the figure, show that the area of the parallelogram is given by the determinant

$$\begin{vmatrix} x_1 - x_0 & y_1 - y_0 \\ x_3 - x_0 & y_1 - y_0 \end{vmatrix}.$$

Here we assume that the orientation of the parallelogram is positive, that is, the interior of the parallelogram is on your left as one proceeds in the order P_0, P_1, P_2, P_3. (If the orientation is negative, the area is the negative of the determinant above.)



(ii) Show that for an affine transformation

$$g : (x, y) \mapsto (a_{11}x + a_{12}y + b, a_{21}x + a_{22}y + c),$$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \det A \neq 0,$$

the area of the parallelogram $g(P_0)g(P_1)g(P_2)g(P_3)$ is $|\det A|$ times the area of the parallelogram $P_0P_1P_2P_3$.

1.2. Carry out a complete affine classification of the quadratic curves (1.8). State conditions under which we can obtain $x^2 + y^2 = 1$, $x^2 - y^2 = 1$, or $y = x^2$ by a suitable affine transformation.

1.3. For a homogeneous polynomial $f(x_0, x_1, \ldots, x_n)$ of degree m, prove that

$$\sum_{i=0}^{n} x_i \frac{\partial f}{\partial x_i} = mf.$$

This is called **Euler's identity**.

1.4. Prove the Taylor formula (1.53) for a homogeneous polynomial.

1.5. For the dual curves C^*

$$(y_0^3 + y_1^3 - y_2^3)^2 - 4y_0^3y_1^3 = 0$$
$$4y_0y_1^3 - 4y_1^4 + y_2^2(27y_0^2 - 36y_0y_1 + 8y_1^2 - 4y_2^2) = 0$$
$$y_0y_2^2 + \frac{4}{27}y_1^3 = 0,$$

that were computed in Examples 1.3, 1.4, and 1.5, show that their dual curves are, respectively,

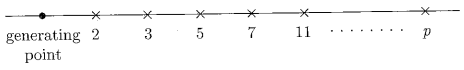
$$x_0^3 + x_1^3 + x_2^3 = 0$$
$$x_0x_1^2 + x_1^3 - x_2^2x_0 = 0$$
$$x_1^3 - x_2^2x_0 = 0.$$

Grothendieck's scheme theory

In scheme theory all commutative rings can be regarded as geometric objects. Prime ideals of a commutative ring are points of a scheme.

(1) Spec **Z**.

The scheme Spec **Z** corresponding to the set of all integers **Z** consists of points corresponding to prime numbers and a generating point.



(2) Frey curves.

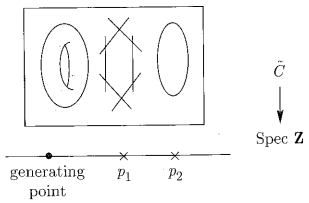
For a certain prime number p assume that there is a triple $(a, b, c) \neq (0, 0, 0)$ of relatively prime integers satisfying

$$a^p + b^p = c^p,$$

that is, the Fermat conjecture is not valid. If we take a cubic curve

$$C : x_0 x_2^2 - x_1(x_1 - a^p x_0)(x_1 + b^p x_0) = 0$$

in the projective plane, we may think of it as a scheme over **Z**, because the defining equation has integral coefficients. (For each prime q, reduce mod q the coefficients of the defining equation of C and get a cubic curve over the finite field $\mathbf{F}_q$. Now by varying q through all primes, we may regard C as a scheme over Z.) This is called the Frey curve. It is a nonsingular cubic curve for primes that do not divide a, b, c, but C has singularities for primes that divide a, b, c. If we remove singularities, we have rings of projective lines. This curve plays an important role in the solution of the Fermat conjecture (Fermat's last theorem).



CHAPTER 2

Projective Spaces and Projective Varieties

In this chapter we introduce the projective spaces in which algebraic geometry is developed, and then introduce algebraic varieties. From the preceding chapter the reader must by now have some idea of the role the complex projective spaces play. In this chapter we shall introduce projective spaces once again. We first define the projective line and clarify its fundamental properties. Then we proceed to introduce two-dimensional projective space, namely, the projective plane, and discuss plane curves again. At this point it will be obvious how to define projective spaces in general. The definition of projective varieties is also intuitively clear, but a rigorous definition will require knowledge from algebra. Here we assume as few prerequisites as possible and provide the minimum. Next, we touch upon the resolution of singularities for plane curves. This is one of the fundamental ideas in algebraic geometry.

For simplicity, we discuss properties over the complex number field **C**. However, the arguments below are essentially algebraic and should not be limited to **C** only.

§ 2.1. The projective line

(a) The Riemann sphere and the projective line. Let us view the well-known Riemann sphere in complex function theory from a different angle. The Riemann sphere is the complex plane **C** with a point at infinity ∞ added. If z is the coordinate in **C**, then $w = 1/z$ is the coordinate around the point at infinity. For example, the polynomial $z^3 + 2z^2 + 3z + 1$ is regular on **C**. In a neighborhood of ∞ it is given by

$$w^{-3} + 2w^{-2} + 3w^{-1} + 1$$

and has a pole of order 3 at ∞. The exponential function

$$e^z = \sum_{n=0}^{\infty} z^n/n!$$

is expressed around ∞ by

$$\sum_{n=0}^{\infty} w^{-n}/n!,$$

which has ∞ as an essential singularity. Introducing a point at infinity may seem a convenient but artificial device to many readers. We shall therefore introduce the Riemann sphere by a different method.

Let us denote by $\mathbf{C}^2$ the set of all pairs (a, b) of complex numbers and set $W = \mathbf{C}^2 - \{(0, 0)\}$. For an arbitrary element (a_0, a_1) in W, we consider the ratio $a_0 : a_1$ and denote by $\mathbf{P}^1(\mathbf{C})$ the set of all ratios. We regard $\mathbf{P}^1(\mathbf{C})$ as a "figure",

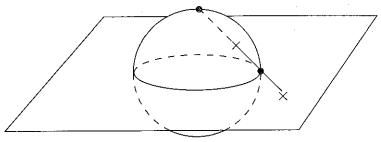


FIGURE 2.1 THE RIEMANN SPHERE AND THE COMPLEX PLANE

called the **complex projective line**, whose "points" are denoted by $(a_0 : a_1)$. If α is a nonzero complex number, we have by definition

$$(a_0 : a_1) = (\alpha a_0 : \alpha a_1);$$

thus it is possible to express a point $(a_0 : a_1)$ in many ways.

Now let us define two subsets U_0 and U_1 by

$$U_0 = \{(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) | a_0 \neq 0\}$$
$$U_1 = \{(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) | a_1 \neq 0\}.$$

Consider U_0. For any point $(a_0 : a_1) \in U_0$, we have

$$(a_0 : a_1) = \left(1 : \frac{a_1}{a_0}\right),$$

that is, an element of U_0 can be written in the form $(1, a)$. Conversely, for any complex number $a \in \mathbf{C}$, we have $(1 : a) \in \mathbf{P}^1(\mathbf{C})$. Moreover, the mapping

$$\phi_0 : (a_0, a_1) \longrightarrow \frac{a_1}{a_0} \in \mathbf{C}$$

is a bijection of U_0 onto $\mathbf{C}$. The inverse mapping ϕ_0^{-1} is given by

$$\phi_0^{-1} : a \in \mathbf{C} \longrightarrow (1 : a) \in U_0.$$

We may identify U_0 with the complex plane $\mathbf{C}$ in this way.

Similarly for U_1, the mapping

$$\phi_1 : (a_0, a_1) \longrightarrow \frac{a_0}{a_1} \in \mathbf{C}$$

is a bijection, and its inverse is given by

$$\phi_1^{-1} : b \in \mathbf{C} \longrightarrow (b : 1) \in U_1.$$

Now a point $(a_0 : a_1)$ belongs to U_0 or U_1 depending on whether $a_0 \neq 0$ or $a_1 \neq 0$. Thus

$$\mathbf{P}^1(\mathbf{C}) = U_0 \cup U_1.$$

Furthermore, for a point $(a_0 : a_1)$ in the intersection $U_0 \cap U_1$, we have

$$(a_0 : a_1) = \left(1 : \frac{a_1}{a_0}\right) = \left(\frac{a_0}{a_1} : 1\right).$$

Hence we see that

$$U_0 - U_0 \cap U_1 = \{(1 : 0)\}$$
$$U_1 - U_0 \cap U_1 = \{(0 : 1)\}$$
$$\mathbf{P}^1(\mathbf{C}) = U_0 \cup \{(0 : 1)\} = U_1 \cup \{(0 : 1)\}.$$

On the other hand, U_0 and U_1 are identified with the complex plane $\mathbf{C}$ by the mappings ϕ_0 and ϕ_1, respectively. In the following, we shall mainly think about U_0, identify it with $\mathbf{C}$, and set

$$\mathbf{P}^1(\mathbf{C}) = \mathbf{C} \cup \{(0 : 1)\}.$$

Then what can we say about the point $(0 : 1)$?

To see this, let us take a sequence of complex numbers $z_1, z_2, z_3, \ldots$ such that

$$\begin{cases} z_\nu \neq 0 \\ \lim\limits_{\nu \to \infty} z_\nu = 0. \end{cases} \tag{2.1}$$

Then we have

$$\lim_{\nu \to \infty} (z_\nu : 1) = (0 : 1).$$

Since $z_\nu \neq 0$, we have $(z_\nu : 1) \in U_0$ and

$$\phi_0((z_\nu : 1)) = \frac{1}{z_\nu}.$$

On the other hand, as $\nu \to \infty$, $1/z_\nu$ approaches the point at infinity ∞ on the Riemann sphere. Moreover, this result holds no matter what sequence $\{z_\nu\}$ we take as long as (2.1) is satisfied. In this way, we may regard the point $(0 : 1)$ of $\mathbf{P}^1(\mathbf{C})$ as the point at infinity ∞ of the Riemann sphere, so that we have

$$\mathbf{P}^1(\mathbf{C}) = \mathbf{C} \cup \{\infty\},$$

that is, we can identify the complex projective line $\mathbf{P}^1(\mathbf{C})$ with the Riemann sphere. Furthermore, we may check that this identification includes the correspondence of coordinates at the point at infinity.

Recall that the identification of $\mathbf{C}$ and U_0 is given by

$$z \in \phi_0^{-1} : \mathbf{C} \longrightarrow (1 : z) \in U_0.$$

As long as $z \neq 0$, we have $(1 : z) = (\frac{1}{z} : 1)$. It follows that $w = 1/z$ is the coordinate around the point at infinity $\infty = (0 : 1)$. Thus we can regard $\mathbf{P}^1(\mathbf{C})$ as the Riemann sphere itself.

What is the significance of the concept of the complex projective line $\mathbf{P}^1(\mathbf{C})$? As long as we are interested in complex function theory, we don't necessarily need this new interpretation of the Riemann sphere. If we look back at the definition of the complex projective line $\mathbf{P}^1(\mathbf{C})$, however, the definition itself is algebraic and it is not essential that we deal with complex numbers. For example, by starting with the pair $\mathbf{R}^2$ of the real number system $\mathbf{R}$, we can define the **real projective line** $\mathbf{P}^1(\mathbf{R})$ as the set of all ratios $(a_0 : a_1)$ other than $(0 : 0)$. Also in this case, we have

$$\mathbf{P}^1(\mathbf{R}) = \mathbf{R} \cup \{(0 : 1)\}$$

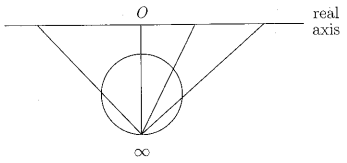


FIGURE 2.2 CORRESPONDENCE BETWEEN $\mathbf{P}^1(\mathbf{R})$ AND $\mathbf{R}$. Whether you go infinitely far to the right or to the left on the real number line, your will reach the point at infinity ∞.

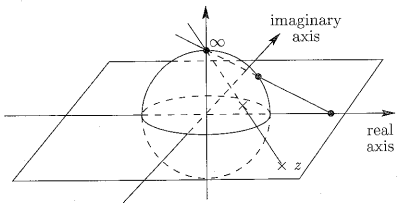


FIGURE 2.3. The points on the Riemann surface corresponding to the real axis and the point at infinity make up $\mathbf{P}^1(\mathbf{R})$.

by the same arguments as before. Here we identify a point a in $\mathbf{R}$ with the point $(1 : a)$ of $\mathbf{P}^1(\mathbf{R})$. For a sequence of real numbers $\{a_\nu\}$ such that $a_\nu > 0$ and $\lim_{\nu\to\infty} a_\nu = +\infty$,

$$(1 : a_\nu) = \left(\frac{1}{a_\nu} : 1\right)$$

implies

$$\lim_{\nu\to\infty} (1 : a_\nu) = (0 : 1);$$

for a sequence of real numbers $\{b_\nu\}$ such that $b_\nu < 0$ and $\lim_{\nu\to\infty} b_\nu = -\infty$, and

$$(1 : b_\nu) = (\frac{1}{b_\nu} : 1)$$

implies once again

$$\lim_{\nu\to\infty} (1 : b_\nu) = (0 : 1).$$

It follows that we may consider $\mathbf{P}^1(\mathbf{R})$ as a circle (see Figure 2.2). Since $\mathbf{R} \subset \mathbf{C}$, we may naturally consider $\mathbf{P}^1(\mathbf{R}) \subset \mathbf{P}^1(\mathbf{C})$. When we identify $\mathbf{P}^1(\mathbf{C})$ with the Riemann sphere, $\mathbf{P}^1(\mathbf{R})$ corresponds to the real axis of the complex plane and the point at infinity (see Figure 2.3).

This being said, we can introduce coordinates and hence the structure of a differentiable or complex manifold in $\mathbf{P}^1(\mathbf{R})$ or $\mathbf{P}^1(\mathbf{C})$, respectively, by means of (U_0, ϕ_0) and (U_1, ϕ_1). This important viewpoint will guide us in developing the theory in Chapter 4. In algebraic geometry, however, we treat $\mathbf{P}^1(\mathbf{R})$ and $\mathbf{P}^1(\mathbf{C})$

from a different viewpoint which allows us to introduce the projective line over an arbitrary commutative field k. Here the arguments necessarily become algebraic.

(b) Projective transformations. Complex function theory tells us that an automorphism (i.e. a conformal transformation) of the Riemann sphere is a linear fractional transformation

$$f(z) = \frac{az+b}{cz+d}, \quad ad - bc \neq 0.$$

Let us note that we can also define it algebraically as follows. That is, for any point $(x_0 : x_1)$ in the complex projective line $\mathbf{P}^1(\mathbf{C})$, consider the correspondence

$$(x_0 : x_1) \longmapsto (a_{00}x_0 + a_{01}x_1 : a_{10}x_0 + a_{11}x_1). \tag{2.2}$$

Since

$$(\alpha x_0 : \alpha x_1) = (x_0 : x_1),$$

we see that

$$\begin{aligned}&(a_{00}(\alpha x_0) + a_{01}(\alpha x_1) : a_{10}(\alpha x_0) + a_{11}(\alpha x_1))\\ &= (\alpha(a_{00}x_0 + a_{01}x_1) : (\alpha(a_{10}x_0 + a_{11}x_1))\\ &= (a_{00}x_0 + a_{01}x_1 : a_{10}x_0 + a_{11}x_1),\end{aligned}$$

which shows that the point $(a_{00}x_0 + a_{01}x_1 : a_{10}x_0 + a_{11}x_1)$ in (2.2) is uniquely defined independently of the expression for the point $(x_0 : x_1)$. However, if

$$\begin{cases} a_{00}x_0 + a_{01}x_1 = 0 \\ a_{10}x_0 + a_{11}x_1 = 0, \end{cases} \tag{2.3}$$

then we do not get any point in $\mathbf{P}^1(\mathbf{C})$. Now for (2.3) to fail at each point $(x_0 : x_1)$ in $\mathbf{P}^1(\mathbf{C})$ it is necessary and sufficient that

$$\begin{vmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{vmatrix} \neq 0. \tag{2.4}$$

When this condition holds, for any element (y_0, y_1) in $W = \mathbf{C}^2 - \{(0,0)\}$ there is a unique element (x_0, x_1) in W satisfying

$$\begin{aligned} y_0 &= a_{00}x_0 + a_{01}x_1 \\ y_1 &= a_{10}x_0 + a_{11}x_1. \end{aligned}$$

Under condition (2.4) it follows that (2.2) defines a bijection of $\mathbf{P}^1(\mathbf{C})$ onto itself. If furthermore $x_0 \neq 0$ and $a_{00}x_0 + a_{01}x_1 \neq 0$, then by setting $z = x_1/x_0$ we may rewrite (2.2) in the form

$$\begin{aligned}(x_0 : x_1) = (1, z) &\longmapsto (a_{00}x_0 + a_{01}x_1, a_{10}x_0 + a_{11}x_1)\\ &= \left(1 : \frac{a_{11}z + a_{10}}{a_{01}z + a_{00}}\right).\end{aligned}$$

Thus (2.2) is nothing but a linear fractional transformation

$$F(z) = \frac{a_{11}z + a_{10}}{a_{01}z + a_{00}}. \tag{2.5}$$

In the case where $x_0 = 0$, we see that the point $(0 : 1)$ goes to the point $(a_{01} : a_{11})$ by (2.2). If $a_{01} \neq 0$, then it corresponds to a_{11}/a_{01}, and if $a_{01} = 0$, then we get

the point $(0:1)$. Needless to say, this is the same point as the image $F(\infty)$ of the point at infinity ∞ by (2.5). Furthermore, when $a_{01} \neq 0$, the image $F(z)$ of $z = -a_{00}/a_{01}$ is ∞. This corresponds to the fact that the image by (2.2) of the point $(a_{01} : -a_{00})$ is equal to $(0:1)$. We have thus seen that (2.2) is a convenient expression for a linear fractional transformation in that it treats all points on an equal basis.

REMARK. In the discussions above we did not use the fact that we are dealing with complex numbers. Thus they are valid for the projective line $\mathbf{P}^1(k)$ over any field k.

DEFINITION 2.1. If complex numbers a_{00}, a_{01}, a_{10}, and a_{11} satisfy

$$\begin{vmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{vmatrix} \neq 0,$$

we call the transformation of $\mathbf{P}^1(\mathbf{C})$ into itself

$$(x_0 : x_1) \longmapsto (a_{00}x_0 + a_{01}x_1 : a_{10}x_0 + a_{11}x_1) \tag{2.6}$$

a **projective transformation**.

We often call the projective transformation given by (2.6) the projective transformation determined by the matrix

$$\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}.$$

For a nonzero element $\alpha \in \mathbf{C}$, it is almost evident that the two matrices

$$\alpha \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} \text{ and } \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}$$

determine the same projective transformation. The importance of using matrices is due to the following fact.

LEMMA 2.1. *Let f_A and f_B be the projective transformations determined by 2×2 complex matrices A and B with $\det A \neq 0$ and $\det B \neq 0$. Then the composition $f_A \circ f_B$ is the projective transformation corresponding to the matrix AB. In other words, to the composition of projective transformations there corresponds the product of matrices.*

PROOF. Set

$$A = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \quad B = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}.$$

Then

$$\begin{aligned} f_A \circ f_B((x_0 : x_1)) &= f_A((b_{00}x_0 + b_{01}x_1 : b_{10}x_0 + b_{11}x_1)) \\ &= (a_{00}(b_{00}x_0 + b_{01}x_1) + a_{01}(b_{10}x_0 + b_{11}x_1) : \\ &\quad a_{10}(b_{00}x_0 + b_{01}x_1) + a_{11}(b_{10}x_0 + b_{11}x_1)) \\ &= ((a_{00}b_{00} + a_{01}b_{10})x_0 + (a_{00}b_{01} + a_{01}b_{11})x_1 : \\ &\quad (a_{10}b_{00} + a_{11}b_{10})x_0 + (a_{10}b_{01} + a_{11}b_{11})x_1). \end{aligned}$$

On the other hand, we have

$$AB = \begin{pmatrix} a_{00}b_{00} + a_{01}b_{10} & a_{00}b_{01} + a_{01}b_{11} \\ a_{10}b_{00} + a_{11}b_{10} & a_{10}b_{01} + a_{11}b_{11} \end{pmatrix}$$

and hence

$$f_A \circ f_B((x_0 : x_1)) = f_{AB}((x_0 : x_1)).$$

From this lemma we see that the inverse ${f_A}^{-1}$ of the projective transformation f_A is equal to $f_{A^{-1}}$. It follows that the set of all projective transformations of $\mathbf{P}^1(\mathbf{C})$ forms a group. We call it the **projective general linear group** of degree 1 and denote it by $PGL(1, \mathbf{C})$. For $A = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$, $\alpha \neq 0$, we have $f_A = \mathrm{id}$ (the identity mapping). From the lemma above we see that

$$PGL(1, \mathbf{C}) \cong GL(2, \mathbf{C})/\mathbf{C}^* I_2.$$

Here

$$GL(2, \mathbf{C}) = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \middle| \alpha, \beta, \gamma, \delta \in \mathbf{C}, \alpha\delta - \beta\gamma \neq 0 \right\},$$

$$\mathbf{C}^* = \mathbf{C} - \{0\}, \quad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For this reason, $PGL(1, \mathbf{C})$ is often denoted also by $PGL(2, \mathbf{C})$. Now we prove

LEMMA 2.2. *Given two triplets of three distinct points*

$$(P_1, P_2, P_3) \text{ and } (Q_1, Q_2, Q_3),$$

there is a unique projective transformation f of $\mathbf{P}^1(\mathbf{C})$ such that

$$f(P_\nu) = Q_\nu, \quad \nu = 1, 2, 3.$$

PROOF. First consider the case where $P_1 = (1:0), P_2 = (1:1), P_3 = (0:1)$. Let $Q_\nu = (a_\nu, b_\nu)$, $\nu = 1, 2, 3$. Now determine α, β so that

$$\begin{aligned} \alpha a_1 + \beta a_3 &= a_2 \\ \alpha b_1 + \beta b_3 &= a_3. \end{aligned}$$

Such α and β are uniquely determined because $a_1 b_3 - b_1 a_3 \neq 0$ due to the assumption $(a_1 : b_1) \neq (a_3 : b_3)$. Since $(\alpha a_1 : \alpha b_1) = (a_1 : b_1)$ and $(\beta a_3 : \beta b_3) = (a_3 : b_3)$, we may assume $\alpha = \beta = 1$ by replacing a_1, b_1 by $\alpha a_1, \alpha b_1$ and a_3, b_3 by $\beta a_3, \beta b_3$. If f is the projective transformation corresponding to

$$\begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix},$$

then $f(P_\nu) = Q_\nu$, $\nu = 1, 2, 3$.

In the general case, let g be the projective transformation that takes $(1:0), (1:1), (0:1)$ into P_1, P_2, P_3 and let h be the projective transformation that takes $(1:0), (1:1), (0:1)$ into Q_1, Q_2, Q_3. Then $f = h \circ g^{-1}$ is the desired projective transformation. The uniqueness follows from the fact that a projective transformation mapping $(1:0), (1:1), (0:1)$ into $(1:0), (1:1), (0:1)$ is the identity transformation (Exercise 2.1).

(c) Function fields. We used $(x_0 : x_1)$ for a general point in the projective line $\mathbf{P}^1(\mathbf{C})$ in the previous subsection. Now we consider x_0, x_1 as variables and call $(x_0 : x_1)$ the **homogeneous coordinates**. More precisely, x_1/x_0 denotes the coordinate in U_1 and x_0/x_1 that in U_0, while only the ratio in $(x_0 : x_1)$ makes sense and is called the homogeneous coordinates. A polynomial $F(x_0, x_1)$ of x_0 and x_1 is called a **homogeneous polynomial** of degree d if it satisfies

$$F(\alpha x_0, \alpha x_1) = \alpha^d F(x_0, x_1), \tag{2.7}$$

where α is regarded as a variable. This means that

$$F(x_0, x_1) = \sum_{i+j=d, i,j\geq 0} a_{ij} x^i x^j.$$

For a homogeneous polynomial of degree d, $F(a_0, a_1)$ for a point $(a_0 : a_1)$ depends on how we represent the point. That is, we have

$$F(\alpha a_0, \alpha a_1) = \alpha^d F(a_0, a_1)$$

if we take $(\alpha a_0, \alpha a_1)$ instead of (a_0, a_1). However, if $F(a_0, a_1) = 0$, we still have $F(\alpha a_0, \alpha a_1) = 0$. Another remark is that if $F(x_0, x_1)$ is not a homogeneous polynomial, then (2.7) does not hold and it is possible to have $F(a_0, a_1) = 0$ while $F(\alpha a_0, \alpha a_1) \neq 0$. Thus when we work in the projective line $\mathbf{P}^1(\mathbf{C})$, it is basic to deal with homogeneous polynomials of two variables; we have seen that the set of zeros of a homogeneous polynomial makes sense in $\mathbf{P}^1(\mathbf{C})$.

Now let us consider a rational expression determined by two homogeneous polynomials $P(x_0, x_1)$ and $Q(x_0, x_1)$, $Q \neq 0$,

$$R(x_0, x_1) = \frac{P(x_0, x_1)}{Q(x_0, x_1)}.$$

When P and Q have the same degree, we get

$$R(\alpha x_0, \alpha x_1) = R(x_0, x_1),$$

where α is a variable. Thus $R(x_0, x_1)$ is meaningful as a function on $\mathbf{P}^1(\mathbf{C})$. We call such a function a **rational function**. Two rational functions

$$\frac{P(x_0, x_1)}{Q(x_0, x_1)}, \quad \frac{S(x_0, x_1)}{T(x_0, x_1)}$$

are equal if and only if

$$PT = QS.$$

Usually we represent a rational function in the form

$$R(x_0, x_1) = \frac{P(x_0, x_1)}{Q(x_0, x_1)},$$

where $P(x_0, x_1)$ and $Q(x_0, x_1)$ have no common factor (by eliminating all common factors, if necessary). Under this convention, a zero of $P(x_0, x_1)$ in $\mathbf{P}^1(\mathbf{C})$ is called a **zero** of the rational function $R(x_0, x_1)$ and a zero of $Q(x_0, x_1)$ a **pole** of $R(x_0, x_1)$.

The set of all rational functions on $\mathbf{P}^1(\mathbf{C})$ is called the **function field** of $\mathbf{P}^1(\mathbf{C})$ and denoted by $\mathbf{C}(\mathbf{P}^1)$. From the discussion above we may say that $\mathbf{C}(\mathbf{P}^1)$ is the set of all $\frac{P(x_0,x_1)}{Q(x_0,x_1)}$, where P, Q are relatively prime homogeneous polynomials of the same degree and $Q \neq 0$.

LEMMA 2.3. *$\mathbf{C}(\mathbf{P}^1)$ is isomorphic to the field of rational functions $\mathbf{C}(z)$ of one variable.*

PROOF. We can easily verify that the mapping

$$\nu : \frac{P(x_0, x_1)}{Q(x_0, x_1)} \in \mathbf{C}(\mathbf{P}^1) \longmapsto \frac{P(1, z)}{Q(1, z)} \in \mathbf{C}(z)$$

is a monomorphism of fields. We shall verify that ν is surjective. Let us take an arbitrary element of $\mathbf{C}(z)$ and represent it in the form

$$\frac{f(z)}{g(z)},$$

where f and g are relatively prime and have degrees m and n, respectively. If $m \geq n$, we set

$$P(x_0, x_1) = x_0^m f\left(\frac{x_1}{x_0}\right), \quad Q(x_0, x_1) = x_0^m g\left(\frac{x_1}{x_0}\right).$$

If $m \leq n$, we set

$$P(x_0, x_1) = x_0^n f\left(\frac{x_1}{x_0}\right), \quad Q(x_0, x_1) = x_0^n g\left(\frac{x_1}{x_0}\right).$$

Then P and Q have the same degree. By setting

$$R(x_0, x_1) = \frac{P(x_0, x_1)}{Q(x_0, x_1)}$$

we get a rational function on $\mathbf{P}^1(\mathbf{C})$ such that

$$\nu(R(x_0, x_1)) = \frac{P(1, z)}{Q(1, z)} = \frac{f(z)}{g(z)}.$$

Thus ν is surjective, and hence it is an isomorphism of fields.

It is known that a **meromorphic function** on the Riemann sphere is a rational function. This is no accident and shows a piece of result stemming from a deeper fact stated in [12] (called GAGA of Serre) that the complex analytic property and algebraic property agree with each other (see Theorem 4.1).

§ 2.2. The projective plane and plane curves

(a) The projective plane. In the preceding section we discussed the projective line in detail. Perhaps many readers already understand how to define the projective plane, that is, the two-dimensional projective space. We shall now define the complex projective plane, since it will work in the same way for any arbitrary field.

Let $\mathbf{C}^3$ denote the set of all triples (a_0, a_1, a_2) of complex numbers and set

$$W = \mathbf{C}^3 - \{(0, 0, 0)\}.$$

For each element (a_0, a_1, a_2) in W, we consider the ratio $(a_0 : a_1 : a_2)$ and denote by $\mathbf{P}^2(\mathbf{C})$ the totality of such ratios. It is called the two-dimensional complex projective space, or **complex projective plane**. The point determined by (a_0, a_1, a_2) is denoted by $(a_0 : a_1 : a_2)$. For any nonzero complex number α we have

$$(\alpha a_0 : \alpha a_1 : \alpha a_2) = (a_0 : a_1 : a_2).$$

We define subsets U_0, U_1, U_2 by

$$U_j = \{(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) | \, a_j \neq 0\}.$$

Just as in the preceding section, for any point $(a_0 : a_1 : a_2) \in U_0$ we have

$$(a_0 : a_1 : a_2) = \left(1 : \frac{a_1}{a_0} : \frac{a_2}{a_0}\right),$$

and the mapping

$$\phi_0 : (a_0 : a_1 : a_2) \in U_0 \longrightarrow \left(\frac{a_1}{a_0}, \frac{a_2}{a_0}\right) \in \mathbf{C}^2$$

is a bijection of U_0 to $\mathbf{C}^2$, whose inverse is given by

$$\phi_0^{-1} : (x, y) \in \mathbf{C}^2 \longrightarrow (1 : x : y) \in U_0.$$

Likewise, the mappings

$$\phi_1 : (a_0 : a_1 : a_2) \in U_1 \longrightarrow \left(\frac{a_0}{a_1}, \frac{a_2}{a_1}\right) \in \mathbf{C}^2$$

$$\phi_2 : (a_0 : a_1 : a_2) \in U_2 \longrightarrow \left(\frac{a_0}{a_2}, \frac{a_1}{a_2}\right) \in \mathbf{C}^2$$

are bijections of U_1 and U_2 to $\mathbf{C}^2$.

Here we study the structure of $\mathbf{P}^2(\mathbf{C}) - U_0$. By definition, each point of $\mathbf{P}^2(\mathbf{C}) - U_0$ is of the form $(0 : a_1 : a_2)$, where, of course, $(a_1, a_2) \neq (0, 0)$. Hence (a_1, a_2) determines a point of the complex projective line $\mathbf{P}^1(\mathbf{C})$. Conversely, for any point $(b_1 : b_2)$, we get a point $(0 : b_1 : b_2)$ of $\mathbf{P}^2(\mathbf{C}) - U_0$. That is, the mapping

$$(0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) \mapsto (a_1 : a_2) \in \mathbf{P}^1(\mathbf{C})$$

is a bijection. In this way, we may identify $\mathbf{P}^2(\mathbf{C}) - U_0$ and $\mathbf{P}^1(\mathbf{C})$.

Identifying U_0 and $\mathbf{C}^2$ by ϕ_0, we may write

$$\mathbf{P}^2(\mathbf{C}) = \mathbf{C}^2 \cup \mathbf{P}^1(\mathbf{C}).$$

$\mathbf{P}^2(\mathbf{C}) - U_0$ is called the **line at infinity** and denoted by ℓ_∞. In the following, we always fix our attention on U_0 and identify it with $\mathbf{C}^2$ by ϕ_0. Namely, we have

$$\mathbf{P}^2(\mathbf{C}) = U_0 \cup \ell_\infty, \quad U_0 \cong \mathbf{C}^2, \quad \ell_\infty \cong \mathbf{P}^1(\mathbf{C}). \tag{2.8}$$

We call $\mathbf{C}^2$ the **complex affine plane** and denote its coordinates by (x, y). (Note that if we choose U_1 or U_2, we have similar arguments only by changing the line at infinity.)

Now using the coordinates (x, y) in the complex affine plane $\mathbf{C}^2$ we may consider $(1 : x : y)$ as coordinates in $\mathbf{P}^2(\mathbf{C})$, which, however, cannot represent the region of the line at infinity. Hence we introduce the homogeneous coordinates

$(x_0 : x_1 : x_2)$ in $\mathbf{P}^2(\mathbf{C})$. Here x_0, x_1, x_2 are variables, but only the ratios are meaningful. Furthermore, when $x_0 \neq 0$, we have

$$x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}. \tag{2.9}$$

In terms of homogeneous coordinates, the line at infinity ℓ_∞ can be expressed by

$$x_0 = 0.$$

A line in the complex affine plane $\mathbf{C}^2$ can be written in the form

$$\alpha + \beta x + \gamma y = 0, \text{ where } (\beta, \gamma) \neq (0, 0), \alpha, \beta, \gamma \in \mathbf{C}.$$

By using (2.9) we get

$$\alpha + \beta \frac{x_1}{x_0} + \gamma \frac{x_2}{x_0} = 0,$$

and thus

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0.$$

This equation makes sense in the complex projective plane $\mathbf{P}^2(\mathbf{C})$. That is, if

$$\alpha a_0 + \beta a_1 + \gamma a_2 = 0,$$

then

$$\alpha(\lambda a_0) + \beta(\lambda a_1) + \gamma(\lambda a_2) = 0.$$

Thus a point $(a_0, : a_1 : a_2)$ in $\mathbf{P}^2(\mathbf{C})$ satisfying the equation

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0$$

has a geometric meaning. On the other hand, the equation

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 1$$

does not make sense in $\mathbf{P}^2(\mathbf{C})$, because for $\lambda \neq 1$ we would get

$$\alpha(\lambda a_0) + \beta(\lambda a_1 + \gamma(\lambda a_2) = \lambda \neq 1.$$

As you see, only the set of zeros of a homogeneous polynomial in x_0, x_1, x_2 makes sense in $\mathbf{P}^2(\mathbf{C})$. We shall later give more detail.

This being said, what is the difference between the equation of a line in the complex affine plane $\mathbf{C}^2$

$$\alpha + \beta x + \gamma y = 0$$

and the equation in $\mathbf{P}^2(\mathbf{C})$

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0?$$

To work in $\mathbf{C}^2$ is to work with points of the form $(1 : a : b)$ in $\mathbf{P}^2(\mathbf{C})$. Let us check whether a point of the form $(0 : c : d)$ satisfies the equation

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0.$$

Substituting $(0 : c : d)$ in the equation, we get

$$\beta c + \gamma d = 0,$$

which implies $c : d = -\gamma : \beta$, because $(\beta, \gamma) \neq (0, 0)$. Hence the point $(0 : -\gamma : \beta)$ satisfies

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0.$$

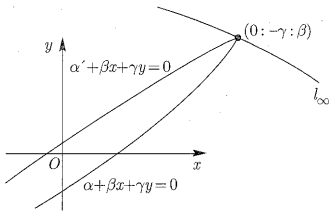


FIGURE 2.4. Two parallel lines in the affine plane meet at a point on the line at infinity.

We may interpret that the line in $\mathbf{C}^2$

$$\alpha+\beta x+\gamma y=0$$

meets the line at infinity ℓ_∞ at the point $(0:-\gamma:\beta)$. This point depends only on the ratio $\beta:\gamma$, that is, on the slope of the line and not on α. Thus if $\alpha\neq\alpha'$, then

$$\alpha+\beta x+\gamma y=0$$
$$\alpha'+\beta x+\gamma y=0$$

are two parallel lines, not intersecting in $\mathbf{C}^2$ but intersecting at the point $(0:-\gamma:\beta)$ on the line at infinity ℓ_∞ in $\mathbf{P}^2(\mathbf{C})$. (See Figure 2.4.)

Conversely, take a point $(0:b:c)$ on the line at infinity ℓ_∞. Then a line

$$a-cx+by=0$$

meets ℓ_∞ at $(0:b:c)$. Here a can be chosen arbitrarily.

We find from the discussions above that the points on the line at infinity and the slopes of the lines

$$\alpha+\beta x+\gamma y=0$$

in $\mathbf{C}^2$ correspond in a one-to-one manner. We may also give the interpretation that a linear homogeneous equation

$$\alpha x_0+\beta x_1+\gamma x_2=0$$

determines a line (projective line, to be exact) in the complex projective plane $\mathbf{P}^2(\mathbf{C})$.

(b) Duality principle and projective transformations. Following the discussions in the preceding subsection, the set of all points in the complex projective plane satisfying a linear homogeneous equation

$$\alpha x_0+\beta x_1+\gamma x_2=0,\quad (\alpha,\beta,\gamma)\neq(0,0,0), \tag{2.10}$$

namely,

$$\ell_{\alpha,\beta,\gamma}=\{(a_0:a_1:a_2)\in\mathbf{P}^2(\mathbf{C})|\alpha a_0+\beta a_1+\gamma a_2=0\},$$

will be called a **line** in $\mathbf{P}^2(\mathbf{C})$. We remark that for any complex number $\lambda\neq0$, the new equation obtained from (2.10) by replacing α,β,γ by $\lambda\alpha,\lambda\beta,\lambda\gamma$, respectively, defines the same line as (2.10).

Actually, we have

LEMMA 2.4.

(i) *The lines $\ell_{\alpha,\beta,\gamma}$ and $\ell_{\alpha',\beta',\gamma'}$ are the same if and only if $(\alpha:\beta:\gamma)=(\alpha':\beta':\gamma')$.*

(ii) *Two distinct lines $\ell_{\alpha,\beta,\gamma}$ and $\ell_{\alpha',\beta',\gamma'}$ intersect at one point.*

PROOF. We first prove (ii). We consider the system of equations

$$\alpha x_0+\beta x_1+\gamma x_2=0$$
$$\alpha' x_0+\beta' x_1+\gamma' x_2=0.$$

From $(\alpha:\beta:\gamma)\neq(\alpha':\beta':\gamma')$, we see that $(\alpha,\beta,\gamma)\neq(\lambda\alpha',\lambda\beta',\lambda\gamma')$ for any complex number $\lambda\neq0$. This implies that the matrix

$$\begin{pmatrix}\alpha&\beta&\gamma\\ \alpha'&\beta'&\gamma'\end{pmatrix}$$

has rank equal to 2, namely,

$$\left(\begin{vmatrix}\beta&\gamma\\ \beta'&\gamma'\end{vmatrix},\begin{vmatrix}\gamma&\alpha\\ \gamma'&\alpha'\end{vmatrix},\begin{vmatrix}\alpha&\beta\\ \alpha'&\beta'\end{vmatrix}\right)\neq(0,0,0).$$

The solution for the system in $\mathbf{P}^2(\mathbf{C})$ is given by

$$\left(\begin{vmatrix}\beta&\gamma\\ \beta'&\gamma'\end{vmatrix}:\begin{vmatrix}\gamma&\alpha\\ \gamma'&\alpha'\end{vmatrix}:\begin{vmatrix}\alpha&\beta\\ \alpha'&\beta'\end{vmatrix}\right),$$

that is, the intersection of the two lines is a single point. The assertion (i) now easily follows.

COROLLARY 2.1. *The set of all lines in $\mathbf{P}^2(\mathbf{C})$ and the complex projective plane correspond one-to-one by*

$$\ell_{\alpha,\beta,\gamma}\longmapsto(\alpha:\beta:\gamma).$$

By this corollary we can identify the set of all lines in $\mathbf{P}^2(\mathbf{C})$ with the complex projective plane. This is called the **dual complex projective plane** and denoted by $\mathbf{P}^2(\mathbf{C})^*$. To a point $(\alpha:\beta:\gamma)$ there corresponds the line $\ell_{\alpha,\beta,\gamma}$ defined by

$$\alpha x_0+\beta x_1+\gamma x_2=0.$$

The following lemma is fundamental in $\mathbf{P}^2(\mathbf{C})$.

LEMMA 2.5. *Given two distinct points in $\mathbf{P}^2(\mathbf{C})$, there is a unique line going through them.*

PROOF. Let $(a_0:a_1:a_2)$ and $(b_0:b_1:b_2)$ be two distinct points. We want to find a line

$$\alpha x_0+\beta x_1+\gamma x_2=0$$

such that

$$\alpha a_0+\beta a_1+\gamma a_2=0$$
$$\alpha b_0+\beta b_1+\gamma b_2=0.$$

Thus we have

$$\alpha:\beta:\gamma=\begin{vmatrix}a_1&a_2\\ b_1&b_2\end{vmatrix}:\begin{vmatrix}a_2&a_0\\ b_2&b_0\end{vmatrix}:\begin{vmatrix}a_0&a_1\\ b_0&b_1\end{vmatrix}.$$

The reader may notice a remarkable similarity in the proofs of Lemmas 2.4 and 2.5. If in Lemma 2.4 (ii) you interchange a line and a point and replace "intersect" by "are gone through", you get Lemma 2.5. Likewise, you may pass from Lemma 2.5 to Lemma 2.4, (ii). This is no coincidence. It comes from the fact that the set of all lines in $\mathbf{P}^2(\mathbf{C})$ is again a projective plane, and finding the intersection of two lines and finding a line going through two points are both reduced to solving a system of two linear equations. That is, we have

THEOREM 2.1. *A valid proposition concerning points and lines in* $\mathbf{P}^2(\mathbf{C})$ *remains valid on interchanging "points" and "lines" and "intersect" and "go through".*

DEFINITION 2.2. Given a 3×3 matrix

$$A = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}, \quad \det A \neq 0,$$

the mapping $f_A : \mathbf{P}^2(\mathbf{C}) \to \mathbf{P}^2(\mathbf{C})$ given by

$$(x_0 : x_1 : x_2) \longmapsto \left(\sum_{j=0}^{2} a_{0j}x_j : \sum_{j=0}^{2} a_{1j}x_j : \sum_{j=0}^{2} a_{1j}x_j \right)$$

is called a **projective transformation determined by** A or simply a **projective transformation**.

In the definition above, it is important that $\det A \neq 0$. Under this assumption, the system of equations

$$\sum_{j=0}^{2} a_{ij}x_j = 0, \quad i = 0, 1, 2,$$

has only $(0, 0, 0)$ as a solution. Thus for any point $(\alpha_0 : \alpha_1 : \alpha_2)$ we have

$$\left(\sum_{j=0}^{2} a_{0j}\alpha_j, \sum_{j=0}^{2} a_{1j}\alpha_j, \sum_{j=0}^{2} a_{2j}\alpha_j \right) \neq (0, 0, 0).$$

Thus $f_A((\alpha_0 : \alpha_1 : \alpha_2))$ is well-defined. On the other hand, the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

would give the "mapping"

$$(x_0 : x_1 : x_2) \longmapsto (x_0 : x_1 : 0);$$

the image of $(0 : 0 : 1)$ would be $(0 : 0 : 0)$, that is, we don't really have a mapping.

LEMMA 2.6.

(i) *For a nonzero complex number* λ, *we have* $f_{\lambda A} = f_A$.

(ii) *For* 3×3 *matrices* A, B *with* $\det A \neq 0, \det B \neq 0$, *we have* $f_A \circ f_B = f_{AB}$. *In particular,* ${f_A}^{-1} = f_{A^{-1}}$.

The proof is similar to that of Lemma 2.1 and hence omitted.

The set of all projective transformations of $\mathbf{P}^2(\mathbf{C})$ forms a group. If we denote it by $PGL(2, \mathbf{C})$, then

$$PGL(2, \mathbf{C}) \cong GL(3, \mathbf{C})/\mathbf{C}^* I_3.$$

This can be proved similarly to the case of $PGL(1, \mathbf{C})$ just after Lemma 2.1.

We can show that a projective transformation maps a line into a line as follows. We consider f_A, where A is a 3×3 matrix with $\det A \neq 0$, and the line $\ell_{\alpha,\beta,\gamma}$ given by

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0.$$

Set

$$(\alpha', \beta', \gamma') = (\alpha, \beta, \gamma)A^{-1}. \tag{2.11}$$

For $a = (a_0 : a_1 : a_2) \in \ell_{\alpha,\beta,\gamma}$, let $b = f_A(a) = (b_0 : b_1 : b_2)$. We easily find

$$\alpha' b_0 + \beta' b_1 + \gamma' b_2 = 0,$$

that is, $b \in \ell_{\alpha',\beta',\gamma'}$. Conversely, for $b \in \ell_{\alpha',\beta',\gamma'}$, set $a = {f_A}^{-1}(b) = (a_0 : a_1 : a_2)$; then we see that $a \in \ell_{\alpha,\beta,\gamma}$. We have thus seen that f_A takes $\ell_{\alpha,\beta,\gamma}$ into $\ell_{\alpha',\beta',\gamma'}$. The relationship between the coefficients α, β, γ and α', β', γ' is given by (2.11).

The discussion above can be more clearly illustrated by using matrices. The line $\ell_{\alpha,\beta,\gamma}$ is given by

$$(\alpha, \beta, \gamma) \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = 0.$$

The projective transformation f_A is expressed by

$$f_A((x_0 : x_1 : x_2)) = (x_0' : x_1' : x_2'),$$

where

$$\begin{pmatrix} x_0' \\ x_1' \\ x_2' \end{pmatrix} = A \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}.$$

Thus by using (2.11) we get

$$(\alpha', \beta', \gamma') \begin{pmatrix} x_0' \\ x_1' \\ x_2' \end{pmatrix} = (\alpha, \beta, \gamma)A^{-1}A \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = (\alpha, \beta, \gamma) \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix},$$

showing that the point $(x_0 : x_1 : x_2)$ of $\ell_{\alpha,\beta,\gamma}$ and the point $(x_0' : x_1' : x_2') = f_A(x_0 : x_1 : x_2)$ of $\ell_{\alpha',\beta',\gamma'}$ correspond to each other. The following lemma should be obvious from the discussion above.

LEMMA 2.7. *An arbitrary line* $\ell_{\alpha,\beta,\gamma}$ *in* $\mathbf{P}^2(\mathbf{C})$ *can be mapped by a projective transformation to the line at infinity* ℓ_∞.

In fact, it is sufficient to find a matrix A such that $(1, 0, 0) = (\alpha, \beta, \gamma)A^{-1}$. From this, it follows that every line is isomorphic with the projective line $\mathbf{P}^1(\mathbf{C})$. To be more precise, take the mapping

$$\phi : \mathbf{P}^1(\mathbf{C}) \longrightarrow \mathbf{P}^2(\mathbf{C}),$$

The reader may notice a remarkable similarity in the proofs of Lemmas 2.4 and 2.5. If in Lemma 2.4 (ii) you interchange a line and a point and replace "intersect" by "are gone through", you get Lemma 2.5. Likewise, you may pass from Lemma 2.5 to Lemma 2.4, (ii). This is no coincidence. It comes from the fact that the set of all lines in $\mathbf{P}^2(\mathbf{C})$ is again a projective plane, and finding the intersection of two lines and finding a line going through two points are both reduced to solving a system of two linear equations. That is, we have

THEOREM 2.1. *A valid proposition concerning points and lines in* $\mathbf{P}^2(\mathbf{C})$ *remains valid on interchanging "points" and "lines" and "intersect" and "go through".*

DEFINITION 2.2. Given a 3×3 matrix

$$A = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}, \quad \det A \neq 0,$$

the mapping $f_A : \mathbf{P}^2(\mathbf{C}) \to \mathbf{P}^2(\mathbf{C})$ given by

$$(x_0 : x_1 : x_2) \longmapsto \left(\sum_{j=0}^{2} a_{0j}x_j : \sum_{j=0}^{2} a_{1j}x_j : \sum_{j=0}^{2} a_{1j}x_j \right)$$

is called a **projective transformation determined by** A or simply a **projective transformation**.

In the definition above, it is important that $\det A \neq 0$. Under this assumption, the system of equations

$$\sum_{j=0}^{2} a_{ij}x_j = 0, \quad i = 0, 1, 2,$$

has only $(0, 0, 0)$ as a solution. Thus for any point $(\alpha_0 : \alpha_1 : \alpha_2)$ we have

$$\left(\sum_{j=0}^{2} a_{0j}\alpha_j, \sum_{j=0}^{2} a_{1j}\alpha_j, \sum_{j=0}^{2} a_{2j}\alpha_j \right) \neq (0, 0, 0).$$

Thus $f_A((\alpha_0 : \alpha_1 : \alpha_2))$ is well-defined. On the other hand, the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

would give the "mapping"

$$(x_0 : x_1 : x_2) \longmapsto (x_0 : x_1 : 0);$$

the image of $(0 : 0 : 1)$ would be $(0 : 0 : 0)$, that is, we don't really have a mapping.

LEMMA 2.6.

(i) *For a nonzero complex number* λ, *we have* $f_{\lambda A} = f_A$.

(ii) *For* 3×3 *matrices* A, B *with* $\det A \neq 0, \det B \neq 0$, *we have* $f_A \circ f_B = f_{AB}$. *In particular,* ${f_A}^{-1} = f_{A^{-1}}$.

The proof is similar to that of Lemma 2.1 and hence omitted.

The set of all projective transformations of $\mathbf{P}^2(\mathbf{C})$ forms a group. If we denote it by $PGL(2, \mathbf{C})$, then

$$PGL(2, \mathbf{C}) \cong GL(3, \mathbf{C})/\mathbf{C}^* I_3.$$

This can be proved similarly to the case of $PGL(1, \mathbf{C})$ just after Lemma 2.1.

We can show that a projective transformation maps a line into a line as follows. We consider f_A, where A is a 3×3 matrix with $\det A \neq 0$, and the line $\ell_{\alpha,\beta,\gamma}$ given by

$$\alpha x_0 + \beta x_1 + \gamma x_2 = 0.$$

Set

$$(\alpha', \beta', \gamma') = (\alpha, \beta, \gamma)A^{-1}. \tag{2.11}$$

For $a = (a_0 : a_1 : a_2) \in \ell_{\alpha,\beta,\gamma}$, let $b = f_A(a) = (b_0 : b_1 : b_2)$. We easily find

$$\alpha' b_0 + \beta' b_1 + \gamma' b_2 = 0,$$

that is, $b \in \ell_{\alpha',\beta',\gamma'}$. Conversely, for $b \in \ell_{\alpha',\beta',\gamma'}$, set $a = f_A^{-1}(b) = (a_0 : a_1 : a_2)$; then we see that $a \in \ell_{\alpha,\beta,\gamma}$. We have thus seen that f_A takes $\ell_{\alpha,\beta,\gamma}$ into $\ell_{\alpha',\beta',\gamma'}$. The relationship between the coefficients α, β, γ and α', β', γ' is given by (2.11).

The discussion above can be more clearly illustrated by using matrices. The line $\ell_{\alpha,\beta,\gamma}$ is given by

$$(\alpha, \beta, \gamma) \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = 0.$$

The projective transformation f_A is expressed by

$$f_A((x_0 : x_1 : x_2)) = (x_0' : x_1' : x_2'),$$

where

$$\begin{pmatrix} x_0' \\ x_1' \\ x_2' \end{pmatrix} = A \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}.$$

Thus by using (2.11) we get

$$(\alpha', \beta', \gamma') \begin{pmatrix} x_0' \\ x_1' \\ x_2' \end{pmatrix} = (\alpha, \beta, \gamma)A^{-1}A \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} = (\alpha, \beta, \gamma) \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix},$$

showing that the point $(x_0 : x_1 : x_2)$ of $\ell_{\alpha,\beta,\gamma}$ and the point $(x_0' : x_1' : x_2') = f_A(x_0 : x_1 : x_2)$ of $\ell_{\alpha',\beta',\gamma'}$ correspond to each other. The following lemma should be obvious from the discussion above.

LEMMA 2.7. *An arbitrary line* $\ell_{\alpha,\beta,\gamma}$ *in* $\mathbf{P}^2(\mathbf{C})$ *can be mapped by a projective transformation to the line at infinity* ℓ_∞.

In fact, it is sufficient to find a matrix A such that $(1, 0, 0) = (\alpha, \beta, \gamma)A^{-1}$. From this, it follows that every line is isomorphic with the projective line $\mathbf{P}^1(\mathbf{C})$. To be more precise, take the mapping

$$\phi : \mathbf{P}^1(\mathbf{C}) \longrightarrow \mathbf{P}^2(\mathbf{C}),$$

where

$$(y_0 : y_1) \longmapsto (a_0y_0 + a_1y_1 : b_0y_0 + b_1y_1 : c_0y_0 + c_1y_1). \tag{2.12}$$

When

$$(\begin{vmatrix} b_0 & b_1 \\ c_0 & c_1 \end{vmatrix} : \begin{vmatrix} c_0 & c_1 \\ a_0 & a_1 \end{vmatrix} : \begin{vmatrix} a_0 & a_1 \\ b_0 & b_1 \end{vmatrix}) \neq (0,0,0),$$

the image of the mapping ϕ in (2.12) is the line defined by

$$\begin{vmatrix} a_0 & a_1 & x_0 \\ b_0 & b_1 & x_1 \\ c_0 & c_1 & x_2 \end{vmatrix} = 0.$$

Furthermore, every line is the image of the mapping ϕ if we choose appropriate $a_0, a_1, b_0, b_1, c_1, c_2$. The proofs of these facts are left as Exercise 2.3. The next lemma can be proved similarly to Lemma 2.2. We make it Exercise 2.1, (iii).

LEMMA 2.8. *Given two sets of four points (P_1, P_2, P_3, P_4) and (Q_1, Q_2, Q_3, Q_4), assume that no three points of each set are on one line. Then there is a unique projective transformation f such that*

$$f(P_\nu) = Q_\nu, \quad \nu = 1, 2, 3, 4.$$

The geometry that studies properties invariant under projective transformations is projective geometry. We can say that algebraic geometry is the geometry that studies properties invariant under **birational transformations**, which are more general than projective transformations. A clearer picture on this will gradually emerge.

(c) The function field of the projective plane. Just as in the case of $\mathbf{P}^1(\mathbf{C})$, a rational function on $\mathbf{P}^2(\mathbf{C})$ is given by

$$f(x_0, x_1, x_2) = \frac{P(x_0, x_1, x_2)}{Q(x_0, x_1, x_2)},$$

where $P(x_0, x_1, x_2)$ and $Q(x_0, x_1, x_2)$ are homogeneous polynomials of the same degree that can be assumed to be relatively prime. It is evident that the set of all rational functions on $\mathbf{P}^2(\mathbf{C})$ forms a field. We call it **the function field** and denote it by $\mathbf{C}(\mathbf{P}^2)$. The following lemma will be obvious.

LEMMA 2.9. *The function field $\mathbf{C}(\mathbf{P}^2)$ of $\mathbf{P}^2(\mathbf{C})$ is isomorphic with the field $\mathbf{C}(x, y)$ of all rational functions of two variables. The isomorphism is given by*

$$\frac{P(x_0, x_1, x_2)}{Q(x_0, x_1, x_2)} \longmapsto \frac{P(1, x, y)}{Q(1, x, y)}.$$

(d) Plane curves. We now study curves in $\mathbf{P}^2(\mathbf{C})$. When a homogeneous polynomial of degree d in three variables with complex coefficients

$$F(x_0, x_1, x_2)$$

is given,

$$C = \{(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) | F(a_0, a_1, a_2) = 0\} \tag{2.13}$$

is called a **plane curve of degree** d, and $F(x_0, x_1, x_2)$ is called the **defining equation**. We may also write

$$C : F(x_0, x_1, x_2) = 0$$

instead of (2.13). In order to emphasize the fact that we deal with the zero set of the homogeneous equation F, we may also write $V(F)$ instead of C. In the following we adopt any of the three notations for convenience.

When $F(x_0, x_1, x_2)$ is an irreducible polynomial, we call $C = V(F)$ an **irreducible plane curve**. If F is reducible, then $V(F)$ is called a **reducible plane curve**. If F is reducible, then we can decompose it into a product of homogeneous polynomials

$$F(x_0, x_1, x_2) = G(x_0, x_1, x_2)H(x_0, x_1, x_2),$$

and by (2.13) we get

$$C = V(F) = V(G) \cap V(H).$$

Thus a plane curve can be expressed as a union of irreducible plane curves. It is therefore sufficient to study irreducible plane curves and the intersection of two irreducible plane curves, as we do in §2.3. We make some preparations.

As before, we take

$$U_0 = \{(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) | a_0 \neq 0\}$$

and the bijection

$$\phi_0 : (a_0 : a_1 : a_2) \in U_0 \longmapsto \left(\frac{a_1}{a_0}, \frac{a_2}{a_0}\right) \in \mathbf{C}^2.$$

We have

$$\mathbf{P}^2(\mathbf{C}) = \mathbf{C}^2 \cup l_\infty.$$

Let $\{x, y\}$ be the coordinates in the affine plane $\mathbf{C}^2$, which are related to the homogeneous coordinates in $\mathbf{P}^2(\mathbf{C})$ by

$$x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}$$

on $U_0 = \mathbf{C}^2$. For a homogeneous polynomial of degree d,

$$\frac{1}{x_0^d} F(x_0, x_1, x_2)$$

is a polynomial in $x = x_1/x_0, y = x_2/x_0$ of degree at most equal to d. Set

$$\frac{1}{x_0^d} F(x_0, x_1, x_2) = f(x, y). \tag{2.14}$$

We may also write

$$f(x, y) = F(1, x, y).$$

If $F(x_0, x_1, x_2)$ is irreducible and if $d \geq 2$, then $f(x, y)$ is also irreducible of degree d, as follows. Writing

$$F(x_0, x_1, x_2) = \sum_{i+j+k=d} a_{ijk} x_0^i x_1^j x_2^k,$$

we get

$$\frac{1}{x_0^d}F(x_0,x_1,x_2)=\sum_{i+j+k=d}a_{ijk}\left(\frac{x_1}{x_0}\right)^j\left(\frac{x_2}{x_0}\right)^k$$

and

$$f(x,y)=\sum_{i+j+k=d}a_{ijk}x^jy^k$$

If the degree of $f(x,y)$ is strictly smaller than d, then we must have

$$a_{0jk}=0,\quad j+k=d,$$

which implies that $F(x_0,x_1,x_2)$ is divisible by x_0. If $F(x_0,x_1,x_2)$ is irreducible, we must have $F=\alpha x_0$. Therefore, if $d\geq 2$, then the degree of $f(x,y)$ must be d. If $f(x,y)$ is reducible, then we write

$$f(x,y)=g(x,y)h(x,y),$$

where g and h have degrees d_1 and d_2. Setting

$$G(x_0,x_1,x_2)=x_0^{d_1}g\left(\frac{x_1}{x_0},\frac{x_2}{x_0}\right)$$

$$H(x_0,x_1,x_2)=x_0^{d_2}h\left(\frac{x_1}{x_0},\frac{x_2}{x_0}\right),$$

we find from $d=d_1+d_2$ and (2.14) that

$$F(x_0,x_1,x_2)=G(x_0,x_1,x_2),$$

contradicting the assumption that $F(x_0,x_1,x_2)$ is irreducible.

By the way, in the case where $d=1$, if $F=\alpha x_0$, then $f=\alpha$ is a constant. In other cases, $f(x,y)$ is linear. For $d\geq 2$, there is a one-to-one correspondence (2.4) between the irreducible polynomials $f(x,y)$ of degree d in x,y and the irreducible homogeneous polynomials $F(x_0,x_1,x_2)$ of degree d in x_0,x_1,x_2.

An irreducible polynomial $f(x,y)$ of degree d determines a curve C_f

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

in the affine plane $\mathbf{C}^2$. We call it an **affine plane curve**. What we have shown above is that an affine plane curve

$$C_f=\{(a,b)\in\mathbf{C}^2|f(a,b)=0\}$$

given by an irreducible polynomial $f(x,y)$ of degree d and an irreducible plane curve $C=V(F)$ of degree d in $\mathbf{P}^2(\mathbf{C})$

$$V(F)=\{(a_0:a_1:a_2)\in\mathbf{C}^2|F(a_0,a_1,a_2)=0\}$$

correspond one-to-one to each other through the correspondence of polynomials

$$f(x,y)=\frac{1}{x_0^d}F(x_0,x_1,x_2)$$

$$F(x_0,x_1,x_2)=x_0^df\left(\frac{x_1}{x_0},\frac{x_2}{x_0}\right).$$

Let us now apply the projective transformation f_A determined by a matrix A to the plane curve $V(F)$ defined by

$$F(x_0,x_1,x_2)=0.$$

We set

$$B=A^{-1}=(b_{ij})_{0\leq i,j\leq 2}$$

and

$$G(x_0,x_1,x_2)=F\left(\sum_{j=0}^2b_{0j}x_j,\sum_{j=0}^2b_{1j}x_j,\sum_{j=0}^2b_{2j}x_j\right).\tag{2.15}$$

Then it is easy to see that $f_A(a)=b\in V(F)$ for every $a=(a_0:a_1:a_2)$. Conversely, for each $b=(b_0:b_1:b_2)\in V(G)$, we have $a=f_A^{-1}(b)\in V(F)$. Hence $f_A(V(F))=V(G)$. We may regard curves that are transformed to each other by a projective transformation as essentially the same. Thus we regard $V(F)$ and $V(G)$ as the same curve. It is therefore convenient to simplify the defining equation by a projective transformation as much as possible.

EXAMPLE 2.1 QUADRATIC CURVES (QUADRICS).

Consider a homogeneous quadratic polynomial

$$F(x_0,x_1,x_2)=\sum_{i,j=)}^2c_{ij}x_ix_j,\quad c_{ij}=c_{ji}.$$

We may express F in matrix form

$$(x_0,x_1,x_2)\begin{pmatrix}c_{00}&c_{01}&c_{02}\\c_{10}&c_{11}&c_{12}\\c_{20}&c_{21}&c_{22}\end{pmatrix}\begin{pmatrix}x_0\\x_1\\x_3\end{pmatrix},\quad C=(c_{ij}).$$

From what we know about quadratic forms (or symmetric matrices) there is a 3×3 matrix A such that ${}^tA^{-1}CA^{-1}$ is one of the forms

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix},\begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix},$$

corresponding to the rank of C being $3,2$, or 1. (See Exercise 2.4; note that we are working with complex entries.) Thus the quadratic form $V(F)$ is changed by the projective transformation f_A into one of the following curves:

$$x_0^2+x_1^2+x_2^2=0$$
$$x_0^2+x_1^2=0$$
$$x_0^2=0.$$

The first curve is irreducible. The second is the union of two lines

$$x_0+ix_1=0,\quad x_0-ix_1=0,\ \text{where } i=\sqrt{-1}.$$

The third is the line at infinity ℓ_∞ counted twice, Thus as far as irreducible quadrics are concerned, they are reduced to

$$x_0^2+x_1^2+x_2^2=0.\tag{2.16}$$

By allowing the projective transformation $(x_0 : x_1 : x_2) \mapsto (ix_0 : x_1; x_2)$ we can change it to

$$-x_0^2 + x_1^2 + z_2^2 = 0. \tag{2.17}$$

The affine curve corresponding to the last quadric is the "unit circle"

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

Or by the projective transformation $(x_0 : x_1 : x_2) \mapsto (ix_0 : x_1 : ix_2)$ we may change (2.16) into

$$x_0^2 + x_1^2 - x_2^2 = 0. \tag{2.18}$$

The corresponding affine curve is the rectangular hyperbola. Furthermore it is possible to change (2.16) by a projective transformation into

$$x_0 x_2 - x_1^2 = 0, \tag{2.19}$$

the corresponding affine curve being the parabola

$$y = x^2.$$

To sum up, in the complex projective plane, the circles, the ellipses, the hyperbolas, and the parabolas are transformable into each other by projective transformations and hence are regarded as the same figure.

(e) Rational mappings and algebraic morphisms. The considerations we have made so far belong to projective geometry rather than algebraic geometry, because they are concerned with projective transformations. In algebraic geometry we deal with a wider class of mappings. Let us first consider mappings from $\mathbf{P}^1(\mathbf{C})$ to $\mathbf{P}^1(\mathbf{C})$. Let $P(x_0, x_1)$ and $Q(x_0, x_1)$ be homogeneous polynomials of degree d and consider the "mapping"

$$\psi : (a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (P(a_0, a_1) : Q(a_0 : a_1)) \in \mathbf{P}^1(\mathbf{C}). \tag{2.20}$$

For ψ to be a real mapping, it is necessary that

$$\psi((\alpha a_0 : \alpha a_1)) = \psi((a_0 : a_1)) \text{ for each nonzero } \alpha \in \mathbf{C} \tag{2.21}$$

and that

$$(P(a_0, a_1), Q(a_0, a_1)) \neq (0, 0). \tag{2.22.}$$

Here (2.21) is valid, since we have

$$\begin{aligned}(P(\alpha a_0, \alpha a_1)) &= (\alpha^d P(a_0, a_1) : \alpha^d Q(a_0, a_1)) \\ &= (P(a_0, a_1) : Q(a_0, a_1)).\end{aligned}$$

As for (2.22), if $P(a_0, a_1) = 0 = Q(a_0, a_1)$, then we can factor

$$\begin{aligned}P(x_0, x_1) &= (a_1 x_0 - a_0 x_1)^m P_1(x_0, x_1) \\ Q(x_0, x_1) &= (a_1 x_0 - a_0 x_1)^n Q_1(x_0, x_1).\end{aligned}$$

Assume $m \geq n$ for simplicity. Then

$$(P(x_0, x_1) : Q(x_0, x_1)) = (a_1 x_0 - a_0 x_1)^{m-n}(P_1(x_0, x_1) : Q_1(x_0, x_1)). \tag{2.23}$$

If $m = n$, then

$$((P(x_0, x_1) : Q(x_0, x_1)) = ((P_1(x_0, x_1) : Q_1(x_0, x_1)).$$

If $m > n$, the same conclusion holds by omitting a common factor. By using the same argument if necessary, we may assume that P_1 and Q_1 have no common factor and satisfy (2.22). Thus we can assume this is the case for P and Q. This illustrates a case in algebraic geometry where, although we cannot define a mapping by an equation like (2.20), we can replace the expression by another form like (2.23) and define a mapping. In algebraic geometry we normally extend the domain of definition of a mapping as much as possible. In this way, we regard (2.20) as defining a mapping from the whole of $\mathbf{P}^1(\mathbf{C})$.

Now in (2.20) P and Q have the same degree, so that

$$f(x_0, x_1) = \frac{Q(x_0, x_1)}{P(x_0, x_1)}, \quad g(x_0, x_1) = \frac{P(x_0, x_1)}{Q(x_0, x_1)}$$

are rational functions on $\mathbf{P}^1(\mathbf{C})$. Thus we may replace (2.20) by

$$(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (1 : f(a_0, a_1)) \in \mathbf{P}^1(\mathbf{C}) \tag{2.24}$$

or

$$(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (g(a_0, a_1) : 1) \in \mathbf{P}^1(\mathbf{C}). \tag{2.25}$$

For example, near a pole of f we change (2.24) into

$$(1 : f(b_0, b_1)) = \left(\frac{1}{f(b_0, b_1)} : 1\right),$$

which is the same as (2.25), since $g = 1/f$. As we see from (2.24) and (2.25), to think of a rational function on $\mathbf{P}^1(\mathbf{C})$ and to think of a rational mapping of $\mathbf{P}^1(\mathbf{C})$ into itself of the form (2.20) amount essentially to the same thing.

Now then, given a rational function on $\mathbf{P}^2(\mathbf{C})$

$$f(x_0, x_1) = \frac{Q(x_0, x_1, x_2)}{P(x_0, x_1, x_2)},$$

is

$$(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) \longmapsto (1 : f(a_0, a_1, a_2)) \in \mathbf{P}^1(\mathbf{C})$$

a mapping? This is the same as

$$(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) \longmapsto (P(a_0, a_1, a_2) : Q(a_0, a_1, a_2)) \in \mathbf{P}^1(\mathbf{C}). \tag{2.26}$$

We may further assume that P and Q have no common divisor. In contrast to the case of $\mathbf{P}^1(\mathbf{C})$ there may be points where

$$P(a_0, a_1, a_2) = Q(a_0, a_1, a_2) = 0$$

and thus the mapping cannot be defined. For instance, suppose $P(x_0, x_1, x_2) = x_0$, $\quad Q(x_0, x_1, x_2) = x_1$. Then (2.26) gives rise to

$$(a_0 : a_1 : a_2) \longmapsto (a_0 : a_1),$$

which is not defined at $(0 : 0 : 1)$. In general, assume P and Q have no common factor. Then (2.26) cannot be defined in the intersection of the plane curves $V(P)$ and $V(Q)$ (which consists of a finite number of points). In such a case, we call the

By allowing the projective transformation $(x_0 : x_1 : x_2) \mapsto (ix_0 : x_1; x_2)$ we can change it to

$$-x_0^2 + x_1^2 + z_2^2 = 0. \tag{2.17}$$

The affine curve corresponding to the last quadric is the "unit circle"

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

Or by the projective transformation $(x_0 : x_1 : x_2) \mapsto (ix_0 : x_1 : ix_2)$ we may change (2.16) into

$$x_0^2 + x_1^2 - x_2^2 = 0. \tag{2.18}$$

The corresponding affine curve is the rectangular hyperbola. Furthermore it is possible to change (2.16) by a projective transformation into

$$x_0 x_2 - x_1^2 = 0, \tag{2.19}$$

the corresponding affine curve being the parabola

$$y = x^2.$$

To sum up, in the complex projective plane, the circles, the ellipses, the hyperbolas, and the parabolas are transformable into each other by projective transformations and hence are regarded as the same figure.

(e) Rational mappings and algebraic morphisms. The considerations we have made so far belong to projective geometry rather than algebraic geometry, because they are concerned with projective transformations. In algebraic geometry we deal with a wider class of mappings. Let us first consider mappings from $\mathbf{P}^1(\mathbf{C})$ to $\mathbf{P}^1(\mathbf{C})$. Let $P(x_0, x_1)$ and $Q(x_0, x_1)$ be homogeneous polynomials of degree d and consider the "mapping"

$$\psi : (a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (P(a_0, a_1) : Q(a_0 : a_1)) \in \mathbf{P}^1(\mathbf{C}). \tag{2.20}$$

For ψ to be a real mapping, it is necessary that

$$\psi((\alpha a_0 : \alpha a_1)) = \psi((a_0 : a_1)) \text{ for each nonzero } \alpha \in \mathbf{C} \tag{2.21}$$

and that

$$(P(a_0, a_1), Q(a_0, a_1)) \neq (0, 0). \tag{2.22.}$$

Here (2.21) is valid, since we have

$$\begin{aligned}(P(\alpha a_0, \alpha a_1)) &= (\alpha^d P(a_0, a_1) : \alpha^d Q(a_0, a_1)) \\ &= (P(a_0, a_1) : Q(a_0, a_1)).\end{aligned}$$

As for (2.22), if $P(a_0, a_1) = 0 = Q(a_0, a_1)$, then we can factor

$$\begin{aligned}P(x_0, x_1) &= (a_1 x_0 - a_0 x_1)^m P_1(x_0, x_1) \\ Q(x_0, x_1) &= (a_1 x_0 - a_0 x_1)^n Q_1(x_0, x_1).\end{aligned}$$

Assume $m \geq n$ for simplicity. Then

$$(P(x_0, x_1) : Q(x_0, x_1)) = (a_1 x_0 - a_0 x_1)^{m-n}(P_1(x_0, x_1) : Q_1(x_0, x_1)). \tag{2.23}$$

If $m = n$, then

$$((P(x_0, x_1) : Q(x_0, x_1)) = ((P_1(x_0, x_1) : Q_1(x_0, x_1)).$$

If $m > n$, the same conclusion holds by omitting a common factor. By using the same argument if necessary, we may assume that P_1 and Q_1 have no common factor and satisfy (2.22). Thus we can assume this is the case for P and Q. This illustrates a case in algebraic geometry where, although we cannot define a mapping by an equation like (2.20), we can replace the expression by another form like (2.23) and define a mapping. In algebraic geometry we normally extend the domain of definition of a mapping as much as possible. In this way, we regard (2.20) as defining a mapping from the whole of $\mathbf{P}^1(\mathbf{C})$.

Now in (2.20) P and Q have the same degree, so that

$$f(x_0, x_1) = \frac{Q(x_0, x_1)}{P(x_0, x_1)}, \quad g(x_0, x_1) = \frac{P(x_0, x_1)}{Q(x_0, x_1)}$$

are rational functions on $\mathbf{P}^1(\mathbf{C})$. Thus we may replace (2.20) by

$$(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (1 : f(a_0, a_1)) \in \mathbf{P}^1(\mathbf{C}) \tag{2.24}$$

or

$$(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (g(a_0, a_1) : 1) \in \mathbf{P}^1(\mathbf{C}). \tag{2.25}$$

For example, near a pole of f we change (2.24) into

$$(1 : f(b_0, b_1)) = \left(\frac{1}{f(b_0, b_1)} : 1\right),$$

which is the same as (2.25), since $g = 1/f$. As we see from (2.24) and (2.25), to think of a rational function on $\mathbf{P}^1(\mathbf{C})$ and to think of a rational mapping of $\mathbf{P}^1(\mathbf{C})$ into itself of the form (2.20) amount essentially to the same thing.

Now then, given a rational function on $\mathbf{P}^2(\mathbf{C})$

$$f(x_0, x_1) = \frac{Q(x_0, x_1, x_2)}{P(x_0, x_1, x_2)},$$

is

$$(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) \longmapsto (1 : f(a_0, a_1, a_2)) \in \mathbf{P}^1(\mathbf{C})$$

a mapping? This is the same as

$$(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) \longmapsto (P(a_0, a_1, a_2) : Q(a_0, a_1, a_2)) \in \mathbf{P}^1(\mathbf{C}). \tag{2.26}$$

We may further assume that P and Q have no common divisor. In contrast to the case of $\mathbf{P}^1(\mathbf{C})$ there may be points where

$$P(a_0, a_1, a_2) = Q(a_0, a_1, a_2) = 0$$

and thus the mapping cannot be defined. For instance, suppose $P(x_0, x_1, x_2) = x_0$, $Q(x_0, x_1, x_2) = x_1$. Then (2.26) gives rise to

$$(a_0 : a_1 : a_2) \longmapsto (a_0 : a_1),$$

which is not defined at $(0 : 0 : 1)$. In general, assume P and Q have no common factor. Then (2.26) cannot be defined in the intersection of the plane curves $V(P)$ and $V(Q)$ (which consists of a finite number of points). In such a case, we call the

"mapping" of the form (2.26) a **rational mapping**. A point where the mapping is not really defined is called a **point of indeterminacy**. If a rational mapping is defined as all points, we call it a **regular mapping** or an **algebraic morphism**, or simply a **morphism**.

Rational mappings can be defined in the same way from $\mathbf{P}^1(\mathbf{C})$ into $\mathbf{P}^2(\mathbf{C})$ as well as from $\mathbf{P}^2(\mathbf{C})$ into $\mathbf{P}^2(\mathbf{C})$. For example, let P, Q and R be homogeneous polynomials in x_0, x_1 of the same degree. We can define a rational mapping from $\mathbf{P}^1(\mathbf{C})$ into $\mathbf{P}^2(\mathbf{C})$ by

$$(a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (P(a_0, a_1) : Q(a_0, a_1) : R(a_0, a_1)) \in \mathbf{P}^2(\mathbf{C}).$$

We can show in the same way as above that this is actually a regular mapping.

EXAMPLE 2.2 (QUADRICS). Consider the rational mapping

$$\phi : (a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (a_0^2 : a_0 a_1 : a_1^2) \in \mathbf{P}^2(\mathbf{C}).$$

It is clear that this is an algebraic morphism. It is also obvious that the image of ϕ is contained in the quadric

$$Q : x_0 x_2 - x_1^2 = 0.$$

We now show that ϕ is a bijection (i.e. surjection and injection) from $\mathbf{P}^1(\mathbf{C})$ onto Q. Let $(b_0 : b_1 : b_2) \in Q$. Since $b_0 b_2 - b_1^2 = 0$, if $b_0 = 0$, then $b_1 = 0$, that is, the point is $(0 : 0 : 1)$. In this case, we have

$$\phi((0 : 1)) = (0 : 0 : 1).$$

If $b_0 \neq 0$, then

$$\frac{b_2}{b_0} = \left(\frac{b_1}{b_0}\right)^2,$$

and hence

$$\begin{aligned}\phi\left(\left(1 : \frac{b_1}{b_0}\right)\right) &= \left(1 : \frac{b_1}{b_0} : (\frac{b_1}{b_0})^2\right) \\ &= \left(1 : \frac{b_1}{b_0} : \frac{b_2}{b_0}\right) = (b_0 : b_1 : b_2).\end{aligned}$$

We have thus proved that ϕ is surjective.

To show that ϕ is injective, let us suppose

$$(a_0)^2 : a_0 a_1 : a_2^2) = (b_0^2 : b_0 b_1 : b_1^2).$$

If $a_0 = 0$, we get $b_0 = 0$, hence $(a_0 : a_1) = (0 : 1), (b_0 : b_1) = (0 : 1)$. If $a_0 \neq 0$, then $b_0 \neq 0$, and

$$\left(1 : \frac{a_1}{a_0} : \left(\frac{a_1}{a_0}\right)^2\right) = \left(1 : \frac{b_1}{b_0} : \left(\frac{b_1}{b_0}\right)^2\right)$$

implies $\frac{a_1}{a_0} = \frac{b_1}{b_0}$, and hence $(a_0 : a_1) = (b_0 : b_1)$. This proves that ϕ is injective.

Continuing the discussions above, we consider the rational "mapping"

$$\psi_1 : (b_0 : b_1 : b_2) \in Q \longmapsto (b_0; b_1) \in \mathbf{P}^1(\mathbf{C}).$$

This is the restriction to the quadric Q of the rational mapping

$$(x_0 : x_1; x_2) \in \mathbf{P}^2(\mathbf{C}) \longmapsto (x_0 : x_1) \in \mathbf{P}^1(\mathbf{C}),$$

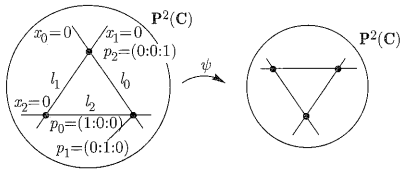


FIGURE 2.5 QUADRATIC TRANSFORMATION. ψ is not defined at the points p_1, p_2, p_3. ℓ_j collapses to p_j by ψ.

and ψ_1 is defined except at $(0 : 0 : 1)$. On the other hand, the "rational mapping"

$$\psi_2 : (b_0 : b_1 : b_2) \in Q \longmapsto (b_1 : b_2) \in \mathbf{P}^1(\mathbf{C})$$

is defined except at the point $(1 : 0 : 0)$. Moreover, at every point of $Q - \{(0 : 0 : 1), (1 : 0 : 0)\}$, where $b_0 b_2 \neq 0$, we have

$$\frac{b_1}{b_0} = \frac{b_2}{b_1}$$

and hence $\psi_1 = \psi_2$. Therefore by pasting the two mappings

$$\begin{aligned}\psi_1 &: Q - \{(0 : 0 : 1)\} \to \mathbf{P}^1(\mathbf{C}) \\ \psi_2 &: Q - \{(1 : 0 : 0)\} \to \mathbf{P}^1(\mathbf{C})\end{aligned}$$

we can get a mapping $\psi : Q \to \mathbf{P}^1(\mathbf{C})$. It is clear from the definition that ψ is the inverse of ϕ. In this way, we have seen that the projective line $\mathbf{P}^1(\mathbf{C})$ and the quadric Q are isomorphic by the algebraic morphism ϕ. As ψ involves a quadratic form, the isomorphism is not generated by a projective transformation. Only in the world of algebraic geometry does such an identification of $\mathbf{P}^1(\mathbf{C})$ and Q arise. As we saw in Example 2.1, every quadratic curve C given by a symmetric matrix of rank 3

$$(x_0, x_1, x_2) C {}^t(x_0, x_1, x_2) = 0$$

is mapped into Q by a projective transformation; it can be identified with $\mathbf{P}^1(\mathbf{C})$.

For geometric interpretations of the identification of Q and $\mathbf{P}^1(\mathbf{C})$, see §1.2 (a), §1.4 (a), and Exercise 2.5.

EXAMPLE 2.3 (QUADRATIC TRANSFORMATION). Let us consider the rational mapping from $\mathbf{P}^2(\mathbf{C})$ into $\mathbf{P}^2(\mathbf{C})$ given by

$$\psi : (a_0 : a_1 : a_2) \longmapsto (a_1 a_2 : a_2 a_0 : a_0 a_1) \in \mathbf{P}^2(\mathbf{C}). \tag{2.27}$$

(See Figure 2.5.) This rational mapping is called a **quadratic transformation** or a **Cremona transformation**. The points of indeterminacy of this rational mapping are such that

$$x_1 x_2 = 0, \quad x_2 x_0 = 0, \quad x_0 x_1 = 0,$$

namely, the three points

$$(0 : 0 : 1), \quad (0 : 1 : 0), \quad (1 : 0 : 0). \tag{2.28}$$

If a point $(a_0 : a_1 : a_2)$ satisfies $a_0 a_1 \neq 0$, then by (2.27) we have

$$\left(1 : \frac{a_1}{a_0} : \frac{a_2}{a_0}\right) \longrightarrow \left(\frac{a_2}{a_0} : \frac{a_2}{a_1} : 1\right) = \left(b : \frac{b}{a} : 1\right).$$

We see from this that ψ is injective **almost everywhere**, that is, one-to-one on the complement of an algebraic set, namely, on the set $\{a_0 : a_1 : a_2 \,|a_0 a_1 \neq 0\}$ in this case. To study the properties of ψ, let us check the images of the lines

$$\ell_j : x_j = 0, \quad j = 0, 1, 2,$$

omitting the three points in (2.28).

We have

$$\begin{aligned}
&\psi(\ell_0') = (1 : 0 : 0) \text{ for } \ell_0' = \ell_0 - \{(0 : 0 : 1), (0 : 1 : 0)\} \\
&\psi(\ell_1') = (0 : 1 : 0) \text{ for } \ell_1' = \ell_1 - \{(0 : 0 : 1), (1 : 0 : 0)\} \\
&\psi(\ell_2') = (0 : 0 : 1) \text{ for } \ell_2' = \ell_2 - \{(1 : 0 : 1), (0 : 1 : 0)\}.
\end{aligned}$$

On the other hand, $\psi^{-1}(\ell_0)$ is equal to the curve $\ell_1 \cup \ell_2$ determined by $x_1 x_2 = 0$ in $\mathbf{P}^2(\mathbf{C})$ from which the three points (2.28) have been removed. As we see from the result above, we have

$$\psi(\psi^{-1}(\ell_0)) = \{(0 : 1 : 0), (0 : 0 : 1)\}$$

and no other point on ℓ_0 is contained in the image of ψ. By similar arguments we find that ψ gives an isomorphism of $\mathbf{P}^2(\mathbf{C}) - \ell_0 \cup \ell_1 \cup \ell_2$ onto itself. In this manner ψ is almost an isomorphism, but there are points of indeterminacy as well as points not contained in the image of ψ. If we iterate ψ, we get

$$\begin{aligned}
\psi \circ \psi((a_0 : a_1 : a_2)) &= \psi((a_1 a_2 : a_2 a_0 : a_0 a_1)) \\
&= ((a_0 a_1 a_2) a_0 : (a_0 a_1 a_2) a_1 : (a_0 a_1 a_2) a_2).
\end{aligned} \tag{2.29}$$

It follows that $\psi \circ \psi$ is the identity mapping (as is the case with rational mappings, we omit a common divisor at the end of (2.29)). Thus we can write $\psi^{-1} = \psi$. When a rational mapping ψ has a rational mapping as its inverse, we say ψ is a **birational mapping**. We have an interpretation of ψ as in Figure 2.6 by using the notion of a blow-up given in §2.4 (f) and §2.5 (a).

§2.3. Plane curves

We touched upon plane curves in the preceding section. Now in this section we shall study the properties of plane curves in more detail.

(a) Tangents and singular points. Let us consider an irreducible plane curve of degree d in $\mathbf{P}^2(\mathbf{C})$

$$C : F(x_0, x_1, x_2) = 0.$$

As before we set

$$\begin{aligned}
&x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}, \\
&f(x, y) = \frac{1}{x_0^d} F(x_0, x_1, x_2).
\end{aligned} \tag{2.30}$$

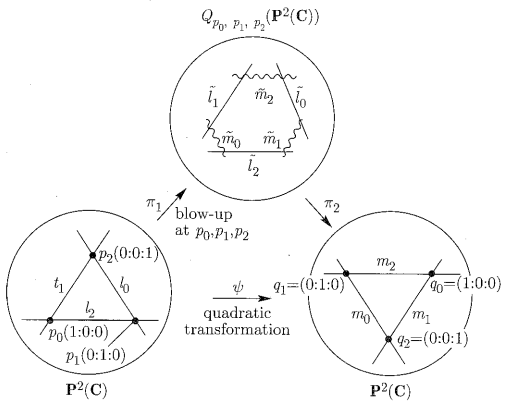


FIGURE 2.6. The quadratic transformation can be explained by using the notion of a blow-up. By the blow-up at the point p_j an exceptional curve $\tilde{m}_j$ appears. The pull-back in the narrow sense (see §2.5) of ℓ_j is $\tilde{\ell}_j$. Since $\tilde{\ell}_0, \tilde{\ell}_1, \tilde{\ell}_2$ are also exceptional curves, their blow-downs are the points q_0, q_1, q_2. The image m_j of $\tilde{m}_j$ is a line.

For the affine curve

$$C_f : f(x, y) = 0$$

in the affine plane $\mathbf{C}^2$, the equation of the tangent line at (a, b) is given by

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) = 0. \tag{2.31}$$

A point (a, b) of C_f is a singular point of C_f if it satisfies

$$\frac{\partial f}{\partial x}(a, b) = 0, \quad \frac{\partial f}{\partial y}(a, b) = 0.$$

We first rewrite the equation (2.31) of the tangent line as the equation of the tangent line for C in terms of homogeneous coordinates. We can rewrite (2.31) as

$$\frac{\partial f}{\partial x}(a, b) x_1 + \frac{\partial f}{\partial y}(a, b) x_2 - \left(a \frac{\partial f}{\partial y}(a, b) + b \frac{\partial f}{\partial y}(a, b)\right) x_0 = 0. \tag{2.32}$$

On the other hand, from (2.30) we get

$$\begin{aligned}\frac{\partial F}{\partial x_0} &= \frac{\partial}{\partial x_0}(x_0^d f(x,y)) \\ &= dx_0^{d-1} f(x,y) - x_0^{d-2}\left(x_1\frac{\partial f}{\partial x} + x_2\frac{\partial f}{\partial y}\right) \\ \frac{\partial F}{\partial x_1} &= x_0^{d-1}\frac{\partial f}{\partial x} \\ \frac{\partial F}{\partial x_2} &= x_0^{d-1}\frac{\partial f}{\partial y}.\end{aligned}$$

By setting $(a_0 : a_1 : a_2) = (1 : a : b)$ and using $f(a,b) = 0$ we get

$$\begin{aligned}\frac{\partial F}{\partial x_0}(a_0 : a_1 : a_2) &= -a_0^{d-2}\left(a_1\frac{\partial f}{\partial x}(a,b) + a_2\frac{\partial f}{\partial y}(a,b)\right) \\ \frac{\partial F}{\partial x_1}(a_0, a_1, a_2) &= a_0^{d-1}\frac{\partial f}{\partial x}(a,b) \\ \frac{\partial F}{\partial x_2}(a_0, a_1, a_2) &= a_0^{d-1}\frac{\partial f}{\partial y}(a,b).\end{aligned}$$

Multiplying (2.32) by a_o^{d-1} and using $a = a_1/a_0, b = a_2/a_0$ we get

$$\frac{\partial F}{\partial x_0}(a_0,a_1,a_2)x_0 + \frac{\partial F}{\partial x_1}(a_0,a_1,a_2)x_1 + \frac{\partial F}{\partial x_2}(a_0,a_1,a_2)x_2 = 0. \tag{2.33}$$

This equation is the same as what we get by rewriting the equation of the tangent line of an affine plane curve in terms of homogeneous coordinates. Thus we define (2.33) as the equation for the tangent line to a plane curve C at the point (a_0, a_1, a_2). It is not entirely clear from (2.33) that the tangent line goes through the point (a_0, a_1, a_2). This geometric fact is clear from Euler's identity (Exercise 1.3) for a homogeneous polynomial $F(x_0, x_1, x_2)$ of degree d:

$$\frac{\partial F}{\partial x_0}(x_0,x_1,x_2)x_0 + \frac{\partial F}{\partial x_1}(x_0,x_1,x_2)x_1 + \frac{\partial F}{\partial x_1}(x_0,x_1,x_2)x_2 = dF(x_0,x_1,x_2). \tag{2.34}$$

DEFINITION 2.3. A point $(a_0 : a_1 : a_2)$ on a plane curve

$$C : F(x_0, x_1, x_2) = 0$$

is called a **singular point** if satisfies

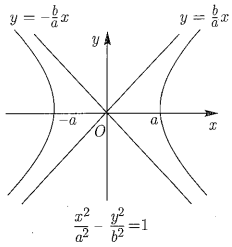
$$\frac{\partial F}{\partial x_0}(a_0,a_1,a_2) = \frac{\partial F}{\partial x_1}(a_0,a_1,a_2) = \frac{\partial F}{\partial x_0}(a_0,a_1,a_2) = 0.$$

A point of C that is not a singular point is called a **nonsingular point** or a **regular point**. A plane curve that has no singular point is called a **nonsingular plane curve**.

As is clear from the defining equation (2.33) for a tangent line, the tangent line makes sense only at a nonsingular point of C. From Euler's identity (2.34) we get the following.

LEMMA 2.10. *A point $(a_0 : a_1 : a_2)$ of $\mathbf{P}^2(\mathbf{C})$ is a singular point of the plane curve*

$$C : F(x_0, x_1, x_2) = 0$$



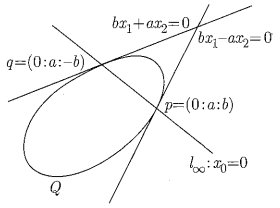
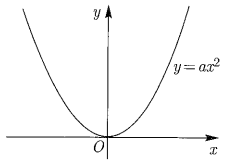


FIGURE 2.7. The asymptotes of the hyperbola are the tangent lines at the points at infinity.



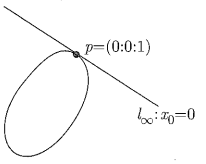


FIGURE 2.8. The tangent lines to the parabola $y = ax^2$ at the point at infinity is the line at infinity.

if and only if it satisfies

$$\frac{\partial F}{\partial x_j}(a_0,a_1,a_2) = 0, \quad j = 0, 1, 2.$$

PROOF. It suffices to show that $F(a_0, a_1, a_2) = 0$, but this is evident from Euler's identity (2.34):

$$dF(a_0,a_1,a_2) = \sum_{j=0}^{2}\frac{\partial F}{\partial x_j}(a_0,a_1,a_2) = 0.$$

EXAMPLE 2.4 (QUADRICS). (i) The plane curve corresponding to the hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

is

$$F = -x_0^2 + \frac{x_1^2}{a^2} - x_2^2 b^2 = 0,$$

as we saw before. Since

$$\frac{\partial F}{\partial x_0} = -2x_0, \quad \frac{\partial F}{\partial x_1} = \frac{2x_1}{a^2}, \quad \frac{\partial F}{\partial x_2} = -\frac{2x_2}{b^2},$$

we see that $Q = V(F)$ is a nonsingular curve. The quadric Q and the line at infinity ℓ_∞ intersect at the points

$$p = (0 : a : b), \quad q = (0 : a : -b).$$

(See Figure 2.7.) The equation of the tangent line at p is

$$bx_1 - ax_2 = 0,$$

and that of the tangent line at q is

$$bx_1 + ax_2 = 0.$$

These lines become

$$bx - ay = 0$$
$$bx + ay = 0$$

in the affine plane; they are nothing but the asymptotes to the hyperbola.

(ii) The quadric corresponding to the parabola $y = ax^2$ is

$$F = x_0x_2 - ax_1^2 = 0.$$

Since

$$\frac{\partial F}{\partial x_0} = x_2, \quad \frac{\partial F}{\partial x_1} = -2ax_1, \quad \frac{\partial F}{\partial x_2} = x_0,$$

the quadric $Q = V(F)$ is a nonsingular curve. The intersection of Q and the line at infinity ℓ_∞ is $p = (0 : 0 : 1)$. (See Figure 2.8.) The tangent line at p is $x_0 = 0$, namely, the line at infinity.

EXAMPLE 2.5. (i) The plane curve corresponding to the affine curve (see Figure 2.9 (a))

$$y^2 = x^2(x + a)$$

is

$$C : F = ax_0x_1^2 + x_1^3 - x_0x_2^2 = 0.$$

From

$$\frac{\partial F}{\partial x_0} = ax_1^2 - x_2^2, \quad \frac{\partial F}{\partial x_1} = 2ax_0x_1 + 3x_1^2, \quad \frac{\partial F}{\partial x_2} = -2x_0x_2$$

we find that C has one singular point $p = (1 : 0 : 0)$.

(ii) The plane curve corresponding to the affine curve

$$y^2 = x^4(1 - x^2)$$

(see Figure 2.9 (c)) is a curve of degree 6

$$C : F = x_0^2x_1^4 - x_1^6 - x_0^4x_2^2 = 0.$$

From

$$\frac{\partial F}{\partial x_0} = 2x_0x_1^4, \quad \frac{\partial F}{\partial x_1} = 4x_0^2x_1^3 - 6x_1^5, \quad \frac{\partial F}{\partial x_2} = -2x_0^4x_2,$$

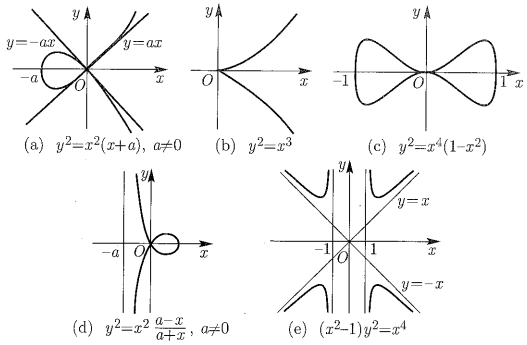


FIGURE 2.9

we find that C has two singular points, $p = (1 : 0 : 0)$ and $q = (0 : 0 : 1)$.

(iii) The strophoid (see Figure 2.9 (d))

$$y^2 = x^2\frac{a - x}{a + x}, \quad a \neq 0,$$

can be regarded as an affine curve

$$(a + x)y^2 = x^2(a - x),$$

and the corresponding plane curve is a cubic curve

$$C : F = ax_0x_2^2 + x_1x_2^2 - ax_0x_1^2 + x_1^3 = 0.$$

From

$$\frac{\partial F}{\partial x_0} = a(x_2^2 - x_1^2), \quad \frac{\partial F}{\partial x_1} = x_2^2 - 2ax_0x_1 + 3x_1^2, \quad \frac{\partial F}{\partial x_2} = 2x_2(ax_0 + x_1)$$

we see that C has one singular point $p = (1 : 0 : 0)$. The intersection of C and the line at infinity consists of the three points $(0 : 0 : 1), (0 : 1 : i), (0 : 1 : -i)$. The tangent line to C at $(0 : 0 : 1)$ is

$$ax_0 + x_1 = 0,$$

which is nothing but the asymptote

$$x = -a$$

of the strophoid. The tangent lines at $(0 : 1 : \pm i)$ are

$$-ax_0 + x_1 \pm ix_2 = 0.$$

This is also an asymptote to the strophoid, but cannot be seen in an illustration on the usual real plane.

(iv) The plane curve corresponding to the affine curve

$$(x^2-1)y^2 = x^4$$

(see Figure 2.9 (e)) is

$$C : F = x_1^2x_2^2 - x_0^2x_2^2 - x_1^4 = 0.$$

From

$$\frac{\partial F}{\partial x_0} = -2x_0x_2^2, \quad \frac{\partial F}{\partial x_1} = 2x_1(x_2^2 - 2x_1^2), \quad \frac{\partial F}{\partial x_0} = 2x_2(x_1^2 - x_0^2)$$

we find that the singular points of C are $p = (1:0:0), q = (0:0:1)$. The intersection of C with the line at infinity ℓ_∞ consists of the three points $q = (0:0:1), (0:1:1), (0:1:-1)$. The tangent lines at $(0:1:\pm 1)$ are

$$x_1 \mp x_2 = 0.$$

These are nothing but the asymptotes of the affine curve

$$x \mp y = 0.$$

On the other hand, the affine lines $x = \pm 1$ are expressed by $x_1 = \pm x_0$ as lines in $\mathbf{P}^2(\mathbf{C})$, and they intersect C at $q = (0:0:1)$. The point q being a singular point, the tangent line cannot be defined there but we can define the tangent cone, as stated in §1.5 (a). In our case, the tangent cone at q is given by

$$x_1^2 - x_0^2 = 0,$$

which is the sum of the two lines $x_1 = \pm x_0$. This is why $x = \pm 1$ appear as asymptotes of the affine plane curve.

Now we shall study the situation around singular points in more detail. First we consider the case where the point $(1:0:0)$ is a singular point of the plane curve $C : F(x_0, x_1, x_2) = 0$. (Actually, we can assume that a given singular point is $(1:0:0)$ by performing a projective transformation.) Let d be the degree of F and let $x = x_1/x_0, y = x_2/x_0$ be affine coordinates. The corresponding affine curve is

$$C_f : f(x,y) = \frac{1}{x_0^d}F(x_0, x_1, x_2).$$

Since the origin $(0,0)$ is a singular point of the affine curve C_f, we can write

$$f(x,y) = \sum_{i+j\geq 2}^{d} a_{ij}x^iy^j.$$

We can easily see that the origin is a singular point of C_f if and only if we have

$$a_{10} = a_{01} = 0.$$

Thus we can write

$$f(x,y) = f_m(x,y) + f_{m+1}(x,y) + \cdots + f_d(x,y), \quad m \geq 2.$$

Here $f_n(x,y)$ is the sum of all the terms of degree n that appear in $f(x,y)$, and we assume $f_m(x,y) \not\equiv 0$. In this case, we say that the singular point has **multiplicity** m. The homogeneous polynomial $f_m(x,y)$ of degree m can be factored into linear factors

$$f_m(x,y) = \prod_{j=1}^{m}(\alpha_j x - \beta_j y).$$

We call

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

the tangent cone of the plane curve C at the singular point $(1:0:0)$. This consists of a finite number of lines. If the origin is a nonsingular point, $f_1(x,y) = 0$ is the equation of the tangent line.

EXAMPLE 2.6. (i) The affine curve

$$f(x,y) = y^2 - x^2(x+a), \quad a \neq 0,$$

has a singular point with multiplicity 2. It is called an **ordinary double point**. The tangent cone at the origin is given by

$$x_2^2 - ax_1^2 = 0,$$

which consists of two lines through the origin. (Incidentally, if we consider the curve over a field of characteristic 2, we get $(x_2 - \sqrt{a}x_1)^2 = 0$. This singular point is an ordinary cusp defined in (ii).) These two lines are regarded as tangent to the curve at the origin. (See Figure 2.9 (a).)

(ii) For the affine curve

$$f(x,y) = y^2 - x^3$$

the origin is a singular point of multiplicity 2 and is called an **ordinary cusp** or a cusp of type $(2,3)$. The tangent cone at the origin is

$$x_2^2 = 0,$$

and is the x-axis counted twice. (See Figure 2.9 (b).)

If a singular point of a plane curve C is elsewhere, we can map the singular point to $(1:0:0)$ by a projective transformation, as we stated earlier. It is, however, possible to deal with the case directly. First, consider the case where a singular point is $(1:a:b)$.. Then the corresponding affine curve C_f has (a,b) as a singular point. We can expand f in the form

$$f(x,y) = \sum_{i+j=2}^{d} c_{ij}(x-a)^i(y-b)^j$$

without linear terms in $x-a$ and $y-b$. We can again write

$$f(x,y) = g_m(x-a, y-b) + g_{m+1}(x-a, y-b) + \cdots + g_d(x-a, y-b), \quad m \geq 2.$$

Here $g_\ell(x-a, y-b)$ is a homogeneous polynomial of degree ℓ in $x-a$ and $y-b$, and $g_m(x-a, y-b) \not\equiv 0$. In this case, the multiplicity of the singular point is m and the tangent cone is defined by

$$g_m(x_1 - ax_0, x_2 - bx_0) = 0.$$

Since we can factor

$$g_m(x_1 - ax_0, x_2 - bx_0) = \prod_{i=1}^{m}\{\alpha_i(x_1 - ax_0) - \beta_j(x_2 - bx_0)\},$$

the tangent cone is the sum of lines including multiplicities.

Next, in the case where the singular point is $(0:1:b)$, we set $u = x_0/x_1$ and $v = x_2/x_1$, and consider

$$g(u,v) = \frac{1}{x_1^d} F(x_0, x_1, x_2).$$

The point $(0,b)$ is a singular point of the affine curve

$$C_g : g(u,v) = 0,$$

and we can write

$$\begin{aligned} g(u,v) &= \sum_{i+j=2}^{d} g_{ij} u^i (v-b)^j \\ &= g_m(u, v-b) + g_{m+1}(u, v-b) + \cdots + g_d(u, v-b), \end{aligned}$$

that is, as the sum of homogeneous polynomials. If $g_m \not\equiv 0$, then the multiplicity of the singular point is m and

$$g_m(x_0, x_2 - bx_1) = 0$$

is the tangent cone.

The case where the singular point is $(0:0:1)$ is left to the reader.

EXAMPLE 2.7. From Example 2.5 (iv), q is a singular point of the plane curve

$$C : F = x_1^2 x_2^2 - x_0^2 x_2^2 - x_1^4 = 0.$$

By setting $w = x_0/x_2, z = x_1/x_2$ we find that the origin $(0,0)$ is a singular point of the plane curve

$$C_h : h(w,z) = \frac{1}{x_2^4} F(x_0, x_1, x_2) = z^2 - w^2 - z^4.$$

This is an ordinary double point with the tangent cone

$$x_0^2 - x_1^2 = 0,$$

consisting of two lines corresponding to $x = \pm 1$. (See Figure 2.9 (e).)

We shall study singular points in more detail. For simplicity, we assume that $p = (1:0:0)$ is a singular point and look at the point $(0,0)$ of the corresponding affine curve

$$C_f : f(x,y) = 0,$$

where $f(x,y)$ is irreducible as a polynomial. However, as a formal power series (even as a convergent power series) it may not be irreducible but can be factored. For instance,

$$f(x,y) = y^2 - x^2(x+a), \quad a \neq 0,$$

is irreducible as a polynomial. But using the power series expansion

$$\sqrt{x+a} = \sqrt{a} \sum_{j=0}^{\infty} \frac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2} \left(\frac{x}{a}\right)^j$$

we can factor $f(x,y)$ as follows:

$$\begin{aligned} &f(x,y) \\ &= \left(y - \sqrt{a} \sum_{j=0}^{\infty} \frac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2 a^j} x^{j+1} \right) \left(y - \sqrt{a} \sum_{j=0}^{\infty} \frac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2 a^j} x^{j+1} \right). \end{aligned}$$

In general, when we have a factorization into irreducible factors

$$f(x,y) = \prod_{j=1}^{n} g_j(x,y) \tag{2.35}$$

using formal power series, $g_j(x,y) = 0$ for each j is called a **branch** of the curve $f(x,y) = 0$ at the point $(0,0)$. In the example above, there are two branches. Each branch $g_j(x,y) = 0$ may no longer be an algebraic curve, but it represents part of the original affine curve C_f in a neighborhood of the origin. Actually, in the example above the curve is branched out in two parts, as is clear from Figure 2.9 (a). (Of course, the whole curve cannot be separated into two parts.)

Now then, how do we find a factorization of the form (2.35)? For this purpose, we first write f in the form

$$f(x,y) = a_0(x) y^n + a_1(x) y^{n-1} + \cdots + a_n(x).$$

Considering the fact that if we substitute a value a in x, then $f(a,y) = 0$ will have n roots, we shall forcefully factor f in the form

$$f(x,y) = a_0(x) \prod_{j=1}^{n} (y - f_j(x)). \tag{2.36}$$

Due to forced factorization, each $f_j(x)$ may be not a power series in x but a power series in x^{1/m_j} or a Laurent series in which a finite number of terms with negative exponents appear. For instance,

$$y^2 - x^3 = (y - x^{3/2})(y + x^{3/2})$$

is a simple example.

We shall write

$$f_j(x) = g_j(x^{1/m_j}), \quad g_j(t) = \sum_{k=k_0}^{+\infty} a_k^j t^k. \tag{2.37}$$

For $m \geq 2$, x^{1/m_j} is a multi-valued function. That is, if $b^{m_j} = a$, then the m_j values of x^{1/m_j} at a are

$$b, \epsilon b, \epsilon^2 b, \cdots, \epsilon^{m_j - 1} b, \text{ where } \epsilon = e^{2\pi/m_j}.$$

The function $f_j(x)$ is as multi-valued as x^{1/m_j} is. We may clarify the multi-valued character by introducing a variable s and setting

$$\begin{cases} x = s^{m_j} \\ f_j(x) = g_j(s). \end{cases} \tag{2.38}$$

We see that the function x of s has the same value for $b, \epsilon b, \ldots, \epsilon^{m_j-1} b$ above, but $g_j(b), g_j(\epsilon b), \ldots, g_j(\epsilon^{m_j-1} b)$ take different values in general. This is the way the multi-valueness of $f_j(x)$ behaves.

The left-hand side of the factorization (2.36) is a polynomial in x, y. If the right-hand side has any factor

$$y - g_j(x^{1/m_j}) \tag{2.39}$$

in the notation of (2.36), then

$$y - g_j(\epsilon^k x^{1/m_j}), \quad \epsilon = e^{2\pi i/m_j},$$

must also appear as factors of the right-hand side, because otherwise terms with fractional powers of x would remain on the right-hand side. It follows that if a factor of the form (2.39) appears on the right-hand side of (2.36), other factors are automatically determined and the set of all these factors determine a branch of the curve at $(x, y) = (0, 0)$. Thus we see that

$$b_0(x) \prod_{k=0}^{m_j-1} (y - g_j(\epsilon^k x^{1/m_j})).$$

Here $b_0(x)$ is a factor of $a_0(x)$. Hence the curve has a finite number of branches at a singular point.

As is obvious from the discussions above, a branch is determined by a parametric representation in (2.38). This means that in a neighborhood of $(0, 0)$ the mapping

$$s \longmapsto (x, y) = (s^{m_j}, g_j(s)) \tag{2.40}$$

gives a parametric representation of the curve. When there are several branches at the singular point $(0, 0)$, this parametric representation expresses one part of the curve (i.e., one branch) and we need other parametric representations for other branches. In other words, a branch of the curve at $(0, 0)$ means that we take out a portion that admits a parametric representation.

We make a remark about the parametric representation (2.38): $g_j(s)$ might contain a term with a negative exponent of s. In this case, (2.38) is not defined at $s = 0$. But this simply means that we picked the wrong inhomogeneous coordinates $x = x_1/x_0, \quad y = x_2/x_0$. If we set $u = s_0/x_2, v = x_1/x_0$, then we have

$$s \longrightarrow (u, v) = \left(\frac{1}{g_j(s)}, \frac{s^{m_j}}{g_j(s)}\right),$$

which is defined in a neighborhood of $(0, 0)$ and gives a parametric representation.

EXAMPLE 2.8. Let us again consider an affine curve with an ordinary cusp

$$y^2 - x^3 = 0.$$

We can factor

$$y^2 - x^3 = (y - x^{3/2})(y + x^{3/2})$$

and the parametrization (2.38) is given by

$$\begin{cases} x = s^2 \\ y = s^3. \end{cases}$$

There is only one branch at the singular point $(0, 0)$. If we interchange x and y and write

$$y^2 - x^3 = -(x - y^{2/3})(x - \omega y^{2/3})(x - \omega^2 y^{2/3}), \quad \omega = e^{2\pi i/3},$$

then we obtain the parametrization

$$\begin{cases} y = t^3 \\ x = t^2, \end{cases}$$

corresponding to the parametrization (2.38).

In the above example, interchanging x and y had no effect on the number of branches (this is geometrically obvious).

EXAMPLE 2.9. Consider the affine curve

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

with an ordinary double point. We already know that we can factor

$$y^2 - x^2(x+1) = \left(y - \sum_{j=0}^{\infty} \frac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2} x^{j+1}\right) \times \left(y + \sum_{j=0}^{\infty} \frac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2} x^{j+1}\right).$$

Since the factors on the right-hand side are power series in x and y, there are two branches. The parametric representations are

$$\begin{cases} x = s \\ y = \pm \sum_{j=0}^{\infty} \frac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2} s^{j+1}. \end{cases}$$

This can be observed in Figure 2.9 (a).

EXAMPLE 2.10. The affine curve

$$y^3 + xy^2 - (x + x^2)y + (x^2 + 2x^3) = 0$$

has the origin $(0, 0)$ as a singular point. (See Figure 2.10.) By factoring the left-hand side we obtain

$$(y - x - 3x^2 - 12x^3 - 84x^4 - \cdots) \times (y - x^{1/2} + x + \frac{3}{2}x^2 + 3x^{5/2} + \cdots)$$
$$\times (y + x^{1/2} + x + \frac{3}{2}x^2 - 3x^{5/2} + \cdots).$$

Thus there are two branches with parametric representations:

$$\begin{cases} x = s \\ y = s + 3s^2 + 12s^3 + 84s^4 + \cdots \end{cases}$$

and

$$\begin{cases} x = s^2 \\ y = s - s^2 - 3/2s^4 - 3s^5 + \cdots. \end{cases}$$

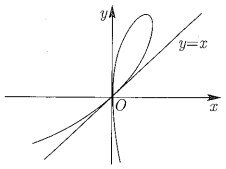


FIGURE 2.10. $y^3 + xy^2 - (x + x^2)y + (x^2 + 2x^3) = 0$

So far we have assumed that a singular point is located at the origin $(0,0)$ in the (x,y) plane, namely, the point $(1:0:0)$ of the projective plane $\mathbf{P}(\mathbf{C})^2$. If a singular point is at a point $(a_0 : a_1 : a_2)$, we can move it to $(1:0:0)$ by a projective transformation. The variable s that appears in the parametric representation (2.38) is called a **local parameter** of the branch of the curve at the singular point. If $(0,0)$ is not a singular point, then $\frac{\partial f}{\partial x}(0,0) \neq 0$ or $\frac{\partial f}{\partial y}(0,0) \neq 0$. If $\frac{\partial f}{\partial x}(0,0) \neq 0$, then by virtue of the inverse function theorem we can find locally a convergent power series

$$x = \sum_{j=1}^{\infty} a_j y^j$$

to solve $f(x,y) = 0$. In this case, we call y a local parameter at $(0,0)$, since the curve can be parametrized by

$$y \longmapsto (\sum_{j=1}^{\infty} a_j y_j, y)$$

in a neighborhood of $(0,0)$. If $\frac{\partial f}{\partial y}(0,0) \neq 0$, then $f(x,y) = 0$ can be solved in the form

$$y = \sum_{i=1}^{\infty} b_i x^i$$

in a neighborhood of $(0,0)$. We get a parametric representation of the curve

$$x \longmapsto (x, \sum_{i=1}^{\infty} b_i x^i)$$

in a neighborhood of $(0,0)$. In this case, x is a local parameter. If $\frac{\partial f}{\partial x}(0,0) \neq 0$, $\frac{\partial f}{\partial y}(0,0) \neq 0$, then x is represented as a power series in y, and vice versa; we can use either of them as a local parameter. In general, there are many ways of choosing a local parameter. If t is a local parameter, then a convergent power series

$$s = \sum_{k=1}^{\infty} c_k t^k$$

is a local parameter if and only if $c_1 \neq 0$; in this case, we can write

$$t = \sum_{\ell=1}^{\infty} d_\ell s^\ell.$$

Therefore we may use a more general form of the parametric representation (2.38) as follows:

$$\begin{cases} x = (\sum_{\ell=1}^{\infty} d_\ell s^\ell)^{m_j}, & d_1 \neq 0, \\ f_j(x) = h_j(s). \end{cases}$$

However, we normally use (2.38) to avoid complexity.

More than one branch appears at a singular point because different parts of the curve intersect at one point. If we regard each branch as a separate part of the genuine curve and introduce a local parameter for each branch and a local parameter at a non-singular point, then we are naturally led to the structure of a **complex manifold**. This is nothing but the idea of a Riemann surface. (We shall discuss Riemann surfaces in Chapter 4). From the viewpoint of algebraic geometry we may, by removing a singular point, determine an algebraic curve corresponding to a Riemann surface. We shall discuss this in §2.5.

(b) The intersection theory of plane curves. In the preceding subsection, we discussed singular points of plane curves in detail. We now use the information to give an outline of the **intersection theory** of plane curves. To find the intersections of the curves

$$C : F(x_0, x_1, x_2) = 0$$
$$D : G(x_0, x_1, x_2) = 0,$$

we solve the system of equations $F = 0, G = 0$. If C and D are different and have degrees m and n, respectively, Bézout's theorem says that the number of intersections is mn including multiplicities (see Theorem 1.1). Finding intersections is relatively simple because it is a matter of solving a system of equations. How do we go about finding the multiplicity of each intersection? First we shall discuss the method of describing the multiplicity by using the parametric representation (2.38) for a branch. For simplicity, assume that the intersection is $p = (1:0:0)$ and set

$$x = x_1/x_0, \qquad y = x_2/x_0,$$

Further setting

$$f(x,y) = \frac{1}{x_0^m} F(x_0, x_1, x_2)$$
$$g(x,y) = \frac{1}{x_0^n} G(x_0, x_1, x_2),$$

we want to find the **intersection multiplicity** of the affine curves $C_f : f(x,y) = 0$ and $C_g : g(x,y) = 0$ at the origin $(0,0)$. We call it the **intersection multiplicity** $I_p(C,D)$ of the plane curves C and D at p. Assume that the affine curve $f(x,y) = 0$ has k branches at $(0,0)$ and that each branch is parametrized by

$$\begin{cases} x = s^{m_j} \\ y = h_j(s). \end{cases}$$

Substituting this parametrization into $g(x,y)$, we find that

$$g(s^{m_j}, h_j(s)) = \alpha s^{\ell_j} + \text{ terms in } s \text{ of degree } \geq \ell_j + 1.$$

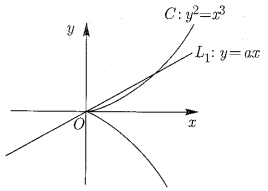


FIGURE 2.11. The intersection number at the origin of the x-axis and the curve $y^2 = x^3$ is 3. The intersection multiplicity of any other line through the origin with $y^2 = x^3$ at the origin is 2

Here we can regard ℓ_j as the intersection multiplicity of the j-th branch and $g(x, y) = 0$ at the origin O. Hence we get

$$I_O(C_f, C_g) = \sum_{j=1}^{k} \ell_j.$$

Thus we define the intersection multiplicity of the plane curves C and D at $p = (1 : 0 : 0)$ to be

$$I_p(C, D) = \sum_{j=1}^{k} \ell_j.$$

This definition was derived by starting with the parametric representation of a branch of C_f. We can show that we arrive at the same conclusion by starting with the parametric representation of a branch of C_g, namely,

$$I_p(C, D) = I_p(D, C).$$

EXAMPLE 2.11. Let us consider the affine curves

$$C : y^2 = x^3$$
$$L_1 : y = ax$$
$$L_2 : x = 0.$$

(See Figure 2.11.) We find the intersection numbers of C and L_1 and of C and L_2. At the origin C has only one branch, and its parametrization is given by

$$\begin{cases} x = s^2 \\ y = s^3. \end{cases}$$

By substituting it in $y - ax$ we get $s^3 - as^2$ and hence

$$I_O(C, L_1) = \begin{cases} 2 & a \neq 0, \\ 3 & a = 0. \end{cases}$$

Similarly, we have

$$I_O(C, L_2) = 2.$$

Conversely, if we start with L_1 and substitute its parametric representation at the origin

$$\begin{cases} x = t \\ y = at \end{cases}$$

into $y^2 - x^3$, we get $a^2t^2 - t^3$. Therefore we have

$$I_O(L_1, C) = \begin{cases} 2, & a \neq 0, \\ 3, & a = 0, \end{cases}$$

and certainly

$$I_O(C, L_1) = I_O(L_1, C).$$

EXAMPLE 2.12. We find the intersection multiplicity at the origin of the affine curves

$$C : y^2 = x^3$$
$$D : y^2 = x^2(x + 1).$$

Substituting the parametric representation of C at the origin

$$\begin{cases} x = s^2 \\ y = s^3 \end{cases}$$

into $y^2 - x^2(x + 1)$, we get $s^6 - s^4(s^2 + 1)$, and thus

$$I_O(C, D) = 4.$$

On the other hand, D has two branches at the origin. If we substitute the parametric representation of D

$$\begin{cases} x = s \\ y = \pm \sum_{j=0}^{\infty} \dfrac{(-1)^{j-1}(2j)!}{(2j-1)2^{2j}(j!)^2} s^{j+1} \end{cases}$$

into $y^2 - x^3$, we get an expansion starting with the term s^2, and hence

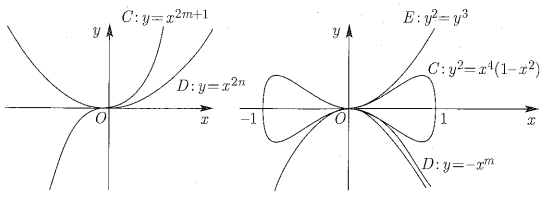
$$I_O(D, C) = 2 + 2 = 4.$$

The computations above are simple and accurate without depending on intuition. What we have defined here as the intersection multiplicity is equal to the notion defined in §1.5 (a), namely, the number of intersections near the origin when C or D is moved a little. Although we are not ready for the details, we can show, by developing the ideas above and using the notion of the formal power series $C[[x, y]]$ (see Example A.20 in the Appendix), that

$$I_O(C_f, C_g) = \dim_{\mathbf{C}} \mathbf{C}[[x, y]]/(f, g),$$

where (f, g) is the ideal of $C[[x, y]]$ generated by $f(x, y)$ and $g(x, y)$. The equation above says that the dimension over $\mathbf{C}$ of the residue ring $\mathbf{C}[[x, y]]/(f, g)$ is equal to the intersection multiplicity at the origin O. For example, for $f(x, y) = y^2 - x^3$ and $g(x, y) = x$ we have

$$\mathbf{C}[[x, y]]/(y^2 - x^3, x) \equiv \mathbf{C}[[y]]/(y^2)$$



(a) $I_O(C, D)=\begin{cases}2m+1, & m<n,\\ 2n, & m\geq n.\end{cases}$ (b) $I_O(C, D)=\begin{cases}2, & m=1,\\ 6, & m=2,\\ 4, & m\geq 3,\end{cases}$

$I_O(C, E)=6.$

FIGURE 2.12. The intersection multiplicities of various curves at the origin.

The right-hand side has a basis $\{\bar{1}, \bar{y}\}$ as a vector space over $\mathbf{C}$, where $\bar{1}$ and $\bar{y}$ are the residue classes of 1 and y. Hence

$$\dim_{\mathbf{C}} \mathbf{C}[[x, y]]/(y^2 - x^3, x) = 2.$$

This certainly coincides with $I_O(C, L_2) = 2$.

So far we have discussed the intersections at the point $(1 : 0 : 0)$. For another point, we may move it to $(1 : 0 : 0)$ by a projective transformation and apply the same method. The intersection number of the plane curves C and D is defined by

$$I(C, D) = \sum_{p \in C \cap D} I_p(C, D).$$

Bézout's theorem states that this is equal to the product of the degrees of C and D. Unfortunately, the proof of this theorem requires further preparations and has to be omitted.

(c) Function fields of plane curves. Given a rational function r on $\mathbf{P}^2(\mathbf{C})$ of the form

$$r = \frac{G(x)}{H(x)},$$

where $G(x), H(x)$ are homogeneous polynomials of degree m, we consider its restriction to an irreducible plane curve of degree d

$$C : F(x) = 0.$$

Unless $H(x)$ is identically 0 on C, the restriction $r|_C$ makes sense and can be regarded as a rational function on C. The set $\mathbf{C}(C)$ of all such rational functions $r|_C$, where r is a rational function on $\mathbf{P}^2(\mathbf{C})$ whose pole does not contain the curve C, is called the **function field** of C. To study the structure of $\mathbf{C}(C)$ it is necessary to find a condition under which a homogeneous polynomial $H(x)$ vanishes identically on C. The following is a special case of the Hilbert zero point theorem, often called by the German name "Nullstellensatz". (See Theorem 2.4.)

THEOREM 2.2. *A homogeneous polynomial $H(x_0, x_1, x_2)$ vanishes identically on the irreducible curve*

$$C : F(x) = 0$$

if and only if $H(x)$ is divisible by $F(x)$.

To say that $H(x)$ is identically 0 on C means that the plane curve $H(x)$ has C as an irreducible component. Thus the meaning of the theorem is intuitively clear.

By this theorem we see that a rational function r determines a rational function $r|_C$ on C provided the denominator $H(x)$ is not divisible by $F(x)$. Moreover, if the numerator $G(x)$ of r is divisible by $F(x)$, then $r|_C \equiv 0$. More generally, for any homogeneous polynomials $G_1(x)$ and $H_1(x)$ of degree m, if both $G_1(x) - G(x)$ and $H_1(x) - H(x)$ are divisible by $F(x)$, the restriction $r_1|_C$ of $r_1 = G_1(x)/H_1(x)$ to C is the same as $r|_C$. This is because for any point $(a_0 : a_1 : a_2)$ on C we have $H_1(a) = H(a)$ and $G_1(a) = G(a)$. In other words, $r|C$ depends only on the remainders in dividing $G(x)$ and $H(x)$ by $F(x)$. This also means that we can get any rational function on C as the restriction $f|_C$ of $f(x) = G(x)/H(x)$, where $G(x)$ and $H(x)$ have degree less than d and $H(x)$ is not a constant multiple of $F(x)$.

In terms of inhomogeneous coordinates $x = x_1/x_0, y = x_2/x_0$ we may write

$$r = \frac{g(x, y)}{h(x, y)}$$
$$g(x, y) = \frac{1}{x_0^m} G(x_0, x_1, x_2)$$
$$h(x, y) = \frac{1}{x_0^m} H(x_0, x_1, x_2).$$

If we also write the defining equation for C in inhomogeneous coordinates

$$f(x, y) = \frac{1}{x_0^d} F(x_0, x_1, x_2),$$

then the function field $\mathbf{C}(C)$ is equal to the set of all $r = \frac{g(x,y)}{h(x,y)}$, where $h(x, y)$ is not divisible by $f(x, y)$, with the relation $f(x, y) = 0$ taken into consideration. In the language of commutative rings, $\mathbf{C}(C)$ is nothing but the quotient field $\mathbf{C}[x, y]/(f(x, y))$. We shall discuss this a little further in the next section.

EXAMPLE 2.13. Let us find the function field $\mathbf{C}(C)$ of the curve

$$C : x_0 x_2^2 - x_1^4 = 0.$$

In inhomogeneous coordinates the defining equation is

$$y^2 - x^3 = 0,$$

and so it is necessary to consider

$$r = \frac{g(x, y)}{h(x, y)},$$

where $h(x, y)$ is not divisible by $y^2 - x^3$, in connection with the relation $y^2 - x^3 = 0$. If we consider the correspondence

$$r = \frac{g(x, y)}{h(x, y)} \longmapsto \bar{r}(t) = \frac{g(t^2, t^3)}{h(t^2, t^3)},$$

we get a mapping from $\mathbf{C}(C)$ into the rational function field $\mathbf{C}(t)$. Since $h(x,y)$ is not divisible by y^2-x^3, we have $h(t^2,t^3)\not\equiv 0$, so $\bar{r}(t)$ is always defined. If $\bar{r}(t)\equiv 0$, then $g(t^2,t^3)\equiv 0$, which means that $g(x,y)$ is divisible by y^2-x^3, and thus $r|_C\equiv 0$. It follows that by the correspondence above we can write $\mathbf{C}(C)\subset\mathbf{C}(t)$. Conversely, if an element $a(t)\in\mathbf{C}(t)$ is written in the form

$$a(t)=\frac{A(t)}{B(t)},$$

where $A(t)$ is a linear polynomial and $B(t)$ is a polynomial of degree m, then

$$r=\frac{A(y/x)}{B(y/x)}$$

is a rational function of two variables. By construction of r, we get $\bar{r}(t)=a(t)$ and hence $\mathbf{C}(C)=\mathbf{C}(t)$. In other words, the function fields of C and $\mathbf{P}^1(\mathbf{C})$ coincide. As we show in §2.5(b), C with singular points removed is isomorphic to $\mathbf{P}^1(\mathbf{C})$.

EXAMPLE 2.14. We shall find the function field of the cubic curve with an ordinary double point

$$C: x_0x_2^2-x_1^2(x_1+x_0)=0.$$

In terms of inhomogeneous coordinates the defining equation is

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

Let us set $u=(y/x)|_C$. Since the rational function on $\mathbf{P}^2(\mathbf{C})$

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

is 0 when restricted to C, we have

$$u^2=x|_C+1.$$

Thus

$$x|_C=u^2-1$$
$$y|_C=x|_C\cdot u=u(u^2-1).$$

Therefore by the correspondence

$$r=\frac{g(x,y)}{h(x,y)}\longmapsto\bar{r}(u)=\frac{g(1-u^2,u(1-u^2))}{h(1-u^2,u(1-u^2))}\in\mathbf{C}(u)$$

we get $\mathbf{C}(C)=\mathbf{C}(u)$. We shall show that C with the singular point removed is the same as $\mathbf{P}^1(\mathbf{C})$.

EXAMPLE 2.15. We wish to find the function field of a nonsingular cubic

$$C: x_0x_2^2-4x_1^3-g_2x_0^2x_1-g_3x_0^3=0,\quad g_2^2-27g_3^2\neq 0.$$

In inhomogeneous coordinates the defining equation for C is

$$y^2-4x^3-g_2x-g_3=0.$$

Here x,y are rational functions on $\mathbf{P}^2(\mathbf{C})$. For simplicity, we use the same letters to denote the restrictions to $\mathbf{C}$. Then on C we have

$$y^2=4x^3+g_2x+g_3$$

as rational functions. By replacing y^2 with $4x^3+g_2x+g_3$, a rational function on C can be written as the restriction to C of a rational function of the form

$$\frac{a(x)+b(x)y}{c(x)+d(x)y}.$$

Since

$$\frac{a(x)+b(x)y}{c(x)-d(x)y}=\frac{(a(x)+b(x)y)(c(x)-d(x)y)}{c(x)^2-d(x)^2y^2}$$
$$=\frac{A(x)+B(x)y}{C(x)},$$

it follows that the rational function on C can be expressed as the restriction to C of

$$\alpha(x)+\beta(x)y,\ \text{where } \alpha(x),\beta(x)\in\mathbf{C}(x).$$

Moreover, if $\alpha(x)+\beta(x)y|_C\equiv 0$, then $\alpha(x)+\beta(x)y\equiv 0$. Therefore

$$\mathbf{C}(C)=\{\alpha(x)+\beta(x)y|\,\alpha(x),\beta(x)\in\mathbf{C}(x)\}.$$

We may compute products and quotients for elements in $\mathbf{C}(C)$ by using $y^2=4x^3+g_2x+g_3$. To indicate this, we also write

$$\mathbf{C}(C)=\mathbf{C}(x,\sqrt{4x^3+g_2x+g_3}).$$

We see that the function field of the nonsingular cubic plane curve is different from that of the projective line. The field $\mathbf{C}(C)$ is often called the **elliptic function field**.

EXAMPLE 2.16. The elliptic function field can be generalized to the **hyperelliptic function field**. For a polynomial $f(x)$ in one variable x with n distinct roots, we consider the curve defined by

$$y^2=f(x)$$

in the inhomogeneous coordinates. The function field can be found to be

$$\mathbf{C}(C)=\{\alpha(x)+\beta(x)y|\,\alpha(x),\beta(x)\in\mathbf{C}(x)\},$$

by a method similar to that in Example 2.15. Or we may write this as

$$\mathbf{C}(C)=\mathbf{C}(x,\sqrt{f(x)}).$$

$\mathbf{C}(C)$ is called the hyperelliptic function field for $n\geq 5$.

§2.4. Projective varieties

(a) Projective space. From the discussions we had so far the reader will have no problem in defining the n-dimensional complex projective space $\mathbf{P}^n(\mathbf{C})$. Denote by $\mathbf{C}^{n+1}$ the set of all $(n+1)$-tuples of complex numbers $(a_0,a_1,\ldots,a_n)$ and set

$$W=\mathbf{C}^{n+1}-\{(0,0,\ldots,0)\}.$$

Take the ratio $a_0:a_1:\cdots:a_n$ determined by an element $(a_0,a_1,\ldots,a_n)\in W$. The set of all such ratios is denoted by $\mathbf{P}^n(\mathbf{C})$ and called the **n-dimensional complex projective space**. For $n=1$, we have the projective line; for $n=2$, we have the

projective plane. Denote by $(a_0 : a_1 :, \cdots : a_n)$ the point in $\mathbf{P}^n(\mathbf{C})$ determined by the ratio $a_0 : a_1 : \cdots : a_n$. Thus

$$(a_0 : a_1 : \cdots : a_n) = (\alpha a_0 : \alpha a_1 : \cdots : \alpha a_n)$$

for any non-zero complex number α.

Just as before, we define for $i = 0, 1, 2, \ldots, n$

$$U_i = \{(a_0 : a_1 : \cdots : a_n) \in \mathbf{P}^n(\mathbf{C}) | a_i \neq 0\}.$$

In the same way as in the cases $n = 1, 2$ we have a bijection

$$\phi_i : (a_0 : a_1 : \cdots : a_n) \in U_i \longmapsto (\frac{a_0}{a_i}, \ldots, \frac{a_{i-1}}{a_i}, \frac{a_{i+1}}{a_i}, \ldots \frac{a_n}{a_i}) \in \mathbf{C}^n \tag{2.41}$$

that identifies U_i with the n-dimensional affine space $\mathbf{C}^n$.

Let us concentrate on U_0 as before. We identify U_0 with $\mathbf{C}^n$. From

$$\mathbf{P}^n(\mathbf{C}) - U_0 = \{(0 : a_1 : \cdots : a_n) \in \mathbf{P}^n(\mathbf{C})\}$$

we may think of $\mathbf{P}^n(\mathbf{C}) - U_0$ as $\mathbf{P}^{n-1}(\mathbf{C})$. In the homogeneous coordinates $(x_0 : x_1 : \cdots : x_n)$ in $\mathbf{P}^n(\mathbf{C})$, $\mathbf{P}^n(\mathbf{C}) - U_0$ can be expressed by $x_0 = 0$. We call $\mathbf{P}^n(\mathbf{C}) - U_0$ the **hyperplane at infinity** and denote it by H_∞. In general, a hyperplane in $U_0 = \mathbf{C}^n$ can be expressed by

$$\alpha_0 + \alpha_1 z_1 + \cdots + \alpha_n z_n = 0$$

in the coordinates $(z_1, z_2, \ldots, z_n)$ of $\mathbf{C}^n$. Using (2.41) and setting

$$z_1 = \frac{x_1}{x_0}, z_2 = \frac{x_2}{x_0}, \ldots, z_n = \frac{x_n}{x_0}, \tag{2.42}$$

we may define the corresponding hyperplane in $\mathbf{P}^n(\mathbf{C})$ by

$$\alpha_0 x_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n = 0. \tag{2.43}$$

We call $(z_1, z_2, \ldots, z_n)$ in (2.42) the inhomogeneous coordinates.

As in the cases of the projective line and projective plane, we define projective transformations of $\mathbf{P}^n(\mathbf{C})$. Given a nonsingular matrix of degree $n+1$

$$A = (a_{ij})_{0 \leq i,j \leq n},$$

the corresponding projective transformation P_A is defined by

$$P_A : (x_0 : x_1 : \cdots : x_n) \longmapsto \left(\sum_{i=0}^{n} a_{0i} x_i : \sum_{i=0}^{n} a_{1i} x_i : \cdots : \sum_{i=0}^{n} a_{ni} x_i \right). \tag{2.44}$$

As is clear from (2.44), we have for any non-zero complex number α

$$P_A = P_{\alpha A}.$$

We have also

$$P_A \circ P_B = P_{AB},$$

as can be easily verified. It follows that the inverse of P_A is given by ${P_A}^{-1} = P_{A^{-1}}$.

Now for any $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_n) \neq (0, 0, \ldots, 0)$, consider the hyperplane

$$H_\alpha : \alpha_0 x_0 + \alpha_1 x_1 \cdots + \alpha_n x_n = 0.$$

We can find $b_{ij}, 1 \leq i \leq n, 0 \leq j \leq n$, such that the matrix

$$A = \begin{pmatrix} \alpha_0 & \alpha_1 & \ldots & \alpha_n \\ b_{10} & b_{11} & \ldots & b_{1n} \\ \vdots & \vdots & & \vdots \\ b_{n0} & b_{n1} & \ldots & b_{nn} \end{pmatrix}$$

is nonsingular. By the projective transformation P_A we have

$$H_\infty = P_A(H_\alpha)$$

because

$$P_A(a_0 : a_1 : \cdots : a_n) = \left(\sum_{j=0}^{n} \alpha_j a_j : \sum_{j=0}^{n} b_{1j} a_j : \cdots : \sum_{j=0}^{n} b_{nj} a_j \right),$$

which implies

$$P_A(a_0 : a_1 : \cdots : a_n) = \left(0 : \sum_{j=0}^{n} b_{1j} a_j : \cdots : \sum_{j=0}^{n} b_{nj} a_j \right)$$

for $(a_0 : a_1 : \cdots : a_n) \in H_\alpha$. Thus $P_A(H_\alpha) \subset H_\infty$. On the other hand, for any point $(0 : b_1 : \cdots : b_n) \in H_\infty$, define $(a_0 : a_1 : \cdots : a_n)$ by

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = A^{-1} \begin{pmatrix} 0 \\ b_1 \\ \vdots \\ b_n. \end{pmatrix}$$

Then it is easy to check that

$$\alpha_0 a_0 + \alpha_1 a_1 + \cdots + \alpha_n a_n = 0$$

and hence $P_{A^{-1}}(H_\infty) \subset H_\alpha$. It follows that $H_\infty \subset P_A(H_\alpha)$ and hence $H_\infty = P_A(H_\alpha)$. We have seen that an arbitrary hyperplane can be mapped onto the hyperplane at infinity by a projective transformation and is hence isomorphic to $\mathbf{P}^{n-1}(\mathbf{C})$.

The duality principle between the points and the hyperplanes in $\mathbf{P}^n(\mathbf{C})$ can be seen from (2.43). We leave the detail to the readers.

(b) Projective sets and varieties. Let $F(x_0, x_1, \ldots, x_n)$ be a given homogeneous polynomial in $x_0, x_1, \ldots, x_n$. The subset of $\mathbf{P}^n(\mathbf{C})$

$$V(F) = \{(a_0 : a_1 : \cdots : a_n) \in \mathbf{P}^n(\mathbf{C}) | F(a_0, a_1, \ldots, a_n) = 0\}$$

is called a **hypersurface of degree** m. We also denote $X = V(F)$ by

$$X : F(x_0, x_1, \ldots, x_n) = 0.$$

In particular, when $n = 3$, we call X a surface of degree m.

EXAMPLE 2.17. Let us make some observations about the quadratic hypersurface

$$Q : \sum_{i=0}^{n}\sum_{j=0}^{n} c_{ij}x_ix_j = 0.$$

We may assume that $c_{ij} = c_{ji}$, because if $c_{ij} \neq c_{ji}$ for some pair $i \neq j$, then we can write

$$c_{ij}x_ix_j + c_{ji}x_jx_i = \left(\frac{c_{ij}+c_{ji}}{2}\right)x_ix_j + \left(\frac{c_{ij}+c_{ji}}{2}\right)x_jx_i.$$

When the rank of (c_{ij}) is $\ell+1$, the theory of symmetric matrices tells us that there is a suitable complex nonsingular matrix M of degree $n+1$ such that

$${}^tM(c_{ij})M = \begin{pmatrix} I_{\ell+1} & 0 \\ 0 & O_{m-\ell} \end{pmatrix}$$

where $I_{\ell+1}$ is the identity matrix of degree $\ell+1$ and $O_{n-\ell}$ the zero matrix of degree $n-\ell$. Note that the standard form is simple because we are using complex numbers. Now setting $A = M^{-1}$ we consider the projective transformation P_A of $\mathbf{P}^n(\mathbf{C})$. The image $P_A(Q)$ turns out to be the quadratic hypersurface defined by

$$x_0^2 + x_1^2 + \cdots + x_\ell^2 = 0.$$

In fact, if $(a_0 : a_1 : \cdots : a_n) \in Q$, then setting

$$\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} = A\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix},$$

we get

$$\begin{aligned}(b_0, b_1, \ldots, b_n)&\begin{pmatrix} I_{\ell+1} & 0 \\ 0 & O_{m-\ell} \end{pmatrix}\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} \\ &= (a_0, a_1, \cdots, a_n)^tA\begin{pmatrix} I_{\ell+1} & 0 \\ 0 & O_{m-\ell} \end{pmatrix}A\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} \\ &= (a_0, a_1, \cdots, a_n)(c_{ij})\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}.\end{aligned}$$

We may also use other standard forms for quadratic hypersurfaces such as

$$x_0x_1 + x_2x_3 + \cdots + x_{\ell-1}x_\ell = 0, \quad \ell : \text{ odd}$$

$$x_0x_1 + x_2x_3 + \cdots + x_{\ell-2}x_{\ell-1} + {x_\ell}^2 = 0, \quad \ell : \text{ even}.$$

In particular, in $\mathbf{P}^2(\mathbf{C})$, when (c_{ij}) has rank 4, we also use the standard form

$$x_0x_3 - x_1x_2 = 0.$$

In general, given ℓ homogeneous polynomials

$$F_1(x_0, x_1, \ldots, x_n), F_2(x_0, x_1, \ldots, x_n), \ldots, F_\ell(x_0, x_1, \ldots, x_n),$$

of possibly different degrees, we define the subset $V((F_1, F_2, \ldots, F_\ell))$ in $\mathbf{P}^n(\mathbf{C})$ by

$$\begin{aligned}&V((F_1, F_2, \ldots, F_\ell)) \\ &\quad = \{(a_0 : a_1 : \cdots : a_n) \in \mathbf{P}^m(\mathbf{C}) | F_i(a_0, a_1, \cdots : a_n) = 0,\ i = 1, \ldots, \ell\}.\end{aligned}$$

This is called a **projective set**, and $F_1 = 0, F_2 = 0, \ldots, F_\ell = 0$ the **defining equations**.

EXAMPLE 2.18. We find the image $\phi(\mathbf{P}^1(\mathbf{C}))$ of the mapping $\phi : \mathbf{P}^1(\mathbf{C}) \to \mathbf{P}^3(\mathbf{C}))$ given by

$$(a_0 : a_1) \longmapsto (a_0^3 : a_0^2a_1 : a_0a_1^2 : a_1^3).$$

From the definition of ϕ it is easy to see that each point of $\phi(\mathbf{P}^1(\mathbf{C}))$ satisfies

$$\begin{aligned}&F(x_0, x_1, x_2, x_3) = x_0x_3 - x_1x_2 = 0 \\ &G(x_0, x_1, x_2, x_3) = x_1^2 - x_0x_2 = 0 \\ &H(x_0, x_1, x_2, x_3) = x_2^2 - x_1x_3 = 0.\end{aligned} \tag{2.45}$$

Thus $\phi(\mathbf{P}^1(\mathbf{C})) \subset V((F, G, H))$. Conversely, let us take a point (b_0, b_1, b_2, b_3) in $V((F, G, H))$. If $b_0 = 0$, we get $b_1 = 0$ from $G(b_0, b_1, b_2, b_3) = 0$ and $b_2 = 0$ from $H(b_0, b_1, b_2, b_3) = 0$. Thus we must have $(b_0 : b_1 : b_2 : b_3) = (0 : 0 : 0 : 1)$, because we know $(b_0, b_1, b_2, b_3) \neq (0, 0, 0, 0)$. On the other hand, since $\phi((0 : 1)) = (0 : 0 : 0 : 1)$, we get $(0 : 0 : 0 : 1) \in \phi(\mathbf{P}^1(\mathbf{C}))$. Next, we consider the case where $b_0 \neq 0, b_1 = 0$. From $H(b_0, b_1, b_2, b_3) = 0$ we get $b_2 = 0$ and from $F(b_0, b_1, b_2) = 0$ we get $b_3 = 0$. Thus $(b_0 : b_1 : b_2 : b_3) = (1 : 0 : 0 : 0)$ and so $\phi((1 : 0)) = (1 : 0 : 0 : 0)$, which means $(1 : 0 : 0 : 0) \in \phi(\mathbf{P}^1(\mathbf{C}))$.

It remains to check the case where $b_0 \neq 0, b_1 \neq 0$. In this case, $G(b_0, b_1, b_2, b_3) = 0$ implies

$$b_2 = \frac{b_1^2}{b_0}$$

and $H(b_0, b_1, b_2, b_3) = 0$ further implies

$$b_3 = \frac{b_2^2}{b_1} = \frac{b_1^3}{b_0^2}.$$

It follows that

$$\begin{aligned}\phi\left(\left(1 : \frac{b_1}{b_0}\right)\right) &= \left(1 : \frac{b_1}{b_0} : \left(\frac{b_1}{b_0}\right)^2 : \left(\frac{b_1}{b_0}\right)^3\right) \\ &= \left(1 : \frac{b_1}{b_0} : \frac{b_2}{b_0} : \frac{b_3}{b_0}\right) \\ &= (b_0 : b_1 : b_2 : b_3),\end{aligned}$$

which shows that $(b_0 : b_1 : b_2 : b_3) \in \phi(\mathbf{P}^1(\mathbf{C}))$. Hence $V((F, G, H)) \subset \phi(\mathbf{P}^1(\mathbf{C}))$, that is, $V((F, G, H)) = \phi(\mathbf{P}^1(\mathbf{C}))$,

Now we show that the mapping ϕ is an injection from $\mathbf{P}^1(\mathbf{C})$ to $\phi(\mathbf{P}^1(\mathbf{C}))$. Suppose

$$\phi((a_0 : a_1) = \phi((a_0' : a_1')),$$

that is,

$$(a_0^3 : a_0^2a_1 : a_0a_1^2 : a_1^3) = ({a'_0}^3 : {a'_0}^2a'_1 : a'_0{a'_1}^2 : {a'_1}^3).$$

If $a_0 = 0$, then $a'_0 = 0$ and hence $(a_0 : a_1) = (0 : 1) = (a'_0 : a'_1)$. If $a_0 \neq 0$, then $a'_0 \neq 0$, which implies

$$(a_0^3 : a_0^2a_1 : a_0a_1^2 : a_1^3) = \left(1 : \frac{a_1}{a_0} : \left(\frac{a_1}{a_0}\right)^2 : \left(\frac{a_1}{a_0}\right)^3\right)$$

$$({a'_0}^3 : {a'_0}^2a'_1 : a'_0{a'_1}^2 : {a'_1}^3) = \left(1 : \frac{a'_1}{a'_0} : \left(\frac{a'_1}{a'_0}\right)^2 : \left(\frac{a'_1}{a'_0}\right)^3\right).$$

Since these two represent the same point, we obtain

$$\frac{a_1}{a_0} = \frac{a'_1}{a'_0},$$

namely, $(a_1 : a_0) = (a'_1 : a'_0)$. Hence ϕ is an injection. We may thus identify $\mathbf{P}^1(\mathbf{C})$ and $V(F, G, H)$ by ϕ. We call $V(F, G, H)$ a **twisted cubic**.

It is not easy to see from the defining equations (2.45) that $V(F, G, H)$ can be identified with the projective line. In the example above we considered the projective set defined as the set of common zeros of the three equations. Now let us see what figures we obtain if we take the set of common zeros of two of the three equations.

EXAMPLE 2.19. We study the relation between the projective set $V((G, H))$, where

$$G = x_1^2 - x_0x_2 = 0, \quad H = x_2^2 - x_1x_3 = 0,$$

and $V((F, G, H))$ in Example 2.18. Obviously,

$$V(F, G, H) \subset V((G, H)).$$

Take a point $(c_0 : c_1 : c_2 : c_3) \in V((G, H))$. From

$$c_1^2 - c_0c_2 = 0, \quad c_2^2 - c_1c_3 = 0$$

we get

$$(c_1c_2)^2 - c_1c_2c_0c_3 = 0.$$

If $c_1c_2 \neq 0$, then $c_1c_2 - c_0c_3 = 0$, that is, $F(c_0, c_1, c_2, c_3) = 0$.

If $c_1c_2 = 0$, then $c_1 = c_2 = 0$ and $c_0 : c_3$ is arbitrary. The equations

$$x_1 = 0, \quad x_2 = 0$$

define a line ℓ_{12} in $\mathbf{P}^3(\mathbf{C})$. Hence $V((G, H)) = \ell_{12} \cup V((F, G, H))$.

In the same fashion, we can study the projective set $V((F, G))$, where

$$F = x_0x_3 - x_1x_2 = 0, \quad G = x_1^2 - x_0x_2 = 0.$$

For any arbitrary point $(c_0 : c_1 : c_2 : c_3) \in V((F, G))$ we have

$$c_0c_1^2c_3 - c_0c_1c_2^2 = 0.$$

If $c_0c_1 \neq 0$, then we get $c_1c_3 - c_2^2 = 0$ and hence $H(c_0, c_1, c_2, c_3) = 0$. If $c_0c_1 = 0$, then we obtain $c_0 = 0, c_1 = 0$ or $c_0 \neq 0, c_1 = c_2 = c_3 = 0$. In the former case, $c_2 : c_3$ can be chosen in an arbitrary way. The equations

$$x_0 = 0, \quad x_1 = 0$$

determine a line ℓ_{01}, and we have

$$V((F, G)) = \ell_{01} \cup V((F, G, H)).$$

DEFINITION 2.4. A projective set V is said to be **reducible** if there exist projective sets V_1 and V_2 such that

$$V = V_1 \cup V_2, \quad V_1 \not\supset V_2, \quad V_2 \not\supset V_1.$$

If V is not reducible, it is said to be **irreducible**. An irreducible projective set is called a **projective variety**.

According to this definition, the projective set $V((G, H))$ in Example 2.19 is reducible. On the other hand, we can show that $V((F, G, H))$ is irreducible and hence a projective variety. (See Problem 2.8.) If a homogeneous polynomial $J(x_0, x_1, \ldots, x_n)$ is reducible, that is, if it can be expressed as a product of homogeneous polynomials

$$J(x_0, x_1, \ldots, x_n) = K(x_0, x_1, \ldots, x_n)L(x_0, x_1, \ldots, x_n),$$

we have

$$V(J) = V(K) \cup V(L),$$

and the hypersurface $V(J)$ is reducible. We shall show in the next subsection that if J is an irreducible polynomial, then $V(J)$ is irreducible.

(c) Projective sets and homogeneous ideals. One of the important tools in today's algebraic geometry is the theory of commutative rings. In this subsection we shall briefly touch on ideal theory. We deal with only the fundamental facts from ideal theory, without proofs. The reader is referred to the existing books as need arises. A summary of the fundamentals is found in the Appendix, "Commutative rings and fields",

Consider the projective set $V((F_1, F_2, \ldots, F_\ell))$ in $\mathbf{P}^n(\mathbf{C})$ defined by homogeneous polynomials $F_1, F_2, \ldots, F_\ell$ in the variables $x_1, x_2, \ldots, x_n$. Let m_j be the degree of F_j and let $m \geq \max(m_1, m_2, \ldots, m_\ell)$. Take an arbitrary homogeneous polynomial g_j of degree $m - m_j$ and form a homogeneous polynomial of degree m

$$H = \sum_{j=1}^{\ell} G_jF_J. \tag{2.46}$$

Then for any point $(a_0 : a_1 : \cdots : a_n)$ in $V((F_1, F_2, \ldots, F_\ell))$ we have

$$H(a_0, a_1, \ldots, a_n) = \sum_{j=1}^{\ell} G_j(a_0, a_1, \ldots, a_n)F_j(a_0, a_1, \ldots, a_n).$$

It follows that

$$V((F_1, F_2, \ldots, F_\ell, H)) = V((F_1, F_2, \ldots, F_\ell)).$$

Thus the set of all homogeneous polynomials of the form (2.46) and their sums (including terms with different degrees) is useful in dealing with the projective set $V((F_1, F_2, \dots, F_\ell))$.

We prepare some terminology. The set of all polynomials in $x_0, x_1, \dots, x_n$ with complex coefficients is denoted by $\mathbf{C}[x_0, x_1, \dots, x_n]$ and called the **polynomial ring** over $\mathbf{C}$. An element of $\mathbf{C}[x_0, x_1, \dots, x_n]$, namely, a polynomial $P(x_0, x_1, \dots, x_n)$, can be written as a sum of homogeneous polynomials of different degrees:

$$P(x_0, x_1, \dots, x_n) = \sum_{d=0}^{m} P_d(x_0, x_1, \dots, x_n),$$

where $P_d(x_0, x_1, \dots, x_n)$ is homogeneous of degree d; we call P_d the **homogeneous component** of degree d.

Now given homogeneous polynomials $F_j(x_0, x_1, \dots, x_n)$ of degree m_j, $1 \le j \le \ell$, we denote by $\mathfrak{a} = (F_1, F_2, \dots, F_\ell)$ the set of all polynomials of the form

$$\sum_{j=1}^{\ell} K_j(x_0, x_1, \dots, x_n) H_j(x_0, x_1, \dots, x_n),$$

where $K_j \in \mathbf{C}[x_0, x_1, \dots, x_n]$. This is clearly a subset of $\mathbf{C}[x_0, x_1, \dots, x_n]$ and has the following properties; the proof is easy and left to the reader.

LEMMA 2.11.

(I1) *If G and H are in $\mathfrak{a}$, then $G \pm H$ is in $\mathfrak{a}$.*

(I2) *If $P \in \mathbf{C}[x_0, x_1, \dots, x_n]$ and G is in $\mathfrak{a}$, then PG is in $\mathfrak{a}$.*

(I3) *If G is in $\mathfrak{a}$, then each homogeneous component of G is in $\mathfrak{a}$.*

DEFINITION 2.5. A subset $\mathfrak{a}$ of $\mathbf{C}[x_0, x_1, \dots, x_n]$ satisfying the conditions (I1) and (I2) is called an **ideal** of the polynomial ring $\mathbf{C}[x_0, x_1, \dots, x_n]$. If it also satisfies I3, it is called a **homogeneous ideal**. The subset (0) containing only the zero polynomial and the ring $\mathbf{C}[x_0, x_1, \dots, x_n]$ itself are also ideals.

According to this definition, $\mathfrak{a} = (F_1, F_2, \dots, F_\ell)$ above is a homogeneous ideal. It is said to be **generated by** $F_1, F_2, \dots, F_\ell$; $F_1, F_2, \dots, F_\ell$ are called the **generators**. As was stated in the beginning, every homogeneous polynomial belonging to this ideal satisfies

$$H(a_0, a_1, \dots, a_n) = 0$$

for an arbitrary point $(a_0 : a_1 : \cdots : a_n)$ in $V((F_1, F_2, \dots, F_\ell))$. Thus we can write

$$\begin{aligned} &V((F_1, F_2, \dots, F_\ell)) \\ &\quad = \{(a_0 : a_1 : \cdots : a_n) \in \mathbf{P}^n(\mathbf{C}) \mid G(a_0, a_1, \dots, a_n) = 0, \\ &\qquad \text{for all homogeneous polynomials } G \in \mathfrak{a}\}. \end{aligned}$$

Conversely, suppose an arbitrary homogeneous ideal $\mathfrak{b}$ of $\mathbf{C}[x_0, x_1, \dots, x_n]$ is given. Let

$$\begin{aligned} V(\mathfrak{b}) = \{&(a_0 : a_1 : \cdots : a_n) \in \mathbf{P}^n(\mathbf{C}) \mid H(a_0, a_1, \dots, a_n) = 0, \\ &\text{for all homogeneous polynomials } H \in \mathfrak{b}\}. \end{aligned} \tag{2.47}$$

This is the set of common zeros for infinitely many homogeneous polynomials in $\mathfrak{b}$ and, as such, it is not a projective set in the proper sense. (Recall that a projective set is defined as the zeros of a finite number of homogeneous polynomials.) However, it turns out that $V(\mathfrak{b})$ is a projective set, thanks to Hilbert's basis theorem.

THEOREM 2.3 (HILBERT'S BASIS THEOREM). *Every (homogeneous) ideal $\mathfrak{b}$ of the polynomial ring $\mathbf{C}[x_0, x_1, \dots, x_n]$ is generated by a finite number of (homogeneous) polynomials $G_1, G_2, \dots, G_k$, that is,*

$$\mathfrak{b} = (G_1, G_2, \dots, G_k).$$

Using this theorem, we find that $V(\mathfrak{b}) = V((G_1, G_2, \dots, G_k))$ is a projective set. In this way, taking the zero set for a finite number of homogeneous polynomials and taking the zero set for all homogeneous polynomials contained in a homogeneous ideal are the same thing. It might appear more complicated to make use of ideals; it is, however, an effective method in studying the properties of projective sets.

Let us consider the projective set $V(\mathfrak{a})$ determined by a homogeneous ideal $\mathfrak{a}$ in the polynomial ring $\mathbf{C}[x_0, x_1, \dots, x_n]$. We now define

$$\begin{aligned} I(V(\mathfrak{a})) = \{&G \in \mathbf{C}[x_0, x_1, \dots, x_n] \mid \\ &\text{each homogeneous component of } G \text{ vanishes on } V(\mathfrak{a})\}. \end{aligned} \tag{2.48}$$

It is obvious from the definition that $I(V(\mathfrak{a}))$ satisfies the conditions (I1)-(I3) for homogeneous ideals. It is also clear that $\mathfrak{a} \subset I(V(\mathfrak{a}))$. Indeed, we have a much stronger result.

THEOREM 2.4 (HILBERT'S ZERO POINT THEOREM). *If a homogeneous polynomial $H(x_0, x_1, \dots, x_n)$ vanishes at each point of $V(\mathfrak{a})$ (that is, $H \in I(V(\mathfrak{a}))$), then there is a positive integer k such that $H^k \in \mathfrak{a}$.*

Here k may not be 1. To give a simple example, let $\mathfrak{a} = (F^m)$ be the ideal generated by the homogeneous polynomial F^m. Then $V(\mathfrak{a}) = V(F)$, because if $F(a_0, a_1, \dots, a_n)^m = 0$, then $F(a_0, a_1, \dots, a_n) = 0$. But $F \notin (F^m)$. As this example shows, it is possible that two different ideals define the same projective set. Hilbert's zero point theorem says that the difference is just that one is a power of the other. To clarify this matter, we define for a homogeneous ideal $\mathfrak{a}$ the **radical** $\sqrt{\mathfrak{a}}$ by

$$\sqrt{\mathfrak{a}} = \{G \in \mathbf{C}[x_0, x_1, \dots, x_n] \mid G^k \in \mathfrak{a} \text{ for some integer } k\}. \tag{2.49}$$

We may prove that $\sqrt{\mathfrak{a}}$ is a homogeneous ideal as follows. If $G, H \in \sqrt{\mathfrak{a}}$, then take positive integers m_1, m_2 such that

$$G^{m_1} \in \mathfrak{a}, \quad H^{m_2} \in \mathfrak{a}.$$

Then

$$(G \pm H)^{m_1+n_2} = \sum_{k=0}^{m_1+m_2} (\pm 1)^k \binom{m_1+m_2}{k} G^{m_1+m_2-k} H^k.$$

Now $k \le m_2$ implies $G^{m_1+m_2-k} \in \mathfrak{a}$, and $k \ge m_2$ implies $H^k \in \mathfrak{a}$. We have hence $(G \pm H)^{m_1+m_2} \in \sqrt{\mathfrak{a}}$, proving (II). If $G^{m_1} \in \mathfrak{a}$, then for any $P \in \mathbf{C}[x_0, x_1, \dots, x_n]$ we have $(PG)^{m_1} = P^{m_1} G^{m_1} \in \mathfrak{a}$ from (I2). Hence $PG \in \sqrt{\mathfrak{a}}$, that is, (I2) holds. Finally, for any $G \in \sqrt{\mathfrak{a}}$, we write

$$G = \sum_{j=1}^{k} G_{d_j}, \quad d_1 < d_2 < \cdots < d_k,$$

namely, the sum of homogeneous polynomials of distinct degrees. If $G^m \in \mathfrak{a}$, the homogeneous component of the highest degree of G^m is equal to $G_{d_k}^m$, which implies that $G_{d_k}^m \in \mathfrak{a}$, because $\mathfrak{a}$ is a homogeneous ideal. Thus $G_{d_k} \in \mathfrak{a}$. Since (I1) holds for $\sqrt{\mathfrak{a}}$, it follows that $G - G_{d_k} \in \sqrt{\mathfrak{a}}$. By the same argument, we obtain $G_{d_k-1} \in \sqrt{\mathfrak{a}}$, and consequently $G_{d_j} \in \sqrt{\mathfrak{a}}$ for $j = k-2, k-3, \ldots, 1$. Hence (I3) holds.

From the discussion above, we see that Hilbert's zero point theorem is equivalent to the following theorem.

THEOREM 2.5. *For any homogeneous ideal $\mathfrak{a}$ of the polynomial ring*

$$\mathbf{C}[x_0, x_1, \ldots, x_n],$$

we have

$$I(V(\mathfrak{a})) = \sqrt{\mathfrak{a}}.$$

An ideal such that $\mathfrak{a} = \sqrt{\mathfrak{a}}$ is called a **reduced** ideal. Thus to consider a projective set as a point set, it is sufficient to deal with a reduced homogeneous ideal. If $\mathfrak{a}$ is such an ideal, then

$$I(V(\mathfrak{a})) = \mathfrak{a}.$$

Now if $\mathfrak{a}$ and $\mathfrak{b}$ are homogeneous ideals of $\mathbf{C}[x_0, x_1, \ldots, x_n]$, then so is $\mathfrak{a} \cap \mathfrak{b}$, as can be easily seen. If $\{\mathfrak{a}_\lambda\}_{\lambda \in \Lambda}$, then $\sum_{\lambda \in \Lambda} \mathfrak{a}_\lambda$ is the ideal generated by $\mathfrak{a}_\lambda$, $\lambda \in \Lambda$, namely, the ideal consisting of all finite sums of the form

$$F_1 + \cdots + F_k, \quad F_i \in \mathfrak{a}_{\lambda_i}.$$

We leave the proof of the following lemma to the reader.

LEMMA 2.12. *Let $\mathfrak{a}$, $\mathfrak{b}$, and $\mathfrak{a}_\lambda$, $\lambda \in \Lambda$, be homogeneous ideals. Then the following relations hold.*

(i) $V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a} \cap \mathfrak{b})$.
(ii) $\bigcap_{\lambda \in \Lambda} V(\mathfrak{a}_\lambda) = V(\sum_{\lambda \in \Lambda} \mathfrak{a}_\lambda)$.
(iii) $V((0)) = \mathbf{P}^n(\mathbf{C})$, $V(\mathbf{C}[x_0, x_1, \ldots, x_n]) = \emptyset$.
(iv) *A necessary and sufficient condition for $V(\mathfrak{a}) \subset V(\mathfrak{b})$ is $\sqrt{\mathfrak{a}} \supset \sqrt{\mathfrak{b}}$.*

By Lemma 2.12 (i), (ii), and (iii) we see that we can put a topology in $\mathbf{P}^n(\mathbf{C})$ by taking projective sets as closed subsets. This is called the **Zariski topology**. It does not satisfy Hausdorff's separation axiom, but plays an important role in the algebraic treatment of algebraic geometry.

Finally, we express irreducibility of a projective set by using the terminology of ideals. If a projective set $V = V(\mathfrak{a})$ is reducible, then we can write

$$V = V_1 \cup V_2, \quad V_1 \not\supset V_2, \quad V_2 \not\supset V_1,$$
$$V_1 = V(\mathfrak{b}),\ V_2 = V(\mathfrak{c}).$$

Since $V_1 \not\supset V_2$, there is a homogeneous polynomial G which is 0 at every point of V_1 and is not 0 at some point of V_2, that is,

$$F \in I(V_1) = \sqrt{\mathfrak{b}}, \quad G \notin I(V_2) = \sqrt{\mathfrak{c}}.$$

Similarly, from $V_2 \not\supset V_1$ we see that there exists a homogeneous polynomial H such that

$$H \notin \sqrt{\mathfrak{b}}, \quad H \in \sqrt{\mathfrak{c}}.$$

Since G and H are not identically 0 on V. we have

$$G \notin I(V) = \sqrt{\mathfrak{a}}, \quad H \notin I(V) = \sqrt{\mathfrak{a}}.$$

On the other hand, GH vanishes on V_1 and V_2, and we have

$$GH \in I(V).$$

Conversely, if there exist homogeneous polynomials G and H such that

$$G \notin I(V), \quad H \notin I(V), \quad GH \in I(V),$$

then for $I(V) = (F_1, F_2, \ldots, F_\ell)$ we set

$$V_1 = V((G, F_1, F_2, \ldots, F_\ell)), \quad V_2 = V((H, F_1, F_2, \ldots, F_\ell)).$$

From the assumption, we have

$$V \neq V_1, \quad V \neq V_2.$$

From $GH \in I(V)$, we get

$$V = V_1 \cup V_2,$$

which shows that V is reducible.

By taking the contrapositive of the assertion above, we get the following important result.

LEMMA 2.13. *For the projective set $V(\mathfrak{a})$ to be irreducible, it is necessary and sufficient that the radical $\sqrt{\mathfrak{a}}$ of $\mathfrak{a}$ satisfies the following condition: If $F, G \in \mathbf{C}[x_0, x_1, \ldots, x_n]$ satisfy*

$$F \notin \sqrt{\mathfrak{a}}, \quad G \notin \sqrt{\mathfrak{a}},$$

then

$$FG \notin \sqrt{\mathfrak{a}}.$$

The last condition may be paraphrased as follows: if $FG \in \sqrt{\mathfrak{a}}$, then $F \in \sqrt{\mathfrak{a}}$ or $G \in \sqrt{\mathfrak{a}}$. In general, an ideal $\mathfrak{b}$ is called a **prime ideal** if $FG \in \mathfrak{b}$ entails $F \in \mathfrak{b}$ or $G \in \mathfrak{b}$. So $\sqrt{\mathfrak{a}}$ is a prime ideal, and we may say that a projective variety is a projective set defined by a prime ideal.

A necessary and sufficient condition for the ideal (F) generated by a homogeneous polynomial F to be a prime ideal is that F is irreducible. If $G \notin (F)$, $H \notin (F)$ and $GH \in (F)$, then $GH = JF$ for some polynomial J. But since F is irreducible, it divides G or H and hence $G \in (F)$ or $H \in (F)$, contrary to the assumption.

EXAMPLE 2.20. The homogeneous ideal

$$\mathfrak{a} = (x_1x_3 - x_2^2, x_0x_2x_3 - x_1^3, x_0x_3^2 - x_1^2x_2)$$

in $\mathbf{C}[x_0, x_1, \ldots, x_n]$ is a prime ideal. $V(\mathfrak{a})$ coincides with the image of the mapping

$$\mathbf{P}^1(\mathbf{C}) \to \mathbf{P}^3(\mathbf{C})$$

defined by

$$(a_0 : a_1) \mapsto (a_0^5 : a_0^2a_1^3 : a_0a_1^4 : a_1^5).$$

(d) Dimension of projective varieties and function fields. Intuitively speaking, the dimension of a projective variety is the number of parameters that can vary freely. For example, for a hypersurface $X = V(F)$ in $\mathbf{P}^n(\mathbf{C})$

$$X : F(x_0, x_1, \ldots, x_{n-1}, x_n) = 0$$

we may fix the ratio of n of the $n+1$ homogeneous coordinates, say,

$$x_0 : x_1 : \cdots : x_{n-1} = a_0 : a_1 : \cdots : a_{n-1}$$

and determine a finite number of values of a_n by solving the equation

$$F(a_0, a_1, \ldots, a_{n-1}, x_n) = 0.$$

Thus X may be regarded as $(n-1)$-dimensional because $n-1$ parameters (corresponding to the ratios of n numbers) can vary freely. We might like to think that as the number of equations increases by one, the number of free parameters decreases by one, but that is not necessarily the case as we have seen in Example 2.18 in §2.4 (b), where $X = V((F, G, H))$, being the image of $\mathbf{P}^1(\mathbf{C})$, admits only one free parameter. It is, however, necessary to use three equations to define X, as is illustrated in Example 2.19.

Now how can we determine the dimension for a projective variety in general? We may think geometrically as follows. Take a projective variety $X = V((F_1, F_2, \ldots, F_\ell))$ in $\mathbf{P}^n(\mathbf{C})$. Pick a point $P = (a_0 : a_1 : \cdots : a_n)$ outside X and fix it. We also take a hyperplane H in $\mathbf{P}^n(\mathbf{C})$ given by

$$\sum_{i=0}^{n} \alpha_i x_i = 0;$$

we choose H in as general position as possible. Pick a point $Q = (b_0 : b_1 : \cdots : b_n)$ in X and take the line $\overline{PQ}$, which we may parametrize as

$$(a_0 s + b_0 t : a_1 s + b_1 t : \cdots : a_n s + b_n t), \quad (s : t) \in \mathbf{P}^1(\mathbf{C}),$$

where $(s : t)$ are homogeneous parameters. The line $\overline{PQ}$ and the hyperplane H intersect at one point, since we can find a unique solution $(s : t)$ by solving

$$\sum_{i=0}^{n} \alpha_i (a_i s + b_i t) = 0.$$

Denote the intersection by $R(Q)$. Then we can define a mapping

$$\phi_P : X \to H$$

by

$$Q \mapsto R(Q).$$

Since we can identify H with the $(n-1)$-dimensional projective space $\mathbf{P}^{n_1}(\mathbf{C})$, we may regard $\phi_P(X)$ as a subset of $\mathbf{P}^{n_1}(\mathbf{C})$. As a matter of fact, we can see that $\phi_P(X)$ is also a projective variety. We call ϕ_P the **projection** with center P.

If X is a hyperplane, we can see that $\phi_P(X) = H$ (see Problem 2.11). If $\phi_P(X) \neq H$, then within $H = \mathbf{P}^{n-1}(\mathbf{C})$ we perform a similar process on $X_1 = \phi_P(X)$ and get the projection with center P_1

$$\phi_{P_1} : X_1 \to \mathbf{P}^{n-2}(\mathbf{C}).$$

By repeating this process a number of times, we finally get a surjective mapping

$$\phi_{P_m} : X_m \to \mathbf{P}^{n-m-1}(\mathbf{C}).$$

In this case, we see that the dimension of X is equal to $n-m-1$ as follows. If we choose an arbitrary point in $\mathbf{P}^{n-m-1}(\mathbf{C})$, there are a finite number of points that are mapped to that point by ϕ_{P_m}. To these points there correspond a finite number of points in X_{m-1}, and so on. By repetition, eventually a finite number of points are determined in X. (This is because $\phi_P, \phi_{P_1}, \ldots, \phi_{P_{m-1}}$ are one-to-one almost everywhere, that is, outside a projective set.) In this way, we find that there are $n-m-1$ free parameters and the dimension of X is $n-m-1$. We denote it by $\dim X$.

EXAMPLE 2.21. The twisted cubic $X = V((F, G, H))$ in Example 1.18, where

$$F = x_0 x_3 - x_1 x_2 = 0$$
$$G = x_1^2 - x_0 x_2 = 0$$
$$H = x_2^2 - x_1 x_3 = 0,$$

has dimension 1, because X can be identified with $\mathbf{P}^1(\mathbf{C})$ by the mapping ϕ in Example 2.18. We indicate another argument by using the idea of projection. The point $P = (0 : 1 : 0 : 0)$ is outside X. Take the hyperplane $H : x_1 = 0$. For $Q = (b_0 : b_1 : b_2 : b_3) \in X$, the line $\overline{PQ}$ is parametrized as

$$(b_0 t : s + b_1 t : b_2 t : b_3 t), \quad (s : t) \in \mathbf{P}^1(\mathbf{C}).$$

The intersection $R(Q)$ of the line with H can be found from $s + b_1 t = 0$, that is, $(s, t) = (-b_1 : 1)$, and hence

$$R(Q) = (b_0 : b_2 : b_3).$$

We have $\phi_P(b_0 : b_1 : b_2 : b_3) = (b_0 : b_2 : b_3)$. We now show that $\phi_P(X)$ is a projective variety, in fact, a cubic plane curve. Since the equation H involves only x_1, x_2, x_3, we can eliminate x_1 from $F = 0, G = 0$ to find that each point of $\phi_P(X)$ satisfies

$$x_0(x_2^3 - x_0 x_3^2) = 0,$$

and eliminate x_1 from $G = 0, H = 0$ to find that each point of $\phi(X)$ also satisfies

$$x_2(x_2^3 - x_0 x_3^2) = 0.$$

From these two equations we see that $\phi_P(X)$ is contained in the cubic plane curve $C = V(x_2^3 - x_0 x_3^2)$. Conversely, for any $(c_0 : c_2 : c_3) \in C$, if $c_2 \neq 0$, set

$$c_1 = \frac{c_0 c_3}{c_2}.$$

Then, since $c_2^3 = c_0 c_3^2$ we get

$$c_1^3 = \frac{(c_0 c_3)^3}{c_2^3} = \frac{(c_0 c_3)^3}{c_0 c_3^2} = c_0^2 c_3,$$

and

$$(c_0 : c_1 : c_2 : c_3) \in X, \quad \phi_P((c_0 : c_1 : c_2 : c_3)) = (c_0 : c_2 : c_3).$$

If $c_2 = 0$, we have $(c_0 : c_2 : c_3) = (0 : 0 : 1)$ or $(1 : 0 : 0)$. Since

$$(0:0:0:1) \in X, \quad \phi_P((0:0:0:1)) = (0:0:1),$$
$$(1:0:0:0) \in X, \quad \phi_P((1:0:0:0)) = (1:0:0),$$

all these points belong to $\phi_P(X)$. Hence $\phi_P(X) = C$. If we take the projection from $\mathbf{P}^2(\mathbf{C})$ to $\mathbf{P}^1(\mathbf{C})$ with center P_1 outside C, we can easily see that we generally get a 3-to-1 mapping onto C. Thus we find that the dimension of X is 1.

The dimension of a projective variety can be seen from its function field $\mathbf{C}(X)$. Rational functions on $\mathbf{P}^n(\mathbf{C})$ are defined, just as in the case where $n = 1, 2$, to be the quotients of homogeneous polynomials of the same degree

$$\frac{F(x_0, x_1, \ldots, x_n)}{G(x_0, x_1, \ldots, x_n)}.$$

If we introduce the inhomogeneous coordinates

$$z_1 = \frac{x_1}{x_0}, z_2 = \frac{x_2}{x_0}, \ldots, z_n = \frac{x_n}{x_0}$$

and assume that F and G have degree m, then we have

$$\begin{aligned}\frac{F(x_0, x_1, \ldots, x_n)}{G(x_0, x_1, \ldots, x_n)} &= \frac{\frac{1}{x_0^m} F(x_0, x_1, \ldots, x_n)}{\frac{1}{x_0^m} G(x_0, x_1, \ldots, x_n)} \\ &= \frac{f(z_1, z_2, \ldots, z_n)}{g(z_1, z_2, \ldots, z_n)},\end{aligned}$$

and this is a rational function in $z_1, \ldots, z_n$. Hence the totality $\mathbf{C}(\mathbf{P}^n(\mathbf{C}))$ of rational functions on $\mathbf{P}^n(\mathbf{C})$ is the totality $\mathbf{C}(z_1, \ldots, z_n)$ of rational functions in the variables $z_1, \ldots, z_n$.

A rational function on the projective variety X is the restriction of a rational function on $\mathbf{P}^n(\mathbf{C})$ to X. For this to make sense, it is necessary that the rational function on $\mathbf{P}^n(\mathbf{C})$ has no pole in X. The set of rational functions on X is denoted by $\mathbf{C}(X)$ and called the **function field** of X.

EXAMPLE 2.22. We find the function field $\mathbf{C}(X)$ of the twisted cubic X in Example 2.18. It is defined by

$$\begin{aligned}F &= x_0x_3 - x_1x_2 = 0 \\ G &= x_1^2 - x_0x_2 = 0 \\ H &= x_2^2 - x_1x_3 = 0.\end{aligned}$$

Denoting the restriction of $z_i = x_i/x_0$ to X also by z_i, we have from the defining equations of X the relations

$$z_3 = z_1z_2, \quad z_1^2 = z_2, \quad z_2^2 = z_1z_3.$$

As functions on X we have

$$z_3 = z_1^3, \quad z_2 = z_1z_2,$$

and the restriction of a rational function in z_1, z_2, z_3 can be expressed as rational functions of z_1, that is, we have $\mathbf{C}(X) = \mathbf{C}(z_1)$.

In general, the function field $\mathbf{C}(X)$ of an n-dimensional projective variety X in $\mathbf{P}^N(\mathbf{C})$ is obtained by restricting to X the quotients

$$\frac{F(x_0, x_1, \ldots, x_N)}{G(x_0, x_1, \ldots, x_N)}$$

of homogeneous polynomials F and G of the same degree. Now if we choose homogeneous polynomials $\tilde{F}(x_0, x_1, \ldots, x_N)$ and $\tilde{G}(x_0, x_1, \ldots, x_N)$ of the same degree as F such that both $F(x_0, x_1, \ldots, x_n) - \tilde{F}(x_0, x_1, \ldots, x_N)$ and $G(x_0, x_1, \ldots, x_N) - \tilde{G}(x_0, x_1, \ldots, x_N)$ vanish on X, then we have

$$\left.\frac{F(x_0, x_1, \ldots, x_N)}{G(x_0, x_1, \ldots, x_N)}\right|_X = \left.\frac{\tilde{F}(x_0, x_1, \ldots, x_N)}{\tilde{G}(x_0, x_1, \ldots, x_N)}\right|_X.$$

In the language of ideals in the preceding subsection (c), we may say that F/G and $\tilde{F}/\tilde{G}$ represent the same element in $\mathbf{C}(X)$, provided $F - \tilde{F} \in I(X)$ and $G - \tilde{G} \in I(X)$.

We state the following important result without a proof.

THEOREM 2.6. *The function field $\mathbf{C}(X)$ of an n-dimensional projective variety X is a finite extension of the field of rational functions $\mathbf{C}(z_1, z_2, \ldots, z_n)$.*

In the language of field theory, $\mathbf{C}(X)$ is what is called a **finite extension** (of degree m) of $\mathbf{C}(z_1, z_2, \ldots, z_n)$, that is, $\mathbf{C}(X)$ is a vector space of finite dimension (m-dimensional). In this case, it follows that each element of $\mathbf{C}(x)$ is algebraic over $\mathbf{C}(z_1, z_2, \ldots, z_n)$, that is, it satisfies a polynomial equation with coefficients in $\mathbf{C}[z_1, z_2, \ldots, z_n]$. From the the theory of fields, we also know that $\mathbf{C}(X)$, being of characteristic 0, is in fact a simple extension of $\mathbf{C}(z_1, z_2, \ldots, z_n)$, that is, there is an element y in $\mathbf{C}(X)$ such that every element in $\mathbf{C}(X)$ can be expressed in the form

$$A_0(z)y^{m-1} + A_1(z)y^{m-2} + \cdots + A_{m-1}(z),$$

where $A_j(z) = A_j(z_1, z_2, \ldots, z_n)$ are in $\mathbf{C}(z_1, z_2, \ldots, z_n)$. In the current case, we say that the **transcendence degree** of $\mathbf{C}(X)$ is n. From Theorem 2.6 we see that the transcendence degree of $\mathbf{C}(X)$ is equal to the dimension of the projective variety X. The following example will better clarify the situation than more explanation would.

EXAMPLE 2.23. We study the function field of the quadratic surface

$$X : x_0^2 + x_1^2 + x_2^2 + x_3^2 = 0$$

in $\mathbf{P}^3(\mathbf{C})$. Denoting by z_1 the restriction of x_1/x_0 to X, we can express each element of $\mathbf{C}(X)$ as a rational function of z_1, z_2, z_3, except that z_1, z_2, z_3 are subject to

$$1 + z_1^2 + z_2^2 + z_3^2 = 0. \tag{2.50}$$

We show that an element of $\mathbf{C}(X)$ can be written in the form

$$A(z_1, z_2)z_3 + B(z_1, z_2).$$

By using (2.50), we may write a rational function of z_1, z_2, z_3 in the form

$$g(z_1,z_2,z_3) = \frac{G(z_1,z_2,z_3)}{H(z_1,z_2,z_3)} = \frac{G_1(z_1,z_2)z_3 + G_2(z_1,z_2)}{H_1(z_1,z_2)z_3 + H_2(z_1,z_2)},$$

where G_j, H_j are polynomials in z_1, z_2. Multiplying the denominator and the numerator by $H_1(z_1,z_2)z_3 - H_2(z_1,z_2)$ and using (2.50), we obtain

$$\begin{aligned} g(z_1,g_2,z_3) &= \frac{(G_1z_3+G_2)(H_1z_3-H_2)}{(H_1z_3+H_2)(H_1z_3-H_2)} \\ &= \frac{G_1H_1z_3^2 + (G_2H_1 - G_1H_2)z_3 - G_2H_2}{H_1^2z_3^2 - H_2^2} \\ &= A(z_1,z_2)z_3 + B(z_1,z_2), \end{aligned}$$

where

$$A(z_1,z_2) = \frac{G_2H_1 - G_1H_2}{-H_1^2(z_1^2+z_2^2+1) - H_2^2}$$
$$B(z_1,z_2) = \frac{G_1H_1(z_1^2+z_2^2+1) + G_2H_2}{H_1^2(z_1^2+z_2^2+1) + H_2^2}.$$

Thus we can set $y = z_3$.

The function field $\mathbf{C}(X)$ has a different representation. Set

$$u = \frac{z_2 + z_3 i}{z_1 + i}, \quad v = \frac{z_2 - z_3 i}{z_1 + i} \quad (i = \sqrt{-1}).$$

They are in $\mathbf{C}(X)$. From (2.50) we get

$$uv + \frac{z_1 - i}{z_1 + i} = 0, \quad \text{that is,} \quad z_1 = -i + \frac{2i}{1+uv}.$$

Therefore by definition of u and v we get

$$z_2 = \frac{(u+v)i}{1+uv}, \quad z_3 = \frac{(u-v)i}{1+uv},$$

which implies $\mathbf{C}(X) = \mathbf{C}(u,v)$. Although $\mathbf{C}(X)$ has various expressions, they are based on the rational functions of two variables. The transcendence degree of $\mathbf{C}(X)$ is 2. We already stated that the dimension of X is 2.

When two projective varieties X_1 and X_2 have the same (to be exact, isomorphic) function fields $\mathbf{C}(X_1)$ and $\mathbf{C}(X_2)$, we say that X_1 and X_2 are **birationally equivalent**. In Example 2.23, the function field $\mathbf{C}(X)$ of the quadratic surface X is isomorphic with $\mathbf{C}(u,v)$, that is, isomorphic with the function field $\mathbf{C}(\mathbf{P}^2(\mathbf{C}))$ of the complex projective plane. So the quadratic surface and the projective plane are birationally equivalent. As we show in subsection (f), the quadratic surface X is isomorphic to $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^1(\mathbf{C})$ and is different from $\mathbf{P}^2(\mathbf{C})$. Thus for projective varieties of dimension ≥ 2, the notion of birational equivalence is more general than isomorphism.

(e) Singular points, nonsingular points, and tangent hyperplanes. We have already discussed singular points of plane curves. Here we briefly discuss singular points of a projective variety. Consider a point $(a_0 : a_1 : \cdots : a_n)$ of the projective variety $X = V((F_1, F_2, \ldots, F_m))$. We take the defining equations $F_1, F_2, \ldots, F_m$ so that

$$I(X) = (F_1, F_2, \ldots, F_m).$$

Consider first the case where $a_0 \neq 0$. Set

$$b_k = \frac{a_k}{a_0}, \quad z_k = \frac{x_k}{x_0}, \quad k = 1,2,\ldots,n,$$

$$f(z_1,z_2,\ldots,z_n) = \frac{1}{x_0^{d_j}} F(x_0,x_1,\ldots,x_n),$$

$$d_j = \deg F_j, \quad j = 1,2,\ldots,m.$$

DEFINITION 2.6. The point $(a_0 : a_1 : \cdots : a_n) = (1 : b_1 : \cdots : b_n)$ is called a **nonsingular point** of the projective variety $V((F_1, F_2, \ldots, F_m))$ if the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial z_1}(b_1,\ldots,b_n) & \frac{\partial f_1}{\partial z_2}(b_1,\ldots,b_n) & \cdots & \frac{\partial f_1}{\partial z_n}(b_1,\ldots,b_n) \\ \frac{\partial f_2}{\partial z_1}(b_1,\ldots,b_n) & \frac{\partial f_2}{\partial z_2}(b_1,\ldots,b_n) & \cdots & \frac{\partial f_2}{\partial z_n}(b_1,\ldots,b_n) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial z_1}(b_1,\ldots,b_n) & \frac{\partial f_m}{\partial z_2}(b_1,\ldots,b_n) & \cdots & \frac{\partial f_m}{\partial z_n}(b_1,\ldots,b_n) \end{pmatrix} \tag{2.51}$$

has rank $n-d$. We say it is a singular point if the rank is $\leq n-d-1$.

We know that the rank of the matrix (2.51) is $\leq n-d$. In the above discussion we assumed that $a_0 \neq 0$. In general, if $a_i \neq 0$, then we set

$$c_1 = \frac{a_0}{a_i}, c_2 = \frac{a_1}{a_i}, \ldots, c_i = \frac{a_{i-1}}{a_i}, c_{i+1} = \frac{a_{i+1}}{a_i}, \ldots, c_n = \frac{a_n}{a_i},$$

$$w_1 = \frac{x_0}{x_i}, w_2 = \frac{x_1}{x_i}, \ldots, w_i = \frac{x_{i-1}}{x_i}, w_{i+1} = \frac{x_{i+1}}{x_i}, \ldots, w_n = \frac{x_n}{x_i},$$

$$g_j(w_1,\ldots,w_n) = \frac{1}{x_i^{d_j}} F_j(x_1,\ldots,x_n), \quad j = 1,2,\ldots,m,$$

and consider the matrix

$$\begin{pmatrix} \frac{\partial g_1}{\partial w_1}(c_1,\ldots,c_n) & \frac{\partial g_1}{\partial w_2}(c_1,\ldots,c_n) & \cdots & \frac{\partial g_1}{\partial w_n}(c_1,\ldots,c_n) \\ \frac{\partial g_2}{\partial w_1}(b_1,\ldots,b_n) & \frac{\partial g_2}{\partial w_2}(c_1,\ldots,c_n) & \cdots & \frac{\partial g_2}{\partial w_n}(c_1,\ldots,c_n) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial w_1}(c_1,\ldots,c_n) & \frac{\partial g_m}{\partial w_2}(c_1,\ldots,c_n) & \cdots & \frac{\partial g_m}{\partial w_n}(c_1,\ldots,c_n) \end{pmatrix}. \tag{2.52}$$

We say that the point $(a_0 : a_1 : \cdots : a_n) = (c_1 : c_2 : \cdots : 1 : c_{i+1} : \cdots, : c_n)$ is a nonsingular point if the rank of (2.52) is $n-d$ and a singular point if the rank is $\leq n-d-1$. We leave it as an exercise for the reader to verify that the definition is independent of the choice of i such that $a_i \neq 0$.

EXAMPLE 2.24. For a plane curve of degree ℓ,

$$C : F(x_0, x_1, x_2) = 0,$$

the point $(a_0 : a_1 : a_2) = (1 : b_1 : b_2) \in C$ is a singular point if

$$\frac{\partial f}{\partial z_1}(b_1, b_2) = 0, \quad \frac{\partial f}{\partial z_2}(b_1, b_2) = 0, \tag{2.53}$$

where

$$f(z_1, z_2) = \frac{1}{x_0^\ell} F(x_0, x_1, x_2).$$

On the other hand, from $F(x_0, x_1, x_2) = x_0^\ell f\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right)$ we get

$$\begin{aligned}
\frac{\partial F}{\partial x_0} &= \ell x_0^{\ell-1} f\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right) - x_0^{\ell-2} x_1 \frac{\partial f}{\partial z_1}\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right) \\
&\quad - x_0^{\ell-2} x_1 \frac{\partial f}{\partial z_1}\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right) \qquad (2.54) \\
\frac{\partial F}{\partial x_1} &= x_0^{\ell-1} \frac{\partial f}{\partial z_1}\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right) \\
\frac{\partial F}{\partial x_2} &= x_0^{\ell-1} \frac{\partial f}{\partial z_2}\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right)
\end{aligned}$$

From (2.53) we obtain

$$\frac{\partial F}{\partial x_0}(a_0, a_1, a_2) = 0, \quad \frac{\partial F}{\partial x_1}(a_0, a_1, a_2) = 0, \quad \frac{\partial F}{\partial x_2}(a_0, a_1, a_2) = 0. \tag{2.55}$$

Conversely, if (2.55) holds, then from (2.54) we get

$$\frac{\partial f}{\partial z_1}(b_1, b_2) = 0, \quad \frac{\partial f}{\partial z_2}(b_1, b_2) = 0.$$

It follows that the two definitions, Definition 2.3 and Definition 2.6, are equivalent.

This example can be generalized to a lemma in the following form, and the proof is left to the reader as an exercise (see Exercise 2.12).

LEMMA 2.14. *For a point $(a_0 : a_1 : \cdots : a_n)$ to be a nonsingular point of the d-dimensional projective variety $X = V((F_1, F_2, \ldots, F_m))$, it is necessary and sufficient that the matrix*

$$\begin{pmatrix}
\frac{\partial F_1}{\partial x_0}(a_0, \ldots, a_n) & \frac{\partial F_1}{\partial x_1}(a_0, \ldots, a_n) & \cdots & \frac{\partial F_1}{\partial x_n}(a_0, \ldots, a_n) \\
\frac{\partial F_2}{\partial x_0}(a_0, \ldots, a_n) & \frac{\partial F_2}{\partial x_1}(a_0, \ldots, a_n) & \cdots & \frac{\partial F_2}{\partial x_n}(a_0, \ldots, a_n) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial F_m}{\partial x_0}(a_0, \ldots, a_n) & \frac{\partial F_m}{\partial x_1}(a_0, \ldots, a_n) & \cdots & \frac{\partial F_m}{\partial x_n}(a_0, \ldots, a_n)
\end{pmatrix}$$

has rank $n - d$.

Now suppose a point $(a_0 : a_1 : \cdots : a_N) = (c_1 : \cdots : c_r : 1 : c_{r+1} : \cdots : c_n)$ is a nonsingular point of the d-dimensional projective variety $X = V((F_1, F_2, \ldots, F_m))$.

Then the rank of the matrix (2.52) is $n - d$, and we may assume, by reindexing $g_1, \ldots, g_m, w_1, \ldots, w_n$ if necessary, that

$$\det\left(\frac{\partial g_i}{\partial w_j}(c_1, \ldots, c_n)\right)_{1 \le i \le j \le n-d} \neq 0.$$

If we set

$$u_i = w_{n-d+i} - c_{n-d+i}, \quad i = 1, 2, \ldots, d,$$

then by the implicit function theorem we may represent a neighborhood of the point $(c_1 : \cdots : c_r : 1 : c_{r+1} : \cdots : c_n)$ of X in terms of the parameters $u_1, \ldots, u_d$, namely,

$$\begin{aligned}
&(\phi_1(u_1, \ldots, u_d) : \cdots : \phi_i(u_1, \ldots, u_d) : 1 : \phi_{i+1}(u_1, \ldots, u_d) : \phi_{n-d}(u_1, \ldots, u_d) : \\
&\qquad u_1 + c_{n-d+1} : \cdots : u_d + c_n).
\end{aligned}$$

Here $\phi_j(u_1, \ldots, u_d)$ is a holomorphic function of $u_1, \ldots, u_d$ in a neighborhood of $(0, 0, \ldots, 0)$. We shall call $u_1, \ldots, u_d$ **local parameters** around the point $(a_0 : a_1 : \cdots : a_N) = (c_1 : \cdots : c_r : 1 : c_{r+1} : \cdots : c_n)$. By using local parameters we can show that a d-dimensional nonsingular projective variety admits the structure of a d-dimensional complex analytic manifold. We shall not go any deeper into this here.

EXAMPLE 2.25. For the quadratic hypersurface

$$Q : \sum_{i,j=0}^{n} c_{ij} x_i x_j = 0, \quad c_{ij} = c_{ji},$$

treated in Example 2.17, a singular point comes from a solution of the equation other than $(0, 0, \ldots, 0)$. Hence Q is a nonsingular projective variety if and only if the rank of (c_{ij}) is maximal, i.e., $n + 1$.

EXAMPLE 2.26. The singularities of the hypersurface

$$S : F(x_0, x_1, x_2, x_3) = x_0^2 x_1^2 + x_2^2 x_3^2 = 0$$

are found by solving

$$\begin{aligned}
\frac{\partial F}{\partial x_0} &= 2x_0 x_1^2 = 0 \\
\frac{\partial F}{\partial x_1} &= 2x_0^2 x_1 = 0 \\
\frac{\partial F}{\partial x_2} &= 2x_2 x_3^2 = 0 \\
\frac{\partial F}{\partial x_3} &= 2x_2^2 x_3 = 0.
\end{aligned}$$

Therefore the singular points consist of four lines in $\mathbf{P}^3(\mathbf{C})$:

$$\ell_{ij} : x_i = 0, \quad x_{2+j} = 0, \quad 0 \le i, j \le 1,$$

that is, every point of S outside ℓ_{ij} is nonsingular.

EXAMPLE 2.27. We show that the twisted cubic we discussed in Example 2.18 is a nonsingular projective variety. The curve C is given by the three equations

$$F(x_0,x_1,x_2,x_3) = x_0x_3 - x_1x_2 = 0$$
$$G(x_0,x_1,x_2,x_3) = x_1^2 - x_0x_2 = 0$$
$$H(x_0,x_1,x_2,x_3) = x_2^2 - x_1x_3 = 0.$$

Given a point $(1 : b_1 : b_2 : b_3)$ on C, we set $z_i = x_i/x_0, i = 1,2,3$, and

$$f(z) = \frac{1}{x_0^2}F(x) = z_3 - z_1z_2$$
$$g(z) = \frac{1}{x_0^2}G(x) = z_1^2 - z_2$$
$$h(z) = \frac{1}{x_0^2}H(x) = z_2^2 - z_1z_3.$$

Then the matrix (2.51) is

$$\begin{pmatrix} -b_2 & -b_1 & 1 \\ 2b_1 & -1 & 0 \\ -b_3 & 2b_2 & -b_1 \end{pmatrix},$$

which has rank 2. The point of C with the coordinates $(0 : a_1 : a_2 : a_3)$ is $(0 : 0 : 0 : 1)$. By setting

$$w_i = \frac{x_{i-1}}{x_3}, \quad i = 1,2,3,$$

we get

$$\tilde{f}(w) = \frac{1}{x_3^2}F(x) = w_1 - w_2w_3$$
$$\tilde{g}(w) = \frac{1}{x_3^2}G(x) = w_2^2 - w_1w_3$$
$$\tilde{h}(w) = \frac{1}{x_3^2}H(x) = w_3^2 - w_2,$$

and the matrix in question turns out to be

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix},$$

whose rank is 2. Therefore C is a nonsingular projective variety.

Finally, we touch upon the **tangent hyperplane**. Assume $(a_0 : a_1 : \cdots : a_n)$ is a nonsingular point of a d-dimensional projective variety $X = V((F_1,\ldots,F_m))$. The tangent hyperplane T_a to X is given by

$$\frac{\partial F_j}{\partial x_i}(a_0,a_1,\ldots,a_n)x_i = 0,\ j = 1,2,\ldots,m. \tag{2.56}$$

It is a direct consequence of Euler's identity that T_a passes through the point $(a_0 : a_1 : \cdots : a_n)$. We shall show that T_a coincides with the totality of lines through the point $(a_0 : a_1 : \cdots : a_n)$ that are tangent to X at the point. The line

going through $(a_0 : a_1 : \cdots : a_n)$ and another point $(b_0 : b_1 : \cdots : b_n)$ admits a parametric representation in a neighborhood of the point $(a_0 : a_1 : \cdots : a_n)$:

$$(sb_0 + (1-s)a_0 : sb_1 + (1-s)a_1 : \cdots : sb_n + (1-s)a_n),$$

with parameter s. A necessary and sufficient condition for this line to be tangent to X at $(a_0 : a_1 : \cdots : a_n)$ is that

$$F_j(sb_0 + (1-s)a_0, sb_1 + (1-s)a_1, \ldots, sb_n + (1-s)a_n) = 0, \quad j = 1,2,\ldots,m,$$

has $s = 0$ as a multiple root. On the other hand, we have

$$\begin{aligned} &F_j(sb_0 + (1-s)a_0, sb_1 + (1-s)a_1, \cdots, sb_n + (1-s)a_n) \\ &\quad = F_j((1-s)a_0, (1-s)a_1, \cdots, (1-s)a_n) \\ &\qquad + \sum_{i=0}^{n} \frac{\partial F_j}{\partial x_i}((1-s)a_0, (1-s)a_1, \ldots, (1-s)a_n)sb_i \\ &\qquad + \text{terms of degree greater than 1 in } s \\ &\quad = (1-s)^{d_j-1}s\sum_{i=0}^{n} \frac{\partial F_j}{\partial x_i}(a_0,a_1,\ldots,a_n)b_i + \text{terms of degree greater than 1 in } s. \end{aligned}$$

Thus the condition for $s = 0$ to be a multiple root is

$$\sum_{i=0}^{n} \frac{\partial F_j}{\partial x_i}(a_0,a_1,\ldots,a_n)b_i = 0, \quad j = 1,2,\ldots,m.$$

This means that $(b_0 : b_1 : \cdots : b_n) \in T_a$ and that the line in question lies on T_a. Conversely, it is obvious from the discussion above that a line on T_a that goes through $(a_0 : a_1 : \cdots : a_n)$ is tangent to X at the point. We cannot define the tangent hyperplane at a singular point. Also, unlike the case of plane curves, it is not in general possible to have a parametric representation by using a number of local parameters. In order to remove a singular point, we would need blowing-up, as stated in 2.5 (c) and its generalization. In 2.5 (c) we shall discuss a simple example concerning removal of an isolated singular point on a surface.

(f) The product of projective spaces. In this subsection we show how we can regard the product $\mathbf{P}^m(\mathbf{C}) \times \mathbf{P}^n(\mathbf{C})$ as a projective variety. For simplicity, we write $\mathbf{P}^n$ for $\mathbf{P}^n(\mathbf{C})$. As a set, $\mathbf{P}^m \times \mathbf{P}^n$ is defined as the set of all pairs of points in $\mathbf{P}^m$ and $\mathbf{P}^n$

$$\{((a_0 : a_1 : \cdots : a_m), (b_0 : b_1 : \cdots : b_n)) \mid (a_0 : a_1 : \cdots : a_m) \in \mathbf{P}^m, \\ (b_0 : b_1 : \cdots : b_n) \in \mathbf{P}^n\}.$$

We begin with the simplest situation with $m = n = 1$. We consider the following mapping:

$$\begin{aligned} \psi : &\mathbf{P}^1 \times \mathbf{P}^1 \longrightarrow \mathbf{P}^3 \\ &((a_0 : a_1), (b_0 : b_1)) \longmapsto (a_0b_0 : a_0b_1 : a_1b_0 : a_1b_1). \end{aligned} \tag{2.57}$$

First we want to make sure that ψ is well-defined; this follows from

$$\begin{aligned}\psi((\alpha a_0, \alpha a_1), (\beta b_0, \beta b_1)) &= (\alpha\beta a_0 b_0 : \alpha\beta a_0 b_1 : \alpha\beta a_1 b_1) \\ &= (a_0 b_0 : a_0 b_1 : a_1 b_o : a_1 b_1) \\ &= \psi((a_0 : a_1), (b_0, b_1)).\end{aligned}$$

Now is the image $\psi(\mathbf{P}^1 \times \mathbf{P}^1)$ the zero set of a number of polynomials? Consider the polynomial $F = x_0 x_3 - x_1 x_2$ of the homogeneous coordinates $(x_0 : x_1 : x_2 : x_3)$. Then

$$F(a_0 b_0, a_0 b_1, a_1 b_0, a_1 b_1) = a_0 b_0 a_1 b_1 - a_0 b_1 a_1 b_0 = 0.$$

Conversely, let $(c_0 : c_1 : c_2 : c_3)$ be a point of $V(F)$. We have

$$c_0 c_3 = c_1 c_2.$$

First take the case where $c_0 \neq 0$. In this case, we have

$$\frac{c_3}{c_0} = \frac{c_1}{c_0} \cdot \frac{c_2}{c_0}. \tag{2.58}$$

Then we have from (2.58)

$$\begin{aligned}\psi\left(\left(1 : \frac{c_2}{c_0}\right), \left(1 : \frac{c_1}{c_0}\right)\right) &= \left(1 : \frac{c_1}{c_0} : \frac{c_2}{c_0} : \frac{c_2}{c_0} \cdot \frac{c_1}{c_0}\right) \\ &= \left(1 : \frac{c_1}{c_0} : \frac{c_2}{c_0} : \frac{c_3}{c_0}\right) \\ &= (c_0 : c_1 : c_2 : c_3),\end{aligned}$$

which shows that the point $(c_0 : c_1 : c_2 : c_3)$ is contained in $\psi(\mathbf{P}^1 \times \mathbf{P}^1)$. We consider next the case where $c_1 \neq 0$. In this case, we have

$$\frac{c_2}{c_1} = \frac{c_0}{c_1} \cdot \frac{c_3}{c_1}. \tag{2.59}$$

We also have from (2.59)

$$\begin{aligned}\psi\left(\left(1 : \frac{c_3}{c_1}\right), \left(\frac{c_0}{c_1} : 1\right)\right) &= \left(\frac{c_0}{c_1} : 1 : \frac{c_3}{c_1} \cdot \frac{c_0}{c_1} : \frac{c_3}{c_1}\right) \\ &= \left(\frac{c_0}{c_1} : 1 : \frac{c_2}{c_1} : \frac{c_3}{c_1}\right) \\ &= (c_0 : c_1 : c_2 : c_3),\end{aligned}$$

which shows that $(c_0 : c_1 : c_2 : c_3) \in \psi(\mathbf{P}^1 \times \mathbf{P}^1)$, The cases $c_2 \neq 0, c_3 \neq 0$ are handled in the same way. Now we know from Example 2.17 that a nonsingular quadratic surface can be mapped by a projective transformation onto the standard form $x_0 x_3 - x_1 x_2 = 0$. Hence we can identify a nonsingular quadratic surface with $\mathbf{P}^1 \times \mathbf{P}^1$.

Next, we study $\mathbf{P}^1 \times \mathbf{P}^2$, which is a little more complicated. We take the mapping

$$\begin{gathered}\psi : \mathbf{P}^1 \times \mathbf{P}^2 \longrightarrow \mathbf{P}^5, \\ ((a_0, : a_1), (b_0 : b_1 : b_2)) \longmapsto (a_0 b_0 : a_0 b_1 : a_0 b_2 : a_1 b_0 : a_1 b_1 : a_1 b_2).\end{gathered} \tag{2.60}$$

This is well-defined, as can be easily verified as before. Using the homogeneous coordinates $(z_0 : z_1 : z_2 : z_3 : z_4 : z_5)$ we find the equations satisfied by the image of ψ. We consider the polynomials

$$\begin{aligned}F &= z_0 z_4 - z_1 z_3 \\ G &= z_0 z_5 - z_2 z_3 \\ H &= z_1 z_5 - z_2 z_4.\end{aligned}$$

It is easy to check that

$$\psi(\mathbf{P}^1 \times \mathbf{P}^2) \subset V((F, G, H)).$$

(Incidentally, we explain how we have found F, G and H. We can write $a_0 a_1 b_0 b_1$ as a product in two ways, $(a_0 b_0)(a_1 b_1)$ and $(a_0 b_1)(a_1 b_0)$, and so we see that the point $(a_0 b_0 : a_0 b_1 : a_0 b_2 : a_1 b_0 : a_1 b_1 : a_1 b_2)$ is annulled by $z_0 z_4 - z_1 z_3$. Similarly, by writing $a_0 a_1 b_0 b_2, a_0 a_1 b_1 b_2$ as a product in two ways we can find G and H.) Now consider a point $(c_0 : c_1 : c_2 : c_3 : c_4 : c_5)$ in $V((F, G, H))$, We have

$$\begin{aligned}c_0 c_4 - c_1 c_3 &= 0 \\ c_0 c_5 - c_2 c_3 &= 0 \\ c_1 c_5 - c_2 c_4 &= 0.\end{aligned}$$

First, assume $c_0 \neq 0$. In this case, we have

$$\frac{c_4}{c_0} = \frac{c_1}{c_0} \cdot \frac{c_3}{c_0}, \quad \frac{c_5}{c_0} = \frac{c_2}{c_0} \cdot \frac{c_3}{c_0}, \quad \frac{c_1}{c_0} \cdot \frac{c_5}{c_0} = \frac{c_2}{c_0} \cdot \frac{c_4}{c_0}. \tag{2.61}$$

We obtain using (2.61)

$$\begin{aligned}\psi\left(\left(1 : \frac{c_3}{c_0}\right), \left(1 : \frac{c_1}{c_0} : \frac{c_2}{c_0}\right)\right) &= \left(1 : \frac{c_1}{c_0} : \frac{c_2}{c_0} : \frac{c_3}{c_0} : \frac{c_3}{c_0} \cdot \frac{c_1}{c_0} : \frac{c_3}{c_0} \cdot \frac{c_2}{c_0}\right) \\ &= \left(1 : \frac{c_1}{c_0} : \frac{c_2}{c_0} : \frac{c_3}{c_0} : \frac{c_4}{c_0} : \frac{c_5}{c_0}\right) \\ &= (c_0 : c_1 : c_2 : c_3 : c_4 : c_5),\end{aligned}$$

which shows that the point $(c_0 : c_1 : c_2 : c_3 : c_4 : c_5)$ is in $\psi(\mathbf{P}^1 \times \mathbf{P}^2)$. The cases $c_i \neq 0, i = 1, \ldots, 5$, are similar. Hence we have shown that $V((F, G, H)) = \psi(\mathbf{P}^1 \times \mathbf{P}^2)$.

The reader will now see how to generalize the mapping ψ. We take

$$\psi : \mathbf{P}^m \times \mathbf{P}^n \longrightarrow \mathbf{P}^{(m+1)(n+1)-1} \tag{2.62}$$

defined by

$$\begin{aligned}((a_0 : \cdots : a_m), (b_0 : \cdots : b_n)) \longmapsto (a_0 b_0 : a_0 b_1 : \cdots : a_0 b_n : a_1 b_0 : a_1 b_1 : \cdots : a_1 b_n : \\ \cdots : a_i b_0 : a_i b_1 : \cdots : a_i b_n : \cdots : a_m b_0 : a_m b_1 : \cdots : a_m b_n).\end{aligned}$$

The mapping is well-defined. To find the equations satisfied by the image of ψ, we take the homogeneous coordinates $(z_0 : z_1 : \cdots : z_{(m+1)(n+1)-1})$ in $\mathbf{P}^{(m+1)(n+1)-1}$. For $0 \leq i < j \leq m, 0 \leq k < \ell \leq n$, we have

$$a_i a_j b_k b_\ell = (a_i b_k)(a_j b_\ell) = (a_i b_\ell)(a_j b_k),$$

which implies that the image of ψ annuls the quadratic homogeneous polynomial

$$F_{(i,j)(k,\ell)} = z_{(n+1)i+k} z_{(n+1)j+\ell} - z_{(n+1)i+\ell} z_{(n+1)j+k}.$$

(Note that, according to (2.62), a_ib_j appears as the $((n+1)i+j+1)$-th coordinate in $\mathbf{P}^{(m+1)(n+1)-1}$, and the first coordinate is z_0.) Arguing as before, we get

$$\psi(\mathbf{P}^m \times \mathbf{P}^n) = V\left((F_{(i,j)(k,\ell)}, 0 \le i < j \le m, 0 \le k < \ell \le n)\right).$$

We state

THEOREM 2.7. *$\mathbf{P}^m \times \mathbf{P}^n$ can be regarded as a projective variety in $\mathbf{P}^{(m+1)(n+1)-1}$ by means of the mapping ψ, and the defining equations are*

$$F_{(i,j)(k,\ell)} = z_{(n+1)i+k}z_{(n+1)j+\ell} - z_{(n+1)i+\ell}z_{(n+1)j+k} = 0,$$
$$\text{where } 0 \le i < j \le m, 0 \le k < \ell \le n.$$

The same kind of discussions can be made for the product of a finite number of projective spaces $\mathbf{P}^{n_1} \times \mathbf{P}^{n_2} \times \cdots \times \mathbf{P}^{n_\ell}$. We leave it as an exercise for the reader.

After learning that $\mathbf{P}^m \times \mathbf{P}^n$ is a projective variety, we want to define its submanifold structure directly. Let $(x_0 : x_1 : \cdots : x_m)$ and $(y_0 : y_1 : \cdots : y_n)$ be the homogeneous coordinates of $\mathbf{P}^m$ and $\mathbf{P}^n$, respectively. We say that a polynomial $F(x_0, x_1, \ldots, x_m, y_0, y_1, \ldots, y_n)$ of x_i, y_j is doubly homogeneous of degree d relative to x_i and of degree e relative to y_j (or doubly homogeneous of degree (d, e)) if

$$F(\alpha x_0, \alpha x_1, \ldots, \alpha x_m, \beta y_0, \beta y_1, \ldots, \beta y_n)$$
$$= \alpha^d\beta^e F(x_0, x_1, \ldots, x_m, y_0, y_1, \ldots, y_n) \text{ for arbitrary } \alpha, \beta \in \mathbf{C}^* = \mathbf{C} - \{0\},$$

that is, F is homogeneous of degree d as a polynomial in $(x_0 : x_1 : \cdots : x_m)$ with $(y_0 : y_1 : \cdots : y_n)$ fixed, and is homogeneous of degree e as a polynomial in $(y_0 : y_1 : \cdots : y_n)$ with $(x_0 : x_1 : \cdots : x_m)$ fixed. For example,

$$x_0^2y_1^3 + 2x_1x_2y_0y_1^2 + 3x_2^2y_0^3$$

is homogeneous of degree 2 in x_0, x_1, x_2 and homogeneous of degree 3 in y_0, y_1, y_2. On the other hand,

$$x_0^2y_1^3 + 2x_1^2y_0 + 3x_2^2y_0^3$$

is homogeneous of degree 2 in x_0, x_1, x_2 but not homogeneous in y_0, y_1.

Let $F(x_0, x_1, \ldots, x_m, y_0, y_1, \ldots, y_n)$ be doubly homogeneous of degree (d, e). Suppose also

$$F(a_0, a_1, \ldots, a_m, b_0, b_1, \ldots, b_n) = 0$$

for $(a_0 : a_1 : \cdots : a_m) \in \mathbf{P}^m$ and $(b_0 : b_1 : \cdots : b_n) \in \mathbf{P}^n$. Then for all $\alpha, \beta \in \mathbf{C}^*$ we have

$$F(\alpha a_0, \ldots, \alpha a_m, \beta b_0, \ldots, \beta b_n) = 0;$$

hence it makes sense to say that a point $((a_0 : a_1 : \cdots : a_m), (b_0 : b_1 : \cdots : b_n)) \in \mathbf{P}^m \times \mathbf{P}^n$ is a zero of F. That is,

$$V(F) = \{((a_0 : a_1 : \cdots : a_m), (b_0 : b_1 : \cdots : b_n)) \in \mathbf{P}^m \times \mathbf{P}^n |$$
$$F(a_0, \ldots, a_m, b_0, \ldots, b_n) = 0\}$$

defines a subset of $\mathbf{P}^m \times \mathbf{P}^n$, usually called a **hypersurface of degree** (d, e). More generally, for a number of doubly homogeneous polynomials $F_1, F_2, \ldots, F_\ell$, we can define

$$V((F_1, \ldots, F_\ell)) = \{((a_0 : \cdots : a_m), (b_0 : \cdots : b_n)) \in \mathbf{P}^m \times \mathbf{P}^n |$$
$$F_i(a_0, \ldots, a_m, b_0, \ldots, b_n) = 0,\ 1 \le i \le \ell\}.$$

We call it a **projective set** in $\mathbf{P}^m \times \mathbf{P}^n$. As with the case of projective spaces, we can define the notion that a projective set is **reducible** or **irreducible**. An irreducible projective set is called a **projective submanifold**. A projective submanifold V in $\mathbf{P}^m \times \mathbf{P}^n$ can be considered by means of the mapping ψ in (2.62) as a subset of $\mathbf{P}^{(m+1)(n+1)-1}$; it turns out that $\psi(V)$ is a projective variety in $\mathbf{P}^{(m+1)(n+1)-1}$.

EXAMPLE 2.28. Consider the hypersurface V of degree (1,1) defined by the equation

$$x_0y_2 - x_1y_1 = 0 \tag{2.63}$$

in $\mathbf{P}^1 \times \mathbf{P}^2$. Since the polynomial $x_0y_2 - x_1y_1$ is irreducible, V is irreducible. Let π denote the restriction to V of the projection of $\mathbf{P}^1 \times \mathbf{P}^2$ onto the second factor:

$$\pi : ((a_0 : a_1)(b_0 : b_1 : b_2)) \in V \longmapsto (b_0 : b_1 : b_2) \in \mathbf{P}^2.$$

We show that π is surjection. If $(b_0 : b_1 : b_2) \neq (1 : 0 : 0)$, then we get from (2.63)

$$\pi^{-1}((b_0 : b_1 : b_2)) = ((b_1 : b_2), (b_0 : b_1 : b_2)).$$

For the point $(1 : 0 : 0)$, (2.63) imposes no condition, so that

$$\pi^{-1}((1 : 0 : 0)) = \mathbf{P}^1 \times \{(1 : 0 : 0)\}.$$

Set $E = \pi^{-1}((1 : 0 : 0))$ and $p = (1 : 0 : 0)$. Then π gives an isomorphism from $V - E$ onto $\mathbf{P}^2 - \{p\}$ while $\pi(E) = p$. That is, only the curve E in V collapses to a single point by π, and elsewhere π is an isomorphism. Thus the projective variety V and the projective plane $\mathbf{P}^2$ differ only slightly. It turns out that their function fields are isomorphic and that V and $\mathbf{P}^2$ are birationally equivalent and π is a birational mapping. In the present case, V is obtained by inserting the projective line $\mathbf{P}^2$ instead of the point p. We call this process a **blow-up** and the curve E an **exceptional curve**. We shall give more detail in the next section. Here we study the significance of points of E. A line going through the point $p = (1 : 0 : 0)$ can be written in the form $\alpha y_2 - \beta y_1 = 0$. Hence the points $(1 : \alpha t : \beta t), t \in \mathbf{C}$, lie on the line. For $t \neq 0$ we have

$$\pi^{-1}((1 : \alpha t : \beta t)) = ((\alpha t, \beta t), (1 : \alpha t : \beta t))$$
$$= ((\alpha : \beta), (1 : \alpha t : \beta t)).$$

As $t \to 0$, this point becomes $((\alpha : \beta), (1 : 0 : 0))$ and determines a point of E. From this it follows that each point $((\alpha : \beta), (1 : 0 : 0))$ corresponds to the slope of the line $\alpha y_2 - \beta y_1 = 0$ that goes through the point $(1 : 0 : 0)$. We may interpret a blow-up as the process of enlarging the point p by taking the slopes of the lines going through it.

Finally, let us consider the mapping

$$\varpi : ((a_0 : a_1), (b_0 : b_1 : b_2)) \longmapsto (a_0 : a_1) \in \mathbf{P}^1.$$

Since $(a_0 : a_1) \neq (0, 0)$, we can solve (2.63) and find

$$\varpi^{-1}((a_0 : a_1)) = ((a_0 : a_1), (b : a_0 : a_1)).$$

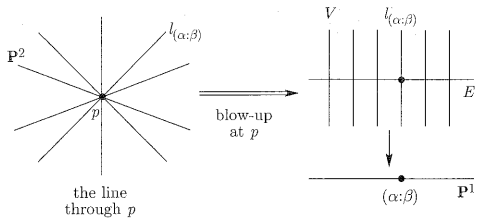


FIGURE 2.13 THE BLOW-UP IN THE PROJECTIVE PLANE

Here b may take any value. Thus we can regard $\varpi^{-1}((a_0 : a_1))$ as $\mathbf{P}^1$. Or we may identify $\varpi^1((a_0 : a_1))$ with the line determined by

$$a_0y_2 - a_1y_1 = 0$$

in $\mathbf{P}^2$, and therefore say that $\varpi^{-1}((a_0 : a_1))$ is isomorphic with $\mathbf{P}^1$. In this way, the mapping ϖ has the structure that the inverse image of each point of $\mathbf{P}^1$ can be identified with $\mathbf{P}^1$. Geometrically, the situation looks like Figure 2.13. Let us think of the totality of lines in $\mathbf{P}^2$ that go through p. Such a line is expressed by

$$\alpha y_2 - \beta y_1 = 0$$

and thus corresponds to the point $(\alpha : \beta)$ in $\mathbf{P}^1$. Denote this line by $\ell_{(\alpha:\beta)}$. As sets, we have

$$\mathbf{P}^2 = \bigcup_{(\alpha:\beta)\in\mathbf{P}^2} \ell_{(\alpha:\beta)}.$$

If $(\alpha : \beta) \neq (\alpha' : \beta')$, we see that $\ell_{(\alpha:\beta)}$ and $\ell_{(\alpha':\beta')}$ meet at p only and that each point of $\mathbf{P}^2 - \{p\}$ lies on only one line $\ell_{(\alpha:\beta)}$. It follows that we can define a mapping

$$\varpi' : q \in \mathbf{P}^2 - \{p\} \longmapsto (\alpha : \beta) \in \mathbf{P}^1,$$

where q is a point on $\ell_{(\alpha:\beta)}$. Although we cannot extend the mapping to p, we may blow-up $\mathbf{P}^2$ to V at p. If we realize that each point of the exceptional curve E corresponds to the slope of a line through p, then ϖ' can be extended to a mapping from V to $\mathbf{P}^1$. The result is nothing but the mapping ϖ.

We can handle the blow-up at a point $(a_0 : a_1 : a_2)$ other than $(1 : 0 : 0)$. In this case, we can take the hypersurface of degree (1,1) in $\mathbf{P}^1 \times \mathbf{P}^2$ defined by

$$x_0(a_0y_2 - a_2y_0) - x_1(a_0y_1 - a_1y_0) = 0$$

instead of (2.63). Since the arguments are similar, we shall leave it as an exercise for the reader.

§ 2.5. The resolution of singularities

In this section we shall discuss the resolution of singularities for a plane curve. We already touched on blowing-up in §2.4 (f); here we give more detail and apply it to the resolution of singularities of a plane curve.

(a) Blowing-up in the projective plane. We discuss blowing-up at a point p in $\mathbf{P}^2(\mathbf{C})$. We consider the slope of each line through p as a point whose totality is then $\mathbf{P}^1(\mathbf{C})$, which we substitute in p to make up a new projective variety V. We denote the result by $Q_p(\mathbf{P}^2(\mathbf{C}))$ and call the curve that has replaced p the exceptional curve E_p. Now having understood this concept, can we proceed to the blow-up of $Q_p(\mathbf{P}^2(\mathbf{C}))$ at one of its points? To answer this question we begin by defining the blow-up of the affine plane $\mathbf{A}^2$.

Let (u, v) be the coordinates on the affine plane $\mathbf{A}^2$ and $(x_0 : x_1)$ the homogeneous coordinates of $\mathbf{P}^1(\mathbf{C})$. Define the subset $\tilde{U}$ of $\mathbf{P}^1(\mathbf{C}) \times \mathbf{A}^2$ by the equation

$$ux_1 - vx_0 = 0, \tag{2.64}$$

that is,

$$\tilde{U} = \{((a_0 : a_1), (b, c)) \in \mathbf{P}^1(\mathbf{C}) \times \mathbf{A}^2 | ba_1 - ca_0 = 0\}.$$

If we further define $\tilde{\mathcal{U}}_i$, $i = 0, 1$, by

$$\tilde{\mathcal{U}}_i = \{((a_0 : a_1), (b, c)) | a_i \neq 0\}, \quad i = 0, 1,$$

then we get

$$\tilde{U} = \tilde{\mathcal{U}}_1 \cup \tilde{\mathcal{U}}_2.$$

For a point $((a_0 : a_1), (b, c)) \in \tilde{U}$ we have

$$ba_1 - ca_0 = 0$$

and hence

$$c = b \cdot \frac{a_1}{a_0},$$

so that

$$((a_0 : a_1), (b, c)) = \left(\left(1 : \frac{a_1}{a_0}\right), \left(b, b \cdot \frac{a_1}{a_0}\right)\right).$$

It follows that there is an isomorphism ϕ_0 from $\mathbf{A}^2$ onto $\tilde{U}_0$,

$$\phi_0 : (u, v) \in \mathbf{A}^2 \longmapsto ((1 : u), (v, uv)) \in \tilde{\mathcal{U}}_0. \tag{2.65}$$

Similarly, we have an isomorphism ϕ_1 from $\mathbf{A}^2$ onto $\tilde{\mathcal{U}}_1$,

$$\phi_1 : (w, z) \in \mathbf{A}^2 \longmapsto ((w : 1), (wz, z)) \in \tilde{\mathcal{U}}_1. \tag{2.66}$$

In this fashion we can identify $\tilde{\mathcal{U}}_0$ and $\tilde{\mathcal{U}}_1$ with the affine plane and regard $\tilde{U}$ as the superimposition of the two affine planes by the map $\phi = \phi_1^{-1} \circ \phi_0$. Here $\phi = \phi_1^{-1} \circ \phi_0$ is an isomorphism of $\mathbf{A}^2 - \{u = 0\}$ onto $\mathbf{A}^2 - \{w = 0\}$ which is given by

$$w = \frac{1}{u} \quad \text{and} \quad z = uv, \tag{2.67}$$

that is, by

$$\phi((x, y)) = (1/x, xy).$$

We show how we may regard $\tilde{U}$ as the blow-up of $\mathbf{A}^2$ at the origin. We take the projection from $\mathbf{P}^1(\mathbf{C}) \times \mathbf{A}^2$ and get the map

$$\tilde{\pi} : ((a_0 : a_1), (b, c)) \in \tilde{U} \longmapsto (b, c) \in \mathbf{A}^2. \tag{2.68}$$

If $(b, c) \neq (0, 0)$, then (2.64) implies

$$\tilde{\pi}^{-1}((b, c)) = ((b : c), (b, c)),$$

and at $(0,0) \in \mathbf{A}^2$ we have

$$\tilde{\pi}^{-1}((0,0)) = \mathbf{P}^1(\mathbf{C}) \times (0,0).$$

By setting $E = \tilde{\pi}^{-1}((0,0))$, we find that it is isomorphic to the projective line and $\tilde{\pi}$ gives an isomorphism from $\tilde{U} - E$ onto $\mathbf{A}^2 - \{(0,0)\}$, The situation is just like that for $\pi : V \to \mathbf{P}^2(\mathbf{C})$ in §2.4 (f). Recall that $\mathbf{P}^2(\mathbf{C})$ can be thought of as made up by three affine planes $\mathbf{A}^2$ pasted by the relations

$$\text{(2.69)} \qquad \begin{cases} u_1 = \dfrac{1}{v_1} \\ u_2 = \dfrac{v_2}{v_1} \end{cases} \quad \begin{cases} v_1 = \dfrac{w_1}{w_2} \\ v_2 = \dfrac{1}{w_2} \end{cases} \quad \begin{cases} w_1 = \dfrac{1}{u_2} \\ w_2 = \dfrac{u_1}{u_2}. \end{cases}$$

(To see this, take the homogeneous coordinates $(x_0 : x_1 : x_2)$ in $\mathbf{P}^2$ and set

$$(u_1, u_2) = (x_1/x_0, x_2/x_0), (v_1, v_2) = (x_0/x_1, x_2/x_1), (w_1, w_2) = (x_0/x_2, x_1/x_2)).$$

Denote by $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ the affine planes with (u_1, u_2), (v_1, v_2), and (w_1, w_2) as coordinates, respectively. We can now observe that $\mathbf{P}^2(\mathbf{C})$ is obtained by pasting $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ by the relations (2.69). Thus

$$\text{(2.70)} \qquad \mathbf{P}^2(\mathbf{C}) = \mathcal{U} \cup \mathcal{V} \cup \mathcal{W}.$$

Now regard the affine plane $\mathbf{A}^2$ that appears in (2.68) as $\mathcal{U}$. If we consider $\tilde{U}_0$ and $\tilde{U}_1$ as affine planes by means of ϕ_0 and ϕ_1 and express the mapping $\tilde{\pi} : \tilde{U} \to \mathcal{U}$, we find from (2.65) and (2.66) that

$$\text{(2.71)} \qquad \begin{aligned} (u, v) &\to (v, uv) \\ (w, z) &\to (wz, z). \end{aligned}$$

We note that in the pasting (2.70) we can paste $\tilde{U}$ instead of U with $\mathcal{V} \cup \mathcal{W}$ as follows. Note that $\tilde{\pi}$ gives an isomorphism between $\tilde{U} - E$ and $\tilde{\mathcal{U}} - \{(0,0)\}$. and we have $u_1 \neq 0$ where $\mathcal{U}$ and $\mathcal{V}$ are pasted and $u_2 \neq 0$ where $\mathcal{U}$ and $\mathcal{W}$ are pasted. Thus the point $(0,0)$ is not contained in any part that is to be pasted. Thus by identifying $\tilde{U} - E$ and $\tilde{\mathcal{U}} - \{(0,0)\}$ by $\tilde{\pi}$ we can paste together $\tilde{U}$, $\mathcal{V}$, and $\mathcal{W}$. Since $\mathcal{V} \cup \mathcal{W} = \mathbf{P}^2(\mathbf{C}) - \{(0,0)\}$, we represent our process above by

$$V = \tilde{U} \cup_{\tilde{\pi}} (\mathbf{P}^2(\mathbf{C}) - \{(1:0:0)\} = \tilde{U} \cup \mathcal{V} \cup \mathcal{W}.$$

We can define a mapping

$$\pi : V \longmapsto \mathbf{P}^2(\mathbf{C}) = \mathcal{U} \cup \mathcal{V} \cup \mathcal{W}$$

by taking $\pi = \tilde{\pi}$ on $\tilde{U}$ and setting π to be the identity mapping on $\mathcal{V}$ and $\mathcal{W}$. It is easily seen that this is the same blow-up of $\mathbf{P}^2(\mathbf{C})$ as that in Example 2.28 in §2.4 (f). That is, we have $V = Q_p(\mathbf{P}^2(\mathbf{C})), p = (1:0:0)$. Furthermore, $\tilde{U}$ can be constructed by pasting two affine planes $\tilde{\mathcal{U}}_0$ and $\tilde{\mathcal{U}}_1$. We have thus shown that $V = Q_p(\mathbf{P}^2(\mathbf{C}))$ can be constructed by pasting four affine planes, and that the essential part of this process is the blow-up $\tilde{U}$ of the affine plane $\mathbf{A}^2$.

In the discussions above we considered blow-ups at the origin of $\mathbf{A}^2$. Blow-ups at other points can be defined in the same way. That is, the blow-up at the point $(a,b) \in \mathbf{A}^2$ is given by the subset

$$\tilde{U}_{(a,b)} : (u-a)x_1 - (v-b)x_0 = 0$$

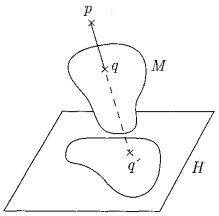


FIGURE 2.14 THE PROJECTION FROM POINT p

of $\mathbf{P}^1(\mathbf{C}) \times \mathbf{A}^2$. We leave it an exercise for the readers to construct the blow-up $Q_p(\mathbf{P}^2(\mathbf{C}))$ of $\mathbf{P}^2(\mathbf{C})$ at its arbitrary point p.

In the following it becomes necessary to further blow-up $Q_p(\mathbf{P}^2(\mathbf{C}))$ at one of its points p_1. As we stated already, $Q_p(\mathbf{P}^2(\mathbf{C}))$ can be constructed by pasting affine planes, one of which, say $\mathcal{X}$, contains p_1. Denote by $\tilde{\mathcal{X}}$ the blow-up of $\mathcal{X}$ at p_1. Then we may define

$$Q_{p_1}(Q_p(\mathbf{P}^2(\mathbf{C}))) = (Q_p(\mathbf{P}^2(\mathbf{C})) - \{p_1\}) \cup \tilde{\mathcal{X}}$$

as the blow-up of $Q_p(\mathbf{P}^2(\mathbf{C}))$ at the point p_1. Since it is also possible to construct this by pasting affine planes, we may repeat this process and show the following result.

LEMMA 2.15. *The result V obtained from $\mathbf{P}^2(\mathbf{C})$ by a finite number of blow-ups is a projective variety and can be obtained by pasting a finite number of affine planes.*

The fact that $Q_p(\mathbf{P}^2(\mathbf{C}))$ obtained by a blow-up of $\mathbf{P}^2(\mathbf{C})$ is a projective variety was proved in §2.4 (f). It is somewhat involved to prove that the result V of a finite number of pastings of affine planes is a projective variety. Intuitively, the blow-up can be constructed as follows.

Consider an m-dimensional projective variety M (see Figure 2.14). We assume that M is not contained in any hyperplane of $\mathbf{P}^N(\mathbf{C})$ and that N is sufficiently large relative to m. We fix a hyperplane H in a most general position in $\mathbf{P}^N(\mathbf{C})$ and further fix a point p. We first treat the case where $p \notin M$. Take an arbitrary point q of M and let q' be the intersection of the line $\overline{pq}$ with H.

Then we get the mapping

$$\phi_p : q \longmapsto q' = H \cap \overline{pq} \in H,$$

and the image $M' = \phi_p(M)$ turns out to be a projective variety in $H = \mathbf{P}^{N-1}(\mathbf{C})$; M is imbedded in a sufficiently general position and ϕ_p is an isomorphism if N is large enough relative to m.

Next, in the case where $p \in M$, ϕ_p is defined on $M - \{p\}$ only. If q_n is a sequence of points in $M - \{p\}$ that converges to p, then the line $\overline{pq_n}$, as $n \to \infty$, will have a limit line, say ℓ, tangent to M at p. Thus the "limit" of $\phi_p(q_n)$ as $n \to \infty$ is the intersection of ℓ and H. Let $\tilde{M}$ be the set obtained from $\phi_p(M - \{p\})$ by adjoining the limits of $\phi_p(q_n)$ for all sequences $q_n \to p$. It turns out that $\tilde{M}$ is a projective

variety of $H = \mathbf{P}^{N-1}(\mathbf{C})$. The set $E = \tilde{M} - \phi_p(M - \{p\}$ coincides with the set of all intersections of the tangents at p with H, and we may consider E as the set of all tangent directions to M at p. In particular, if p is a nonsingular point of M, the set of all tangent directions to M at p is $\mathbf{P}^{m-1}(\mathbf{C})$, and thus $E = \mathbf{P}^{m-1}(\mathbf{C})$. For the case where $m = 2$, this gives the same result as the blow-up of $\mathbf{P}^2(\mathbf{C})$.

The preceding intuitive arguments can be made into a rigorous discussion, which we have to omit since it is beyond the scope of this book. We hope that the geometric significance of blow-ups is clear to a certain extent. It is also possible to describe blow-ups by using local parameters introduced in §2.4 (e). By doing so, the blow-up of a projective variety at a nonsingular point can be rigorously defined.

(b) Resolution of singularities of plane curves. Consider an algebraic curve $C = V(F)$ in $\mathbf{P}^2(\mathbf{C})$ defined by an irreducible homogeneous polynomial $F(x_0, x_1, x_2)$ of degree m. Recall that a point $p = (a_0 : a_1 : a_2)$ is a singular point of C if

$$\frac{\partial F}{\partial x_i}(a_0, a_1, a_2) = 0, \quad i = 0, 1, 2.$$

Let us assume that $p = (1 : 0 : 0)$ is a singular point. If we take inhomogeneous coordinates

$$u_1 = x_1/x_0, \quad u_2 = x_2/x_0,$$

then in the affine plane $\mathcal{U} = \{(a_0 : a_1 : a_2) \in \mathbf{P}^2(\mathbf{C}) | a_0 \neq 0\}$ the curve C is defined as the set of zeros of the polynomial

$$f(u_1, u_2) = x_0^{-m} F(x_0, x_1, x_2) = \sum_{i+j \geq n} a_{ij} u_1^i u_2^j. \tag{2.72}$$

Since p is a singular point, we have $n \geq 2$ (in fact, it is a singularity of order n).

Now let us blow-up $\mathbf{P}^2(\mathbf{C})$ at the point $p = (1 : 0 : 0)$. What is affected is the affine plane $\mathcal{U}$ that contains p. In the notation of the preceding subsection (a), we have a mapping $\tilde{\pi}$ from the blow-up $\tilde{\mathcal{U}}$ of $\mathcal{U} = \mathbf{A}^2$ at the origin to $\mathcal{U}$. We study the pull-back by $\tilde{\pi}$ of C, or more precisely, of $C \cap \mathcal{U}$.

Pulling-back (2.72) by the mapping $\tilde{\pi}$ and using the correspondence (2.71), we see that $\tilde{\pi}^{-1}(C \cap \mathcal{U})$ in $\tilde{\mathcal{U}}$ is the zero set of the polynomial

$$\begin{aligned} g(u, v) &= \sum_{i+j \geq n} a_{ij} v^i (uv)^j \\ &= \sum_{i+j \geq n} a_{ij} u^j v^{i+j} \\ &= v^n \sum_{j \geq 0, k \geq 0} a_{n+k-j,j} u^j v^k \end{aligned} \tag{2.73}$$

and in $\tilde{\mathcal{U}}_1$ the zero set of the polynomial

$$\begin{aligned} h(w, z) &= \sum_{i+j \geq n} a_{ij} (wz)^i z^j \\ &= \sum_{i+j \geq n} a_{ij} w^i z^{i+j} \\ &= \sum_{i \geq 0, k \geq 0} a_{i,n+k-1} w^i z^k. \end{aligned} \tag{2.74}$$

Even though $f(u_1, u_2)$ is an irreducible polynomial, $g(u, v)$ and $h(z, w)$ are reducible. Let us consider the meaning of the factors v^n and z^n. The equations

$$v = 0, \quad z = 0$$

determine the exceptional curve E in $\tilde{U}$. Hence $g = 0, h = 0$ contain the part that defines E n-fold. On the other hand, if we write

$$g(u, v) = v^n \tilde{g}(u, v), \quad h(z, w) = z^n \tilde{h}(z, w), \tag{2.75}$$

we obtain

$$\begin{aligned} \tilde{g}(u, v) &= \sum_{j \geq 0, k \geq 0} a_{n+k-j,j} u^j v^k \\ \tilde{h}(z, w) &= \sum_{i \geq 0, k \geq 0} a_{i,n+k-i} w^i z^k, \end{aligned} \tag{2.76}$$

so that

$$\tilde{g}(u, v) = 0, \quad \tilde{h}(u, v) = 0$$

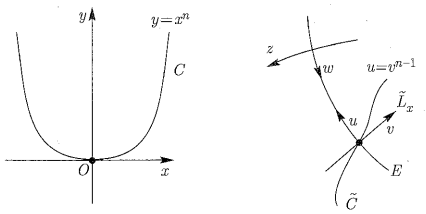
determine a curve $\tilde{C}$ on $\tilde{U}$. In $\tilde{U}_0 \cap \tilde{U}_1$, the equations $\tilde{g}(u, v) = 0$ and $\tilde{h}(u, v) = 0$ define the same curve as follows. From $w = 1/u, z = uv$ in (2.67), we get

$$\begin{aligned} \tilde{h}(z, w) = \tilde{h}(uv, \frac{1}{u}) &= \sum_{i \geq 0, k \geq 0} a_{i,n+k-i} u^{-i} (uv)^k \\ &= \sum_{i \geq 0, k \geq 0} a_{i,n+k-i} u^{k-i} v^k \\ &= u^{-n} \sum_{i \geq 0, k \geq 0} a_{i,n+k-i} u^{n+k-i} v^k \\ &= u^{-n} \sum_{i \geq 0, k \geq 0} a_{n+k-j,j} u^j v^k \\ &= u^{-n} \tilde{g}(u, v) \end{aligned}$$

on $\tilde{U}_0 \cap \tilde{U}_1$ (where $u \neq 0$).

Now by means of the mapping $\pi : Q_p(\mathbf{P}^2(\mathbf{C})) \to \mathbf{P}^2(\mathbf{C})$, in $\mathcal{V} \cup \mathcal{W}$, $\pi^{-1}(\mathcal{V})$ and $\mathcal{V}$ are isomorphic and $\pi^{-1}(\mathcal{W})$ and $\mathcal{W}$ are isomorphic. In $\tilde{U} \cap \pi^{-1}(\mathcal{V} \cup \mathcal{W})$, $\tilde{C}$ and $\pi^{-1}(C - \{p\})$ coincide. Hence $\tilde{C}$ and $\pi^{-1}(C - \{p\})$ can be pasted together, giving rise to a curve in $Q_p(\mathbf{P}^2(\mathbf{C}))$. Let us denote this curve also by $\tilde{C}$

The observations above show that if we take a curve C in $\mathbf{P}^2(\mathbf{C})$ and a point $p = (1 : 0 : 0)$, blow-up $\mathbf{P}^2(\mathbf{C})$ at p, and get the mapping $\pi : Q_p(\mathbf{P}^2(\mathbf{C})) \to \mathbf{P}^2(\mathbf{C})$, then $\pi^{-1}(C)$ is, as a set, the union of the curve $\tilde{C}$ and the exceptional curve E. More precisely, the curve defined by the pull-back by π of the defining equation of



(a) (b)

FIGURE 2.15. (a) The curve $C : y = x^n$ is tangent to the x-axis at the origin with multiplicity n. (b) A neighborhood of the exceptional curve after the blow-up. $\tilde{C}$ and $\tilde{L}_x$ are tangent with multiplicity $n-1$.

C is, by virtue of (2.75) and (2.76), given by E taken n-fold and $\tilde{C}$. We express this as

$$\pi^* C = nE + \tilde{C}$$

and call it the **total transform** of C by π. We denote the curve $\tilde{C}$ by $\pi^{-1}[C]$ and call it the **strict transform** of C by π. (Do not confuse the expressions $\pi^{-1}[C]$ and $\pi^{-1}(C)$.) From the definition of blow-up it follows that $\tilde{C} \cap E$ and $C - \{p\}$ are isomorphic. This implies that, intuitively speaking, $\tilde{C}$ is the curve we get by closing $\pi^{-1}(C - \{p\})$ by adjoining a finite number of points. In fact, $\tilde{C}$ is the closure $\overline{\pi^{-1}(C - \{p\})}$ of $\pi^{-1}(C - \{p\})$ relative to the usual topology of $Q_p(\mathbf{P}^2(\mathbf{C}))$.

To study the properties of $\tilde{C}$ let us examine a few examples. Since it is sufficient to study $C \cap \mathcal{U}$, we set $\pi : \tilde{U} \mapsto \mathcal{U} = \mathbf{A}^2$ and denote $C \cap \mathcal{U}$ by C. Using coordinates (u, v), (w, z), and (x, y) in the affine planes $\tilde{\mathcal{U}}_0$, $\tilde{\mathcal{U}}_1$, and $\mathcal{U}$, respectively, we can express the mapping $\pi : \tilde{U} = \mathcal{U}_0 \cup \mathcal{U}_1 \to \mathcal{U}$

$$\pi : (u, v) \mapsto (v, uv) = (x, y)$$
$$\pi : (w, z) \mapsto (wz, z) = (x, y)$$

and on $\tilde{\mathcal{U}}_0 \cap \tilde{\mathcal{U}}_1$ by

$$\begin{cases} w = \dfrac{1}{u} \\ z = uv. \end{cases}$$

(See (2.67).)

EXAMPLE 2.29. Let us consider the curve

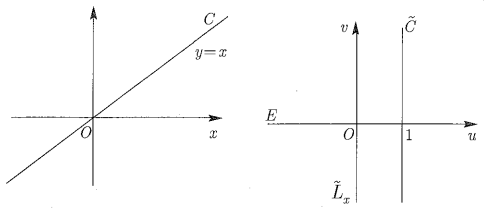
$$C : y - x^n = 0,$$

which is nonsingular at the origin $(0, 0)$. The total transform $\pi^*(C)$ of C is given by

$$v(u - v^{n-1}) = 0 \text{ on } \tilde{\mathcal{U}}_0$$

and by

$$z(w^n z^{n-1} - 1) = 0 \text{ on } \tilde{\mathcal{U}}_1.$$



(a) (b)

FIGURE 2.16. (a) Line C: $y = x$. (b) $\tilde{C}$ and $\tilde{L}_x$ do not intersect.

Thus we have

$$\pi^*(C) = E + \tilde{C},$$

where $\tilde{C}$ is the strict transform of C. In our case, the origin $(0, 0)$ is a nonsingular point of C, and $\tilde{C}$ and C are isomorphic by π. Incidentally, if we denote by L_x and L_y the x-axis ($y = 0$) and the y-axis ($x = 0$), then the total transform $\pi^*(L_x)$ is defined by

$$uv = 0 \text{ on } \tilde{\mathcal{U}}_0$$

and by

$$z = 0 \text{ on } \tilde{\mathcal{U}}_1.$$

Hence we get

$$\pi^* L_x = E + (v\text{-axis}),$$

and the strict transform $\tilde{L}_x$ turns out to be the v-axis. It also follows that the total transform $\pi^* L_y$ of L_y is given by

$$v = 0 \text{ on } \tilde{\mathcal{U}}_0$$

and by

$$wz = 0 \text{ on } \tilde{\mathcal{U}}_1.$$

Hence

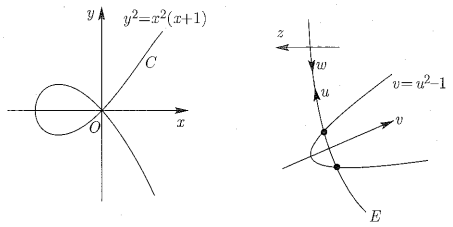
$$\pi^* L_y = E + (z\text{-axis}),$$

which shows that the strict transform $\tilde{L}_y$ of L_y is equal to the z-axis. In $\mathcal{U}$ the curve C and the y-axis L_y intersect at the origin, and the tangent lines to both curves at the origin are the x-axis and the y-axis, respectively. By blowing-up at the origin, the strict transforms $\tilde{C}$ and $\tilde{L}_y$ of C and L_y do not intersect. This can be seen because $\tilde{C} \cap E$ and $\tilde{L}_y \cap E$ correspond to the directions of the tangents to C and L_y, respectively, at the origin.

On the other hand, if $n \geq 2$, then C and L_x are tangent with multiplicity n at the origin, and $\tilde{C}$ and $\tilde{L}_x$ are expected to intersect because they have a common tangent line. In fact, they intersect with multiplicity $n - 1$ at the origin of $\tilde{\mathcal{U}}_0$. If $n = 1$, then C and L_x are two distinct lines intersect at the origin, and $\tilde{C}$ and $\tilde{L}_x$ do not intersect (see Figure 2.16).

EXAMPLE 2.30. The curve

$$C : y^2 - x^2(x + 1) = 0$$



(a) (b)

FIGURE 2.17. (a) $y^2 = x^2(x+1)$ has an ordinary double point at the origin. (b) The blow-up at the origin.

has the origin as a singular point (an ordinary double point). As can be seen from Figure 2.17, C has two distinct tangent lines at $(0,0)$. The term of lowest degree in x,t of the defining equation is

$$y^2 - x^2,$$

and the equation

$$y^2 - x^2 = 0$$

determines the tangent cone of C at the origin, as was discussed in §2.3 (a), Example 2.6. It should follow that $\tilde{C} \cap E$ consists of two points.

The total transform of C is defined by

$$v^2(u^2 - (v+1)) = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_0$$

and by

$$z^2(1 - w^2(wz+1)) = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_1.$$

The strict transform $\tilde{C}$ is defined by

$$v = u^2 - 1 \quad \text{on} \quad \tilde{\mathcal{U}}_0,$$

and intersects the exceptional curve E at two points $(u,v) = (\pm 1, 0)$. Obviously, $\tilde{C}$ has no singular point.

EXAMPLE 2.31. The curve

$$C : y^2 - x^3 = 0$$

has a singularity (an ordinary cusp) only at the origin. As can be imagined from Figure 2.18, there is only one tangent line to C at the origin. The lowest term in the defining equation of C is

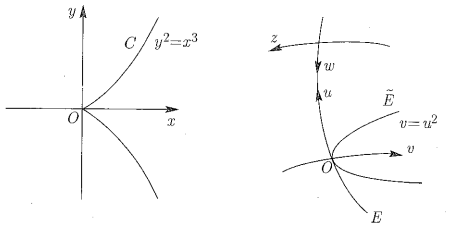
$$y^2$$

and such a singular point is called an ordinary cusp or a cusp of type (2,3). Here $\tilde{C} \cap E$ should consist of one point. The total transform of C is defined by

$$v^2(u^2 - v) = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_0$$

and by

$$z^2(1 - w^3 z) = 0 \quad \text{on} \quad \mathcal{U}_1.$$



(a) (b)

FIGURE 2.18. (a) $y^2 = x^3$ has a cusp at the origin. (b) The blow-up at the origin.

Therefore the strict transform $\tilde{C}$ is given by

$$v = u^2 \quad \text{on} \quad \tilde{\mathcal{U}}_0$$

and by

$$w^3 z = 1 \quad \text{on} \quad \tilde{\mathcal{U}}_1,$$

and it has no singular point. The curve $\tilde{C}$ and the exceptional curve E are tangent at $(u,v) = (0,0)$, and the map

$$\varpi = \tilde{\pi}|_{\tilde{C}} : \tilde{C} \longrightarrow C,$$

as a mapping of point set, is one-to-one and onto. However, $\tilde{C}$ is nonsingular and C has a singular point at the origin; they are different as algebraic curves, and ϖ is not an isomorphism. Let us find a concrete form of the mapping ϖ in a neighborhood of $\tilde{C}$ around $(u,v) = (0,0)$. Since $\tilde{C}$ is defined by $v = u^2$ on $\tilde{\mathcal{U}}_0$, the mapping

$$\mu : \mathbf{A}^1 \longrightarrow \tilde{C} \cap \tilde{\mathcal{U}}_0$$

given by

$$t \in \mathbf{A}^1 \longmapsto (t, t^2) \in \tilde{C} \cap \tilde{\mathcal{U}}_0$$

is an isomorphism of the affine line. Since $\tilde{C}$ is defined by $v = u^2$ on $\tilde{\mathcal{U}}_0$, we see that

$$\tilde{\varpi} = \varpi \circ \mu : \mathbf{A}^1 \longrightarrow C$$

gives

$$t \longmapsto (t^2, t^3).$$

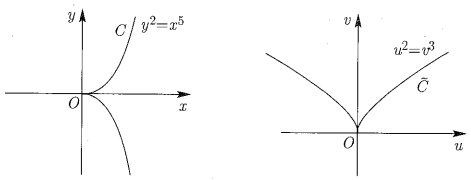
Also noteworthy is the fact that we have

$$\tilde{C} \subset \tilde{\mathcal{U}}_0$$

because $\tilde{C}$ is defined by $w^3 z = 1$ in $\tilde{\mathcal{U}}_1$ and $w \neq 0$ in $\tilde{C}$, and thus $\tilde{C} \cap \tilde{\mathcal{U}}_1$ is contained in the part of $\tilde{\mathcal{U}}_1$ where $w \neq 0$. But this part is nothing but $\tilde{\mathcal{U}}_0 \cap \tilde{\mathcal{U}}_1$. Therefore we see that μ gives an isomorphism of $\mathbf{A}^1$ onto $\tilde{C}$, and the mapping $\tilde{\varpi} : \mathbf{A}^1 \to C$ shows how the singular points of C arise.

More generally, let p and q be two relative prime numbers greater than 2. Consider the mapping

$$\nu : \mathbf{A}^1 \to \mathbf{A}^2$$



(a) (b)

FIGURE 2.19. (a) The curve $y^2 = x^5$ has a cusp of type (2,5) at the origin. (b) The blow-up at the origin brings the strict transform $\tilde{C}$ with an ordinary cusp at the origin.

such that

$$t \mapsto (t^p, t^q).$$

The image by ν is contained in the curve

$$C : x^q - y^p = 0.$$

It is also easy to see that $\nu(\mathbf{A}^1) = C$. The curve C has a cusp of type (p, q) at the origin, and

$$\nu : \mathbf{A}^1 \to C$$

gives a resolution of the singular point of C.

We shall now derive these results by using blow-ups. For simplicity, let us treat cusps of type $(2, 5)$.

EXAMPLE 2.32. The curve

$$C = x^5 - y^2 = 0$$

has a cusp of type $(2, 5)$ at the origin and is nonsingular elsewhere. The total transform of C is defined by

$$v^2(v^3 - u^2) = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_0$$

and by

$$z^2(w^5 z^3 - 1) = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_1.$$

Thus the strict transform $\tilde{C}$ of C is defined by

$$v^3 - u^2 = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_0$$

and by

$$w^5 z^3 - 1 = 0 \quad \text{on} \quad \tilde{\mathcal{U}}_1.$$

It is easy to show that $\tilde{C} \subset \tilde{\mathcal{U}}_0$. In contrast to the previous examples, $\tilde{C}$ in this case has an ordinary cusp at $(u, v) = (0, 0)$. By further blowing up the origin of $\tilde{\mathcal{U}}_0$ we get the strict transform $\tilde{\tilde{C}}$. From the preceding example we see that $\tilde{\tilde{C}}$ is nonsingular. We leave it to the reader to verify that $\tilde{\tilde{C}} \to C$ is given by the mapping in the preceding example $\nu : \mathbf{A}^1 \to \mathbf{A}^2$ defined by $t \mapsto (f^2, t^5)$. (See Exercise 2.13.)

As is clear from the example above, if we blow-up $\mathbf{P}^2(\mathbf{C})$ at a singular point p of a plane curve C, and think about the strict transform $\tilde{C}$ of C and the naturally defined mapping $\pi : \tilde{C} \to C$, $\pi^{-1}(p)$ may not be a single point, and every point of it may be a singular point of $\tilde{C}$. What is important is that the behavior of singular points in π^{-1} is generally better than the behavior of the singular point p. That is, if we blow-up each singular point in π^{-1} and get the strict transform $\tilde{\tilde{C}}$ of $\tilde{C}$, then the behavior of singular points is improved over that for $\tilde{C}$. By iterating this process a finite number of times we finally get a curve without singular points. The end result is no longer a plane curve but a curve imbeddable in a higher-dimensional projective space. Thus we have

THEOREM 2.8 (RESOLUTION OF SINGULARITIES FOR PLANE CURVES). *By repeating blowing-ups a finite number of times, an irreducible plane curve C gives rise to the strict transform that is a nonsingular projective curve $\tilde{C}$. Furthermore, if $\{p_1, \ldots, p_m\}$ are singular points of C, the natural mapping $\pi : \tilde{C} \to C$ induces an isomorphism from $\tilde{C} - \pi^{-1}\{p_1, p_2, \ldots, p_m\}$ onto $C - \{p_1, \ldots, p_m\}$.*

The theorem above can be extended to algebraic varieties by defining general blow-ups along a submanifold. This is Hironaka's theorem on the resolution of singularities.

(c) Resolution of singularities for a surface. We shall now extend the argument in the preceding subsection to the case of a surface. For simplicity we shall discuss only one example.

Let us consider the affine surface

$$S : x^{n+1} + y^2 + z^2 = 0 \tag{2.77}$$

in the three-dimensional affine space $\mathbf{A}^3$. The surface has a singular point at the origin $(0, 0, 0)$, so we blow-up $\mathbf{A}^3$ a the origin in the same way as in §2.5 (a). In $\mathbf{P}^2(\mathbf{C}) \times \mathbf{A}^3$ we consider a submanifold $\tilde{U}$ defined by the equations

$$\begin{aligned} x_0 y - x_1 x &= 0 \\ x_0 z - x_2 x &= 0 \\ x_1 z - x_2 y &= 0, \end{aligned} \tag{2.78}$$

where $(x_0 : x_1 : x_2)$ are the homogeneous coordinates in $\mathbf{P}^2(\mathbf{C})$. Then $\tilde{U}$ is the union of the three affine spaces

$$\tilde{U}_i = \{((x_0 : x_1 : x_2), (x, y, z)) \in \tilde{U} | x_i \neq 0\}, \quad i = 0, 1, 2.$$

Here $\tilde{U}_i \cong \mathbf{A}^2$. To see this, take $i = 0$ for example. From (2.78) we get

$$y = \frac{x_1}{x_0}, \quad z = \frac{x_2}{x_0}.$$

Thus we can take as coordinates in $\tilde{U}_0$

$$x, \quad u_1 = \frac{x_1}{x_0}, \quad u_2 = \frac{x_2}{x_0},$$

and we get an isomorphism

$$(x, u_1, u_2) \in \mathbf{A}^2 \longmapsto ((1 : u_1 : u_2), (x, u_1 x, u_2 x)) \in \tilde{U}_0. \tag{2.79}$$

Similarly, for $\tilde{U}_1$ we have coordinates

$$y, \quad v_1 = \frac{x_0}{x_1}, \quad v_2 = \frac{x_2}{x_1},$$

and an isomorphism

$$(y, v_1, v_2) \in \mathbf{A}^2 \longmapsto ((v_1 : 1 : v_2), (v_1 y, y, v_2 y)) \in \tilde{U}_1. \tag{2.80}$$

Likewise, for $\tilde{U}_2$ we have coordinates

$$z, \quad w_1 = \frac{x_0}{x_2}, \quad v_2 = \frac{x_1}{x_2},$$

and an isomorphism

$$(z, w_1, w_2) \in \mathbf{A}^2 \longmapsto ((w_1 : w_2 : 1), (w_1 z, w_2 z, z)) \in \tilde{U}_2. \tag{2.81}$$

Furthermore, using the projection from $\mathbf{P}^2(\mathbf{C}) \times \mathbf{A}^3$ to $\mathbf{A}^3$, we get a mapping

$$\pi : ((x_0 : x_1 : x_2), (x, y, z)) \in \tilde{U} \longmapsto (x, y, z) \in \mathbf{A}^3,$$

which can be expressed in terms of the coordinates in $\tilde{U}_0, \tilde{U}_1$, and $\tilde{U}_2$ by

$$\begin{aligned} (x, u_1, u_2) &\longmapsto (x, u_1 x, u_2 x) \\ (y, v_1, v_2) &\longmapsto (v_1 y, y, v_2 y) \\ (z, w_1, w_2) &\longmapsto (w_1 z, w_2 z, z). \end{aligned} \tag{2.82}$$

We have $\pi^{-1}((0,0,0)) = \mathbf{P}^2(\mathbf{C}) \times \{(0,0,0)\}$, which we denote by E and call the **exceptional surface**. The mapping π gives an isomorphism between $\tilde{U} - E$ and $\mathbf{A}^2 - \{(0,0,0)\}$ in the same way as in §2.5 (a).

By pulling back the equation (2.77) by π we get

$$\begin{aligned} x^2(x^{n-1} + u_1^2 + u_2^2) = 0 \quad &\text{on} \quad \tilde{U}_0 \\ y^2(v_1^{n+1} y^{n-1} + 1 + v_2^2) = 0 \quad &\text{on} \quad \tilde{U}_1 \\ z^2(w_1^{n+1} z^{n-1} + w_2^2 + 1) = 0 \quad &\text{on} \quad \tilde{U}_2. \end{aligned}$$

The defining equation for E is

$$x = 0 \text{ on } \tilde{U}_0, \quad y = 0 \text{ on } \tilde{U}_1, \quad z = 0 \text{ on } \tilde{U}_2,$$

and the strict transform $\tilde{S} = \pi^{-1}(S)$ is given by

$$\begin{aligned} x^{n-1} + u_1^2 + u_2^2 = 0 \quad &\text{on} \quad \tilde{U}_0 \\ v_1^{n+1} y^{n-1} + 1 + v_2^2 = 0 \quad &\text{on} \quad \tilde{U}_1 \\ w_1^{n+1} z^{n-1} + w_2^2 + 1 = 0 \quad &\text{on} \quad \tilde{U}_2. \end{aligned}$$

By simple computation we see that this surface has no singular point on $\tilde{U}_1$ and $\tilde{U}_2$, and it is sufficient to study

$$x^{n-1} + u_1^2 + u_2^2 = 0 \tag{2.83}$$

on $\tilde{U}_0$. If $n = 1$ or 2, then there is no singular point. If $n \geq 3$, we may repeat the same process on $\tilde{U}_0$ a finite number of times and get a nonsingular surface. For the

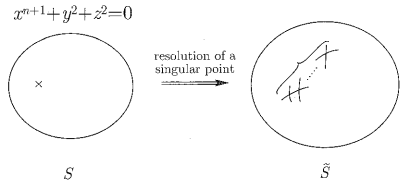


FIGURE 2.20. By resolving the singularity of $x^{n+1} + y^2 + z^2 = 0$, n copies of $\mathbf{P}^1(\mathbf{C})$ appear with a tree-like intersection.

case where $n = 1$, let us check the difference between $\tilde{S} = \pi^{-1}[S]$ and S. Since π is an isomorphism on $\tilde{U} - E$, it is sufficient to find the intersection of $\tilde{S}$ and E. Now $\tilde{S} \cap E$ is determined by

$$\begin{aligned} x = 0, \quad &1 + u_1^2 + u_2^2 = 0 \quad \text{on } \tilde{U}_0 \\ y = 0, \quad &v_1^{n+1} y^{n-1} + 1 + v_2^2 = 0 \quad \text{on } \tilde{U}_1 \\ z = 0, \quad &1 + w_1^2 + w_2^2 = 0 \quad \text{on } \tilde{U}_2. \end{aligned}$$

Since $E = \mathbf{P}^2(\mathbf{C}) \times \{(0,0,0\} \subset \mathbf{P}^2(\mathbf{C}) \times \mathbf{A}^3$, we may express $E \cap \tilde{S}$ in terms of the homogeneous coordinates $(x_0 : x_1 : x_2)$ in $\mathbf{P}^2(\mathbf{C})$ by making use of (2.79), (2.80), (2.81). We obtain the quadric

$$x_0^2 + x_1^2 + x_2^2 = 0,$$

that is, when we resolve the singular point $(0,0,0)$ of S and get the surface $\tilde{S}$, there appears the quadric, namely, $\mathbf{P}^1(\mathbf{C})$, corresponding to the singular point. The situation here is quite different from the case of resolution of singularities for curves.

A singular point is called an **isolated singularity** if it is the only singular point in its neighborhood, just like the singular point $(0,0,0)$ in the surface S above. In resolving an isolated singularity for a surface or higher-dimensional variety there always appears a projective set of a certain dimension. This situation is entirely different from the case of singularity of a curve for which each branch can be parametrized. This makes it difficult to analyze a singularity of a projective variety of dimension at least equal to two. The other side of the coin is that the singularity has a much richer structure.

Going back to the surface S above, we check on $\tilde{S} \cap E$ for the case $n = 2$. This time the defining equation is given by

$$\begin{aligned} x = 0, \quad &u_1^2 + u_2^2 = 0 \text{ on } \tilde{U}_0, \\ y = 0, \quad &1 + v_2^2 = 0 \text{ on } \tilde{U}_1, \\ z = 0, \quad &1 + w_2^2 = 0 \text{ on } \tilde{U}_2. \end{aligned}$$

Since $E = \mathbf{P}^2(\mathbf{C}) \times \{0,0,0\}$, we may express it in terms of the coordinates in $\mathbf{P}^2(\mathbf{C})$ by

$$x_1^2 + x_2^2 = 0,$$

that is, two projective lines. In other words, two $\mathbf{P}^1(\mathbf{C})$ meeting at one point appear as a result of the resolution of singularity. Analogous results hold for general n. (See Figure 2.20.)

Problems

2.1. (i) If a projective transformation g of $\mathbf{P}^1(\mathbf{C})$ satisfies

$$g((1:0)) = (1:0), \quad g((0:1)) = (0:1), \quad g((1:1)) = (1:1)),$$

then prove that g is the identity mapping.

(ii) If a projective transformation h satisfies

$$h((1:0:0)) = (1:0:0), \quad h((0:1:0)) = (0:1:0),$$
$$h((0:0:1)) = (0:0:1), \quad h((1:1:1)) = (1:1:1),$$

then show that h is the identity mapping.

(iii) Given a pair of four distinct points (P_1, P_2, P_3, P_4) and (Q_1, Q_2, Q_3, Q_4) in $\mathbf{P}^2(\mathbf{C})$ such that no three points among P_1, P_2, P_3, P_4 are collinear and the same is true for Q_1, Q_2, Q_3, Q_4, then show that there is a unique projective transformation f of $\mathbf{P}^2(\mathbf{C})$ such that $f(P_j) = Q_j$ for $1 \leq j \leq 4$.

2.2. Prove that a meromorphic function on the Riemann sphere is a rational function. [Hint: If a meromorphic function f has poles $a_1, a_2, \ldots, a_N$ of order $m_1, m_2, \ldots, m_N$, then

$$f(z)/\prod_{j=1}^{N}(z-a_j)^{m_j}$$

is holomorphic on the Riemann sphere and hence is a constant function. If a_N happens to be a point at infinity, consider

$$f(z)/z^{m_N}\prod_{j=1}^{N-1}(z-a_j)^{m_j}.]$$

2.3. (i) If $a_0, a_1, b_0, b_1, c_0, c_1$ satisfy

$$\left(\begin{vmatrix} b_0 & b_1 \\ c_0 & c_1 \end{vmatrix}, \begin{vmatrix} c_0 & c_1 \\ a_0 & a_1 \end{vmatrix}, \begin{vmatrix} a_0 & a_1 \\ b_0 & b_1 \end{vmatrix}\right) \neq (0,0,0),$$

then the point $(a_0 s + a_1 t : b_0 s + b_1 t : c_0 s + c_1 t)$, $(s,t) \in \mathbf{P}^1(\mathbf{C})$, lies on the line of $\mathbf{P}^2(\mathbf{C})$ given by

$$\begin{vmatrix} a_0 & a_1 & x_0 \\ b_0 & b_1 & x_1 \\ c_0 & c_1 & x_2 \end{vmatrix} = 0.$$

(ii) For the line

$$\ell_{\alpha,\beta,\gamma} : \alpha x_0 + \beta x_1 + \gamma x_2 = 0$$

show that there exist $a_0, a_1, b_0, b_1, c_0, c_1$ such that

$$\alpha = \begin{vmatrix} b_0 & b_1 \\ c_0 & c_1 \end{vmatrix}, \quad \beta = \begin{vmatrix} c_0 & c_1 \\ a_0 & a_1 \end{vmatrix}, \quad \gamma = \begin{vmatrix} a_0 & a_1 \\ b_0 & b_1 \end{vmatrix}.$$

Using this fact and (i), show that $\ell_{\alpha,\beta,\gamma}$ is the image of the mapping

$$(y_0 : y_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (a_0 y_0 + a_1 y_1 : b_0 y_0 + b_1 y_1 : c_0 y_0 + c_1 y_1) \in \mathbf{P}^2(\mathbf{C}).$$

2.4. (i) For any 3×3 real symmetric matrix $A \neq O$ show that there is a regular matrix B such that ${}^tBAB$ is one of the following matrices

$$\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

[Hint: We can find an orthogonal matrix X such that

$${}^tXAX = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}.]$$

(ii) For any 3×3 complex symmetric matrix $C \neq O$, show that there is a complex regular matrix B such that ${}^tBCB$ is one of the following matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(The results in (i) and (ii) can be generalized to $n \times n$ real or complex matrices for an arbitrary n.)

2.5. Show that the correspondence between the unit circle and the line (Figure 1.2) discussed in §1.2 (a) can be interpreted as an algebraic morphism from the line

$$\ell : x_0 - x_1 = 0 \text{ in } \mathbf{P}^2(\mathbf{C})$$

onto the quadric

$$C : x_0^2 - x_1^2 - x_2^2 = 0$$

given by

$$(u : u : v) \in \ell \longmapsto (4u^2 + v^2 : 4u^2 - v^2 : 4uv) \in C \subset \mathbf{P}^2(\mathbf{C}),$$

where we set $x = x_1/x_0, y = x_2/x_0$.

2.6. Show that a homogeneous polynomial $f(x, y)$ of degree m in two variables can be decomposed as a product of linear polynomials

$$f(x, y) = \prod_{j+1}^{m} (\alpha_j x - \beta_j y).$$

[Hint: Set $u = y/x$ and $g(u) = \frac{1}{x^m} f(x, y)$. Then $g(u)$ has degree m and m roots.]

2.7. Following Example 2.19, study the projective set $V((F, H))$, where

$$F = x_0x_3 - x_1x_2, \quad H = x_2^2 - x_1x_3.$$

2.8. (i) Show that an irreducible projective variety in $\mathbf{P}^1(\mathbf{C})$ is either one point or $\mathbf{P}^1(\mathbf{C})$.

(ii) Using the fact that a twisted cubic $V((F, G, H))$ with

$$F = x_0x_1 - x_1x_2, \quad G = x_1^2 - x_0x_2, \quad H = x_2^2 - x_1x_3$$

is the image of the mapping (see Example 2.1)

$$\phi : (a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (a_0^3 : a_0^2a_1 : a_0a_1^2 : a_1^3) \in P^3(\mathbf{C})$$

show that it is irreducible. [Hint: If $V((F, G, H))$ is written as a sum $W_1 \cup W_2 \cup \cdots \cup W_m$ of projective varieties, then $\phi^{-1}(W_j)$ is an irreducible projective variety and we can write

$$\mathbf{P}^1(\mathbf{C}) = \phi^{-1}(W_1) \cup \phi^{-1}(W_2) \cup \cdots \cup \phi^{-1}(W_m),$$

which contradicts (i) above.]

2.9. Show that for any projective sets V and W in $\mathbf{P}^n(\mathbf{C})$ we have

$$I(V \cup W) = I(V) \cap I(W), \quad I(V \cap W) = I(V) \cup I(W).$$

(See (2.48) for the definition of $I(V)$, etc.)

2.10. Show that the homogeneous ideal

$$\mathfrak{a} = (x_0x_3 - x_1x_2, x_1^3 - x_0^2x_2, x_2^3 - x_1x_3^2, x_1^2x_3 - x_0x_2^2)$$

in $\mathbf{C}[x_0, x_1, x_2, x_3]$ is a prime ideal. [Hint:$V(\mathfrak{a})$ coincides with the image of the mapping

$$\phi : (a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \mapsto (a_0^4 : a_0^3a_1 : a_0a_1^3 : a_1^4) \in \mathbf{P}^3(\mathbf{C}).]$$

2.11. Consider an irreducible hypersurface $V(F)$, where we assume that the homogeneous polynomial of degree d satisfies the conditions

$$F(0, x_1, x_2, \ldots, x_n) \not\equiv 0, \quad F(1, 0, 0, \ldots, 0) \neq 0.$$

Let $P = (1 : 0 : 0 : \cdots : 0)$. For a point $Q \in V(F)$, let $R(Q)$ be the intersection of the line $\overline{PQ}$ with the hyperplane at infinity $H_\infty : x_0 = 0$. Show that the mapping

$$\phi_P : Q \in V(F) \mapsto R(Q) \in H_\infty$$

is generally a d-to-1 mapping. Also show that for a point $R \in H_\infty$, there is a point $Q \in V(F)$ such that $R(Q) = R$.

2.12. Let $V = I(F)$ in $\mathbf{P}^n(\mathbf{C})$. Show that a necessary and sufficient condition for a point $(a_0 : a_1 : \cdots : a_n)$ on V to be a singular point is that

$$\frac{\partial F}{\partial x_j}(a_0, a_1, \ldots, a_n) = 0, \quad 0 \leq j \leq n.$$

Also prove Lemma 2.14.

2.13. Let ℓ and m be relatively prime integers greater than or equal to 2. Show that the resolution of a singularity for the curve

$$C : x^\ell - y^m = 0$$

is given by the mapping

$$\nu : t \in \mathbf{A}^1 \longmapsto (t^m, t^\ell) \in \mathbf{A}^2,$$

that is, $\nu(\mathbf{A}^1) = C$, and ν gives an isomorphism from $\mathbf{A}^1 - \{(0, 0)\}$ to $C - \{0, 0)\}$.

CHAPTER 3

Algebraic Curves

Algebraic curves are the most fundamental object of research in algebraic geometry, and have been studied intensively. In §§3.1 and 3.2, we explain the theory of divisors, particularly, the Riemann-Roch theorem, that is basic to the study of algebraic curves. After applications of the Riemann-Roch theorem we deal with the theory of elliptic curves in §3.3. Although we have so far treated the theory over the complex number field, we now develop the theory over an arbitrary field k. This is because the theory of elliptic curves over a finite field becomes necessary in applications to coding theory. In §3.4 we give important results and examples on congruence zeta functions for curves defined over a finite field. Projective varieties over a finite field are discrete sets whose points are represented by elements of the finite field and yet have beautiful properties, as we have come to realize. We also anticipate more importance in their applications. We hope §§3.3 and 3.4 can serve as an introduction to these theories.

In this chapter, unfortunately, we cannot give proofs to important theorems and lemmas due to limitation of space. Our description may not be fully detailed either, but we try to offer concrete images of the objects. The proofs of theorems and lemmas are given in the standard textbooks in algebraic geometry, to which our book is meant to be an introduction.

§ 3.1. The Riemann-Roch theorem

(a) Divisors. A 1-dimensional nonsingular projective variety is called an **algebraic curve**. It is known that an arbitrary algebraic curve can be obtained by resolving singular points of a plane algebraic curve. In §2.5 (b) we discussed the resolution of singular points of a plane curve. As stated in §2.3 (a), if we know the branches at a singular point and find their parametric representations, this has the same effect as resolving the singular point. In the following we adopt this point of view.

Now when a number of points $P_1, P_2, \ldots, P_k$ are chosen on an algebraic curve C, a linear combination with integral coefficients

$$m_1P_1 + m_2P_2 + \cdots + m_kP_k$$

is called a **divisor** of C. Given two divisors

$$D_1 = m_1P_1 + m_2P_2 + \cdots + m_kP_k$$
$$D_2 = n_1Q_1 + n_2Q_2 + \cdots + n_\ell Q_\ell,$$

we define $D_1 + D_2$ and $D_1 - D_2$ by

$$D_1 + D_2 = m_1P_1 + m_2P_2 + \cdots + m_kP_k + n_1Q_1 + n_2Q_2 + \cdots + n_\ell Q_\ell$$
$$D_1 - D_2 = m_1P_1 + m_2P_2 + \cdots + m_kP_k - n_1Q_1 - n_2Q_2 - \cdots - n_\ell Q_\ell.$$

Here, of course, we use the conventions:

$$\underbrace{P + P + \cdots + P}_{n} = nP$$

$$nP - mP = (n-m)P,$$

etc. If all the coefficients are 0, we simply write 0 for this special divisor. The set of all divisors on an algebraic curve turns out to be an Abelian group with 0 as the zero element. For a divisor

$$D_1 = m_1P_1 + m_2P_2 + \cdots + m_kP_k$$

its degree, $\deg D$, is defined to be

$$\deg D = m_1 + m_2 + \cdots + m_k.$$

By definition of addition and subtraction we get

$$\deg(D_1 \pm D_2) = \deg D_1 \pm \deg D_2.$$

For a given divisor

$$D = m_1P_1 + m_2P_2 + \cdots + m_kP_k - n_1Q_1 - n_2Q_2 - \cdots - n_\ell Q_\ell, \quad m_i, n_j \geq 2,$$

we consider the subset $\mathbf{L}(D)$ of the function field $\mathbf{C}(C)$ of the curve C as follows:

(3.1)
$$\mathbf{L}(D) = \{f \in \mathbf{C}(C) | f = 0 \text{ or } f \text{ has poles of order at most } m_i \text{ at } P_i, \text{ has zero points of order at least } n_j \text{ at } Q_j, \text{ and is regular elsewhere}\}.$$

Here to say that a rational function f on C has a pole (a zero point) of order m at P means that, with a local parameter t at P for C, f can be expressed in the form

$$f = at^{-m} + (\text{terms of higher order than } -m+1), a \neq 0$$

$$(f = bt^m + (\text{terms of highter order than } m+1), b \neq 0).$$

We also say f is **regular at a point** P if it has no pole at P .

EXAMPLE 3.1. Let $C = \mathbf{P}^1(\mathbf{C}), P = (0:1), Q = (1:0)$ and let $(x_0 : x_1)$ be the homogeneous coordinates of $\mathbf{P}^1(\mathbf{C})$. If we set $x = x_1/x_0$, then we have $\mathbf{C}(C) = \mathbf{C}(x)$. We can take $t = 1/x$ and $t = x$ as local parameters for P and Q, respectively. Therefore the rational function x^k on C has a pole of order k at P and a zero of order k at Q, and is regular elsewhere. Similarly, $1/x^k$ has a zero of order ℓ at P, a pole of order ℓ at Q, and is regular elsewhere. Hence we have

$$\mathbf{L}(nP) = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n | a_i \in \mathbf{C}, 0 \leq i \leq n\}$$
$$\mathbf{L}(nQ) = \{b_0 + b_1/x + b_2/x^2 + \cdots + b_n/x^n | b_i \in \mathbf{C}, 0 \leq i \leq n\}$$

$$\mathbf{L}(mP - nQ)$$
$$= \begin{cases} \{c_0x^n + c_1x^{n+1} + \cdots + c_{m-n}x^m | c_j \in \mathbf{C}, 0 \leq j \leq m-n\} \text{ for } m \geq n, \\ \{0\} \text{ for } m < n. \end{cases}$$

These are vector spaces over $\mathbf{C}$ with

$$\dim \mathbf{L}(nP) = n+1, \quad \dim \mathbf{L}(nQ) = n+1,$$

and

$$\dim \mathbf{L}(mP - nQ) = \begin{cases} m-n+1, \text{ if } m \geq n, \\ 0, \text{ if } m < n. \end{cases}$$

It is no accident that $\mathbf{L}(D)$ in Example 3.1 is a vector space. Suppose $f, g \in \mathbf{L}(D)$. If $\alpha = \beta = 0$, then $\alpha f + \beta g$ is 0 and belongs to $\mathbf{L}(D)$. If $\alpha \neq 0$ or $\beta \neq 0$, then $\alpha f + \beta g$ has poles of order at most m_i at $P_i, 1 \leq i \leq k$, and has zeros of order at least n_j at Q_j, $k+1 \leq j \leq n$, and is regular elsewhere. Hence $\alpha f + \beta g \in \mathbf{L}(D)$. As a matter of fact, we have a stronger result.

LEMMA 3.1. *For any divisor D, $\mathbf{L}(D)$ is a finite-dimensional vector space over* $\mathbf{C}$.

Suppose a rational function f on an algebraic curve C has $Q_1, Q_2, \ldots, Q_s$ as zeros of order $m_1, m_2, \ldots, m_s$, respectively, and $P_1, P_2, \ldots, P_\ell$ as poles of order $n_1, n_2, \ldots, n_\ell$, respectively. Set

$$(f)_0 = m_1Q_1 + m_2Q_2 + \cdots + m_sQ_s$$
$$(f)_\infty = n_1P_1 + n_2P_2 + \cdots + n_\ell P_\ell$$
$$(f) = (f)_0 - (f)_\infty.$$

We call (f) the **principal divisor** determined by f. For any rational functions f, g on $\mathbf{C}$, we have

$$(fg) = (f) + (g),$$
$$\left(\frac{f}{g}\right) = (f) - (g)$$

by definition.

LEMMA 3.2. $\deg(f) = 0$.

In §3.2 (a) we shall discuss the geometric reason why this result holds.

EXAMPLE 3.2. Consider the plane curve

$$C : F = x_0^{n-2}x_2^2 - x_1^n - a_1x_0x_1^{n-1} - a_2x_0^2x_1^{n-2} - \cdots - a_{n-1}x_0^{n-1}x_1 - a_nx_0^n = 0.$$

In the following we assume that

(3.2)
$$x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$$

has n distinct roots $\alpha_1, \alpha_2, \ldots, \alpha_n$. The singular points of C can be found by solving

$$\frac{\partial F}{\partial x_0} = (n-2)x_0^{n-3}x_0^{n_2}x_2^2 - a_1x_1^{n-1} - 2a_2x_0x_1^{n-1} - \cdots$$
$$- (n-1)a_{n-1}x_0^{n-2}x_1 - na_nx_0^{n-1} = 0$$
$$\frac{\partial F}{\partial x_1} = -nx_1^{n-1} - (n-1)a_1x_0x_1^{n-2} - \cdots - 2a_{n-2}x_0^{n-2}x_1 - a_{n-1}x_0^{n-1} = 0$$
$$\frac{\partial F}{\partial x_2} = 2x_0^{n-2}x_2 = 0.$$

If $b = (b_0, b_1, b_2)$ is a solution of this system of equations, then $F_{x_2}(b) = 0$ implies $b_0 = 0$ or $b_2 = 0$. In the case where $b_0 = 0$, $F_{x_1}(b) = 0$ implies $b_1 = 0$, and hence $(b_0 : b_1 : b_2) = (0 : 0 : 1)$. If $n \geq 4$, then $F_{x_0}(0 : 0 : 1) = 0$, so that $(0 : 0 : 1)$ is certainly a singular point. But in the case where $n = 3$ we have $F_{x_0}(0 : 0 : 1) = 1$, so that $(0 : 0 : 1)$ is a nonsingular point. (Here we wrote $\partial F/\partial x_0 = F_{x_0}$, etc., as we shall do for convenience in the following.)

In the case where $b_0 \neq 0$ and $b_2 = 0$, $F(b) = 0$ implies that b_1/b_0 is a solution of (3.2), and thus $(b_0 : b_1) = (1 : \alpha_i)$. Since we assume that (3.2) has no double root, we get $F_{x_1}(b) \neq 0$.

From the observation above, we find that C is a nonsingular cubic curve for $n = 3$ and that, for $n \geq 4$, the point at infinity $P_\infty = (0 : 0 : 1)$ is a singular point. At this point, we find a local parameter as follows. Set

$$u = \frac{x_0}{x_2}, \quad v = \frac{x_1}{x_2},$$
$$g(u, v) = \frac{1}{x_2^n}F(x_0, x_1, x_2).$$

Then

$$g(u,v) = u^{n-2} - (v^n + a_1uv^{n-1} + a_2u^2v^{n-2} + \cdots + a_{n-1}u^{n-1}v + a_nu^n)$$
$$= u^{n-2} - \prod_{j=1}^{n}(v - \alpha_j u).$$

Regarding

$$\prod_{j=1}^{n}(v - \alpha_j u) - u^{n-2} = 0$$

as an equation of degree n for v, we find that it has a solution of the form

$$v = u(u^{-2/n} - \frac{a_1}{n} + \beta_1 u^{2/n} + \beta_2 u^{4/n} + \cdots) \tag{3.3}$$
$$= uh(u^{2/n}),$$

where

$$h(z) = z^{-1} - \frac{a_1}{n} + \beta_1 z + \beta_2 z^2 + \cdots.$$

Therefore, writing

$$\omega = e^{2\pi i/n}, \quad i = \sqrt{-1},$$

we find that for odd n we have a factorization

$$g(u, v) = -\prod_{j=1}^{n}(v - uh(\omega^{2j}u^{2/n})). \tag{3.4}$$

For even n, the equation $g(u, v) = 0$ has a root of the form

$$v = -u\tilde{h}(u^{2/n}),$$

where

$$\tilde{h}(z) = z^{-1} + a_1/n + \cdots.$$

If $n = 2g + 1$, (3.4) shows that the curve C has just one branch at the point at infinity P_∞, and can be parametrized by a local parameter s in the form

$$\begin{cases} u = s^n \\ v = s^n h(s^2). \end{cases} \tag{3.5}$$

If $n = 2g + 2$, we set

$$h_+(z) = h(z), \quad h_-(z) = \tilde{h}(z)$$

and find that $-g(u, v)$ can be decomposed as the product of two power series in u, v:

$$\prod_{j=1}^{g+1}(v - uh_+(\omega^{2j}u^{1/(g+1)})) \quad \text{and} \quad \prod_{j=1}^{g+1}(v - uh_-(\omega^{2j}u^{1/(g+1)})),$$

thus resulting in two branches, which can be represented by a local parameter t in the form

$$\begin{cases} u = t^{g+1} \\ v = \pm t^{g+1}h_\pm(t). \end{cases} \tag{3.6}$$

It also follows that the curve $\tilde{C}$ arising when the singularity of C is resolved has two points $P_\infty^{(+)}, P_\infty^{(-)}$ corresponding to the two branches at P_∞.

We shall also find a parametric representation at a point $(1 : a : b)$ other than the point at infinity P_∞. We have

$$b^2 = a^n + a_1a^{n-1} + \cdots + a_{n-1}a + a_n.$$

If $b = 0$, then $a = \alpha_j, 1 \leq j \leq n$. We set

$$Q_j = (1 : \alpha_j : 0).$$

By further setting

$$x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}$$

and

$$h(x, y) = \frac{1}{x_0^n}F(x_0, x_1, x_2) = y^2 - f(x)$$
$$f(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n,$$

we have at the point $(\alpha_j, 0)$

$$h_x(\alpha_j, 0) = -f'(\alpha_j) \neq 0.$$

The implicit function theorem says that $h(x, y)$ can be parametrized in a neighborhood of $(\alpha_j, 0)$ in the form

$$x = \alpha_j + c_2y^2 + c_3y^3 + \cdots, \tag{3.7}$$

where $t = y$ is a local parameter. (Note that

$$y^2 = (x - \alpha_j)\prod_{k\neq j}(x - \alpha_k)$$

implies that there is no linear term in y and $c_2 \neq 0$.)

On the other hand, if $b \neq 0$ at the point $(1 : a : b)$, then

$$h_y(a, b) = 2b \neq 0.$$

By the implicit function theorem again, we obtain a parametric representation

$$y = b + d_1(x - a) + d_2(x - a)^2 + d_3(x - a)^3 + \cdots, \tag{3.8}$$

where $t = x - a$ is a local parameter.

With these preparations we are now in a position to study the algebraic curve $\tilde{C}$ obtained by resolving the singularity of C. The curve $\tilde{C}$ is called an **elliptic curve** for $n = 3, 4$ and a **hyperelliptic curve** for $n \geq 5$. We see that $x = x_1/x_0$ defines a rational function on $\tilde{C}$. The zeros of x are $(1 : 0 : \pm\sqrt{f(0)})$; there are two zeros unless $\alpha_j = 0$. On the other hand, $x = u/v$ at P_∞. For $n = 2g + 1$, (3.3) and (3.5) give rise to

$$x = \frac{v}{u} = s^{-2} + \cdots,$$

which has a pole of order 2 at the point $\tilde{P}_\infty$ in $\tilde{C}$ corresponding to P_∞. For $n = 2g + 2$, (3.3) and (3.6) give rise to

$$x = \frac{v}{u} = \pm t^{-1},$$

which has poles of order 1 at $P_\infty^{(+)}$ and at $P_\infty^{(-)}$. Therefore we get

$$(x)_0 = \begin{cases} (1 : 0 : \sqrt{f(0)}) + (1 : 0 : -\sqrt{f(0)}) \text{ if no } \alpha_j = 0, \\ 2Q_j \text{ if } \alpha_j = 0, \end{cases}$$

$$(x)_\infty = \begin{cases} 2\tilde{P}_\infty \text{ when } n = 2g + 1, \\ P_\infty^{+} + P_\infty^{(-)} \text{ when } n = 2g + 2. \end{cases}$$

Hence we get $\deg(x) = 0$. Next, we shall study the rational function y. It has zeros at the points $Q_j = (1 : 0 : \alpha_j)$, $1 \leq j \leq n$, and admits y as a local parameter at Q_j. Thus y has a zero of order 1 at Q_j. On the other hand, at the point at infinity we have

$$y = \frac{1}{u} = \begin{cases} \dfrac{1}{s^n} \text{ when } n = 2g + 1, \\ \pm\dfrac{1}{t^{g+1}} \text{ when } n = 2g + 2. \end{cases}$$

For $n = 2g + 1$, $\tilde{P}_\infty$ is a pole of order n; for $n = 2g + 2$, $P_\infty^{(+)}$ and $P_\infty^{(-)}$ are poles of order $n/2$. To sum up, we have

$$(y)_0 = Q_1 + Q_2 + \cdots + Q_n$$

$$(y)_\infty = \begin{cases} n\tilde{P}_\infty \text{ when } n = 2g + 1, \\ \dfrac{n}{2}P_\infty^{(+)} + \dfrac{n}{2}P_\infty^{(-)} \text{ when } n = 2g + 2, \end{cases}$$

and thus $\deg(y) = 0$.

By a **positive divisor** we mean a divisor D that can be written in the form $m_1P_1 + \cdots + m_kP_k$, where $m_i \geq 1, 1 \leq i \leq k$. In this case, we write $D > 0$ and, obviously, $\deg D > 0$. If $D = 0$ or $D > 0$, we write $D \geq 0$. In this case $\deg D \geq 0$.

We can simplify the definition of $\mathbf{L}(D)$ in (3.1) by making use of the notion of positive divisor, that is, we have

$$\mathbf{L}(D) = \{f \in \mathbf{C}(C) | f = 0 \text{ or } (f) + D \geq 0\}, \tag{3.9}$$

as can be easily verified by the following arguments. Suppose a divisor $D = \sum_{i=1}^{k} m_iP_i - \sum_{j=1}^{\ell} n_jQ_j$, where $m_i \geq 1, n_j \geq 1$, is given. Let f be a function in $\mathbf{C}(C)$ that has poles of order $\hat{m}_i$ at P_i and zeros of order $\hat{n}_j$ at Q_j and is elsewhere regular. Then we get

$$(f) = \sum_{i=1}^{k}(-\hat{m}_i)P_i + \sum_{j=1}^{\ell}(\hat{n}_j)Q_j$$

$$(f) + D = \sum_{i=1}^{k}(m_i - \hat{m}_i) + \sum_{j=1}^{\ell}(\hat{n}_j - n_j)Q_j$$

Then we have $(f) + D \geq 0$ if $m_i \geq \hat{m}_i$ with inequality for at least one i, and $\hat{n}_j \geq n_j$ with inequality for at least one j. Of course, $(f) + D = 0$ if $m_i = \hat{m}_i$ and $n_j = \hat{n}_j$ for all i, j. To complete the argument, we have only to note that if $(f) + D \geq 0$ for some $f \in \mathbf{C}(C)$, then f has no poles other than P_i's and no zeros other than Q_j's, and $m_i \geq \hat{m}_i$ and $\hat{n}_j \geq n_j$ for all i, j.

Suppose a divisor D and a function $f \in \mathbf{C}(C)$ satisfy $(f) + D \geq 0$. Then we have

$$0 \leq \deg((f) + D) = \deg(f) + \deg D = \deg D$$

because $\deg(f) = 0$ by Lemma 3.2. Therefore, we get

COROLLARY 3.1. *If a divisor D has $\deg D < 0$, then $\mathbb{L}(D) = \{0\}$.*

We have

DEFINITION 3.1. If two divisors D_1 and D_2 on an algebraic curve C are related by

$$D_1 = D_2 + (f)$$

for some rational function f, we say that they are **linearly equivalent** and write $D_1 \sim D_2$.

From Lemma 3.2 we get

COROLLARY 3.2. *If $D_1 \sim D_2$, then $\deg D_1 = \deg D_2$.*

LEMMA 3.3. *If $D_1 \sim D_2$, then the vector spaces $\mathbf{L}(D_1)$ and $\mathbf{L}(D_2)$ are isomorphic by the mapping*

$$\psi : h \in \mathbf{L}(D_1) \longmapsto hf \in \mathbf{L}(D_2).$$

PROOF. If $h \in \mathbf{L}(D_1)$ and $f \neq 0$, then

$$(h) + D_1 > 0, \quad \text{and} \quad (hf) + D_2 = (h) + (f) + D_2 = (h) + D_1 > 0,$$

which implies $(hf) + D_2 \in \mathbf{L}(D_2)$. Thus ψ is a mapping from $\mathbf{L}(D_1)$ into $\mathbf{L}(D_2)$.

Conversely, if $\tilde{h} \in \mathbf{L}(D_2)$ and $\tilde{h} \neq 0$, then

$$(\tilde{h}/f) + D_1 = (\tilde{h}) - (f) + D_1 = (\tilde{h}) + D_2 > 0,$$

showing that $\tilde{h}/f \in \mathbf{L}(D_1)$. Thus $\tilde{h} \to \tilde{h}/f$ is the inverse mapping of ψ.

(b) Differential forms and the genus of algebraic curves. We need several preparations to develop the theory of differential forms. Here we shall explain them mostly from the standpoint of convenience. We introduce the symbol ds for a local parameter s. For another parameter t such that

$$s = \phi(t)$$

we assume that we have the relation

$$ds = \phi'(t)dt.$$

Also for a rational (or meromorphic) function $\psi(s)$ of s, we mean

$$d\psi(s) = \psi'(s)ds.$$

For f, g in the function field $\mathbf{C}(C)$ of an algebraic curve C, fdg means

$$\alpha(t)d\beta(t) = \alpha(t)\beta'(t)dt, \tag{3.10}$$

where $f = \alpha(t)$, $g = \beta(t)$ with a local parameter t at a point P on C. Thus interpreted, fdg is called a **rational differential form** on C. If $\alpha(t)\beta'(t)$ has a zero (or pole) of order m at $t = 0$, we say that the rational differential form dg has a zero (or pole) of order m at P. A rational differential form is called a **regular differential form** if it has no poles. To any rational differential form $\omega = fdg$ we may associate a divisor, just as in the case of a rational function. If ω has zeros of order $m_1, m_2, \ldots, m_k$ at points $Q_1, Q_2, \ldots, Q_k$, respectively, poles of order $n_1, n_2, \ldots, n_\ell$ at points $P_1, P_2, \ldots, P_\ell$, respectively, and no other zero or pole, we define

$$\begin{aligned}(\omega)_0 &= m_1Q_1 + m_2Q_2 + \cdots + m_kQ_k\\ (\omega)_\infty &= n_1P_1 + n_2P_2 + \cdots + n_\ell P_\ell\\ (\omega) &= (\omega_0) - (\omega)_\infty\end{aligned}$$

and call (ω) the **canonical divisor**. If ω is a regular differential form, then its canonical divisor (ω) is a positive divisor.

LEMMA 3.4. *Let ω_1 and ω_2 be rational differential forms of an algebraic curve C. Then the corresponding canonical divisors are linearly equivalent.*

We shall outline the proof. Let $\omega_1 = f_1dg_1, \omega_2 = f_2dg_2$. Then

$$(\omega_1) = (f_1) + (dg_1), \quad (\omega_2) = (f_2) + (dg_2).$$

It suffices to show that there is a rational function h such that $dg_1 = hdg_2$. This follows naturally by developing the theory of differential forms, but here we only show that h is independent of the choice of a local parameter at P on C. Suppose s and t are two local parameters such that $s = \phi(t)$. Suppose the representations of g_j relative to s and t are $\alpha_j(s)$ and $\beta_j(t)$, respectively. Then we get

$$\beta_j(t) = \alpha_j(\phi(t))$$

and hence

$$dg_j = \beta_j'(t)dt = \alpha_j'(\phi(t))\phi'(t)dt,$$

from which we have

$$\frac{\beta_1'(t)}{\beta_2'(t)} = \frac{\alpha_1'(\phi(t))\phi'(t)}{\alpha_2'(\phi(t))\phi'(t)} = \frac{\alpha_1'(s)}{\alpha_2'(s)}.$$

We see from this that in a neighborhood of the point P the function h is determined uniquely. We have yet to show that h is a rational function on C, but we skip the proof.

EXAMPLE 3.3. If we set $x = x_1/x_0$ for the homogeneous coordinates $(x_0 : x_1)$ of $\mathbf{P}^1(\mathbf{C})$, then $\mathbf{C}(\mathbf{P}^1) = \mathbf{C}(x)$, as we recall from Lemma 2.3. For the point $(1 : a)$ we may take $t = x - a$ as a local parameter and for $(0 : 1)$ we take $s = 1/x$ as a local parameter. Thus we get

$$dx = \begin{cases} dt & \text{at } (1 : a)\\ -\dfrac{ds}{s^2} & \text{at } P_\infty = (0, 1),\end{cases}$$

which implies

$$(dx) = -2P_\infty.$$

For two elements $f(x)$ and $g(x)$ in $\mathbf{C}(\mathbf{P}^1)$ we have

$$\begin{aligned}\omega &= f(x)dg(x) = f(x)g'(x)dx\\ (\omega) &= (f(x)g'(x)) + (dx),\end{aligned}$$

which implies

$$(\omega) \sim (dx) \quad \text{and} \quad \deg(\omega) = \deg(dx) = -2.$$

EXAMPLE 3.4. Let us consider an elliptic curve and a hyperelliptic curve $\tilde{C}$ in Example 3.2. By setting $x = x_1/x_0, y = x_2/x_0$, we have

$$\begin{aligned}\mathbf{C}(\tilde{C}) = \mathbf{C}(C) &= \{\alpha(x) + \beta(x)y | \alpha(x), \beta(x) \in \mathbf{C}(x)\}\\ y^2 = f(x) &= x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n,\end{aligned}$$

as we saw in Example 2.15 and Example 2.16. Let us examine $\omega = dx/y$. Near the point at infinity we use

$$x = \frac{v}{u}, \quad y = \frac{1}{u},$$

(3.5), (3.6), and the local parameters s, t for $\tilde{P}_\infty$ (when $n = 2g + 1$) or $P_\infty^{(+)}$, $P_\infty^{(-)}$ (when $n = 2g + 2$) and obtain

$$\frac{dx}{y} = \begin{cases} 2s^{n+1}h'(s^2)ds = 2s^{n+1}(-\dfrac{1}{s^4} + \beta_1 + \ldots)ds, & \text{when } n = 2g + 1,\\ \pm t^{g+1}h_\pm'(t)dt = \pm t^{g+1}(-\dfrac{1}{t^2} + \beta_1 + \ldots)dt, & \text{when } n = 2g + 2.\end{cases}$$

Furthermore, at the points where $y = 0$, namely, $Q_j = (1 : \alpha_j : 0)$, we use the local parameter y and get from (3.7)

$$\frac{dx}{y} = (2c_2 + 3c_3y + \cdots)dy, \quad c_2 \neq 0,$$

which shows that dx/y has neither a zero nor a pole at Q_j. At other points $(1 : a : b)$, $b \neq 0$, we use the local parameter $t = x - a$. Then (3.8) gives rise to

$$\frac{dx}{y} = \frac{dt}{b + d_1t + d_2t^2 + \cdots},$$

showing that dx/y has neither a zero nor a pole at $t = 0$. Hence we obtain

$$\left(\frac{dx}{y}\right) = \begin{cases} (2g-2)\tilde{P}_\infty & \text{when} \quad n = 2g+1, \\ (g-1)P_\infty^{(+)} + (g-1)P_\infty^{(-)} & \text{when} \quad n = 2g+2, \end{cases}$$

and hence

$$\deg\left(\frac{dx}{y}\right) = 2g-2.$$

Likewise, the result on $x^k dx/y$ is as follows. From (3.4), (3.5), (3.6) we find that in $\tilde{P}_\infty$, $P_\infty^{(+)}$, $P_\infty^{(-)}$

$$\frac{x^k dx}{y} = \frac{v^k d(\frac{v}{u})}{u^{k-1}} = \begin{cases} 2s^{n+1}h(s^2)^k h'(s^2)ds \text{ when } n = 2g+1 \\ \pm t^{g+1}h_\pm(t)^k h_\pm{}'(t)dt \text{ when } n = 2g+2 \end{cases}$$

with the expansions in (s,t) as follows:

$$\frac{x^k dx}{y} = \begin{cases} 2s^{n+1}\left(\dfrac{1}{s^2} + \dfrac{a_1}{n} + \cdots\right)^k \left(-\dfrac{1}{s^4} + \cdots\right) ds & \text{when } n = 2g+1 \\ \pm\, t^{g+1}\left(\dfrac{1}{t} \mp \dfrac{a_1}{n} + \cdots\right)^k \left(-\dfrac{1}{t^2} + \cdots\right) dt & \text{when } n = 2g+2. \end{cases}$$

As with $\frac{dx}{y}$, $\frac{x^k dx}{y}$ has no other pole. It has a zero of order k at

$$P_0^{(\pm)} = (1 : 0 : \pm\sqrt{f(0)})$$

if $\alpha_j \neq 0$, $1 \leq j \leq n$, that is, if $f(0) \neq 0$. If $\alpha_j = 0$, then using the local parameter y for $Q_j = (1:0:0)$ we get from (3.7)

$$\frac{x^k dx}{y} = (2c_2 + 3c_3 y + \cdots)(c_2 y^2 + c_3 y^3 + \cdots)^k dy,$$

so that it has a zero of order $2k$ at Q_j. Therefore, when $n = 2g+1$, we conclude that

$$\left(\frac{x^k dx}{y}\right) = \begin{cases} 2(g-k-1)\tilde{P}_\infty + kP_0^{(\pm)} + kP_0^{(-)} & \text{if } f(0) \neq 0, \\ 2(g-k-1)\tilde{P}_\infty + 2kQ_j & \text{if } \alpha_j = 0, \end{cases}$$

and, when $n = 2g+2$,

$$\left(\frac{x^k dx}{y}\right) = \begin{cases} (g-k-1)P_\infty^{(+)} + (g-k-1)P_\infty^{(-)} + kP_0^{(+)} + kP_0^{(-)} & \text{if} \quad f(0) \neq 0, \\ (g-k-1)P_\infty^{(+)} + (g-k-1)P_\infty^{(-)} + 2kQ_j & \text{if} \quad \alpha_j = 0. \end{cases}$$

It follows that if $k \leq g-1$, then $x^k dx/y$ has no pole on $\tilde{C}$ and is a regular differential form. We have also

$$\deg\left(\frac{x^k dx}{y}\right) = 2g-2 \quad \text{for every} \quad k.$$

Now let K_C be the canonical divisor of an algebraic curve C. Thus there exists a rational differential form ω such that $K_C = (\omega)$. Consider the vector space $\mathbf{L}(K_C)$. For $f \in K_C$ we have by (3.9)

$$(f) + K_C = (f) + (\omega) = (f\omega) \geq 0,$$

which implies that ω is a regular differential form. Conversely, if τ is a regular differential form on C, Lemma 3.4 says that we can write $\tau = f\omega$ and thus $f \in \mathbf{L}(K_C)$. Hence we have

LEMMA 3.5. *For the canonical divisor K_C of an algebraic curve C the vector space $\mathbf{L}(K_C)$ is isomorphic to the vector space formed by all regular differential forms on C.*

DEFINITION 3.2. $g(C) = \dim_{\mathbf{C}} \mathbf{L}(K_C)$ is called the **genus** of the algebraic curve C.

For $C = \mathbf{P}^1(\mathbf{C})$, Example 3.3 implies that $\deg K_C = -2$. Hence by Corollary 3.1 we get

$$\dim_{\mathbf{C}} \mathbf{L}(K_C) = 0,$$

that is, the genus of the projective line is 0. In the next subsection and in §3.3 (a) we shall explain how to compute the genus of algebraic curves.

(c) The Riemann-Roch theorem. For a given divisor of an algebraic curve C we set

$$\ell(D) = \dim_{\mathbf{C}} \mathbf{L}(D).$$

By Lemma 3.1, whose proof was omitted, $\ell(D)$ has finite value. It is generally difficult to compute $\ell(D)$. However for the canonical divisor K_C of C it is possible to compute $\ell(D) - \ell(K_C - D)$.

THEOREM 3.1 (THE RIEMANN-ROCH THEOREM). *For a divisor D of an algebraic curve C of genus g we have*

$$\ell(D) - \ell(K_C - D) = \deg D - g + 1.$$

We unfortunately have to omit the proof of this theorem. From $\ell(K_C - D) \geq 0$ we get

COROLLARY 3.3 (RIEMANN'S INEQUALITY). *For a divisor on a curve of genus g, we have*

$$\ell(D) \geq \deg D - g + 1.$$

Since $\ell(D) \geq 0$, the inequality makes sense only when the right-hand side is positive.

Now there is something that should have been stated in (a) of this section. $\mathbf{L}(0)$ consists of all rational functions f that have no pole on C. From $\deg(f) = 0$, it follows that f has no zero either. It is known that such a rational function has to be a constant function. (We can prove this by using Theorem 2.6. Another method is the following. From the definition of rational function in §2.4 (d), unless the restriction of

$$f = \frac{G(x_0, x_1, \ldots, x_n)}{H(x_0, x_1, \ldots, x_n)}$$

to $C \subset \mathbf{P}^N(\mathbf{C})$ is constant, the zeros of f appear from the intersection of $G = 0$ and C, and the poles appear from the intersection of $H = 0$ and C, although it is necessary to show that the latter intersection is not empty.)

Hence

$$\dim_{\mathbf{C}} \mathbf{L}(0) = 1.$$

We also have by definition

$$\dim_{\mathbf{C}} \mathbf{L}(K_C) = g(C).$$

By setting $D = K_C$ in the Riemann-Roch theorem we get the following result.

COROLLARY 3.4. $\deg K_C = 2g(C) - 2$.

In certain cases, we can use this formula and compute the genus of an algebraic curve C. Take the curve treated in Examples 3.2 and 3.4, which is the algebraic curve $\tilde{C}$ determined by the affine curve $y^2 = f(x)$, where $f(x)$ has no multiple roots. Write the degree n of $f(x)$ as $n = 2g+1$ or $n = 2g+2$. By computations done in Example 3.4, we have

$$\deg K_{\tilde{C}} = 2g - 2.$$

Hence the genus of $\tilde{C}$ is g, and as a basis of the vector space of all regular regular differential forms we may take

$$\frac{dx}{y}, \frac{xdx}{y} \cdots, \frac{x^{g-1}dx}{y}$$

Another application of Corollary 3.4 is the following.

LEMMA 3.6. *The genus $g(C)$ of a nonsingular plane curve of degree n is given by*

$$g(C) = \frac{1}{2}(n-1)(n-2).$$

PROOF. Let

$$F(x_0, x_1, x_2) = 0$$

be the defining equation of C and set

$$x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}$$
$$f(x, y) = \frac{1}{x_0^n} F(x_0, x_1, x_2).$$

Since the point $(1 : a : b)$ of C is nonsingular, we have $f_x(a, b) \neq 0$ or $f_y(a, b) \neq 0$. Denote the restrictions of x, y to C also by x, y and consider them as rational functions on C. From $f(x, y) = 0$ we get

$$0 = df(x, y) = f_x(x, y) + f_y(x, y)dy,$$

and hence

$$\frac{dx}{f_y(x, y)} = -\frac{dy}{f_x(x, y)};$$

we denote this rational differential form by ω. Since $f_y(a, b) \neq 0$ or $f_x(a, b) \neq 0$, we see that ω is regular on $\{x_0 \neq 0\} \cap C$. Take a point where $x_0 = 0$, say, $(0 : a : b)$ with $a \neq 0$, that is, at $(0 : 1 : c)$. Set

$$u = \frac{x_0}{x_1}, \quad v = \frac{x_2}{x_1}$$
$$g(u, v) = \frac{1}{x_1^n} F(x_0, x_1, x_2).$$

Denoting the restrictions of u, v to C by the same letters, we have

$$g(u, v) = 0$$
$$u = \frac{1}{x}, \quad v = \frac{1}{y}$$
$$f(x, y) = \frac{1}{x_0^n} F(x_0, x_1, x_2) = \frac{x_1^n}{x_0^n} \frac{1}{x_1^n} F(x_0, x_1, x_2)$$
$$= \frac{1}{u^n} g(u, v)$$
$$f_y(x, y) = \frac{1}{u^n} g_v(u, v) \frac{1}{x} = \frac{1}{u^{n-1}} g_v(u, v)$$

and

$$\omega = \frac{dx}{f_y(x, y)} = \frac{d\left(\frac{1}{u}\right)}{\frac{1}{u^{n-1}} g_v(u, v)} = -\frac{u^{n-3} du}{g_v(u, v)},$$

which can be written as

$$\omega = \frac{u^{n-3} dv}{g_u(u, v)}$$

by using $g_u du + g_v dv = 0$. Since the point $(0 : 1 : c)$ corresponds to $(u, v) = (0, c)$, ω has a zero of order $n-3$ at $(0 : 1 : c)$. Similarly, we see that at the point $(0 : d : 1)$ it has a zero of order $n - 3$. (The arguments above assume that the line $x_0 = 0$ and C are not tangent. If they are tangent, both u and $v - c$ are local parameters for $(0 : 1 : c)$ We leave the discussions in this case to the reader. As a matter of fact, by performing a projective transformation if necessary, we may assume that the line $x_0 = 0$ and C are not tangent.) Now $x_0 = 0$ and C intersect at n points. Hence

$$\deg(\omega) = n(n-3).$$

This being equal to $2g(C) - 2$, we conclude the proof.

By using the Riemann-Roch theorem and Corollaries 3.2 and 3.4, we obtain an important result.

LEMMA 3.7. *If $\deg D > 2g(C) - 2$, then*

$$\ell(D) = \deg D - g(C) + 1.$$

Lemma 3.7 has many applications, one of which will be given in §3.2.

§ 3.2. Geometry of algebraic curves

(a) The Hurwitz formula. Given a rational function f on an algebraic curve C, consider the mapping

$$\phi : P \in C \longmapsto (1 : f(P)) \in \mathbf{P}^1(\mathbf{C}). \tag{3.11}$$

If P is a pole of f, (3.11) does not make sense, but we can define ϕ by

$$P \in C \longmapsto \left(\frac{1}{f(P)} : 1\right).$$

The mapping so defined is an algebraic morphism. In the following we use the term "**regular mapping**" instead of algebraic morphism. We had the case where

$C = \mathbf{P}^1(\mathbf{C})$ in §2.2 (e). Unless f is a constant function, ϕ is surjective, that is, $\phi(C) = \mathbf{P}^1(\mathbf{C})$.

Conversely, given an algebraic morphism

$$\psi : C \longrightarrow \mathbf{P}^1(\mathbf{C}),$$

we take the rational function x obtained from the homogeneous coordinates $(x_0 : x_1)$ in $\mathbf{P}^1(\mathbf{C})$ and set

$$f(P) = x(\psi(P)).$$

Then f is a rational function on C. It has a pole at $\psi^{-1}((0:1))$. In this way, there is a one-to-one correspondence between the set of all rational functions on C and the set of all regular mappings from C to $\mathbf{P}^1(\mathbf{C})$ through (3.11).

Now we get to more detail about the mapping (3.11) from the curve C to $\mathbf{P}^1(\mathbf{C})$. The inverse image $\phi^{-1}((1:a))$ is nothing but the set of points such that

$$f(P) = 0.$$

For such a point P, represent f relative to a local parameter t as $\alpha(t)$. The mapping ϕ can be expressed in the form

$$x = \alpha(t)$$

from a neighborhood of P to a neighborhood of $(1:a)$. Since $u = x - a$ is a local parameter for the point $(1:a)$, we can write the equation above as

$$u = \beta(t), \quad \beta(t) = \alpha(t) - a.$$

Then $\beta(0) = 0$, so that

$$\beta(t) = t^{e_P}\gamma(t), \quad \gamma(0) \neq 0.$$

Here e_P is a positive integer. If $e_P \geq 2$, we say that ϕ is **ramified** and call P a **ramification point** and e_P the **ramification index** at P. We see that e_P is the multiplicity of the solution P of the equation

$$f(z) - a = 0. \tag{3.12}$$

For the point at infinity $(0:1)$ we define a ramification point and the ramification index by taking $1/f$ instead of f. In analogy to the theory of algebraic equations, the number of solutions counting multiplicities would not vary when a is moved slightly, that is, we would have

$$\sum_{P \in C, f(P)=a} e_P = \sum_{Q \in C, f(Q)=b} e_Q. \tag{3.13}$$

Using the terminology of divisors, we have

$$(f-a)_0 = \sum_{P \in C, f(P)=a} e_P P$$
$$(f-b)_0 = \sum_{Q \in C, f(Q)=b} e_Q Q,$$

so that

$$\deg(f-a)_0 = \deg(f-b)_0$$

and, in particular,

$$\deg(f)_0 = \deg(f-a)_0.$$

Applying similar arguments to $1/f$ instead, we get for $a \neq 0$

$$\deg(1/f)_0 = \deg(1/f - 1/a)_0.$$

Using

$$(1/f - 1/a)_0 = (f-a)_0$$
$$(1/f)_0 = (f)_\infty,$$

we finally get

$$\deg(f) = \deg(f)_0 - \deg(f)_\infty = 0.$$

This gives a geometric meaning to Lemma 3.2. Of course, even though (3.13) is intuitively clear, it calls for a rigorous proof.

The constant

$$w = \sum_{P \in C, f(P)=a} e_P$$

that appears in (3.13) is called the **degree** of the regular mapping ϕ. For general a, it is intuitively clear that the equation $f = a$ has w distinct solutions. As a matter of fact, we can prove that there are only a finite number of ramification points for a regular mapping ϕ. If we know the ramification indices at the ramification points and the degree of ϕ, then we can compute the genus of the algebraic curve. Let ω be a rational differential form on $\mathbf{P}^1(\mathbf{C})$. We can then define the pull-back $\phi^*\omega$ of ω by ϕ, which is a rational differential form on C. Let us compute the degree $\deg(\phi^*\omega)$. Let u be a local parameter for a point Q on $\mathbf{P}^1(\mathbf{C})$ and suppose that ϕ can be represented by using a local parameter t for P on C in the form

$$u = \beta(t), \quad \beta(t) = t^{e_P}\gamma(t), \quad \gamma(0) \neq 0.$$

If

$$\omega = A(u)du,$$

then

$$\phi^*(\omega) = A(\beta(t))\beta'(t)dt = e_P t^{e_P - 1}\left(\gamma(t) + \frac{1}{e_P}t\gamma'(t)\right)A(\beta(t))dt.$$

If $A(u)$ has a zero of order m at u (with the interpretation that if m is a negative integer $-\ell$, this means that $A(u)$ has a pole of order ℓ), then $\phi^*\omega$ has a zero of order $me_P + e_P - 1$. Hence we get from (3.13)

$$\deg(\phi^*\omega) = w\deg(\omega) + \sum_Q (e_Q - 1),$$

where the sum is taken over the ramification points Q. By Example 3.3, we know $\deg(\omega) = -2$. Hence we get the following result.

THEOREM 3.2 (HURWITZ'S THEOREM). *Let $\phi : C \to \mathbf{P}^1(\mathbf{C})$ be the regular mapping in (3.11) determined by a non-constant rational function. If the degree is w and the ramification points are $R_1, R_2, \ldots, R_\ell$ with ramification indices $e_1, e_2, \ldots, e_\ell$, respectively, then with $g(C)$ denoting the genus of C we have*

$$2g(C) - 2 = -2w + \sum_{j=1}^{\ell}(e_j - 1). \tag{3.14}$$

The equality (3.14) is called the **Hurwitz formula**. It can be further generalized as we state in the following.

EXAMPLE 3.5. Let us consider a regular mapping

$$\phi : (a_0 : a_1) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (a_0^n, a_1^n) \in \mathbf{P}^1(\mathbf{C}).$$

This is a regular mapping determined by the rational function $x^n = (x_1/x_0)^n$. To distinguish the two projective lines, we denote by $(x_0 : x_1)$ the homogeneous coordinates of the left $\mathbf{P}^1(\mathbf{C})$ and by $(y_0 : y_1)$ the homogeneous coordinates of the right $\mathbf{P}^1(\mathbf{C})$. As local parameters for the point $(1 : 0)$ we take $x = x_1/x_0, y = y_1/y_0$, respectively, and represent ϕ by

$$y = x^n.$$

Thus $(1 : 0)$ is a ramification point of index n. For the point $(1 : a)$, $a \neq 0$, take b such that $b^n = a$. Using a local parameter $s = x - b$ for $(1 : b)$ and a local parameter $y = y - a$ for $(1 : a)$, we represent ϕ by $t = a + (s + b)^n$, so $(1 : b)$ is not a ramification point. At the point $(0 : 1)$, set $u = x_0/x_1$ and $v = y_0/y_1$, and represent ϕ by

$$v = u^n,$$

which shows that $(0 : 1)$ is a ramification point of index n. Clearly, the degree of ϕ is n. The right-hand side of (3.14) is

$$-2n + 2(n - 1) = -2,$$

so that the Hurwitz formula is certainly valid.

EXAMPLE 3.6. Let us use the Hurwitz formula and compute the genus of an elliptic curve and of a hyperelliptic curve $\tilde{C}$ determined from an affine curve

$$y^2 = f(x)$$

as treated in Example 3.2. We assume that $f(x)$ is a polynomial of degree n without multiple root. Consider the regular mapping from $\tilde{C}$ into $\mathbf{P}^1(\mathbf{C})$

$$\pi : P \in \tilde{C} \longmapsto (1 : x(P)) \in \mathbf{P}^1(\mathbf{C}).$$

In the computations in Example 3.2, we see from (3.7) that $Q_j = (1 : \alpha_j : 0)$, $1 \leq j \leq n$, are ramification points of index 2. Furthermore, if $n = 2g + 1$, $\tilde{P}_\infty$ is a ramification point of index 2 because $(x)_\infty = 2\tilde{P}_\infty$. There are no more ramification points. The degree of ϕ being 2, the Hurwitz formula says

$$2g(\tilde{C}) - 2 = -2 \times 2 + 2g + 2 = 2g - 2,$$

where $n = 2g + 2$ or $2g + 1$. It follows that

$$g(\tilde{C}) = g,$$

which coincides with the result obtained by Corollary 3.4.

Looking back at the proof of Theorem 3.2, we realize that we have not essentially used the fact that we are dealing with $\mathbf{P}^1(\mathbf{C})$. Thus for a regular mapping from an algebraic curve $\tilde{C}$ into an algebraic curve C, we can use local parameters and define the degree of ϕ, the ramification points and indices in just the same way. We obtain the following theorem.

THEOREM 3.3 (THE GENERALIZED HURWITZ THEOREM). *For a regular mapping*

$$\phi : \tilde{C} \longrightarrow C,$$

where $\tilde{C}$ and C are algebraic curves of genus $\tilde{g}$ and g, respectively, we have

$$2\tilde{g} - 2 = w(2g - 2) + \sum_{j=1}^{\ell} (e_j - 1), \tag{3.15}$$

where w is the degree of ϕ , and $e_1, e_2, \ldots, e_\ell$ are the ramification indices at ramification points $R_1, R_2, \ldots, R_\ell$.

(3.15) is also called the Hurwitz formula.

(b) Imbedding into the projective space. For a divisor D on an algebraic curve C, suppose $\dim_{\mathbf{C}} \mathbf{L}(D) = N + 1 \geq 2$. If we take a basis $\{f_0, f_1, \ldots, f_N\}$ in $\mathbf{L}(D)$ over $\mathbf{C}$, then we have a mapping

$$\psi_{|D|} : P \in C \longmapsto (f_0(P) : f_1(P) : \cdots : f_N(P)) \in \mathbf{P}^N(\mathbf{C}). \tag{3.16}$$

If P is a pole of f_j, take f_k that has a maximal order at P among $f_0, f_1, \ldots, f_N$. Then $h_j = f_j/f_k$ is regular at P, and we define

$$\psi_{|D|}(P) = (h_0(P) : \cdots : h_{k-1}(P) : 1 : h_{k+1}(P) : \cdots : h_N(P)).$$

If Q is a common zero of $f_0, f_1, \ldots, f_N$, let f_j have a minimal order of zero at Q and, using a local parameter t at Q, express f_j as $\alpha_j(t)$. We then define in a neighborhood of Q

$$\psi_{|D|} : t \longmapsto \left(\frac{\alpha_0(t)}{t^m} : \frac{\alpha_1(t)}{t^m} : \cdots : \frac{\alpha_N(t)}{t^m} \right).$$

By interpreting (3.16) in this fashion, we see that $\psi_{|D|}$ becomes a regular mapping. It depends on the choice of a basis in $\mathbf{L}(D)$ but differs only by a projective transformation of $\mathbf{P}^N(\mathbf{C})$.

First, consider the case of an algebraic curve of genus g. Since $\deg K_C = -2$ by Corollary 3.4, we have by Lemma 3.7

$$\ell(P) = 2$$

for a certain point P on C. Although $\mathbf{L}(D)$ contains constant functions, $\ell(D) = 2$ means that there is a non-constant rational function f in $\mathbf{L}(D)$. Since $f \notin \mathbf{L}(0)$, f has a pole of order 1 at P and is regular elsewhere. Since $\{1, f\}$ is a basis, we get a regular mapping

$$\psi_{|D|} : Q \in C \longmapsto (1 : f(Q)) \in \mathbf{P}^1(\mathbf{C}).$$

Since f has a pole of order 1 only at one point and is regular elsewhere, we see that $\psi_{|D|}$ has degree 1. As such, there is no ramification point and $\psi_{|D|}$ is an isomorphism. Hence we obtain

PROPOSITION 3.1. *An algebraic curve of genus 0 is (isomorphic to) the projective line $\mathbf{P}^1(\mathbf{C})$.*

The proof above also gives

PROPOSITION 3.2. *If an algebraic curve C admits a rational function that has a pole of order 1 only at one point and is regular elsewhere, then C is (isomorphic to) the projective line $\mathbf{P}(\mathbf{C})$.*

In the following, we identify algebraic curves that are isomorphic. By the results above, algebraic curves of genus 0 are the same as $\mathbf{P}^1(\mathbf{C})$. Take $P = (0:1) \in \mathbf{P}^1(\mathbf{C})$ and consider $\mathbf{L}(nP)$. If we set $x = x_1/x_0$, where $(x_0 : x_1)$ are homogeneous coordinates in $\mathbf{P}^1(\mathbf{C})$, then x has a pole of order 1 at P. Hence $x^j \in \mathbf{L}(nP)$, $0 \le j \le n$, and $\ell(nP) = n+1$. It follows that we can take $\{1, x, x^2, \dots, x^n\}$ as a basis of $\mathbf{L}(nP)$. Thus the regular mapping $\psi_{|D|}$ can be given by

$$\psi_{|D|} : (1:x) \in \mathbf{P}^1(\mathbf{C}) \longmapsto (1 : x : x^2 : \cdots : x^n) \in \mathbf{P}^n(\mathbf{C}).$$

For $n = 2$, this is nothing but the regular mapping ϕ discussed in Example 2.2, and the image is a non-singular quadratic curve. For $n = 3$, it is the regular mapping π treated in Example 2.18; its image is a twisted cubic. For general n, the image of $\psi_{|D|}$ is called a **normal rational curve** of degree n.

By the way, for a divisor D of degree n in $\mathbf{P}^1(\mathbf{C})$, we have by Lemma 3.7

$$\ell(D - nP) = 1, \quad \ell(nP - D) = 1,$$

that is, there exist rational functions f, h such that

$$(f) + D - nP \ge 0$$
$$(h) + nP - D \ge 0.$$

By adding these inequalities we get

$$(fh) = (f) + (h) \ge 0,$$

which shows that fh has no pole and hence it is a constant function. We may multiply by a constant and assume $hf = 1$. Thus we get

$$0 \le (h) + nP - D = -(f) + nP - D$$

and

$$(f) + D - nP \le 0.$$

Together with the first inequality we have

$$(f) + D = nP.$$

We have proved

LEMMA 3.8. *A divisor D of degree n on $\mathbf{P}^1(\mathbf{C})$ is linearly equivalent to nP.*

This lemma and Lemma 3.3 imply that for a curve of genus 0, $\psi_{|D|}$ and $\psi_{|nP|}$ play the same role. By making use of the latter, we can imbed the projective line $\mathbf{P}^1(\mathbf{C})$ in $\mathbf{P}^n(\mathbf{C})$ as a nonsingular curve. This fact can be generalized as follows.

THEOREM 3.4. *Let C be an algebraic curve of genus g and let D be a divisor of degree n on C. If $n \ge 2g+1$, then*

$$\psi_{|D|} : C \longmapsto \mathbf{P}^{n-g}(\mathbf{C})$$

gives an imbedding of C into $\mathbf{P}^{n-g}(\mathbf{C})$.

PROOF. By Lemma 3.7 we have $\ell(D) = n - g + 1$. Take two distinct points P, Q. Again by Lemma 3.7 we get

$$\ell(D - P - Q) = n - g - 1$$
$$\ell(D - P) = n - g$$
$$\ell(D - Q) = n - g.$$

Therefore we have

$$\mathbf{L}(D - P - Q) \subset \mathbf{L}(D - P) \subset \mathbf{L}(D)$$
$$\mathbf{L}(D - P - Q) \subset \mathbf{L}(D - Q) \subset \mathbf{L}(D).$$

We see that there is a rational function $f \in \mathbf{L}(D)$ that has a zero at P but not at Q and a rational function $h_1 \in \mathbf{L}(D)$ that has a zero at Q but not at P. If $\{h_2, h_3, \dots, h_{n-g}\}$ is a basis of $\mathbf{L}(D - P - Q)$, then $\{f, h_1, h_2, h_3, \dots, h_{n-g}\}$ is a basis of $\mathbf{L}(D)$. Relative to this basis, we express $\psi_{|D|}$ by

$$\psi_{|D|} : R \in C \longmapsto (f(R) : h_1(R) : h_2(R) : \cdots : h_{n-g}(R)) \in \mathbf{P}^{n-g}(\mathbf{C}),$$

and find that

$$\psi_{|D|}(P) = (0:1:0:0:\cdots:0)$$
$$\psi_{|D|}(Q) = (1:0:0:0:\cdots:0),$$

which show that $\psi_{|D|}$ is one-to-one.

In the same way, we find that

$$\mathbf{L}(D - 2P) \subset \mathbf{L}(D - P) \subset \mathbf{L}(D),$$

which implies that there exist a rational function $f \in \mathbf{L}(D)$ with no zero at P and a rational function $h_1 \in \mathbf{L}(D)$ with a zero of order 1 at P. If $\{h_2, \dots, h_{n-g}\}$ is a basis of $\mathbf{L}(D - 2P)$, then $\{f, h_1, h_2, \dots, h_{n-g}\}$ is a basis of $\mathbf{L}(D)$. We construct $\psi_{|D|}$ using this basis. Since h_1 has a zero of order 1 at P, it can be taken as a local parameter at P of C. But this is also a local parameter at the point $\psi_{|D|}(P)$ of $\psi_{|D|}(C)$. It follows that C and $\psi_{|D|}(C)$ are isomorphic in neighborhoods of P and $\psi_{|D|}(P)$. We have thus completed the proof of Theorem 3.4.

EXAMPLE 3.7. Let P be a point on a curve C of genus 1. Then the image of

$$\psi_{|3P|} : C \longmapsto \mathbf{P}^2(\mathbf{C})$$

is a nonsingular plane curve of degree 3. As a basis of $\mathbf{L}(3P)$ we have $\{1, f, h\}$, where f has a pole of order 2 at P and h has a pole of order 3 at P, and furthermore we can write

$$\psi_{|3P|} : Q \in C \longmapsto (1 : f(Q) : h(Q)) \in \mathbf{P}^2(\mathbf{C}).$$

The intersection of $\psi_{|3P|}(C)$ and the line $x_0 = 0$ is obtained by solving $1/h = 0$, which has a zero of order 3 at P and no other zero. It follows that the degree of the plane curve is 3.

Next consider the case where a given divisor D equals the canonical divisor K_C. We assume that the genus g of C is at least 2. If $\{f_1, f_2, \dots, f_g\}$ is a basis of $\mathbf{L}(K_C)$ and if $K_C = (\omega)$, then $\{f_1\omega, f_2\omega, \dots, f_g\omega\}$ is a basis for the regular differential forms on C.

LEMMA 3.9. *$\omega_j = f_j\omega$, $1 \le j \le g$, have no common zero.*

PROOF. Suppose Q is a common zero. Then

$$\mathbf{L}(K_C) = \mathbf{L}(K_C - Q).$$

By the Riemann-Roch theorem (Theorem 3.1)

$$\ell(Q) - \ell(K_C - Q) = 2 - g,$$

we get $\ell(Q) = 2$. This means that there is a rational function that has a pole of order 1 at Q and is regular elsewhere. By Proposition 3.2, the genus of C is 0, which is a contradiction. Hence the ω_j's have no common zero.

We now consider conditions under which the **canonical map**

$$\psi_{|K_C|} : C \longrightarrow \mathbf{P}^{g-1}$$

is an imbedding. As can be seen from the proof of Theorem 3.4, it is sufficient if we have $\ell(K_C - P - Q) = g - 2$ for points P, Q on C (including the case where $P = Q$), because then

$$\mathbf{L}(K_C - P) \subset \mathbf{L}(K_C), \quad \mathbf{L}(K_C - Q) \subset \mathbf{L}(K_C)$$

by Lemma 3.9. Since

$$\ell(P + Q) - \ell(K_C - P - Q) = 3 - g$$

by the Riemann-Roch theorem, we see that $\ell(K_C - P - Q) = g - 2$ and $\ell(P + Q) = 1$ are equivalent. Therefore, if $\ell(K_C - P - Q) \neq g - 2$, then $\ell(K_C - P - Q) = g - 1$, in which case we have $\ell(P + Q) = 2$.

If $\ell(P + Q) = 2$, then there exists a rational function f that has a pole of order 1 at P and at Q and is regular elsewhere, and

$$\psi_{|P+Q|} : R \in C \longmapsto (1 : f(R)) \in \mathbf{P}^1(C)$$

is a surjective mapping of degree 2.

DEFINITION 3.3. An algebraic curve C of genus $g \ge 2$ is called a **hyperelliptic curve** if it has a regular mapping of degree 2 onto $\mathbf{P}^1(\mathbf{C})$.

THEOREM 3.5. *If a curve C of genus $g \ge 2$ is not a hyperelliptic curve, then the canonical mapping $\psi_{|K_C|}$ is an imbedding. If C is a hyperelliptic curve, then the image of the canonical mapping $\psi_{|K_C|}$ is a normal rational curve C_0 of degree $g - 1$, and the degree of $\psi_{|K_C|} : C \to C_0$ is 2.*

In fact, all curves C of genus 2 are hyperelliptic curves. This is because $\deg K_C = 2$ implies that $(\omega) = P + Q$ for a regular differential form ω, and hence $\ell(P + Q) = \ell(K_C) = 2$.

Now let

$$\pi : C \longrightarrow \mathbf{P}^1(\mathbf{C})$$

be a mapping of degree 2 and let $R_1, R_2, \ldots, R_\ell$ be the ramification points of π. The ramification indices must be all equal to 2. Hence the Hurwitz formula (3.14) gives

$$2g(C) - 2 = -4 + \ell$$

and hence

$$\ell = 2g(C) + 2.$$

We may assume that $\pi(R_j) = (1 : \alpha_j)$ by composing π with a projective transformation of $\mathbf{P}^1(\mathbf{C})$ if necessary. Then we see that C can be identified with the algebraic curve determined by

$$y^2 = \prod_{j=1}^{2g(C)+2} (x - \alpha_j).$$

We considered this curve in Examples 3.2 and 3.4.

When the curve C is not a hyperelliptic curve, the image of V by the canonical mapping is called a **canonical curve**.

EXAMPLE 3.8. Consider the canonical mapping

$$\psi_{|K_C|} : C \longrightarrow \mathbf{P}^2(\mathbf{C}),$$

where C is a curve of genus 3 which is not a hyperelliptic curve. The canonical curve, the image of this mapping, is a nonsingular plane curve, which is of degree 4 because $\deg K_C = 4$. Conversely, we can show that a nonsingular plane curve of degree 4 is a canonical curve of genus 3 as follows. By using the method of proof for Lemma 3.6 we show that if the affine form of a nonsingular curve of degree 4 is given by $f(x, y) = 0$; then

$$\frac{dx}{f_y}, \frac{xdx}{f_y}, \frac{ydx}{f_y}$$

can be chosen as a basis for the regular differential forms.

§3.3. Elliptic curves

Elliptic curves are the most studied among algebraic curves. Besides, they have applications to criteria of primes, cryptography, coding theory, etc. Because they are so important, we discuss them in more detail. In the following, k will be an algebraically closed field of an arbitrary characteristic. It is known that what we discussed in the case of the complex number field $\mathbf{C}$ is valid for an arbitrary algebraic closed field k.

(a) Curves of genus 1. Let C be a nonsingular projective curve of genus 1, and let us fix a point O on it. By the Riemann-Roch theorem we have

$$\ell(mO) = m$$

for any natural number m. In particular, since $\ell(O) = 1$, we get

$$\mathbf{L}(O) = k \cdot 1,$$

that is, the only rational function on C that has a pole of order at most 1 at O and is regular elsewhere must be a constant function. Since $\ell(2O) = 2$ and $\mathbf{L}(O) \subset \mathbf{L}(2O)$, we see that there is a rational function x on C that has a pole of order 2 at O and is regular elsewhere. Thus we have

$$L(2O) = k \cdot 1 \oplus k \cdot x \quad \text{(direct sum as vector space)}.$$

Since $\ell(3O) = 3$ and $\mathbf{L}(2O) \subset \mathbf{L}(3O)$, there is a rational function y that has a pole of order 3 at O and is regular elsewhere, so that

$$\mathbf{L}(3O) = k \cdot 1 \oplus k \cdot x \oplus k \cdot y.$$

From Example 3.7, we see that the mapping

$$\psi = \psi_{|3O|} : P \in C \longmapsto (1 : x(P) : y(P)) \in \mathbf{P}^2(k)$$

is an imbedding of C into $\mathbf{P}^2(k)$ and its image is a nonsingular plane curve. To find the defining equation of this plane curve, let us examine $\mathbf{L}(4O), \mathbf{L}(5O), \mathbf{L}(6O)$ successively. We have $x^2 \in \mathbf{L}(4O)$ and it has a pole of order 4 at the point O. Since $\ell(4O) = 4$, we have

$$\mathbf{L}(4O) = k \cdot 1 \oplus k \cdot x \oplus k \cdot y \oplus k \cdot x^2.$$

Likewise, $xy \in \mathbf{L}(5O)$, xy has a pole of order 5 at O, and $\ell(5O) = 5$. Hence

$$\mathbf{L}(5O) = k \cdot 1 \oplus k \cdot x \oplus k \cdot y \oplus k \cdot x^2 \oplus k \cdot xy.$$

Furthermore, $x^3 \in \mathbf{L}(6O)$, x^3 has a pole of order 6, and $\ell(6O) = 6$. Therefore

$$\mathbf{L}(6O) = k \cdot 1 \oplus k \cdot x \oplus k \cdot y \oplus k \cdot x^2 \oplus k \cdot xy \oplus k \cdot x^3.$$

But y^2 also has a pole of order 6 at O and is regular elsewhere, so it is in $\mathbf{L}(6O)$. Thus we see that we have a relation

$$y^2 = a_0' x^3 - a_1' xy + a_2' x^2 - a_3' y + a_4' x + a_6',$$

where $a_0', a_1', a_2', a_3', a_4', a_6' \in k$. Now take $\beta \in k$ such that $a_0' = \beta^3$ and write x for βx. Then the equation above can be written in the form

$$y^2 = x^3 - a_1 xy + a_2 x^2 - a_3 y + a_4 x + a_6, \text{ where } a_1, a_2, a_3, a_4, a_6 \in k. \tag{3.17}$$

From the construction above, it is almost obvious that the polynomial

$$f(x, y) = y^2 + a_1 xy + a_3 y - x^3 - a_2 x^2 - a_4 x - a_6$$

is irreducible, because otherwise factoring will produce a linear relation among the basis elements of $\mathbf{L}(5O)$.

By the considerations above we find that the image $\psi(C)$ of C by ψ coincides with the zero set in $\mathbf{P}^2(k)$ of the homogeneous polynomial

$$\begin{aligned} F(x_0, x_1, x_2) = {} & x_0 x_2^2 + a_1 x_0 x_1 x_2 + a_3 x_0^2 x^2 \\ & - x_1^3 - a_2 x_0 x_1^2 - a_4 x_0^2 x_1 - a_6 x_0^3. \end{aligned} \tag{3.18}$$

To sum up, we have the following.

PROPOSITION 3.3. *A nonsingular projective curve C of genus 1 is isomorphic to a plane cubic curve*

$$x_0 x_2^2 + a_1 x_0 x_1 x_2 + a_3 x_0^2 x_2 - x_1^3 - a_2 x_0 x_1^2 - a_4 x_0^2 x_1 - a_6 x_0^3 = 0. \tag{3.19}$$

We already know that a nonsingular plane cubic curve has genus 1, from the preceding section. What the proposition above asserts is that a nonsingular plane cubic curve is isomorphic to the curve (3.19). In fact, we could directly show that a nonsingular plane cubic curve can be transformed to (3.19) by a projective transformation.

If the characteristic of k is other than 2 and 3, we can further simplify (3.19) by a projective transformation. For this, rewrite (3.17) in the form

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6. \tag{3.20}$$

Since the characteristic of k is not 2, we may change (3.20) to

$$\begin{aligned} \left(y + \frac{1}{2}a_1 x + \frac{a_3}{2}\right)^2 = {} & x^3 + \left(a_2 + \frac{1}{4}a_1^2\right) x^2 + \left(a_4 + \frac{a_1}{2}a_3\right) x \\ & + a_6 + \frac{a_3^2}{4}. \end{aligned}$$

Note that we have

$$y + \frac{1}{2}a_1 x + \frac{a_3}{2} \in \mathbf{L}(3O),$$

and this rational function has a pole of order 3 at O and is regular elsewhere. Rewrite y for $y + \frac{1}{2}a_1 x + \frac{a_3}{2}$. Also write a_2, a_4, a_6 for the coefficients of x^2, x, and the constant term on the right-hand side. Then we get

$$y^2 = x^3 + a_2 x^2 + a_4 x + a_6. \tag{3.21}$$

Thus a nonsingular projective line of genus 1 is isomorphic to the plane cubic curve

$$x_0 x_2^2 - x_1^3 - a_2 x_0 x_1^2 - a_4 x_0^2 x_1 - a_6 x_0^3 = 0. \tag{3.22}$$

Since the transition from (3.20) to (3.21) is to replace $y + \frac{1}{2}a_1 x + \frac{a_3}{2}$ by y, the transition from (3.19) to (3.22) is achieved by a projective transformation (although the coefficients a_j may change).

Now if the characteristic of k is not 3 either, then (3.21) can be changed to

$$y^2 = \left(x + \frac{a_2}{3}\right)^3 + \left(a_4 - a_2^2 3\right) x + a_0 - \frac{a_2^3}{27}.$$

By writing x for $x + \frac{a_2}{3} \in \mathbf{L}(2O)$, the equation above changes to

$$y^2 = x^3 + a_4 x + a_6 \tag{3.23}$$

(although with different a_4, a_6). Or by replacing y by $\frac{y}{2}$ we can change (3.23) to

$$y^2 = 4x^3 - g_2 x - g_3 \tag{3.24}$$

(3.24) is called the **Weierstrass canonical form**.

By the observations above, if the characteristic of k is not equal to 2 or 3, a nonsingular algebraic curve of genus 1 is isomorphic to a plane cubic curve

$$x_0 x_2^2 - x_1^3 - a_4 x_0^2 x_1 - a_0 x_0^3 = 0 \tag{3.25}$$

as well as to a plane cubic curve

$$x_0 x_2^2 - 4x_1^3 + g_2 x_0^2 x_1 + g_3 x_0^3. \tag{3.26}$$

Now as the image of the imbedding $\psi : C \hookrightarrow \mathbf{P}^2(k)$, the plane curves (3.19), (3.22), (3.25), and (3.26) are nonsingular curves. This fact implies certain conditions among the a_j or among the q_j. We shall check on this for the plane cubic curve defined by (3.25):

$$F(x_0, x_1, x_2) = x_0 x_2^2 - x_1^3 - a_4 x_0^2 x_1 - a_0 x_0^3 = 0.$$

If there is a singular point, the system of equations

$$\frac{\partial F}{\partial x_0} = x_2^2 - 2a_4x_0x_1 - 3a_6x_0^2 = 0$$
$$\frac{\partial F}{\partial x_1} = -3x_1^2 - a_4x_0^2$$
$$\frac{\partial F}{\partial x_2} = 2x_0x_2$$

has a solution other than $(0,0,0)$. From the last equation we get $x_0 = 0$ or $x_2 = 0$. If $x_0 = 0$, then the second equation gives $x_1 = 0$ and the first equation $x_2 = 0$; we forget about this case. So we may assume $x_0 \neq 0, x_2 = 0$. Hence the system is reduced to

$$2a_4x_1 + 3a_6x_0 = 0$$
$$3x_1^2 + a_4x_0^2 = 0.$$

It follows that a necessary and sufficient condition for the existence of solutions other than $(0,0)$ is

$$4a_4^3 + 27a_6^2 = 0.$$

The quantity

$$\triangle = -16(4a_4^3 + 27a_6^2)$$

is called the **discriminant** of the equation (3.23) or of the plane cubic curve (3.25). Likewise, the discriminant of (3.24) or of (3.26) is defined as

$$\triangle = g_2^3 - 27g_3^2.$$

It follows that a necessary and sufficient condition for the plane cubic curve (3.26) to be nonsingular is

$$\triangle \neq 0.$$

In the following the canonical forms of plane cubic curves, the discriminant $\triangle$, the j**−invariant**, and regular differential forms become important. So we list a summary (with more detail on j-invariant later).

[The case where the characteristic of k is general]

(3.27)

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$
$$x_0x_2^2 + a_1x_0x_1x_2 + a_3x_0^2x_2 - x_1^3 - a_2x_0x_1^2 - a_4x_0^2x_1 - a_6x_0^3 = 0$$
$$b_2 = a_1^2 + 4a_2, \quad b_4 = a_1a_3 + 2a_4, \quad b_6 = a_3^2 + 4a_6$$
$$b_8 = a_1^2a_6 - a_1a_3a_4 + 4a_2a_6 + a_2a_3^2 - a_4^2$$
$$c_4 = b_2^2 - 24b_4, \quad c_6 = -b_2^3 + 36b_2b_4 - 216b_6$$
$$\triangle = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$$
$$j = \frac{c_4^3}{\triangle} = 12^3\frac{c_4^3}{c_4^3 - c_6^2}$$
$$\omega = \frac{dx}{2y + a_1x + a_3} = \frac{dy}{3x^2 + 2a_2x + a_4 - a_1y}.$$

[The case where the characteristic of k is not 2]

$$y^2 = x^3 + a_2x^2 + a_4x + a^6$$
$$x_0x_2^2 - x_1^3 - a_2x_0x_1^2 - a^4x_0^2x_1 - a_6x_0^3 = 0$$
$$\triangle = a_2^2a_4^2 - a_2^3a_6 - a_4^3$$
$$j = \frac{c_4^3}{\triangle}$$
$$\omega = \frac{dx}{2y} = \frac{dy}{3x^2 + 2a_2x + a_4}.$$

[The case where the characteristic is not 2, 3]

$$y^2 = x^3 + a_4x + a_6$$
$$x_0x_2^2 - x_1^3 - a_4x_0^2x_1 - a_6x_0^3$$
$$c_4 = -48a_4, \quad c_6 = -864a_6$$
$$\triangle = -16(4a_4^3 + 27a_6^2)$$
$$j = \frac{c_4^3}{\triangle} = 12^3\frac{c_4^3}{c_4^3 - c_6^2} = 12^3\frac{4a_4^3}{4a_4^3 + 27a_6^2}$$
$$\omega = \frac{dx}{2y} = \frac{dy}{3x^2 + a_4}.$$

In this case, we can take the Weierstrass canonical form as follows.

$$y^2 = 4x^3 - g_2x - g_3$$
$$x_0x_2^2 - 4x_1^3 + g_2x_0^2x_1 + g_3x_0^3 = 0$$
$$\triangle = g_2^3 - 27g_3^2$$
$$j = 12^3\frac{g_2^3}{\triangle}.$$

With these preparations we can get the following theorem.

THEOREM 3.6. *Let O be a point on a nonsingular algebraic curve C of genus 1. Then the mapping defined by using $\mathbf{L}(3O)$*

$$\psi = \psi_{|3O|} : C \longrightarrow \mathbf{P}^2(k)$$

is an imbedding and gives an isomorphism of C to a nonsingular plane cubic curve. Furthermore, by choosing an appropriate basis in $\mathbf{L}(3O)$ the plane cubic curve is of the form

$$x_0x_2^2 + a_1x_0x_1x_2 + a_3x_0^2x_2 - x_1^3 - a_2x_0x_1^2 - a_4x_0^2x_2 - a_6x_0^3 = 0, \tag{3.28}$$

in which case $\psi(O) = (0:0:1)$ and $\triangle \neq 0$.

When the characteristic of k is not 2, ψ gives an isomorphism of C to the plane cubic curve

$$x_0x_2^2 - x_1^3 - a_2x_0x_1^2 - a_4x_0^2x_1 - a_6x_0^3 = 0. \tag{3.29}$$

In this case we have $\psi(O) = (0:0:1)$ and $\triangle \neq 0$.

If furthermore the characteristic of k is not 3 either, C is isomorphic to the plane cubic curve

$$x_0x_2^2 - x_1^3 - a_4x_0^2x_1 - a_6x_0^3 = 0, \tag{3.30}$$

and $\psi(O) = (0:0:1)$ and $\triangle \neq 0$. By ψ, C is also isomorphic to the plane cubic curve

$$x_0x_2^2 - 4x_1^3 + g_2x_0^2x_1 + g_3x_0^3 = 0, \tag{3.31}$$

and $\psi(O) = (0:0:1)$ and $\triangle \neq 0$.

DEFINITION 3.4. By an **elliptic curve** we mean a pair $E = (C, O)$, where C is a nonsingular algebraic curve of genus 1 and O is a point on C. An elliptic curve (C, O) is said to be **defined over a subfield** k_0 of k if all the coefficients a_j of the plane cubic curve (3.28) belong to k_0. We also say that k_0 is a **field of definition** of (C, O). Two elliptic curves $E = (C, O)$ and $E' = (C', O')$ are said to be **isomorphic** if there is an isomorphism $\phi : C \simeq C'$ such that $\phi(O) = O'$.

By the preceding theorem we see that an elliptic curve is identified with a nonsingular plane cubic curve defined by (3.28) together with a point $(0:0:1)$ (point at infinity). If the characteristic of k is not 2, then consider the nonsingular plane cubic curve (3.29) and the point at infinity $(0:0:1)$; if the characteristic of k is neither 2 nor 3, then consider the nonsingular plane cubic curve (3.30) or (3.31) together with the point at infinity. Since these plane curves have only $(0:0:1)$ as point at infinity, we often call

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$
$$y^2 = x^3 + a_2x^2 + a_4x + a_6$$
$$y^2 = x^3 + a_4x + a_6 \quad \text{or} \quad y^2 = 4x^3 - g_2x - g_3$$

the defining equation for each of these elliptic curves. We also call the j-invariant of each of these plane curve the j-invariant of the corresponding elliptic curve $E = (C, O)$, and denote it by $j(E)$.

THEOREM 3.7. *A necessary condition for two elliptic curves E and E' to be isomorphic is that $j(E) = j(E')$.*

The j-invariant is an element of k by definition. Given an arbitrary element $a \in k$, there exists an elliptic curve E such that $j(E) = a$. The following proposition can be shown by direct computation.

PROPOSITION 3.4. *If $a \neq 0, 12^3 (= 1728)$, the elliptic curve*

$$E : y^2 + xy = x^3 - \frac{36}{a - 1728}x - \frac{1}{a - 1728}$$

has $j(E) = a$. Further, the elliptic curve

$$y^2 + y = x^3$$

has j-invariant equal to 0, and if the characteristic of k is not 3, the elliptic curve

$$y^2 = x^3 + x$$

has j-invariant equal to 1728.

PROPOSITION 3.5 (LEGENDRE'S CANONICAL FORM). *Assume that the characteristic of k is not* 2.

(i) *An arbitrary elliptic curve is isomorphic to*

$$E_\lambda : y^2 = x(x-1)(x-\lambda), \quad \lambda \in k, \lambda \neq 0, 1,$$

and

$$j(E_\lambda) = 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}.$$

(ii) *The mapping*

$$j : \lambda \in k - \{0, 1\} \longmapsto j(E_\lambda) \in k$$

is $6:1$ on the complement of $j^{-1}\{0, 1728\}$; the inverse image of $j = 0$ has two points, and the inverse image of $j = 1728$ has three points.

(b) The group structure on an elliptic curve. The reason why elliptic curves play an important role in various fields is that they admit a group structure. This group structure is nothing but a geometric representation of the sum formula of elliptic functions.

We consider the elliptic curve defined by the canonical form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

that is, the nonsingular elliptic cubic curve

$$C : F = x_0x_2^2 + a_1x_0x_1x_2 + a_3x_0^2c_2 - x_1^3 - a_2x_0x_1^2 - a_4x_0^2x_1 - a_6x_0^3 = 0$$

coupled with a point $O = (0:0:1)$. A line in the projective plane $\mathbf{P}^2(k)$ intersects C at three points including multiplicities. First we find the tangent line to C at the point O. From the system

$$\frac{\partial F}{\partial x_0}(0,0,1) = 1, \quad \frac{\partial F}{\partial x_1}(0,0,1) = 0, \quad \frac{\partial F}{\partial x_2}(0,0,1) = 0$$

we find the tangent to be

$$\ell_\infty : x_0 = 0.$$

From the defining equation of C we immediately see that O is the only intersection of C with l_∞. Therefore C and ℓ_∞ are tangent with multiplicity 3. (Such a point is called an **inflection point** of C.) We can check on this directly. Set

$$u = \frac{x_0}{x_2}, \quad v = \frac{x_1}{x_2}.$$

The defining equation for C in these coordinates is

$$f(u, v) = u + a_1uv + a_3u^2 - v^3 - a_2uv^2 - a_4u^2v - a_6u^3 = 0.$$

Hence the multiplicity $I_O(C, \ell_\infty)$ of the intersection of C and ℓ_∞ at O turns out, as expected, to be

$$I_O(C, \ell_\infty) = 3$$

by $f(0, v) = -v^3$ and by the definition in §2.3 (b).

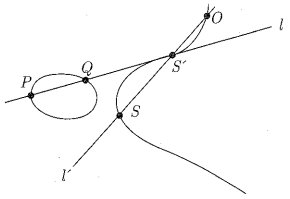


FIGURE 3.1

Now for any two points P, Q on C, take the line joining P and Q (Figure 3.1). If $P = Q$, take the tangent line at P to C as ℓ.

Let $\{P, Q, S'\}$ be the intersection of ℓ and C, let ℓ' be the line joining O and S', and let $\{O, S', S\}$ be the intersection of ℓ' and C. Then we define

$$P + Q = S.$$

PROPOSITION 3.6. *For a plane cubic curve C, we have the following for the points on C.*

(i) *If P, Q, R on C are on a line, then*

$$(P + Q) + R = O.$$

(ii)
$$P + O = P.$$

(iii)
$$P + Q = Q + P.$$

(iv) *For any P on C, there is a point P' on C such that*

$$P + P' = 0.$$

In the following, denote P' by $-P$.

(v) *For any three points P, Q, R on C*

$$(P + Q) + R = P + (Q + R).$$

That is, all the points of the elliptic curve $E = (C, O)$ form an Abelian group with O as the zero.

PROOF. (i) If ℓ is the line joining P and Q, then $\ell \cap C = \{P, Q, R\}$. Let ℓ' be the line joining O and R and let $\ell' = \{O, R, R'\}$; then

$$P + Q = R.$$

The line joining R and R' is just ℓ'. $R' + R$ is determined by the intersection of the tangent ℓ_∞ at O and C. Since ℓ_∞ and C intersect at O with multiplicity 3, we have

$$R' + R = O.$$

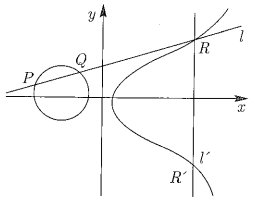


FIGURE 3.2. In x, y-coordinates, O is a point at infinity, and the line joining O and R' is parallel to the y-axis.

(See Figure 3.2.)

(ii) If the line ℓ joining P and Q intersects C at P, Q, S', then the line joining O and S' is equal to ℓ, and hence $P + O = P$.

(iii) Obvious from the definition of $P + Q$.

(iv) This can be shown by computation by using the representation of $P + Q$ in coordinates given in Proposition 3.7 below.

(v) We leave it as an exercise for the reader.

Given an integer m and a point P of an elliptic curve $E = (C, O)$, we define

$$[m]P = \underbrace{P + P + \cdots + P}_{m} \text{ for } m > 0$$
$$[0]P = O$$
$$[m]P = [-m](-P) \text{ for } m < 0.$$

From the proposition above, we have

$$[m]O = O.$$

The next proposition can be easily shown by direct computation and is left to the reader.

PROPOSITION 3.7. *Consider the elliptic curve defined by the canonical form*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

(i) *For $P = (x_0, y_0) \in E$,*

$$-P = (x_0, -y_0 - a_1x_0 - a_3).$$

(ii) *Let $P_i = (x_i, y_i) \in E, i = 1, 2$. If $x_1 = x_2$ and $y_1 + y_2 + a_1x_2 + a_3 = 0$, then*

$$P_1 + P_2 = O.$$

Otherwise, set

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1}, \quad \nu = \frac{y_1x_2 - y_2x_1}{x_2 - x_1}$$

in the case where $x_1 \neq x_2$, *and*

$$\lambda = \frac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3}, \quad \nu = \frac{-x_1^3 + a_4x_1 + 2a_6 - a_3y_1}{2y_1 + a_1x_1 + a_3}$$

in the case where $x_1 = x_2$. *Then* $P_3 = P_1 + P_2 = (x_3, y_3)$ *is given by*

$$x_3 = \lambda^2 + a_1\lambda - a_2 - x_1 - x_2$$
$$y_3 = -(\lambda + a_1)x_3 - \nu - a_3.$$

In particular, if $P_1 \neq \pm P_2$, *we have*

$$x(P_1 + P_2) = x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 + a_1\left(\frac{y_2 - y_1}{x_2 - x_1}\right) - a_2 - x_1 - x_2$$

(*where generally* $x(Q)$ *denotes the* x*-coordinate of* $Q \in E$) *and for* $P = (x, y)$ *we have*

$$x([2]P) = \frac{x^4 - b_4x^2 - 2b_6x - b_8}{4x^3 + b_2x^2 + 2b_4x + b_6},$$

where b_2, b_4, b_6, b_8 *are defined in* (3.27).

EXAMPLE 3.9. Take the elliptic curve

$$E : y^2 = x^3 + 17$$

defined over the rational number field $\mathbf{Q}$. As $\mathbf{Q}$-rational points (points whose coordinates are rational numbers) we easily find

$$P_1 = (-2, 3), P_2 = (-1, 4), P_3 = (2, 5), P_4 = (4, 9), P_5 = (8, -23).$$

We can find more Q-rational points:

$$P_6 = (43, 282), P_7 = (52, 375), P_8 = (5234, 378661),$$

and so on. We can further check

$$[2]P_1 = P_5, \quad P_4 = P_1 - P_3, \quad [3]P_1 - P_3 = P_7,$$
$$[2]P_2 = \left(\frac{137}{64}, -\frac{2651}{512}\right), \quad P_2 + P_3 = \left(-\frac{8}{9}, -\frac{109}{27}\right).$$

Indeed, it is known that an arbitrary $\mathbf{Q}$-rational point P can be expressed uniquely in the form

$$P = [m]P_1 + [n]P_2, \quad m, n \in \mathbf{Z}.$$

Let us denote by $E(\mathbf{Q})$ the set of all $\mathbf{Q}$-rational points of an elliptic curve E defined over $\mathbf{Q}$. Since we can show that $P, Q \in E(\mathbf{Q})$ implies $P + Q \in E(\mathbf{Q})$ by Proposition 3.7, we see that $E(\mathbf{Q})$ is a group. In the example above, $E(\mathbf{Q})$ is a free Abelian group of rank 2 generated by P_1, P_2. In general, $E(\mathbf{Q})$ has elements of finite order. We have the following result on this matter.

PROPOSITION 3.8. *Consider an elliptic curve*

$$E : y^2 = x^3 + a_4x + a_6$$

defined over the rational field $\mathbf{Q}$ *such that the coefficients* a_4, a_6 *are integers. Suppose* $P \in E(\mathbf{Q})$ *has finite order (but is not the point at infinity* O*). Then for the* x*-coordinate* $x(P)$ *of* P *and the* y*-coordinate* $y(P)$ *of* P *we have*

$$x(P), y(P) \in \mathbf{Z}.$$

Furthermore, $[2]P = 0$ *or* $y(P)^2$ *is divisible by* $4a_4^3 + 27a_6^2$.

EXAMPLE 3.10. For the elliptic curve

$$E : y^2 = x^3 - 43x + 166$$

we have

$$4a_4^3 + 27a_6^2 = 425984 = 2^{15} \cdot 13.$$

If we take $P = (3, 8) \in E(\mathbf{Q})$, we find that P has a point of order 7. (From Proposition 3.7 we get

$$x(P) = 3, \quad x([2]P) = -5, \quad x([4]P) = 11, \quad x([8]P) = 3$$

and $[8]P = \pm P$. Indeed, we can show that $[8]P = P$.)

Among the elements in $E(\mathbf{Q})$ of an elliptic curve E over $\mathbf{Q}$, the set of elements of finite order form a subgroup, which we denote by $E(\mathbf{Q})_{\text{tor}}$.

EXAMPLE 3.11. For the elliptic curve

$$E : y^2 = x^3 + x^2 - x$$

the points $(x, y) = (0, 0), (1, 1), (1, -1), (-1, 1), (-1, -1)$ are $\mathbf{Q}$-rational. These five elements and the point at infinity O form a cyclic group of order 6. Furthermore, $E(\mathbf{Q})_{\text{tor}}$ consists of these six points.

THEOREM 3.8 (THE MORDELL-WEIL THEOREM). *For an elliptic curve* E *defined over* $\mathbf{Q}$, $E(\mathbf{Q})$ *is a finitely generated Abelian group.*

According to this theorem we can write

$$E(\mathbf{Q}) = E(\mathbf{Q})_{\text{tor}} \oplus \mathbf{Z}^r \quad \text{(direct sum as group)},$$

where $E(\mathbf{Q})_{\text{tor}}$ is a finite Abelian group. We call r the **rank** of the elliptic curve.

EXAMPLE 3.12 (BREMNER-CASSELS). The elliptic curve $y^2 = x^3 + 877x$ has rank 1, and the x-coordinate of the generator P is given by

$$x(P) = \left(\frac{612776083187947368101}{7884153586063900210}\right)^2.$$

In spite of extensive research on elliptic curves, there are many unsolved problems. For example, the following conjecture, although expected to be valid, has yet to be proved.

CONJECTURE. *For an arbitrary natural number n, there is an elliptic curve defined over $\mathbf{Q}$ whose rank is greater than or equal to n.*

Now let us go back to elliptic curves defined over an arbitrary algebraically closed field. For any positive integer m, set

$$E[m] = \{P \in E \mid [m]P = 0\},$$

which forms a subgroup called the m-**torsion subgroup**.

THEOREM 3.9. *If a positive integer $m \geq 2$ is not divisible by the characteristic of the field k, then*

$$E[m] \simeq \mathbf{Z}/(m) \oplus \mathbf{Z}/(m).$$

If the characteristic p of k is ≥ 2 and if e is a positive integer, it is known that $E[p^e]$ is isomorphic to $Z/(p^e)$ or consists of the point O only.

Now consider a point $P \in E[3]$. Recall that the point $[2]P = P + P$ is defined as follows. Take the tangent line ℓ at P and denote by $\{P, P, S'\}$ the intersection of ℓ and C including multiplicities. Take the line ℓ' joining O and S' and denote its intersection with C by $\{O, S', S\}$. Then S is, by definition, $[2]P$. Next, $[3]P = S + P$. Let m be the line joining S and P and write its intersection with C as $\{S, P, T'\}$. The line m' joining O and T' intersects C at $\{O, T', T\}$, resulting in $[3]P = T$. By assumption, $[3]P = O$ and hence $T = O$. This shows that m' is tangent to C at O, that is, m' is the tangent line at O. But, as stated earlier, the tangent line at O has multiplicity 3, that is, O is an inflection point of C. Therefore, $T' = O$ and S, P, Q are on one line. Since O, S', S are on one line, it follows that $S' = P$. The tangent line ℓ at P has multiplicity 3 at P, that is, P is an inflection point of C. Conversely, if P is an inflection point of C, then we have $[3]P = O$, as is clear from the discussion above. Combined with the preceding theorem, this gives rise to

COROLLARY 3.5. *A point P of an elliptic curve E such that $[3]P = O$ is an inflection point of the plane cubic curve* (3.28)*, and conversely. If the characteristic of k is not* 3*, there are nine inflection points.*

§3.4. Congruence zeta functions on algebraic curves

In this section we assume an elementary knowledge on finite fields (see the Appendix). Let us consider a finite field $k = GF(q)$ with q elements. We denote an algebraically closed field containing k by $\bar{k}$. An extension of degree n of $k = GF(q)$ is denoted by k_n. Thus $k_n = GF(q^n)$, that is, k_n is a finite field with q^n elements.

Now consider a projective variety $V = V((F_1, F_2, \ldots, F_\ell))$ in $\mathbf{P}^N(\bar{k})$ defined as the common zero set of homogeneous polynomials

$$F_j(x_0, x_1, \ldots, x_N), \qquad 1 \leq j \leq \ell.$$

If we can take all the coefficients of F_j, $1 \leq j \leq \ell$, from k, or more precisely, if we can find homogeneous polynomials with coefficients in k for the ideal $(F_1, F_2, \ldots, F_\ell)$ in $\bar{k}[x_0, x_1, \ldots, x_N]$, we say that the projective variety V is **defined over the field** k. In this definition, the finiteness of k is not necessary.

In the following, we assume for simplicity that the defining homogeneous polynomials for V have coefficients in k. A point $(a_0 : a_1 : \cdots : a_N)$ of V is called a k-**rational point** if all a_j's can be taken from k, that is, whenever $a_k \neq 0$, all

$a_j/a_k \in k$ — this condition being independent of the choice of $x_k \neq 0$. If we denote by $V(k)$ the set of all k-rational points of V, then the number of elements $\#V(k)$ is finite, since k is finite. We set

$$N(V) = \#V(k).$$

In particular, for a positive integer n, we take an extension k_n of degree n of $k = GF(q)$ and set

$$N_n(V) = \#V(k_n).$$

DEFINITION 3.5. For a projective variety defined over a finite field k, the **congruence zeta function** or simply **zeta function** is the function $Z(V, u)$ defined as follows.

(i)
$$Z(V, 0) = 1$$

(ii)
$$\frac{d}{du} \log Z(V, u) = \sum_{n=1}^{\infty} N_n(V) u^{n-1}.$$

EXAMPLE 3.13. We consider the projective space $\mathbf{P}^N(k)$ as defined over k, since there is no defining equation except for the zero polynomial. We have

$$N_n(\mathbf{P}^N(k_n)) = \frac{q^{n(N+1)} - 1}{q^n - 1} = \sum_{j=0}^{N} q^{nj}.$$

($\mathbf{P}^N(k_n)$ is the totality of ratios of the elements in $k_n^{N+1} - \{(0, 0, \ldots, 0)\}$.) Hence

$$\begin{aligned} \frac{d}{du} \log \mathbf{Z}(P^N, u) &= \sum_{n=1}^{\infty} \left(\sum_{j=0}^{N} q^{nj} \right) u^{n-1} \\ &= \sum_{j=0}^{N} \sum_{n=1}^{\infty} q^{jn} u^{n-1} = \sum_{j=0}^{N} \frac{q^j}{1 - q^j u}. \end{aligned}$$

In view of condition (i) we get

$$Z(\mathbf{P}^N, u) = \frac{1}{(1-u)(1-qu)(1-q^2u)\cdots(1-q^Nu)}.$$

Although condition (ii) in Definition 3.5 above defines $\log Z(V, u)$ only as a formal power series, observe that for $V \subset \mathbf{P}^N(\bar{k})$, we have

$$N_n(V) \leq N_n(\mathbf{P}^N)$$

and the example above shows that the series

$$\sum_{n=1}^{\infty} N_n(\mathbf{P}^N) u^{n-1}$$

converges uniformly and absolutely in a neighborhood of $u = 0$, showing that $Z(V, u)$ is a regular function in a neighborhood of $u = 0$. It is no accident that in the example above $Z(V, u)$ is a rational function of u. In fact, the following result is known.

THEOREM 3.10 (DWORK). *The zeta function $Z(V,u)$ of a projective variety V defined over a finite field k is a rational function of u with rational coefficients.*

The zeta function, even in the simplest case, that is, the case of a hypersurface $V(F)$, counts the number of solutions of a homogeneous equation

$$F(x_0, x_1, \ldots, x_N) = 0$$

over a finite field k_n, and is complicated. However, when V is a nonsingular projective variety, $Z(V,u)$ has a remarkably beautiful property, as was conjectured by A. Weil in the 1940s. The effort to solve this Weil conjecture became a driving force for further developments in algebraic geometry. Since it takes too much preparation to state Weil's conjecture, we shall give results on the zeta function in the case where V is a nonsingular algebraic curve.

THEOREM 3.11. *The zeta function $\mathbf{Z}(C,u)$ of a nonsingular projective line C of genus g defined over a finite field $k = GF(q)$ has the following properties.*

(i) *$Z(C,u)$ can be expressed in the form*

$$Z(C,u) = \frac{F(u)}{(1-u)(1-qu)},$$

where $F(u)$ is a polynomial of degree $2g$ in u with integral coefficients and takes the form

$$F(u) = 1 + (N_1(C) - q - 1)u + \cdots + q^g u^{2g}.$$

(ii) *$Z(C,u)$ satisfies the functional equation*

$$Z\left(C, \frac{1}{qu}\right) = q^{1-g}u^{2-2g}Z(C,u). \tag{3.32}$$

(iii) *If $F(u) = \prod_{j=1}^{2g}(1-\omega_j u)$ is a factorization, then*

$$|\omega_j| = \sqrt{q}, \quad 1 \le j \le 2g.$$

Part (iii) of Theorem 3.11 is the most important part of the theorem and is called Riemann's conjecture for the zeta function. If we replace the variable u by

$$u = q^{-s},$$

then $H(s) = F(q^{-s})$ has $2g$ zeros $s_1, s_2, \ldots, s_{2g}$, and the relations

$$q^{-s_j} = \frac{1}{\omega_j}$$

hold. Hence $|\omega_j| = \sqrt{q}$ is equivalent to

$$\Re(s_j) = \frac{1}{2},$$

thus taking a similar form as the Riemann conjecture on the zeros of zeta functions, which explains the name of Riemann's conjecture. From (ii) and (iii) of the theorem above we see that $F(1/\omega_j) = 0$ implies $F(\omega_j/q) = 0$. Therefore we get

COROLLARY 3.6. *Under the same assumption as in Theorem 3.11, if*

$$F(u) = \prod_{j=1}^{2g}(1-\omega_j u)$$

is a factorization, then we can re-number $\omega_1, \omega_2, \ldots, \omega_{2g}$ in such a way that

$$\omega_j \omega_{g+j} = q.$$

Incidentally, the proofs for parts (i) and (ii) of Theorem 3.11 are not too difficult; (i) can be shown by using the theory of k-rational divisors for the algebraic curve C, and the functional equation in (ii) is a direct consequence of the Riemann-Roch theorem. For an elementary proof, see C.J. Moreno, *Algebraic Curves over Finite Fields*, Cambridge Tracts in Math. 97, Cambridge University Press, 1991, Chapter 3. Here we only show that the degree 1 term of $F(u)$ is $N_1(C) - q - 1$. Set $F(u) = a_0 + a_1 u + a_2 u^2 + \ldots$ and expand $Z(C,u)$ in a power series in u in a neighborhood of $u = 0$, getting

$$\begin{aligned}\frac{F(u)}{(1-u)(1-qu)} &= \left(\sum_{i=0}^{\infty} u^i\right)\left(\sum_{j=0}^{\infty} q^j u^j\right)(a_0 + a_1 u + \cdots) \\ &= a_0 + \{a_0(1+q) + a_1\}u + \cdots.\end{aligned}$$

On the other hand, we have by definition of $Z(C,u)$

$$\log Z(C,u) = \sum_{m=1}^{\infty} \frac{N_m(C)}{m} u^m$$
$$Z(C,0) = 1,$$

from which we get

$$a_0 = 1, \quad 1 + q + a_1 = N_1(C).$$

Furthermore, since $a_1 = -\sum_{j=1}^{2g} \omega_j$, we get the following result by using (iii) of the theorem.

COROLLARY 3.7. *For a nonsingular projective curve C defined over a finite field $k = GF(q)$, the number $N_1(C)$ of k-rational points satisfies an estimate*

$$|N_1(C) - q - 1| \le 2g\sqrt{q}.$$

We now extend this estimate to the case of the number $N_m(C)$ of k_m-rational points. For this purpose, it is sufficient to prove the following.

LEMMA 3.10. *Let $Z(C,u)$ be the zeta function of a nonsingular projective curve C defined over a finite field $k = GF(q)$. For any positive integer m, let $Z_m(C,u)$ be the zeta function of C as a curve defined over k_m. If*

$$Z(C,u) = \frac{\prod_{j=1}^{2g}(1-\omega_j u)}{(1-u)(1-qu)},$$

then

$$Z_m(C,u) = \frac{\prod_{j=1}^{2g}(1-\omega_j^m u)}{(1-u)(1-q^m u)}.$$

PROOF. By definition we have

$$\log Z_m(C,u) = \sum_{j=1}^{\infty} \frac{N_{mj}(C)}{j} u^j.$$

By setting $\epsilon = e^{2\pi i/m}$, $i = \sqrt{-1}$, we have

$$x^m - 1 = \prod_{k=0}^{m-1} (x - \epsilon^k).$$

By comparing the coefficients of x^{m-1} we get

$$\sum_{k=0}^{m-1} \epsilon^k = 0.$$

If $j \not\equiv 0 \mod m$, then for any $k, 0 \le k \le m-1$, there is a unique $\ell, 0 \le \ell \le m-1$, such that $\ell j \equiv k \mod m$ (see Theorem A.2 in the Appendix). Therefore we have

$$\sum_{\ell=0}^{m-1} \epsilon^{\ell j} = \begin{cases} m, & j \equiv 0 \mod m \\ 0, & j \not\equiv 0 \mod m. \end{cases}$$

Now the function

$$L(u) = \prod_{\ell=0}^{m-1} Z(C, \epsilon^\ell u)$$

satisfies

$$\log L(u) = \sum_{i=0}^{m-1} \log Z(C, \epsilon^\ell u) = \sum_{\ell=0}^{m-1} \sum_{j=1}^{\infty} \frac{N_j(C)}{j} \epsilon^{\ell j} u^j \tag{3.33}$$
$$= \sum_{j=1}^{\infty} \frac{N_j(C)}{j} \left(\sum_{\ell=0}^{m-1} \epsilon^{\ell j} \right) u^j = \sum_{k=1}^{\infty} \frac{N_{mk}(C)}{k} u^{mk}.$$

On the other hand, we have

$$\log Z_m(C, u^m) = \sum_{k=1}^{\infty} \frac{N_{mk}(V)}{k} u^{mk},$$

from which we obtain

$$Z_m(C, u^m) = L(u)$$

By definition, we have

$$L(u) = \prod_{\ell=0}^{m-1} \frac{\prod_{j=1}^{2g}(1 - \omega_j \epsilon^\ell u)}{(1 - \epsilon^\ell u)(1 - q\epsilon^\ell u)} = \frac{\prod_{j=1}^{2g}(1 - \omega_j^m u^m)}{(1 - u^m)(1 - q^m u^m)},$$

and hence

$$Z_m(C,u) = \frac{\prod_{j=1}^{2g}(1 - \omega_j^m u)}{(1-u)(1-q^m u)}.$$

COROLLARY 3.8. *If we denote by $N_m(C)$ the number of k_m-rational points on a nonsingular projective curve C of genus g defined over a finite field $k = GF(q)$, then we have an estimate*

$$|N_m(C) - q^m - 1| \le 2g\sqrt{q^m} \tag{3.34}$$

In fact, the converse also holds. The following proposition is used in the proof of Riemann's conjecture (iii) in Theorem 3.11.

PROPOSITION 3.9. *Let C be a nonsingular projective curve of genus g defined over a finite field $k = GF(q)$. For its zeta function*

$$Z(V,u) = \frac{\prod_{j=1}^{2g}(1 - \omega_j u)}{(1-u)(1-qu)}$$

the following propositions are equivalent.

(i) $|\omega_j| = \sqrt{q}, \quad 1 \le j \le 2g.$

(ii) *There is a constant a such that the estimate*

$$|N_m(C) - q^m - 1| \le a\sqrt{q^m}$$

holds for all natural numbers m.

PROOF. We have already shown that (i) implies (ii). As the constant a, we can take $2g$. In order to show that (ii) implies (i), let

$$Z(C,u) = \frac{F(u)}{(1-u)(1-qu)}.$$

We first show that if there is u_0 such that

$$F(u_0) = 0, \quad |u_0| > 1/\sqrt{q},$$

then there is u_1 such that

$$F(u_1) = 0, \quad |u_1| < 1/\sqrt{q}.$$

By setting $u_1 = 1/qu_0$ in the functional equation

$$Z\left(C, \frac{1}{qu}\right) = q^{1-g} u^{2-2g} Z(C,u)$$

we get $|u_1| < 1/\sqrt{q}$ and $Z(C,u_1) = 0$, proving our assertion. Since we have

$$\frac{d}{du} \log F(u) = \sum_{m=1}^{\infty} (N_m(C) - q^m - 1) u^{m-1}, \tag{3.35}$$

we obtain from the estimate in (ii)

$$\sum_{m=1}^{\infty} |N_m(C) - q^m - 1| \cdot |u|^{m-1} \le a \sum_{m=1}^{\infty} q^{m/2} |u|^{m-1} = a\sqrt{q} \sum_{\ell=0}^{\infty} (\sqrt{q}u)^\ell,$$

where the last power series is convergent for $|u| < 1/\sqrt{q}$. Hence the right-hand side of (3.35) converges for $|u| < 1/\sqrt{q}$, and $(d/du) \log F(u)$ is a regular function in $|u| < 1/\sqrt{q}$. Suppose $F(u)$ has a zero such that $|u| \ne 1/\sqrt{q}$, then it has a zero at

$|u| < 1/\sqrt{q}$, as we have shown above. This means that $(d/du)\log F(u)$ has a pole in $|u| < \sqrt{q}$, which is a contradiction to the result above. Hence

$$|\omega_j| = \sqrt{q}.$$

Thanks to this proposition, in order to show Riemann's conjecture it is sufficient to show the relatively weak estimate in (ii) of Proposition 3.9. Such an estimate was first obtained by S.A. Stepanov. A proof by Weil uses an entirely different method based on the theory of Jacobi varieties.

Estimating the number of k_m-rational points for an algebraic curve C is important in applications. Since $N_m(C) - q^m - 1$ is an integer, the estimate in (3.34) gives

$$|N_m(C) - q^m - 1| \le [2g\sqrt{q^m}],$$

where the Gauss symbol $[x]$ means

$$[x] = \text{the largest integer not exceeding } x.$$

The estimate above is not best. The following estimate due to Serre is known.

THEOREM 3.12. *For a nonsingular projective curve C of genus g over the finite field $k = GF(q)$ the estimate*

$$|N_m(C) - q^m - 1| \le g[2\sqrt{q^m}]$$

holds for all positive integers m.

This estimate by Serre is better than the one given above for $g \ge 2$. For example, for $g = 2, q = 23, m = 1$ we have

$$[2g\sqrt{q}] = [4\sqrt{23}] = 19$$
$$g[2\sqrt{q}] = 2[2\sqrt{23}] = 18.$$

We have so far discussed a general theory of zeta functions $Z(C,u)$. Now we compute their concrete forms in various examples. Their definition involves the number $N_m(C)$ of k_m-rational points for all natural numbers m. What Theorem 3.11 means is that if we know $N_m(C)$ for a finite number of m's we can determine $Z(C,u)$. For example, when $g = 1$, we have

$$F(u) = 1 + (N_1(C) - q - 1)u + qu^2,$$

which is determined as soon as we know $N_1(C)$. In this case, we get

$$F(u) = (1 - \omega u)(1 - \bar{\omega}u),$$

where $\bar{\omega}$ is the complex conjugate. When $g = 2$, we have

$$\begin{aligned} F(u) &= \prod_{j=1}^{4}(1 - \omega_j u) \\ &= 1 + (N_1(C) - q - 1)u + a_2u^2 + a_3u^3 + q^2u^4. \end{aligned}$$

On the other hand, Lemma 3.10 implies

$$\begin{aligned} N_1(C) - q - 1 &= -\sum_{j=1}^{4}\omega_j \\ N_2(C) - q^2 - 1 &= -\sum_{j=1}^{4}\omega_j^2 \\ N_3(C) - q^3 - 1 &= -\sum_{j=1}^{4}\omega_j^3. \end{aligned} \tag{3.36}$$

Thus $N_1(C), N_2(C), N_3(C)$ determine the coefficients a_2, a_3 of $F(u)$ and the zeta function. (In reality, $N_1(C), N_2(C)$ and the functional equation (3.32) determine $F(u)$.) When the zeta function is known, then

$$N_m(C) = q^m + 1 - \sum_{j=1}^{2g}\omega_j^m$$

gives the number of k_m-rational points.

EXAMPLE 3.14. We consider the plane cubic curve

$$C : x_0^3 - x_1^3 - x_2^3 = 0.$$

If the characteristic of k is not 3, C is nonsingular with genus 1. Let $k = GF(5)$. First, assume $x_0 = 0$; then

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

Since $x_1x_2 \ne 0$, the equation is reduced to

$$X^3 + 1 = 0,$$

which has a solution $-1 = 4$ in $GF(5)$. Thus $(0:1:4)$ is a k-rational point. Next, assume $x_0 \ne 0$. Setting

$$x = x_1/x_0, \quad y = x_2/x_0,$$

we get

$$x^3 + y^3 = 1.$$

the number of solutions in k of this equation is $N_1(C) - 1$. Since the solutions are $(1,0),(0,1),(2,2),(3,4),(4,3)$, we have $N_1(C) - q - 1 = 6 - 5 - 1 = 0$, and the zeta function is

$$Z(C,u) = \frac{1 + 5u^2}{(1-u)(1-5u)}.$$

If we write

$$F(u) = (1 - \omega u)(1 - \bar{\omega}u),$$

we have

$$\omega, \bar{\omega} = \pm\sqrt{-5},$$

that is,

$$|\omega| = |\bar{\omega}| = \sqrt{5},$$

showing that Riemann's conjecture is certainly valid. We have further

$$N_2(C) = 5^2 + 1 - (\sqrt{-5})^2 - (-\sqrt{-5})^2 = 26 + 10 = 36$$
$$N_3(C) = 5^3 + 1 - (\sqrt{-5})^3 - (-\sqrt{-5})^3 = 126 + 0 = 126.$$

We recommend that the reader compute directly the number of rational points over $GF(5^2)$ and $GF(5^3)$ by using a computer.

By the way, computation is simpler for $GF(2)(=\mathbf{Z}/(2)=\{0,1\})$. If $a\in GF(2)$, then $a^3=a$ and hence the solutions in $GF(2)$ of $x_0^3-x_1^3-x_2^3=0$ coincide with the solutions of $x_0-x_1-x_2=0$. The three solutions are $(1:1:0),(1:0:1),(0:1:1)$. Hence

$$N_1(C)-q-1=3-2-1=0,$$

and the zeta function is

$$Z(C,u)=\frac{1+2u^2}{(1-u)(1-2u)}.$$

Next let $k=GF(7)(=\mathbf{Z}/(7))$. In this case $X^3+1=0$ has three solutions, $3,5,6$, in k. Thus $(0:1:3),(0:1:5),(0:1:6)$ are the k-rational points with $x_0=0$. On the other hand, $x^3+y^3=1$ has the six solutions $(0,1),(0,2),(0,4),(1,0),(2,0),(4,0)$ in k, so $(1:0:1),(1:0:2),(1:0:4),(1:1:0),(1:2:0),(1:4:0)$ are the k-rational points. Hence $N_1(C)=9$ and the zeta function is

$$Z(C,u)=\frac{1+u+7u^2}{(1-u)(1-7u)}.$$

If we set $F(u)=(1-\omega u)(1-\bar{\omega}u)$, then

$$\omega,\bar{\omega}=\frac{-1\pm 3\sqrt{-3}}{2}$$

and

$$|\omega|=|\bar{\omega}|=\frac{\sqrt{1+27}}{2}=\sqrt{7}.$$

We have also

$$\begin{aligned}N_2(C)&=7^2+1-\left(\frac{-1+3\sqrt{-3}}{2}\right)^2-\left(\frac{-1-3\sqrt{-3}}{2}\right)^2\\&=50+13.\end{aligned}$$

We shall mention a few more examples, and hope that the readers will try computation themselves.

EXAMPLE 3.15. The plane cubic curve

$$x_0x_2^2+x_1^3-x_0^3=0$$

is a nonsingular curve of genus 1 except for the case of characteristic 2 or 3. For $GF(5)$, the zeta function is

$$\frac{1+5u^2}{(1-u)(1-5u)},$$

and for $GF(7)$ it is

$$\frac{1+4u+7u^2}{(1-u)(1-7u)}.$$

EXAMPLE 3.16. The plane cubic curve

$$x_0x_2^2-x_0^2x_1+x_1^3=0$$

is a nonsingular algebraic curve of genus 1 except for the case of characteristic 2 or 3. For $GF(5)$ the zeta function is

$$\frac{1+2u+5u^2}{(1-u)(1-5u)},$$

and for $GF(7)$ it is

$$\frac{1+7u^2}{(1-u)(1-7u)}.$$

EXAMPLE 3.17. The plane cubic curve

$$x_0^3+x_1^3+x_2^3=0$$

is a nonsingular algebraic curve of genus 1 except for the case of characteristic 3. For $GF(2)$ the zeta function is

$$\frac{1+2u^2}{(1-u)(1-2u)},$$

and for $GF(13)$ it is

$$\frac{1-5u+13u^2}{(1-u)(1-13u)}.$$

This plane curve is isomorphic to the plane cubic curve in Example 3.14 by a projective transformation

$$(x_0:x_1:x_2)\longmapsto(x_0:-x_1:x_2).$$

Finally, we consider curves of genus 2.

EXAMPLE 3.18. Excluding the characteristics 2 and 5, consider the nonsingular projective curve $\tilde{C}$ of genus 2 determined by

$$y^2=x^5+1. \tag{3.37}$$

We can realize $\tilde{C}$ as a nonsingular projective curve over $k=GF(p)$ (p a prime not equal to 2 or 5) with only one point at infinity that is a k-rational point. For simplicity, consider the case $k=GF(3)$. The solutions of the equation (3.37) in k are $(x,y)=(0,1),(0,2),(2,0)$, and

$$N_1(\tilde{C})=4.$$

On the other hand, k_2 is equal to $k(\alpha)$, where α is a solution of $X^2+1=0$, and k_2 admits six other solutions of (3.37), that is, $(x,y)=(1,\pm\alpha),(1+\alpha,\pm(1+2\alpha))$, $(1+2\alpha,\pm(2+\alpha))$, in addition to the three solutions in k. We have

$$N_2(\tilde{C})=10.$$

In $k_3=k(\beta)$, where β is a solution of $X^3+X+1=0$, we have

$$N_3(\tilde{C})=28.$$

More generally, if 5 is not a divisor of 3^m-1, the number of k_m-rational numbers is given by

$$N_m(\tilde{C}) = 3^m + 1, \tag{3.38}$$

for the following reason. The mapping

$$a \in k_m \longmapsto a+1 \in k_m$$

is surjective as set mapping, and since $k_m^* = k_m - \{0\}$ is a cyclic group of order 3^m-1 (see Theorem A.7 in the Appendix), the mapping

$$b \in k_m \longmapsto b^5 \in k_m$$

is surjective as a set mapping provided 5 is not a divisor of 3^m-1. In this case, we can write $k_m^* = \{1, \gamma, \gamma^2, \ldots, \gamma^{3^m-2}\}$; $y^2 = \gamma^\ell$ has a solution only when ℓ is even, and the solution is $y = \pm\gamma^{\ell/2}$. Since $\ell \leq 3^m-2$, we have $\gamma^{\ell/2} \neq -\gamma^{\ell/2}$. Hence if we take a such that $a^5+1 = \gamma^{2s}, 0 \leq s \leq (3^m-2)/2$ (such a is unique), then $(x,y) = (a, \pm\gamma^s)$ gives a solution of $y^2 = x^5+1$ in k_m. Further, there is a solution $(a,0)$ of y^2+x^5+1 corresponding to a^5+1. Thus the number of solutions in k_m of $y^2 = x^5+1$ is $(3^m-1)/2\times 2+1 = 3^m$, and since the point at infinity is a k-rational point, we get (3.38). From this and from (3.36) we see that the zeta function is

$$\frac{1+9u^4}{(1-u)(1-3u)}.$$

This argument can be generalized, and for characteristic different from $2, 5$ and provided $q-1, q^2-1, q^3-1$ are not divisible by 5, the zeta function is given by

$$\frac{1+q^2u^4}{(1-u)(1-qu)}.$$

Problems

3.1. For a divisor D on an algebraic curve C, we define the **complete linear system** $|D|$ by

$$|D| = \{E \mid E \geq 0, E \sim D\}.$$

Prove that the mapping

$$f \in \mathbf{L}(D) - \{0\} \mapsto D + (f) \in |D|$$

is surjective, and that if

$$D + (f_1) = D + (f_2),$$

then there exists a constant α such that

$$f_1 = \alpha f_2.$$

3.2. Consider the complete linear system $|D|$ determined by a divisor D on an algebraic curve C. A point Q on C is called a **base point** of $|D|$ if

$$E \geq Q$$

for every element $E \in |D|$. Show that if Q is a base point of $|D|$, then the mapping

$$E' \in |D_Q| \mapsto E' + Q \in |D|$$

is surjective and injective. Show also that $\mathbf{L}(D-Q)$ and $\mathbf{L}(D)$ coincide when we view $\mathbf{L}(D-Q)$ as a subset of $\mathbf{L}(D)$.

3.3. Prove Lemma 3.6 by using the Hurwitz formula. [Hint: For a nonsingular plane curve

$$C : F(x_0, x_1, x_2) = 0, \quad \deg F = n,$$

we may assume

$$F(x_0, x_1, x_2) \not\equiv 0, \quad F(1,0,0) \neq 0$$

by applying a projective transformation if necessary. Let Q be a point of C and denote by $R(Q)$ the intersection of the line joining $(1,0,0)$ and Q with the line at infinity $\ell_\infty : x_0 = 0$. The mapping

$$\psi_P : Q \in C \mapsto R(Q) \in \ell_\infty \simeq \mathbf{P}^1(C)$$

is an n to 1 algebraic morphism admitting ramification points that occur when the line $\overline{PQ}$ is tangent to the curve C. If $\overline{PQ}$ is tangent to C at Q, we may assume that the multiplicity of intersection is 2 by applying a projective transformation. In this case, the index of ramification of ψ_P at Q is 2. Using this fact, count the number of points Q such that $\overline{PQ}$ is tangent to C at Q, and apply the Hurwitz formula.]

3.4. Show that the discriminant of an elliptic curve

$$y^2 = x(x-a^p)(x+b^p)$$

is given by

$$\triangle = 2^{-8}a^{2p}b^{2p}c^{2p},$$

where p is a prime and a, b, c are positive integers such that $a^p+b^p = c^p$. Also show that the j-invariant is given by

$$j = 2^8(a^{2p}+b^{2p}+a^pb^p)^3/a^{2p}b^{2p}c^{2p}$$

(If the Fermat conjecture were disproved, there would exist a prime p and positive integers a, b, c satisfying the condition above. The elliptic curve in this case is called the Frey curve. Also see the box at the end of Chapter 1.)

3.6. Show that the zeta function of the singular cubic curve

$$C : x_0 x_1^2 - x_2^3 = 0$$

over a finite field $GF(q)$ of characteristic different from 3 is given by

$$Z(C, u) = \frac{1}{(1-qu)^2}.$$

CHAPTER 4

The Analytic Theory of Algebraic Curves

We have so far discussed only the algebraic aspects of algebraic geometry. As the study of properties of figures defined algebraically, algebraic geometry naturally has geometric and analytic tools. We have slightly touched upon them in the theory of elliptic curves. Here we shall treat algebraic curves as closed Riemann surfaces. In this chapter we work with the complex number field $\mathbf{C}$. For the sake of contrast to the treatment in the preceding chapter, we redo, for example, the proof of Theorem 4.4, which is almost the same as that of Theorem 3.4. On the whole, however, we omit proofs for many important results. Although somewhat hurriedly, we wanted to show the first step in the analytic theory of closed Riemann surfaces, and recommend the reader to refer to a standard treatise. In this chapter we frequently use i as an index and hence use $\sqrt{-1}$ for the imaginary unit in order to avoid confusion.

§4.1. Closed Riemann surfaces

In § 2.1 we started from the Riemann sphere, considered it as the projective line, generalized it to n-dimensional projective spaces, and defined projective sets and projective varieties as the set of common zeros of a finite number of homogeneous polynomials. Now let us try to reverse this process. We can define the n-dimensional complex projective space $\mathbf{P}^n(\mathbf{C})$ by pasting together $n+1$ copies of n-dimensional affine space $\mathbf{C}^n$. For pasting we use simple rational functions, which we may consider as holomorphic functions where the pasting is carried out. From this point of view, we may consider $\mathbf{P}^n(\mathbf{C})$ as an n-dimensional complex manifold. For $n = 1$ this means that we regard the projective line $\mathbf{P}^1(\mathbf{C})$ as the Riemann sphere. An n-dimensional complex manifold is a "figure" obtained by pasting open subsets in $\mathbf{C}^n$ by biholomorphic mappings. For a rigorous definition, we refer the readers to K. Kodaira: *Complex Manifolds and Deformation of Complex Structures*, Springer, 1986.

Now if an m-dimensional projective variety V in $\mathbf{P}^n(\mathbf{C})$ is nonsingular, we can regard it as an m-dimensional complex submanifold. Suppose V is defined as the set of common zeros of a finite number of homogeneous polynomials

$$f_i(x_0 : x_1 : \cdots : x_n) = 0, \quad 1 \leq i \leq \ell.$$

To say that V is nonsingular at a point $a = (a_0 : a_1 \cdots : a_n) \in V$ means by Lemma 2.14 that

$$\operatorname{rank} \begin{pmatrix} \frac{\partial F_1}{\partial x_0}(a) & \cdots & \frac{\partial F_1}{\partial x_n}(a) \\ \frac{\partial F_2}{\partial x_0}(a) & \cdots & \frac{\partial F_2}{\partial x_n}(a) \\ \cdots & \cdots & \cdots \\ \frac{\partial F_\ell}{\partial x_0}(a) & \cdots & \frac{\partial F_\ell}{\partial x_n}(a) \end{pmatrix} = n - m.$$

For simplicity, assume $a_0 \neq 0$ and write $(1 : b_1 : b_2 : \cdots : b_n)$ for $(a_0 : a_1 : \cdots : a_n)$. Set

$$N_i = \deg F_i, \quad z_i = \frac{x_i}{x_0}, \quad f_i(z_1, z_2, \ldots, z_n) = \frac{1}{x_0^{N_i}} F_i(x_0, x_1, \ldots, x_n).$$

Then the condition above can be rewritten in the form

$$\operatorname{rank} \begin{pmatrix} \frac{\partial f_1}{\partial z_1}(b) & \cdots & \frac{\partial f_1}{\partial z_n}(b) \\ \cdots & \cdots & \cdots \\ \frac{\partial f_\ell}{\partial z_1}(b) & \cdots & \frac{\partial f_\ell}{\partial z_n}(b) \end{pmatrix} = n - m,$$

which we adopted as the definition of "nonsingular" in Definition 2.6. By reordering the polynomials f_i and coordinates z_j, we may rewrite the condition above in the form

$$\begin{vmatrix} \frac{\partial f_1}{\partial z_1}(b) & \cdots & \frac{\partial f_1}{\partial z_{n-m}(b)} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_{n-m}}{\partial z_1}(b) & \cdots & \frac{\partial f_{n-m}}{\partial z_{n-m}}(b) \end{vmatrix} \neq 0.$$

Under this assumption, the implicit function theorem says that a neighborhood of $(1 : b_1 : \cdots : b_n)$ in V can be described by

$$z_j = g_j(z_{n-m+1}, z_{n-m+2}, \cdots, z_n), \quad 1 \leq j \leq n - m,$$

where g_j are holomorphic functions in a neighborhood of $(b_{n-m+1}, \ldots, b_n)$. In other words, a neighborhood, say U, of the point $(1 : b_1 : \cdots : b_n)$ admits a parametric representation on a certain open subset of $\mathbf{C}^m$. We call U together with the parameters $(z_{n-m+1}, \ldots, z_n)$ a **coordinate neighborhood**. Since this situation takes place at each point of V, we see that V is an m-dimensional complex submanifold.

Since $\mathbf{P}^n(\mathbf{C})$ is a compact complex manifold, so is V. In particular, if V is a nonsingular algebraic curve, we may regard it as a compact 1-dimensional complex manifold. A 1-dimensional complex manifold is usually called a **Riemann surface**. Thus a nonsingular projective curve has a structure of Riemann surface. Conversely, in fact, a closed Riemann surface admits the structure of a nonsingular projective curve. For nonsingular projective varieties, certain analytic properties and algebraic properties are equivalent in many ways (Serre's GAGA). The following is also an important theorem.

THEOREM 4.1 (CHOW'S THEOREM). *A closed analytic subset V in $\mathbf{P}^n(\mathbf{C})$ has the structure of a projective set. In particular, a closed complex submanifold of $\mathbf{P}^n(\mathbf{C})$ has the structure of a nonsingular projective variety.*

Here by an **analytic set** is meant a subset that can locally be expressed as the set of common zeros of a number of holomorphic functions. What Chow's theorem asserts is that a closed subset that is locally the set of common zeros of holomorphic functions is in fact the set of common zeros of a number of homogeneous polynomials. The discussion above is probably sufficient to give the reader at least a vague understanding of why we consider nonsingular projective curves as closed Riemann surfaces. Therefore, in the sequel, we shall study closed Riemann surfaces from the viewpoints of geometry and analysis. A closed Riemann surface, as can be imagined from the Riemann sphere, is a closed 2-dimensional surface that is orientable. Such surfaces are well understood from the topological viewpoint and look like a floating bag with g holes (see Figure 4.1). This number of holes is called the **genus of the**

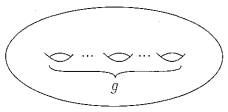


FIGURE 4.1

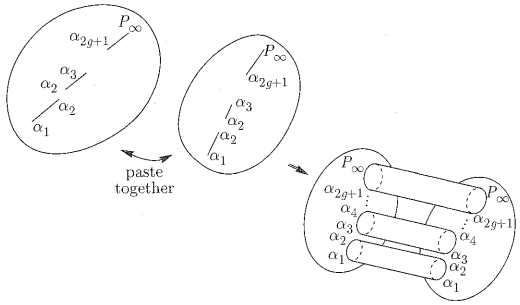


FIGURE 4.2. For $x \neq a_1, a_2, \ldots, a_{2g+1}$, $y^2 = f(x)$ has two distinct roots $\pm\sqrt{f(x)}$. The closed Riemann surface is obtained by adjoining the point at infinity to the set of all $(x, \sqrt{f(x)}), (x, \sqrt{-f(x)})$.

Riemann surface. Thus the Riemann sphere has genus 0 and conforms with the fact that the genus of the projective line is 0.

EXAMPLE 4.1. A hyperelliptic curve

$$y^2 = f(x), \quad \deg f = 2g + 1,$$

has genus g as a closed Riemann surface, as is geometrically visible from Figure 4.2.

Let us define meromorphic differential forms on a closed Riemann surface R. On a local coordinate neighborhood (U, z), where U is an open neighborhood of a certain point and z is a parameter on U, a form that can be written $f(z)dz$ with a meromorphic function $f(z)$ is called a **meromorphic differential form**. When we have a covering of R by local coordinate neighborhoods $\{(U_\lambda, z_\lambda)\}$ each with a meromorphic differential form $g_\lambda dz_\lambda$ such that in the non-empty intersection of any two coordinate neighborhood U_λ and U_μ we have

$$g_\lambda \frac{dz_\lambda}{dz_\mu} = g_\mu(z_\mu),$$

we say that $\{g_\lambda(z_\lambda)dz_\lambda\}$ is a **meromorphic differential form** on R. We often denote it simply by ω. When every g_λ is holomorphic, we call ω a **holomorphic differential form**. For the closed Riemann surface R determined by a nonsingular

projective curve, it is known from GAGA that every meromorphic differential form on R is a rational differential form, and therefore holomorphic differential forms in both senses are the same.

THEOREM 4.2. *On a closed Riemann surface of genus g, there exist g linearly independent holomorphic differential forms.*

This is obtained as a corollary to the Riemann-Roch theorem (Theorem 4.3) to be stated later.

EXAMPLE 4.2 (THE ONE-DIMENSIONAL COMPLEX TORUS). Let τ be an arbitrarily fixed complex number such that $\Im\tau > 0$. Let $\Lambda = \{m+n\tau | m, n \in \mathbf{Z}\}$, which can be displayed as a set of lattice points on the complex plane. Identifying two points z_1 and z_2 in $\mathbf{C}$ such that

$$z_1 - z_2 \in \Lambda,$$

we get the quotient space $E_\tau = \mathbf{C}/\Lambda$. An arbitrary point $z \in \mathbf{C}$ can be identified with an interior point or a boundary point of the shaded parallelogram (Figure 4.3). Furthermore, points on the parallel sides of the parallelogram are identified as illustrated in Figure 4.4. Thus E_τ is topologically a torus with genus equal to 1 (see Figure 4.5). On the other hand, for each point of E_τ, we may take the coordinates in $\mathbf{C}$ as a local coordinate system. Thus E_τ has the structure of a complex manifold. We call E_τ a **one-dimensional complex torus**. A meromorphic function f on E_τ can be identified with a meromorphic function $\tilde{f}$ on $\mathbf{C}$ with double periodicity, that is,

$$\tilde{f}(z+m+n\tau) = \tilde{f}(z), \quad m, n \in \mathbf{Z}.$$

Such a function $\tilde{f}(z)$ is called an **elliptic function** with **fundamental periods** $1, \tau$. Meromorphic differential forms ω on E_τ can be identified with meromorphic differential forms

$$\tilde{g}(z)$$

that are invariant under the transformation $z \mapsto z + m + n\tau$, $m, n \in \mathbf{Z}$, and thus

$$\tilde{g}(z+m+n\tau) = \tilde{g}(z), \quad m, n \in \mathbf{Z},$$

that is, $g(z)$ is an elliptic function with fundamental periods $1, \tau$. It follows that a holomorphic differential form ω on $E\tau$ is determined by a holomorphic doubly periodic function $\tilde{g}(z)$ and that $\omega = cdz$ (c is a constant), since a holomorphic elliptic function is a constant function. We have shown that there is only one linearly independent holomorphic differential form, confirming Theorem 4.2 in the case of the one-dimensional complex torus.

We shall continue to deal with meromorphic functions on a closed Riemann surface R. Just as in § 3.1, a **divisor** means a linear combination $D = \sum_{i=1}^{k} m_i P_i$, where $P_1, \ldots, P_k$ are points in R and the coefficients $m_1, \ldots, m_k$ are integers. As before, the set of all divisors forms an Abelian group. When all $m_i > 0$, we say D is a **positive divisor** and write $D > 0$. (As before, we also write $D \geq 0$ if $D > 0$ or $D = 0$.) Suppose a meromorphic function f on R has poles of order n_i at P_i,

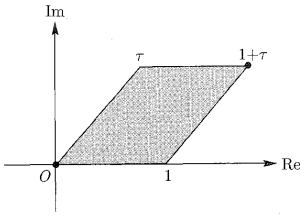


FIGURE 4.3

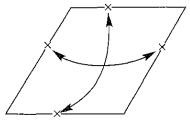


FIGURE 4.4

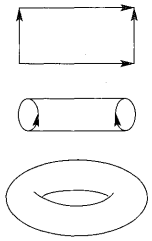


FIGURE 4.5

$1 \leq i \leq t$, and zeros of order m_j at q_j, $1 \leq j \leq s$. We define, as in § 3.1,

$$(f)_0 = \sum_{j=1}^{s} m_j Q_j$$

$$(f)_\infty = \sum_{i=1}^{t} n_i P_i$$

$$(f) = (f)_0 - (f)_\infty.$$

We call (f) the **principal divisor** determined by f. Two divisors D_1, D_2 are said to be **linearly equivalent** if there is a meromorphic function f such that

$$D_1 - D_2 = (f).$$

In this case, we write $D_1 \sim D_2$. For any divisor $D = \sum m_i P_i$, we define the **degree** by

$$\deg D = \sum_{i=1}^{k} m_i.$$

We may show, as in § 3.1, that $\deg(f) = 0$ for the principal divisor of a meromorphic function f. We also rewrite

$$D = \sum_{i=1}^{k} m_i P_i - \sum_{j=1}^{\ell} n_j Q_j, \quad m_i \geq 1, n_j \geq 1,$$

and define

$$\begin{aligned}\mathbf{L}(D) = \{f : &\text{ meromorphic on } R | f = 0 \text{ or} \\ &f \text{ has poles of order at most } m_i \text{ at } P_i, \\ &\text{and has zero points of order at least } n_j \text{ at } Q_j\}.\end{aligned}$$

This is a vector space over $\mathbf{C}$. It can be shown that it is finite-dimensional when R is a closed Riemann surface. Set

$$\ell(D) = \dim_{\mathbf{C}} \mathbf{L}(D).$$

We could have defined

$$\mathbf{L}(D) = \{f : \text{ meromorphic on } R | f \equiv 0 \text{ or } (f) + D \geq 0\},$$

as we did in § 3.1. We also see that if divisors D and F are linearly equivalent, then $\mathbf{L}(D)$ and $\mathbf{L}(F)$ are isomorphic as vector spaces. To be specific, let $D - F = (h)$. Then

$$f \in \mathbf{L}(D) \longmapsto fh \in \mathbf{L}(F)$$

is an isomorphism from $\mathbf{L}(D)$ onto $\mathbf{L}(F)$.

Suppose a meromorphic differential form ω on a closed Riemann surface R has zeros of order m_i at Q_i, $1 \leq i \leq k$, and poles of order n_j at P_j, $1 \leq j \leq \ell$. We set

$$\begin{aligned}(\omega)_0 &= \sum_{i=1}^{k} m_i Q_i \\ (\omega)_\infty &= \sum_{j=1}^{\ell} n_j P_j \\ (\omega) &= (\omega)_0 - (\omega)_\infty,\end{aligned}$$

and call (ω) the **canonical divisor**. Suppose τ is another rational differential form on R. Then there exists a meromorphic function f such that

$$\omega = f\tau$$

and

$$(\omega) = (f) + (\tau).$$

This means that the canonical divisors of meromorphic forms are all linearly equivalent (cf. Lemma 3.4). We denote the canonical divisor of the closed Riemann surface R by K_R or simply by K.

With these preparations the Riemann-Roch theorem for closed Riemann surfaces of genus g takes the same form as in the case of algebraic curves.

THEOREM 4.3 (RIEMANN-ROCH THEOREM). *For a closed Riemann surface R of genus g, we have*

$$\ell(D) - \ell(K - D) = \deg D - g + 1,$$

where K is the canonical divisor of R.

To prove this theorem, it is convenient to use the sheaf cohomology theory. The Riemann-Roch theorem has now been proved for compact complex manifolds in general and can be considered as a special case of the Atiyah-Singer index theorem. This is one of the great achievements of mathematics in the 20th-century, ranging from topology to analysis.

Now the Riemann-Roch theorem has various applications. First, we shall derive Theorem 4.2. By applying the Riemann-Roch theorem to the zero divisor, we get

$$\ell(0) - \ell(K) = 1 - g.$$

Here $\mathbf{L}(0)$ is the totality of holomorphic functions on the closed Riemann surface R. By the maximum-value theorem, there are only constant functions in $\mathbf{L}(0)$, hence $\ell(0) = 1$ so that

$$\ell(K) = g.$$

On the other hand, if we can write $K = (\omega)$ with a meromorphic form ω, the condition $(h) + K \geq 0$ for $h \in \mathbf{L}(K)$ means that $h\omega$ is a holomorphic differential form. Conversely, if τ is a holomorphic differential form, then there exists a meromorphic function h on R such that

$$\tau = h\omega,$$

and hence $(h) + K \geq 0$ and $h \in \mathbf{L}(K)$. Therefore

$$h \in \mathbf{L}(K) \longmapsto h\omega \in \{\text{holomorphic differential forms on } R\}$$

is an isomorphism of vector spaces. From $\dim_{\mathbf{C}} \mathbf{L}(K) = g$ we see that there are g linearly independent holomorphic differential forms. This proves Theorem 4.2. We shall further prove

$$\deg K = 2g - 2.$$

By the Riemann-Roch theorem we have

$$\ell(K) - \ell(0) = \deg K - g + 1$$

and hence $\deg K = 2g - 2$.

Next we are going to use the theorem and show that a closed Riemann surface has the structure of a nonsingular projective curve. Just as in § 3.2, for a divisor D on R such that $\mathbf{L}(D) \neq \{0\}$, we take a basis $\{\phi_0, \phi_1, \ldots, \phi_N\}$ of the vector space $\mathbf{L}(D)$ and define a mapping

$$\psi_{|D|} : P \in R \longmapsto (\phi_0(P), \phi_1(P), \ldots, \phi_N(P) \in \mathbf{P}^N(\mathbf{C}).$$

When P is a common zero or a common pole of $\phi_0, \phi_1, \ldots, \phi_N$, we cannot define $\psi_{|D|}$ directly. If P is a common zero, let ϕ_j have the lowest order of zero at P and define

$$\Psi_{|D|}(P) = \left(\frac{\phi_0}{\phi_j}(P) : \cdots : \frac{\phi_{j-1}}{\phi_j}(P) : 1 : \frac{\phi_{j+1}}{\phi_j}(P) : \cdots : \frac{\phi_N}{\phi_j}(P)\right),$$

so that the mapping $\psi_{|D|}$ is defined at P. If P is a pole, $\psi_{|D|}(P)$ can be similarly defined.

THEOREM 4.4. *Assume* $\deg D \geq 2g+1$. *Then* $\psi_{|D|}$ *defines an imbedding of* R *into* $\mathbf{P}^{\deg D-g}(\mathbf{C})$ *whose image is a closed complex submanifold. Hence* $\psi_{|D|}(R)$ *is a nonsingular projective variety by Chow's theorem (Theorem 4.1).*

PROOF. Let $N = \deg D - g$ in the following. First we show that $\psi_{|D|}$ is injective, that is, $\psi_{|D|}(P) \neq \psi_{|D|}(Q)$ if $P \neq Q$ in R. As we showed in §3.1 (a), if a divisor F on R has $\deg F < 0$, then $\mathbf{L}(F) = \{0\}$ for the following reason. If there is an $f \neq 0$ in $\mathbf{L}(F)$, then $(f) + F \geq 0$. On the other hand, we have by assumption

$$\deg((f) + F) < 0,$$

which contradicts $(f) + F \geq 0$. Applying this to the divisor $K - (D - P - Q)$, we get

$$K - (D - P - Q) = 2g - 2 - \deg D + 2 \leq -1,$$

from which we have $\mathbf{L}(K - (D - P - Q)) = \{0\}$. Similarly, we have

$$\mathbf{L}(K - (D - P)) = \{0\}, \quad \mathbf{L}(K - (D - Q)) = \{0\}, \quad \mathbf{L}(K - D) = \{0\}.$$

From these we obtain by the Riemann-Roch theorem

$$\begin{aligned} \ell(D) &= \deg D - g + 1 = N + 1 \\ \ell(D_P) &= \deg D - g = N \\ \ell(D - P - Q) &= \deg D - g - 1 = N - 1. \end{aligned}$$

Clearly, by definition, we have

$$\mathbf{L}(D - P - Q) \subset \mathbf{L}(D - P) \subset \mathbf{L}(D),$$

where dimension increases by 1. Thus there exist two meromorphic functions $\psi_0 \in \mathbf{L}(D), \psi_0 \notin \mathbf{L}(D - P)$, and $\psi_1 \in \mathbf{L}(D - P), \psi_1 \notin \mathbf{L}(D - P - Q)$. From the choice of ψ_0 and ψ_1 we have $\psi_0(P) = 0, \psi_1(P) = 0, \psi_1(Q) \neq 0$. By choosing a basis $\{\psi_2, \psi_3, \ldots, \psi_N\}$ of the complex vector space $\mathbf{L}(D - P - Q)$ we have

$$\begin{aligned} &(\psi_0(P) : \psi_1(P) : \psi_2(P) : \cdots : \psi_N(P)) = (1 : 0 : 0 \cdots : 0) \\ &(\psi_0(Q) : \psi_1(Q) : \psi_2(Q) : \cdots : \psi_N(Q)) = (a : 1 : 0 \cdots : 0). \end{aligned} \tag{4.1}$$

Since $\{\psi_0, \psi_2, \ldots, \psi_N\}$ can be written as linear combinations with complex coefficients of the basis elements $\{\psi_0, \psi_1, \psi_2, \ldots, \psi_N\}$, (4.1) shows that $\psi_{|D|}(P) \neq \psi_{|D|}(Q)$.

Next we show that $\psi_{|D|}$ is locally an imbedding as a submanifold. We have as before

$$\begin{aligned} \ell(D) &= N \\ \ell(D_P) &= N - 1 \\ \ell(D - 2P) &= N - 2. \end{aligned}$$

If we take a basis $\{\psi_2, \psi_3, \ldots, \psi_N\}$ of $\mathbf{L}(D - 2P)$ and ψ_1, ψ_0 such that

$$\psi_1 \in \mathbf{L}(D - P), \quad \psi_1 \notin \mathbf{L}(D - 2P), \quad \psi_0 \in \mathbf{L}(D), \quad \psi_0 \notin \mathbf{L}(D - P),$$

then $\{\psi_0, \psi_1, \ldots, \psi_N\}$ is a basis of $\mathbf{L}(D)$. Now ψ has a zero of order 1 at P and $\psi_0(P) \neq 0$, and hence ψ_1/ψ_0 has a zero of order 1 at P. On the other hand, each

$\psi_i, i \geq 2$, has a zero of order at least 2 at P , and so does $f_i = \psi_/\psi_0$ for each $i \geq 2$. Therefore the mapping

$$\Psi : z \in U \longmapsto (f_1(z), f_2(z), \ldots, f_N(z)) \in \mathbf{C}^N,$$

defined in a neighborhood U of P, has non-zero differential $d\Psi_P$ at P, since $df_1(P) \neq 0$ and $df_i(P) = 0, i \geq 2$. Hence Ψ is locally an imbedding and $\Psi(U)$ is a 1-dimensional closed complex submanifold in a closed neighborhood of $\Psi(P)$. Since $\{\phi_0, \phi_1, \ldots, \phi_N\}$ are linear combinations with complex coefficients of $\{\psi_0, \psi_1, \ldots, \psi_N\}$, we see that $d_P\psi_{|D|} \neq 0$, and hence $\psi_{|D|}(R)$ has the structure of a 1-dimensional complex submanifold.

From the two observations above, we see that $\psi_{|D|}(R)$ is a closed complex submanifold of $\mathbf{P}^N(\mathbf{C})$, and complete the proof of the theorem.

EXAMPLE 4.3 (ONE-DIMENSIONAL COMPLEX TORUS). We use the notation in Example 4.2. We denote by $[z]$ the point of the one-dimensional complex torus E_τ determined by z in the complex plane. The lattice points determine the point $[0]$ of E_τ, that is, $[m + n\tau] = [0], m, n \in \mathbf{Z}$. Since the holomorphic differential form dz on E_τ has no zero, we can take 0 as the canonical divisor. Recall also that the genus of E_τ is 1. Hence for any divisor D with $\deg D > 0$ the Riemann-Roch theorem says

$$\ell(D) = \deg D.$$

In particular, for any positive integer n, we get

$$\ell(n[0]) = n.$$

On the other hand, from the theory of elliptic functions we get

$$\mathbf{L}([0]) = \mathbf{C} \cdot 1.$$

(There is no elliptic function that has only one pole of order 1 in the fundamental parallelogram.) We define the Weierstrass $\mathfrak{p}$ function by

$$\mathfrak{p}(z) = \frac{1}{z^2} + \sum_{(m,n)\in\mathbf{Z}^2,(m,n)\neq(0,0)} \left\{ \frac{1}{(z + m + n\tau)^2} - \frac{1}{(m + n\tau)^2} \right\}.$$

It is known from the theory of elliptic functions that

$$\begin{aligned} \mathbf{L}(2[0]) &= \mathbf{C} \cdot 1 \oplus \mathbf{C} \cdot \mathfrak{p}(z) \\ \mathbf{L}(3[0]) &= \mathbf{C} \cdot 1 \oplus \mathbf{C} \cdot \mathfrak{p}(z) \oplus \mathbf{C} \cdot \mathfrak{p}'(z), \end{aligned}$$

which agree with the results from the Riemann-Roch theorem, as naturally expected. According to Theorem 4.4, $\psi_{|D|}$ imbeds E_τ into the projective space. This can be seen as follows. As a basis of $\mathbf{L}(3[0])$ take $\{1, \mathfrak{p}(z), \mathfrak{p}'(z)\}$ and consider the mapping $\phi = \psi_{|D|}$,

$$\phi : [z] \in E_\tau \longmapsto (1 : \mathfrak{p}(z) : \mathfrak{p}'(z)) \in \mathbf{P}^2(\mathbf{C}).$$

In a neighborhood of $[z] = [0]$ define ϕ by

$$\phi : [z] \in E_\tau \longmapsto \left(\frac{1}{\mathfrak{p}'(z)} : \frac{\mathfrak{p}(z)}{\mathfrak{p}'(z)} : 1 \right).$$

Since $\mathfrak{p}'(z)$ has a pole of order 3 at $z = 0$ and $\mathfrak{p}(z)$ has a pole of order 2 at $z = 0$, ϕ is well-defined in a neighborhood of $[0]$ and $\phi([0]) = (0 : 0 : 1)$. Now the Weierstrass function $\mathfrak{p}$ and its derivative $\mathfrak{p}'(z)$ satisfy the relation

$$\mathfrak{p}'(z)^2 = 4\mathfrak{p}(z)^3 - g_2(\tau)\mathfrak{p}(z) - g_3(\tau),$$

where

$$g_2(\tau) = 60 \sum_{(m,n)\in\mathbf{Z}^2,(m,n)\neq(0,0)} \left\{\frac{1}{(m+n\tau)^4}\right\}$$

$$g_3(\tau) = 140 \sum_{(m,n)\in\mathbf{Z}^2,(m,n)\neq(0,0)} \left\{\frac{1}{(m+n\tau)^6}\right\}.$$

If we set

$$\triangle(\tau) = g_2(\tau)^3 - 27g_3(\tau)^2,$$

it is known that $\triangle(\tau) \neq 0$. Further, if we set

$$F_\tau = x_0x_2^2 - 4x_1^3 + g_2(\tau)x_0^2x_1 + g_3(\tau)x_0^3,$$

we can verify that

$$\phi(E_\tau) \subset V(F_\tau).$$

In fact, ϕ is surjective and injective. In order to prove this, we first show that $\phi(E_\tau) = V(F_\tau)$. The point $(0 : 0 : 1) \in V(F_\tau)$ is the image $\phi([0])$. This is the only intersection of $V(E_\tau)$ and $x_0 = 0$. So we now consider $V(E_\tau) - \{(0 : 0 : 1)\}$. If we set $x = x_1/x_0, y = x_2/x_0$, then a point $(1 : x : y) \in V(F_\tau) - \{(0 : 0 : 1)\}$ satisfies

$$y^2 = 4x^3 - g_2(\tau)x - g_3(\tau).$$

In this case, there exists a $z \in \mathbf{C}$ such that

$$x = \mathfrak{p}(z).$$

(See Exercise 4.1 (i).) Since

$$\mathfrak{p}'(z)^2 = 4\mathfrak{p}(z)^3 - g_2(\tau)\mathfrak{p}(z) - g_3(\tau),$$

it follows that $y = \mathfrak{p}'(z)$ or $y = -\mathfrak{p}'(z)$. Since $\mathfrak{p}(-z) = \mathfrak{p}(z), \mathfrak{p}'(-z) = -\mathfrak{p}'(z)$, we have $\phi([z]) = (1 : x : y)$ or $\phi([-z]) = (1 : x : y)$, and thus $\phi(E_\tau) = V(F_\tau)$. Next we show that $\phi([z]) = \phi([z'])$ implies $[z] = [z']$. If $[z] \neq [0]$, then we have

$$(1 : \mathfrak{p}(z) : \mathfrak{p}'(z)) = (1 : \mathfrak{p}(z') : \mathfrak{p}'(z')).$$

From $\mathfrak{p}(z) = \mathfrak{p}(z')$ we obtain

$$z = \pm z' + m + n\tau. \quad m, n \in \mathbf{Z}.$$

Since $\mathfrak{p}'(z) = \mathfrak{p}'(z')$ and since $\mathfrak{p}'(z)$ is an odd function, we conclude that

$$z = \pm z' + m + n\tau.$$

Therefore $[z] = [z']$. (In the case $\mathfrak{p}'(z) = \mathfrak{p}'(z')$, the argument above does not apply. But in the fundamental parallelogram in Figure 4.3, $\mathfrak{p}'(z) = 0$ is valid only at three points: $1/2, \tau/2, (1 + \tau)/2$; this fact implies $[z] = [z']$. Note that $[1/2] = [-1/2]$, etc.) Finally, if $[z] = [0]$, then $\phi([0]) = (0 : 0 : 1)$ implies $[z'] = [0]$. We can thus conclude that ϕ is injective. We can also show, by a similar argument, that the mapping ϕ has non-zero differential $d\phi_{[z]} \neq 0$ at each point $[z]$ of E_τ. Hence we can identify E_τ with the cubic curve $V(F_\tau)$ in $\mathbf{P}^2(\mathbf{C})$. As can be seen from this example, Theorem 4.4 gives a generalization of one aspect of the theory of elliptic functions. Incidentally, the cubic curve $V(F_\tau)$ is nonsingular because $\triangle(\tau) \neq 0$, as was shown in §3.3.

§4.2. Period matrices

A closed Riemann surface R of genus g is topologically shaped like a floating bag with g holes. Further R has g linearly independent holomorphic differential forms. We shall derive interesting properties of the surface from these facts.

First, we define the integral

$$\int_\gamma \omega$$

of a differential form ω along a curve γ. Take a covering of the Riemann surface R with local coordinate neighborhoods $\{(U_\lambda, z_\lambda)\}$. By a piecewise smooth curve γ we understand a continuous mapping γ from the interval $[a, b]$ into R with the following properties: we can subdivide $[a, b]$ into a finite number of intervals $[a_j, a_{j+1}]$, $0 \leq j \leq n-1$, $a_0 = a < a_1 < a_2 \cdots < a_n = b$ such that $\gamma([a_j, a_{j+1}]) \subset U_{\lambda_j}$ and, using the local coordinate, we write γ as

$$z_\lambda(t) = x_{\lambda_j}(t) + iy_{\lambda_j}(t)$$

($x_{\lambda_j}(t)$, $y_{\lambda_j}(t)$ are infinitely differentiable). We now write the meromorphic differential form ω in the form $f_\lambda(z_\lambda)dz_\lambda$. If there is no pole of ω on $\gamma([a, b])$, we define the integral of ω along γ by

$$\int_\gamma \omega = \sum_{j=0}^{n-1} \int_{a_j}^{a_{j+1}} f_{\lambda_j}(z_{\lambda_j}(t))z'_{\lambda_j}(t)dt. \tag{4.2}$$

That this definition does not depend on the way we subdivide $[a, b]$ follows from the theory of the change of variables in integration.

We shall be concerned with closed curves γ, that is, curves on R such that $\gamma(a) = \gamma(b)$. According to Cauchy's integral theorem, if a closed curve γ is continuously deformable to a single point, then

$$\int_\gamma \omega = 0$$

for any differential form ω that is holomorphic inside γ and in a neighborhood of γ. The formula does not necessarily hold if γ is not continuously deformable to a single point. On a closed surface of genus greater than or equal to 1, there are closed curves like α, β as in Figure 4.7 that cannot be continuously shrunk to a single point. (In contrast, every closed curve on the sphere can be continuously shrunk to a single point.)

In the following, we consider only closed curves γ on a Riemann surface R. A continuous closed curve can be approximated by a piecewise smooth curve. In order to pick up meaningful closed curves on R, we need the 2-dimensional homology group $H_1(R, \mathbf{Z})$. For the definition, we refer the reader to *Algebraic Topology: An Introduction* by W.S. Massey, Springer, 1977. For a Riemann surface R of genus

FIGURE 4.6. γ can be continuously shrunk to a single point

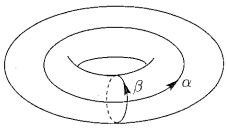


FIGURE 4.7

g, each element of $H_1(R, \mathbf{Z})$ can be uniquely written as a linear combination with integral coefficients of the closed curves $\alpha_1, \alpha_2, \ldots, \alpha_g, \beta_1, \beta_2, \ldots, \beta_g$, that is,

$$H_1(R, \mathbf{Z}) = \mathbf{Z}\alpha_1 \oplus \cdots \oplus \mathbf{Z}\alpha_g + \mathbf{Z}\beta_1 \oplus \cdots \oplus \mathbf{Z}\beta_g$$

(we regard $-\alpha_j$, $-\beta_j$ as curves with reversed orientation). Thus $\alpha_1, \alpha_2, \ldots, \alpha_g$, $\beta_1, \beta_2, \ldots, \beta_g$, form a basis for $H_1(R, \mathbf{Z})$ over $\mathbf{Z}$. We can define the notion of **intersection number** so that

$$\begin{aligned} \alpha_i \cdot \alpha_j &= 0 \\ \alpha_i \cdot \beta_j &= \delta_{ij} \\ \beta_i \cdot \beta_j &= 0. \end{aligned} \tag{4.3}$$

Here $\delta_{ij} = 0, i \neq j$, and $\delta_{ii} = 1$. (We have, for example, $\beta_j \cdot \alpha_i = -\delta_{ij}$, reflecting the orientation of the closed surface as well as the orientation of the curves at the intersection. For details, see Massey's book cited above.) A $\mathbf{Z}$-basis of $H^1(R, \mathbf{Z})$ satisfying (4.3) is called a **symplectic basis**. There are many choices; one that is used frequently is described in Figure 4.9. For this basis, if we cut open the Riemann surface along each of these closed curves, then we get a $4g$-gon that can be used for various kinds of computation, as we shall see in a moment.

Now on the Riemann surface R we choose two arbitrary curves γ_1 and γ_2 from P to Q and consider the integrals of a holomorphic differential form ω on R,

$$\int_{\gamma_1} \omega, \quad \int_{\gamma_2} \omega.$$

Take the composite curve $\gamma_1\gamma_2^{-1}$, where γ_2^{-1} is the curve γ_2 with reversed orientation. The closed curve $\gamma_1\gamma_2^{-1}$ determines an element of the 1-dimensional homology

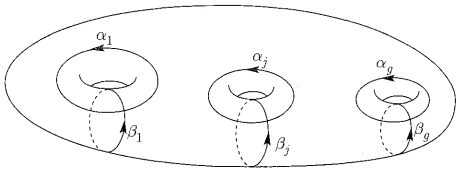


FIGURE 4.8

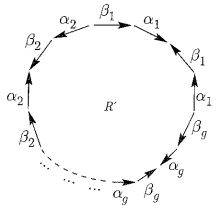


FIGURE 4.9. By pasting along the arrow directions of α_j, β_j a closed surface of genus g results.

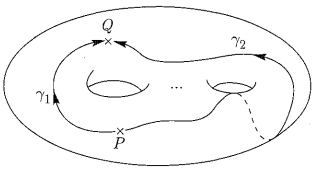


FIGURE 4.10

group $H_1(R, \mathbf{Z})$, denoted by $[\gamma_1\gamma_2^{-1}]$. (See Figure 4.10.) We have a unique representation of the form

$$[\gamma_1\gamma_2^{-1}] = \sum_{j=1}^{g} m_j\alpha_j + \sum_{k=1}^{g} n_k\beta_k, \quad m_j, n_k \in \mathbf{Z},$$

$\{\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g\}$ being a basis of the 1-dimensional homology group $H_1(R, \mathbf{Z})$. For the time being we use a symplectic basis as in Figure 4.9. From the representation of $[\gamma_1\gamma_2^{-1}]$ above we get

$$\int_{\gamma_1} \omega - \int_{\gamma_2} \omega = \sum_{j=1}^{g} m_j \int_{\alpha_j} \omega + \sum_{k=1}^{g} n_k \int_{\beta_k} \omega.$$

In other words, if two closed curves γ and γ' starting at P determine the same element of $H_1(R, \mathbf{Z})$, then for any holomorphic differential form ω on R we have

$$\int_\gamma \omega = \int_{\gamma'} \omega.$$

This is a generalization of Cauchy's integral theorem. What is important here is that if we take an arbitrary curve γ from a point P to another Q, then, although $\int_\gamma \omega$ depends on γ, it remains the same modulo a linear combination with integral coefficients of

$$\int_{\alpha_j} \omega, \quad \int_{\beta_j} \omega.$$

Suppose we pick a linearly independent system of g holomorphic differential forms $\omega_1, \omega_2, \ldots, \omega_g$ on R. Since an arbitrary holomorphic differential form is a linear combination with complex coefficients of $\omega_1, \omega_2, \ldots, \omega_g$, we may easily imagine an important role in integration theory for the $2g \times g$ matrix

$$\Omega = \begin{pmatrix} \int_{\alpha_1} \omega_1 \cdots \int_{\alpha_1} \omega_g \\ \vdots \quad\quad \vdots \\ \int_{\alpha_g} \omega_1 \cdots \int_{\alpha_g} \omega_g \\ \int_{\beta_1} \omega_1 \cdots \int_{\beta_1} \omega_g \\ \vdots \quad\quad \vdots \\ \int_{\beta_g} \omega_1 \cdots \int_{\beta_g} \omega_g \end{pmatrix}.$$

We call it the **period matrix** of the closed Riemann surface. The period matrix Ω varies depending on the choice of a symplectic basis and of a linearly independent set of holomorphic forms. To see how Ω changes, first observe that for any curve γ on R and for any holomorphic form ω we have

$$\overline{\int_\gamma \omega} = \int_\gamma \bar{\omega},$$

where the bar means the complex conjugate, and $\int_\gamma \bar{\omega}$ is defined, similarly to (4.2), by

$$\int_\gamma \bar{\omega} = \sum_{j=0}^{n-1} \int_{a_j}^{a_{j+1}} \overline{f_{\lambda_j}(z_{\lambda_j}(t)) z'_{\lambda_j}(t)} dt.$$

In order to describe properties of the period matrix Ω we introduce the matrix

$$J = \begin{pmatrix} 0 & I_g \\ -I_g & 0 \end{pmatrix},$$

where I_g stands for the $g \times g$ unit matrix.

THEOREM 4.5 (RIEMANN'S RELATION). *The period matrix Ω has the following properties:*

(i) ${}^t\Omega J \Omega = 0$.

(ii) $\sqrt{-1}\,{}^t\Omega J \bar{\Omega}$ *is a positive-definite Hermitian matrix.*

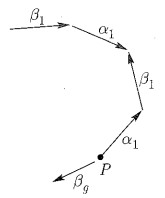


FIGURE 4.11

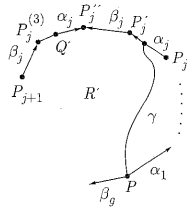


FIGURE 4.12

PROOF. For the proof we use surface integration on R and the Stokes theorem. We also use those closed curves $\{\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g\}$ in Figure 4.9. The complement R' of the union of these closed curves in the Riemann surface R is the interior of a $4g$-gon, and the integral of a holomorphic differential form ω along a curve from P, the initial point of α_j, β_j, to an arbitrary point Q of R', written

$$h(Q) = \int_P^Q \omega,$$

is uniquely determined provided the integration path is taken within R'. Thus h is a holomorphic function on R', which can be continuously extended to the boundary of R'. The boundaries α_j and β_j appear as pairs twice (see Figure 4.11), and at points Q, Q' (which coincide as a point on R) h has different values. Suppose P is the initial point of α_1 and β_g on the boundary of R', and we name the points P'_j, etc. as in Figure 4.12. All these points represent one and the same point P on R, but when R is cut out along the closed curves α_j, β_j, $1 \le j \le g$, they are different points in the resulting $4g$-gon. As in Figure 4.12, take Q, Q' on α_j that represent the same point on R. Choose a curve γ joining P and Q within R'. Then we get

$$\begin{aligned} \int_P^Q \omega &= \int_\gamma \omega \\ \int_P^{Q'} &= \int_\gamma \omega + \int_Q^{P'_j} \omega + \int_{\beta_j} \omega + \int_{P''_j}^{Q'} \omega \end{aligned}$$

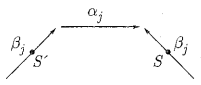


FIGURE 4.13

Here integrating ω from Q to P_j' along α_j and integrating ω from Q' to P_j'' along α_j give the same result as line integrals in R. Hence

$$\int_{P_j''}^{Q'} \omega = -\int_{Q'}^{P_j''} \omega = -\int_{Q}^{P_j'} \omega$$

and

$$\int_P^{Q'} \omega = \int_P^Q \omega + \int_{\beta_j} \omega.$$

Similarly, if we take S, S' on β_j (see Figure 4.13) that represent one and the same point in R, then

$$\int_P^{S'} \omega = \int_P^S \omega - \int_{\alpha_j} \omega.$$

With these preparations we now derive (i) of the theorem. For holomorphic differential forms ω_j, ω_k on the Riemann surface R already mentioned before, we have

$$\omega_j \wedge \omega_k \equiv 0,$$

where $\wedge$ is exterior product. Hence

$$0 = \int_R \omega_j \wedge \omega_k.$$

On the other hand, we can find a holomorphic function h_j on R' such that

$$dh_j = \omega_j.$$

Therefore, using the Stokes theorem, we have

$$\begin{aligned}
0 &= \int_R dh_j \wedge \omega_k = \int_{R'} d(h_j\omega_k) \\
&= \sum_{i=1}^{g} \left(\int_{\alpha_i^+} h_j\omega_k + \int_{\beta_i^+} h_j\omega_k - \int_{\alpha_i^-} h_j\omega_k - \int_{\beta_i^+} h_j\omega_k \right) \\
&= \sum_{i=1}^{g} \left\{ \int_{\alpha_i^+} h_j\omega_k + \int_{\beta_j^+} h_j\omega_k - \int_{\alpha_i^+} \left(h_j + \int_{\beta_i} \omega_j \right) \omega_k - \int_{\beta_i^+} \left(h_j - \int_{\alpha_i} \omega_j \right) \omega_k \right\} \\
&= \sum_{i=1}^{g} \left\{ \int_{\alpha_i} \omega_j \int_{\beta_i} \omega_k - \int_{\beta_i} \omega_j \int_{\alpha_i} \omega_k \right\}.
\end{aligned}$$

Here α_j^+ denotes α_j with Q and α_j^- denotes α_j with Q' in Figure 4.12. In Figure 4.13, β_j^+ denotes β_j with S and β_j^- denotes β_j with S'. The last term represents

the (j,k)-component of the matrix $-{}^t\Omega J\Omega$, and hence (i) is proved. Using the fact that

$$\sqrt{-1} \int_R \omega \wedge \bar{\omega} > 0$$

for any holomorphic form ω on R, we can prove (ii) in a similar fashion.

Using this theorem, we may rewrite the period matrix in a nice form. Setting

$$A = \begin{pmatrix} \int_{\alpha_1} \omega_1 \cdots \int_{\alpha_1} \omega_g \\ \cdots\cdots\cdots\cdots \\ \int_{\alpha_g} \omega_1 \cdots \int_{\alpha_g} \omega_g \end{pmatrix}, \quad B = \begin{pmatrix} \int_{\beta_1} \omega_1 \cdots \int_{\beta_1} \omega_g \\ \cdots\cdots\cdots\cdots \\ \int_{\beta_g} \omega_1 \cdots \int_{\beta_g} \omega_g \end{pmatrix}$$

we show that

$$\det A \neq 0.$$

If we had $\det A = 0$, there would be a vector $\mathbf{a} = (a_1, a_2, \cdots, a_g) \neq (0, 0, \ldots, 0)$ such that

$$\mathbf{a}A = 0.$$

On the other hand, since

$$\sqrt{-1}\,{}^t\Omega J\bar{\Omega} = \sqrt{-1}\,({}^tA\bar{B} - {}^tB\bar{A}),$$

we get

$$\sqrt{-1}\,{}^t\Omega J\bar{\Omega}\,{}^t\bar{\mathbf{a}} = 0.$$

Since $\sqrt{-1}\,{}^t\Omega J\bar{\Omega}$ is a positive-definite Hermitian matrix, we have $\mathbf{a} = 0$, a contradiction. Hence $\det A \neq 0$. If we set

$$\tau = BA^{-1},$$

then $\begin{pmatrix} I_g \\ \tau \end{pmatrix}$ is also a period matrix, because if we define $\tilde{\omega}_i, 1 \leq i \leq g$, by

$$(\tilde{\omega}_1, \tilde{\omega}_2, \ldots, \tilde{\omega}_g) = (\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_g)A^{-1},$$

they are g linearly independent holomorphic differential forms on R such that

$$\begin{pmatrix} \int_{\alpha_1} \tilde{\omega}_1 \cdots \int_{\alpha_1} \tilde{\omega}_g \\ \vdots \quad\quad \vdots \\ \int_{\alpha_g} \tilde{\omega}_1 \cdots \int_{\alpha_g} \tilde{\omega}_g \\ \int_{\beta_1} \tilde{\omega}_1 \cdots \int_{\beta_1} \tilde{\omega}_g \\ \vdots \quad\quad \vdots \\ \int_{\beta_g} \tilde{\omega}_1 \cdots \int_{\beta_g} \tilde{\omega}_g \end{pmatrix} = \begin{pmatrix} I_g \\ \tau \end{pmatrix}.$$

We call $\begin{pmatrix} I_g \\ \tau \end{pmatrix}$ the **normalized period matrix**. From Riemann's relation (i) in Theorem 4.5 we obtain

$${}^t\tau = \tau,$$

that is, τ is a symmetric matrix. From (ii) we find that the imaginary part $\operatorname{Im}\tau$ is a positive-definite symmetric matrix. Summarizing, we obtain

COROLLARY 4.1. *Let $\omega_1, \omega_2, \ldots, \omega_g$ be g linearly independent holomorphic forms such that*

$$\int_{\alpha_i} \omega_j = \delta_{ij} \tag{4.4}$$

and define the $g \times g$ matrix τ by

$$\tau = \begin{pmatrix} \int_{\beta_1} \omega_1 \cdots \int_{\beta_1} \omega_g \\ \vdots \quad \vdots \\ \int_{\beta_g} \omega_1 \cdots \int_{\beta_g} \omega_g \end{pmatrix}.$$

Then τ is a complex symmetric matrix and $Im\ \tau$ is a positive-definite symmetric matrix.

A basis $\{\omega_1, \omega_2, \ldots, \omega_g\}$ satisfying (4.4) is called a **normalized basis** for holomorphic differential forms. It is uniquely determined if we fix a symplectic basis. The matrix τ is often called a **period matrix** of R. It depends only on R and its symplectic basis $\{\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g\}$, and not on the choice of holomorphic forms.

The discussions so far — in particular, the proofs for Theorem 4.5 and Corollary 4.1 — were given with a particular choice of symplectic basis; in fact, we can show that these results hold for an arbitrary symplectic basis.

Let $\{\alpha'_1, \ldots, \alpha'_g, \beta'_1, \ldots, \beta'_g\}$ be another choice of symplectic basis. Then there exists a unique $2g \times 2g$ integral matrix with $g \times g$ blocks

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

such that

$$\begin{pmatrix} \beta'_1 \\ \vdots \\ \beta'_g \\ \alpha'_1 \\ \vdots \\ \alpha'_g \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_g \\ \alpha_1 \\ \vdots \\ \alpha_g \end{pmatrix}.$$

Since the inverse matrix of M should also be integral, we must have

$$\det M = \pm 1.$$

(In fact, we can show that it is 1.) Because

$$\begin{aligned} &\alpha_i \cdot \alpha_j = 0, \quad \beta_i \cdot \beta_j = 0, \quad \alpha_i \cdot \beta_j = \delta_{ij}, \\ &\alpha'_i \cdot \alpha'_j = 0, \quad \beta'_i \cdot \beta'_j = 0, \quad \alpha'_i \cdot \beta'_j = \delta_{ij}, \end{aligned}$$

we have

$$MJ^tM = J.$$

We denote by $Sp(g, \mathbf{Z})$ the set of all $2g \times 2g$ integral matrices satisfying the condition above and call it the **symplectic group**, namely,

$$Sp(g, \mathbf{Z}) = \{M \in GL(2g, \mathbf{Z}) \mid MJ^tM = J\}.$$

We have then

$$\Omega' = \begin{pmatrix} \int_{\alpha'_1} \omega_1 \cdots \int_{\alpha'_1} \omega_g \\ \vdots \quad \vdots \\ \int_{\alpha'_g} \omega_1 \cdots \int_{\alpha'_g} \omega_g \\ \int_{\beta'_1} \omega_1 \cdots \int_{\beta'_1} \omega_g \\ \vdots \quad \vdots \\ \int_{\beta'_g} \omega_1 \cdots \int_{\beta'_g} \omega_g \end{pmatrix} = \begin{pmatrix} D & C \\ B & A \end{pmatrix} \begin{pmatrix} \int_{\alpha_1} \omega_1 \cdots \int_{\alpha_1} \omega_g \\ \vdots \quad \vdots \\ \int_{\alpha_g} \omega_1 \cdots \int_{\alpha_g} \omega_g \\ \int_{\beta_1} \omega_1 \cdots \int_{\beta_1} \omega_g \\ \vdots \quad \vdots \\ \int_{\beta_g} \omega_1 \cdots \int_{\beta_g} \omega_g \end{pmatrix}.$$

This implies that Ω' also satisfies Theorem 4.5. By switching to normalized period matrices we have

$$\begin{pmatrix} I_g \\ \tau' \end{pmatrix} = \begin{pmatrix} I_g \\ (A\tau + B)(C\tau + D)^{-1} \end{pmatrix}.$$

Since τ' also satisfies Lemma 4.1, it is symmetric. We set

$$\mathfrak{S}_g = \{\tau \mid \tau : g \times g \text{ complex symmetric}; \Im\tau : \text{positive-definite}\},$$

and call it the **Siegel upper half-space of degree** g. For $g = 1$, we have the ordinary upper half-plane. For an element $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp(g, \mathbf{Z})$ and for $\tau \in \mathfrak{S}_g$, we define

$$M \cdot \tau = (A\tau + B)(C\tau + D)^{-1}.$$

Then $M \cdot \tau \in \mathfrak{S}_g$. Each element $M \in Sp(g, \mathbf{Z})$ induces an isomorphism of $\mathfrak{S}_g$ onto itself. For $g = 1$, we have $Sp(1, \mathbf{Z}) = SL(2, \mathbf{Z})$, which is known to act on the upper half-plane as linear fractional transformations.

§4.3. Jacobian varieties

On a closed Riemann surface R we fix a symplectic basis $\{\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g\}$ of the first homology group $H_1(R, \mathbf{Z})$. By using a normalized basis $\{\omega_1, \ldots, \omega_g\}$ for holomorphic forms, the normalized period matrix $\begin{pmatrix} I_g \\ \tau \end{pmatrix}$ is given by

$$\tau = (\tau_{ij}), \quad \tau_{ij} = \int_{\beta_i} \omega_j, \quad 1 \le i, j \le g.$$

We denote by $e_1, e_2, \ldots, e_n$ the basis of the g-dimensional complex vector space $\mathbf{C}^g$ determined by the rows of $\begin{pmatrix} I_g \\ \tau \end{pmatrix}$. Fix a point P on the Riemann surface R. For any curves γ_1 and γ_2 on R from P to Q, we have seen that

$$\left(\int_{\gamma_2} \omega_1, \int_{\gamma_2} \omega_2, \ldots, \int_{\gamma_2} \omega_g\right) = \left(\int_{\gamma_1} \omega_1, \int_{\gamma_1} \omega_2, \ldots, \int_{\gamma_1} \omega_g\right) + \sum_{j=1}^{2g} n_j e_j, \quad n_j \in \mathbf{Z}.$$

The values of integrals vary with the choice of curves from P to Q, but they differ only up to a linear combination with integral coefficients of the $2g$ vectors determined by the period matrix. With this in mind, let us define the g-dimensional **complex torus** $J(R)$ following Example 4.2, in which the 1-dimensional torus was defined.

Two elements $(z_1, z_2, \dots, z_g)$ and $(z_1', z_2', \dots, z_g')$ in the g-dimensional complex vector space $\mathbf{C}^g$ are now identified when

$$(z_1', z_2', \dots, z_g') = (z_1, z_2, \dots, z_g) + \sum_{j=1}^{2g} n_j e_j, \quad n_j \in \mathbf{Z}.$$

Then each point in $\mathbf{C}^g$ can be identified with a unique element of the form

$$\sum_{j=1}^{2g} n_j e_j, \quad 0 \le a_j < 1. \tag{4.5}$$

In this identification space we introduce a complex manifold structure by using the complex coordinates naturally determined by those in $\mathbf{C}^g$. We denote this complex manifold by $J(R)$ and call it the **Jacobian variety** of the closed Riemann surface. As can be seen from (4.5), $J(R)$ is topologically the direct product of $2g$ circles S^1, but its properties as a complex manifold vary depending on the Riemann surface R, and more precisely, on the period matrix $\begin{pmatrix} I_g \\ \tau \end{pmatrix}$.

For a point $(z_1, z_2, \dots, z_g)$ of $\mathbf{C}^g$ let us write $[z_1, z_2, \dots, z_g]$ for the point of $J(R)$ it determines. For two points $[z_1, z_2, \dots, z_g]$ and $[w_1, w_2, \dots, w_g]$ in $J(R)$, we define the sum by

$$[z_1, z_2, \dots, z_g] + [w_1, w_2, \dots, w_g] = [z_1 + w_1, z_2 + w_2, \dots, z_g + w_g].$$

In this way, $J(R)$ becomes a group (Abelian group) with zero $[0, 0, \dots, 0]$ and $-[z_1, z_2, \dots, z_g] = [-z_1, -z_2, \dots, -z_g]$. Furthermore, we have a mapping

$$\phi : Q \in R \longmapsto \Big[\int_P^Q \omega_1, \int_P^Q \omega_2, \dots, \int_P^Q \omega_g\Big] \in J(R),$$

where the integral $\int_P^Q \omega_j$ depends on a curve γ from P to Q. That the image point in $J(R)$ is well-determined follows from the definition of $J(R)$. On the other hand, the mapping ϕ varies with the initial point P. If we choose another starting point P' and get the mapping $\phi' : R \to J(R)$, we obtain

$$\begin{aligned}&\Big[\int_P^Q \omega_1, \int_P^Q \omega_2, \dots, \int_P^Q \omega_g\Big] \\ &\quad = \Big[\int_{P'}^Q \omega_1, \int_{P'}^Q \omega_2, \dots, \int_{P'}^Q \omega_g\Big] + \Big[\int_P^{P'} \omega_1, \int_P^{P'} \omega_2, \dots, \int_P^{P'} \omega_g\Big]\end{aligned}$$

and

$$\phi(Q) = \phi'(Q) + \Big[\int_P^{P'} \omega_1, \int_P^{P'} \omega_2 \dots, \int_P^{P'} \omega_g\Big],$$

so that the difference is not essential, because it can be expressed as a sum of elements in $J(R)$.

THEOREM 4.6. *The mapping $\phi : R \to J(R)$ is an imbedding as a complex manifold.*

When $g = 1$, this means that ϕ is an isomorphism and that a closed Riemann surface of genus 1 is nothing but a one-dimensional complex torus. The fact that ϕ is a monomorphism (that is, $\phi(Q) = \phi(Q')$ implies $Q = Q'$) follows from the following theorem of Abel.

THEOREM 4.7 (ABEL'S THEOREM). *Let $\{P_1, P_2, \dots, P_N\}$ and $\{Q_1, Q_2, \dots, Q_N\}$ be a pair of N-tuples of (possibly not all distinct) points on a closed Riemann surface. For the divisor $\sum_{j=1}^N P_j - \sum_{j=1}^N Q_j$ to be a principal divisor, that is, for there to be a meromorphic function f such that*

$$(f) = \sum_{j=1}^N P_j - \sum_{j=1}^N Q_j,$$

it is necessary and sufficient that

$$\sum_{j=1}^N \phi(P_j) = \sum_{j=1}^N \phi(Q_j). \tag{4.6}$$

The condition (4.6) can be rewritten as

$$\Big[\sum_{j=1}^N \int_{P_1}^{Q_1} \omega_1, \sum_{j=1}^N \int_{P_1}^{Q_1} \omega_2, \dots, \sum_{j=1}^N \int_{P_1}^{Q_1} \omega_g\Big] = [0, 0, \dots, 0].$$

By Proposition 3.2 we see that Abel's theorem implies that ϕ is injective.

Now let us just prove that (4.6) is a necessary condition. We may, without loss of generality, assume that $\{P_1, P_2, \dots, P_N\} \cap \{Q_1, Q_2, \dots, Q_N\} = \emptyset$ and thus

$$(f)_0 = \sum_{j=1}^N P_j$$
$$(f)_\infty = \sum_{j=1}^N Q_j.$$

Now for an arbitrary complex number t, $f = t$ is also a meromorphic function on R and we can write

$$(f - t)_0 = \sum_{j=1}^N P_j(t).$$

We define a mapping $\psi : \mathbf{C} \to J(R)$ by

$$\psi : t \in \mathbf{C} \longmapsto \Big[\sum_{j=1}^N \int_P^{P_j(t)} \omega_1, \sum_{j=1}^N \int_P^{P_j(t)} \omega_2, \dots, \sum_{j=1}^N \int_P^{P_j(t)} \omega_g\Big],$$

which may be seen to be a holomorphic function. We can further extend the mapping to a mapping from $\mathbf{P}^1$ to $J(R)$. For this purpose, let us introduce a new variable s such that $s = 1/t$ when $s \neq 0, t \neq 0$. Then for $s \neq 0, t \neq 0$, we have

$$(f - t)_0 = \Big(f - \frac{1}{s}\Big)_0.$$

On the other hand, we have for $s \neq 0$

$$\Big(f - \frac{1}{s}\Big)_0 = (sf - 1)_0 = \Big(s - \frac{1}{f}\Big).$$

So if for $s \neq 0$ we set

$$\left(f - \frac{1}{s}\right)_0 = \sum_{j=1}^{N} Q_j(s),$$

we get

$$\sum_{j=1}^{N} Q_j(s) = \sum_{j=1}^{N} P_j\left(\frac{1}{s}\right).$$

Furthermore, if $s \neq 0$, then

$$\left(f - \frac{1}{s}\right)_0 = \left(s - \frac{1}{f}\right)_0,$$

and if $s = 0$, we may write

$$\sum_{j=1}^{N} Q_j(0) = \sum_{j=1}^{N} Q_j.$$

Hence the mapping

$$\hat{\psi} : s \in \mathbf{C} \longmapsto \Big[\sum_{j=1}^{N}\int_P^{Q_j(s)} \omega_1, \sum_{j=1}^{N}\int_P^{Q_j(s)} \omega_2, \ldots, \sum_{j=1}^{N}\int_P^{Q_j(s)} \omega_g\Big] \in J(R)$$

is also holomorphic. Now for $s \neq 0$ we have

$$\hat{\psi}(s) = \psi\left(\frac{1}{s}\right),$$

which induces a holomorphic mapping $\tilde{\psi} : \mathbf{P}^1(\mathbf{C}) \to J(R)$. Since $J(R)$ is a g-dimensional complex torus, there exist g linearly independent holomorphic differential forms, say, $dz_1, dz_2, \ldots, dz_g$. The pullbacks $\tilde{\psi}^*\omega_1, \tilde{\psi}^*\omega_2, \ldots, \tilde{\psi}^*\omega_g$ of these forms by $\tilde{\psi}$ are holomorphic differential forms on $\mathbf{P}^1(\mathbf{C})$. However, $\mathbf{P}^1(\mathbf{C})$ has no holomorphic form except for 0. Hence $\tilde{\psi}^*\omega_j = 0, 1 \leq j \leq n$, which implies that $\tilde{\psi}(\mathbf{P}^1(\mathbf{C}))$ is a single point. Thus we have $\tilde{\psi}(0) = \tilde{\psi}(\infty)$. Now in view of

$$\tilde{\psi}(0) = \psi(0) = \Big[\sum_{j=1}^{N}\int_P^{P_j} \omega_1, \sum_{j=1}^{N}\int_P^{P_j} \omega_2, \ldots, \sum_{j=1}^{N}\int_P^{P_j} \omega_g\Big]$$

$$\tilde{\psi}(\infty) = \hat{\psi}(0) = \Big[\sum_{j=1}^{N}\int_P^{Q_j} \omega_1, \sum_{j=1}^{N}\int_P^{Q_j} \omega_2, \ldots, \sum_{j=1}^{N}\int_P^{Q_j} \omega_g\Big],$$

we get

$$\sum_{j=1}^{N} \phi(P_j) = \sum_{j=1}^{N} \phi(Q_j).$$

The proof that the condition is sufficient requires some more preparation, and so is omitted.

Now we define a holomorphic function $\theta(\tau, z)$ on $\mathbf{C}^g$ by

$$\theta(\tau, z) = \sum_{n=(n_1,n_2,\ldots,n_g)\in\mathbf{Z}^g} e^{\pi\sqrt{-1}(n\tau^t n + 2n^t z)}$$

and call it the **theta function**.

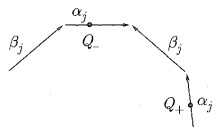


FIGURE 4.14

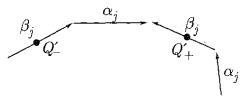


FIGURE 4.15

For $\ell = (\ell_1, \ell_2, \ldots, \ell_g), m = (m_1, m_2, \ldots, m_g) \in \mathbf{Z}^g$ we have

$$\theta(\tau, z + \ell + m\tau) = e^{-\pi\sqrt{-1}(m\tau^t m + 2m^t z)}\theta(\tau, z) \tag{4.7}$$

The theta function has this pseudo-periodicity but not periodicity, so we cannot say it is a holomorphic function on $J(R)$. But we may observe that

$$e^{-\pi\sqrt{-1}(m\tau^t m + 2m^t z)}$$

never vanishes, so that $\theta(\tau, z) = 0$ and $\theta(\tau, z + \ell + m\tau) = 0$ are equivalent. Therefore we set

$$\Theta = \{[z] \in J(R) | \theta(\tau, z) = 0\}$$

and call it the **theta divisor** of the Jacobian variety. For $e = (e_1, e_2, \ldots, e_g)$ and $\phi : R \to J(R)$, $\theta(\tau, \phi(Q) - e)$ has no meaning as a function on R in view of (4.7). (If $z_0 \in \mathbf{C}^g$ represents $\phi(Q)$, then so does $z_0 + \ell + m\tau$.) We might say that it is a multi-valued function on R, but whether it is 0 or not is meaningful. We have

LEMMA 4.1. *If $\theta(\tau, \phi(Q) - e) \not\equiv 0$, then $\theta(\tau, \phi(Q) - e)$ has g zeros on R, counting multiplicities.*

PROOF. We make a $4g$-gon R' by cutting open the Riemann surface R along the closed curves $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$ in Figure 4.9. Then θ is a single-valued holomorphic function on R'. We may assume that none of the zeros of θ lies on the closed curves $\alpha_j, \beta_j, 1 \leq j \leq g$, by slightly deforming the curves in the beginning if necessary. Then we have

$$\text{the number of zeros} = \frac{1}{2\pi\sqrt{-1}}\int_{\partial R'} d\log\theta$$

Let us compute the integral on the right-hand side. For a point $Q \in R$ on α_j, denote by Q_+, Q_- the corresponding points on $\partial R'$ as in Figure 4.14.

Then we obtain

$$\begin{aligned}\phi(Q_-) &= \phi(Q_+) + \Big[\int_{\beta_j}\omega_1, \int_{\beta_j}\omega_2, \ldots, \int_{\beta_j}\omega_g\Big] \\ &= \phi(Q_+) + (\tau_{j1}, \tau_{j2}, \ldots, \tau_{jg}).\end{aligned}$$

In the same manner, let Q'_+, Q'_- be the two points on the boundary of R' that correspond to a point Q' on β_j, as in Figure 4.15. Then we have

$$\begin{aligned}\phi(Q'_-) &= \phi(Q'_+) - \Big[\int_{\alpha_j}\omega_1, \int_{\alpha_j}\omega_2, \dots, \int_{\alpha_j}\omega_g\Big] \\ &= \phi(Q'_+) - [0, \dots, 1, 0, \dots, 0]\end{aligned}$$

with 1 as the j-th component.

Therefore we obtain

$$\begin{aligned}&\int_{\partial R'} d\log\theta(\tau, \phi(Q) - e) \\ &= \frac{1}{2\pi\sqrt{-1}}\sum_{j=1}^{g}\Big\{\int_{\alpha_j}(d\log\theta(\tau,\phi(Q_+) - e) - d\log\theta(\tau,\phi(Q_-) - e)) \\ &\qquad + \int_{\beta_j}(d\log\theta(\tau,\phi(Q'_+) - e - d\log\theta(\tau,\phi(Q'_-) - e))\Big\} \\ &= \frac{1}{2}\sum_{j=1}^{g}\int_{\alpha_j} d\Big(\tau_{jj} + 2\int_P^Q \omega_j\Big) \\ &= \sum\int_{\alpha_j}\omega_j = g\end{aligned}$$

Next, we find the relation between the g zeros $Q_1, Q_2, \dots, Q_g$ of $\theta(\tau, \phi(Q) - e)$ and $e = (e_1, e_2, \dots, e_g)$. For this purpose, we use the holomorphic function on R'

$$h_j(Q) = \int_P^Q \omega_j$$

and compute

$$\frac{1}{2\pi\sqrt{-1}}\int_{\partial R'} h_j(Q) d\log\theta(\tau, \phi(Q) - e).$$

Leaving the details of computation to the reader, we get

THEOREM 4.8. *Assuming $\theta(\tau, \phi(Q) - e) \not\equiv 0$, let $Q_1, Q_2, \dots, Q_n$ be the zeros of the function $\theta(\tau, \phi(Q) - e)$. Then*

$$\sum_{j=1}^{g}\phi(Q_j) + [k_1, k_2, \dots, k_g] = [e_1, e_2, \dots, e_g]$$

is valid on the Jacobian variety. Here $k = (k_1, k_2, \dots, k_g)$ is called the **Riemann constant**; *it is a vector that depends only on the choice of a symplectic basis $\{\alpha_1, \alpha_2, \dots, \alpha_g, \beta_1, \beta_2, \dots, \beta_g\}$.*

COROLLARY 4.2. *If $\theta(\tau, \phi(Q) - \sum_{j=1}^{g}\phi(Q_j) - k) \not\equiv 0$, then this multi-valued function has $Q_1, Q_2, \dots, Q_g$ as its zeros.*

By definition, the theta function is an even function, that is,

$$\theta(\tau, -z) = \theta(\tau, z).$$

Using this property, we obtain

COROLLARY 4.3. *For any $g-1$ points $P_1, \dots, P_g$ on the Riemann surface, we have*

$$\theta(\tau, \sum_{j=1}^{g-1}\phi(P_j) + k) \equiv 0.$$

PROOF. Taking arbitrary points $P_1, \dots, P_{g-1}, Q'$ on the Riemann surface R, we have

$$\theta(\tau, \phi(Q) - \phi(Q') - \sum_{j=1}^{g-1}\phi(P_j) - k) \not\equiv 0$$

as a function of Q. As a multi-valued function of Q, its zeros are $Q', P_1, \dots, P_{g-1}$ by Corollary 4.2 just above. Setting $Q = Q'$, we obtain

$$\theta(\tau - \sum_{j=1}^{g-1}\phi(P_j) - k) = 0$$

$$\theta(\tau - \sum_{j=1}^{g-1}\phi(P_j) + k) = 0.$$

Since $P_1, \dots, P_{g-1}$ are arbitrary, the last equation holds for $g-1$ arbitrary points.

We may extend the mapping $\phi : R \to J(R)$ to a mapping of the ℓ-fold direct product $R^\ell = R \times \cdots \times R$ into $J(R)$ by

$$\phi_\ell : (Q_1, Q_2, \dots, Q_\ell) \in R^\ell \longmapsto \Big[\sum_{j=1}^{\ell}\int_P^{Q_j}\omega_1, \sum_{j=1}^{\ell}\int_P^{Q_j}\omega_2, \dots, \sum_{j=1}^{\ell}\int_P^{Q_j}\omega_g\Big].$$

Furthermore, to an arbitrary divisor $D = \sum_{i=1}^{N} n_i P_i$ we may associate the point of $J(R)$ given by

$$\phi(D) = \Big[\sum_{i=1}^{N} n_i\int_P^{P_i}\omega_1, \sum_{i=1}^{N} n_i\int_P^{P_i}\omega_2, \dots, \sum_{i=1}^{N} n_i\int_P^{P_i}\omega_g\Big].$$

THEOREM 4.9. *Set $W_{g-1} = \phi_{g-1}(R^{g-1})$. Then*

$$\Theta = W_{g-1} + [k_1, k_2, \dots, k_g],$$

where Θ is the theta divisor of $J(R)$ and $k = [k_1, k_2, \dots, k_g]$ is the Riemann constant.

This theorem shows that the theta divisor and the image $\phi_{g-1}(W_{g-1})$ coincide up to translation. Furthermore, the following results on the Riemann constant can be obtained as corollaries

COROLLARY 4.4. *For any canonical divisor K on a Riemann surface R and the Riemann constant $(k_1, k_2, \dots, k_g)$ there is the relation*

$$2[k_1, k_2, \dots, k_g] = \phi(K).$$

There is a further, closer relation between the theta function and divisors.

THEOREM 4.10. *For points $Q_1, \dots, Q_g, Q$ on a Riemann surface R, regard*

$$\theta(\tau, \phi(Q) - \sum_{j=1}^{g} \phi(Q_j) - k)$$

as a multi-valued function of Q on R.

(i) *If $\ell(K - \sum_{j=1}^{g} Q_j) = 0$, then*

$$\theta(\tau, \phi(Q) - \sum_{j=1}^{q} \phi(Q_j) - k) \not\equiv 0$$

and this multi-valued function has $Q_1, \dots, Q_g$ as its zeros.

(ii) *If $\ell(K - \sum_{j=1}^{g} Q_j) = r \geq 1$, then for any*

$$1 \leq i_1 \leq i_2 \leq \cdots \leq i_s \leq g, \quad s \leq r - 1$$

we have

$$\theta^{(i_1, i_2, \dots, i_s)}(\tau, \phi(Q) - \sum_{j=1}^{g} \phi(Q_j) - k) \equiv 0.$$

Moreover, for some $1 \leq j_1 \leq j_2 \leq \cdots \leq j_r \leq g$, we have

$$\theta^{(j_1, j_2, \dots, j_r)}(\tau, \phi(Q) - \sum_{j=1}^{g} \phi(Q_j) - k) \neq 0$$

at each $Q = Q_j$. Here we have set

$$\theta^{(a_1, a_2, \dots, a_\ell)}(\tau, z) = \frac{\partial^\ell}{\partial z_{a_1} \partial z_{a_2} \cdots \partial z_{a_\ell}} \theta(\tau, z).$$

From this theorem we can get the following important result, but the proof will be omitted.

COROLLARY 4.5 (RIEMANN'S SINGULARITY THEOREM). *Let x be a point of the theta divisor Θ of a Riemann surface R:*

$$x = \sum_{j=1}^{g-1} \phi(Q_j) + [k_1, k_2, \dots, k_g].$$

Then we have

$$\mathrm{mult}_x \Theta = \ell(\sum_{j=1}^{g-1} Q_j),$$

where $\mathrm{mult}_x \Theta$ denotes the multiplicity of Θ at the point x.

We add the following explanation. To say $\mathrm{mult}_x \Theta = m$ means that

$$\frac{\partial^s \theta(\tau, x)}{\partial z_1^{s_1} \cdots \partial z_g^{s_g}} = 0, \quad s_1 + s_2 + \cdots + s_g = s \leq m - 1,$$

always holds and that there exist $m_1, \dots, m_g, m_1 + \cdots + m_g = m$, such that

$$\frac{\partial^m \theta(\tau, x)}{\partial z_1^{m_1} \cdots \partial z_g^{m_g}} \neq 0.$$

We hope that the discussions above have helped the reader to imagine how deeply and mysteriously a closed Riemann surface, its Jacobian variety $J(R)$ and the theta function $\theta(\tau, z)$ are interrelated. Indeed, here begins the theory of special divisors of a closed Riemann surface and the theory of moduli, for which we refer the reader to the standard literature.

Problems

4.1. Consider the one-dimensional complex torus E_τ in Example 4.2.
(i) Show that $\psi_{|2[0]|}$ can be expressed by

$$[z] \in E_\tau \longmapsto (1 : \mathfrak{p}) \in \mathbf{P}^1(\mathbf{C}),$$

that it is a $(2 : 1)$-holomorphic mapping, and that it has 4 ramification points $[0], [1/2], [\tau/2], [1/2+\tau/2]$. Using these facts, deduce that the genus of E_τ is 1 from Hurwitz's formula.

(ii) Show that one can choose $1, \mathfrak{p}(z), \mathfrak{p}'(z), \mathfrak{p}(z)^2$ as a basis of $\mathbf{L}(4[0])$, and determine the defining equation for the image of

$$\psi_{|4[0]|} : [z] \in E_\tau \longrightarrow (1 : \mathfrak{p}(z) : \mathfrak{p}'(z) : \mathfrak{p}(z)^2) \in \mathbf{P}^3(\mathbf{C}).$$

4.2. Prove that the indefinite integral

$$\int \frac{dx}{\sqrt{1-x^3}}$$

cannot be represented by elementary functions and their inverse functions, by using the fact that we may consider

$$\frac{dx}{\sqrt{1-x^3}}$$

as a holomorphic differential form on the elliptic curve

$$y^2 = 1 - x^3.$$

[Hint: Among the elementary functions, periodic functions of trigonometric and exponential functions are simply periodic. On the other hand, integration over an elliptic curve produces a double period along the closed curves α and β. Also,

$$y = \int_0^x \frac{dt}{\sqrt{1-t^2}} = \arcsin x$$

can be expressed by

$$x = \sin y,$$

and x is a function of y which is simply periodic with fundamental period 2π. Similarly, if we set

$$y = \int_0^x \frac{dt}{\sqrt{1-t^3}},$$

then x is doubly periodic as a function of y.]

Appendix

Commutative Rings and Fields

In these Appendices we offer some elementary knowledge on commutative rings and fields that is required for understanding this book. Since the material is abstract, we try to explain by taking concrete examples as much as possible. As a result, the exposition might have become too long, contrary to the best feature of abstract algebra. Here we explain in detail the idea of a residue ring, which beginners find difficult to understand, by starting with the case of integers. If we have not covered all fundamental facts about commutative rings and fields, we hope the reader will follow up with a standard reference on algebra and get to understand the theories of commutative rings and fields necessary for algebraic geometry without too much difficulty.

§ A.1. Integers and congruence

We denote by $\mathbf{Z}$ the set of all integers. We now fix an integer $n \geq 2$. Two integers a and b are said to be **congruent mod** n, and we write

$$a \equiv b \,(\mathrm{mod}\ n),$$

if $a - b$ is a multiple of n. In this case, $b - a$ is also a multiple of n, and thus

$$b \equiv a \,(\mathrm{mod}\ n).$$

If furthermore

$$b \equiv c \,(\mathrm{mod}\ n),$$

then $a - c = (a - b) + (b - c)$ is also a multiple of n, so that

$$a \equiv c \,(\mathrm{mod}\ n).$$

If for $a_1, a_2, b_1, b_2 \in \mathbf{Z}$

$$a_1 \equiv b_1 \,(\mathrm{mod}\ n)$$
$$a_2 \equiv b_2 \,(\mathrm{mod}\ n),$$

then we have

$$a_1 + a_2 \equiv b_1 + b_2 \,(\mathrm{mod}\ n)$$
$$a_1 - a_2 \equiv b_1 - b_2 \,(\mathrm{mod}\ n)$$
$$a_1 a_2 \equiv b_1 b_2 \,(\mathrm{mod}\ n).$$

Indeed, if $a_1 - b_1, a_2 - b_2$ are multiples of n, then

$$(a_1 + a_2) - (b_1 + b_2),\ (a_1 - a_2) - (b_1 - b_2),\ a_1a_2 - b_1b_2 = a_1(a_2 - b_2) + (a_1 - b_1)b_2$$

are also multiples of n.

For an integer a, we set

$$\bar{a} = \{m \in \mathbf{Z} |\ m \equiv a \pmod{n}\}$$

and call it the **residue class** of a mod n.

LEMMA A.1. *If $a \equiv b \pmod{n}$, then the residue classes $\bar{a}$ and $\bar{b}$ mod n coincide.*

PROOF. Under the assumption $a \equiv b \pmod{n}$, we see that

$$m \equiv a \pmod{n} \quad \text{implies} \quad m \equiv b \pmod{n},$$

that is,

$$\bar{a} \subset \bar{b}.$$

Interchanging the roles of a and b (and also of $\bar{a}$ and $\bar{b}$), we get

$$\bar{b} \subset \bar{a}.$$

Hence $\bar{a}$ and $\bar{b}$ coincide.

LEMMA A.2. *If $\bar{a} \cap \bar{b} \neq \emptyset$ for $a, b \in \mathbf{Z}$. then*

$$\bar{a} = \bar{b} \text{ and } a \equiv b \pmod{n}.$$

PROOF. Suppose $c \in \bar{a} \cap \bar{b}$. Then

$$c \equiv a \pmod{n}$$
$$c \equiv b \pmod{n}.$$

Hence

$$a \equiv b \pmod{n},$$

which implies $\bar{a} = \bar{b}$ by Lemma A.1.

From Lemma A.2 we see that the set $\mathbf{Z}$ can be partitioned into mutually exclusive residue classes mod n. The set of all residue classes is denoted by $\mathbf{Z}/(n)$. Since division of integers by n produces one of the remainders $0, 1, 2, \ldots, n-1$, we can write

$$\mathbf{Z}/(n) = \{\bar{0}, \bar{1}, \bar{2}, \ldots, \overline{n-1}\}.$$

DEFINITION A.1. For $\bar{a}, \bar{b} \in \mathbf{Z}/(n)$. we define the sum $\bar{a} + \bar{b}$ and the product $\bar{a} \cdot \bar{b}$ by

$$\bar{a} + \bar{b} = \overline{a+b}$$
$$\bar{a} \cdot \bar{b} = \overline{ab}.$$

Suppose $\bar{a}_1 = \bar{a}_2$ and $\bar{b}_1 = \bar{b}_2$. By Lemma A.2, we get

$$a_1 \equiv a_2 \pmod{n}, \quad b_1 \equiv b_2 \pmod{n},$$

and hence

$$a_1 + b_1 \equiv a_2 + b_2 \pmod{n}$$
$$a_1b_1 \equiv a_2b_2 \pmod{n},$$

that is,

$$\overline{a_1 + b_1} = \overline{a_2 + b_2}$$
$$\overline{a_1b_1} = \overline{a_2b_2}.$$

We have thus verified that both $\bar{a} + \bar{b}$ and $\bar{a} \cdot \bar{b}$ are uniquely determined by the residue classes $\bar{a}$ and $\bar{b}$. This means that we now have defined sum and product in the set $\mathbf{Z}/(n)$.

EXAMPLE A.1. Let $n = 5$. We show examples of computation for sum and product in $\mathbf{Z}/(5)$.

$$\bar{3} + \bar{4} = \overline{7} = \bar{2}, \quad \bar{2} + \bar{3} = \bar{5} = \bar{0}$$
$$\bar{4} + \bar{4} = \bar{8} = \bar{3}, \quad \bar{6} + \bar{7} = \overline{13} = \bar{3}$$
$$\bar{3} \cdot \bar{4} = \overline{12} = \bar{2}, \quad \bar{3} \cdot \bar{2} = \bar{6} = \bar{1}$$
$$\bar{4} \cdot \bar{4} = \overline{16} = \bar{1}, \quad \bar{6} \cdot \bar{7} = \overline{42} = \bar{2}.$$

From Definition A.1 we see that for an arbitrary $\bar{a} \in \mathbf{Z}/(n)$ we have

$$\bar{a} + \bar{0} = \overline{a+0} = \bar{a}$$
$$\bar{a} \cdot \bar{1} = \overline{a \cdot 1} = \bar{a},$$

showing that $\bar{0}$ and $\bar{1}$ behave like ordinary 0 and ordinary 1, respectively. We find that sum and product in $\mathbf{Z}/(n)$ have the following properties.

THEOREM A.1. *Sum and product in $\mathbf{Z}/(n)$ given in Definition A.1 have the following properties.*

I. Properties of addition.

(i) *(commutativity) For any $\bar{a}, \bar{b} \in \mathbf{Z}/(n)$*

$$\bar{a} + \bar{b} = \bar{b} + \bar{a}.$$

(ii) *(associativity) For any $\bar{a}, \bar{b}, \bar{c} \in \mathbf{Z}/(n)$*

$$(\bar{a} + \bar{b}) + \bar{c} = \bar{a} + (\bar{b} + \bar{c}).$$

(iii) *(existence of the zero) For any $\bar{a} \in \mathbf{Z}/(n)$*

$$\bar{a} + \bar{0} = \bar{a}.$$

($\bar{0}$ is called the **zero** *of $\mathbf{Z}/(n)$.)*

(iv) *(Existence of additive inverse) For any $\bar{a} \in \mathbf{Z}/(n)$, there is an element $\bar{b} \in \mathbf{Z}/(n)$ such that*

$$\bar{a} + \bar{b} = \bar{0}.$$

(Such $\bar{b}$ is unique and is denoted by $-\bar{a}$.)

II. Properties of multiplication.

(i) *(commutativity) For any* $\bar{a}, \bar{b} \in \mathbf{Z}/(n)$

$$\bar{a} \cdot \bar{b} = \bar{b} \cdot \bar{a}.$$

(ii) *(associativity) For any* $\bar{a}, \bar{b}, \bar{c} \in \mathbf{Z}/(n)$

$$(\bar{a} \cdot \bar{b}) \cdot \bar{c} = \bar{a} \cdot (\bar{b} \cdot \bar{c}).$$

(iii) *(existence of the identity) For any* $\bar{a} \in \mathbf{Z}/(n)$

$$\bar{a} \cdot \bar{1} = \bar{a}.$$

($\bar{1}$ *is called the* **identity** *of* $\mathbf{Z}/(n)$.)

III. Distributive law.

For any $\bar{a}, \bar{b}, \bar{c} \in \mathbf{Z}/(n)$ *we have*

$$\bar{a} \cdot (\bar{b} + \bar{c}) = \bar{a} \cdot \bar{b} + + \bar{a} \cdot \bar{c}.$$

The proofs are obvious from Definition A.1. For example, commutativity in (i) follows from

$$\bar{a} + \bar{b} = \overline{a + b} = \overline{b + a} = \bar{b} + \bar{a},$$

and associativity in (ii) from

$$\begin{aligned}(\bar{a} + \bar{b}) + \bar{c} &= \overline{\overline{(a + b)} + \bar{c}} \\ &= \overline{\overline{a + (b + c)}} = \bar{a} + \overline{b + c} = \bar{a} + (\bar{b} + \bar{c}).\end{aligned}$$

For (iv) in I, we have

$$-\bar{a} = \overline{-a},$$

since

$$\bar{a} + \overline{-a} = \overline{a + (-a)} = \overline{a - a} = \bar{0}.$$

We can finally derive III from

$$\begin{aligned}\bar{a} \cdot (\bar{b} + \bar{c}) &= \bar{a} \cdot \overline{b + c} = \overline{a(b + c)} \\ &= \bar{a} + \overline{b + c} = \bar{a} \cdot \bar{b} + \bar{a} \cdot \bar{c}.\end{aligned}$$

What is essential in the proofs above is that the set of integers $\mathbf{Z}$ itself has the properties I (i)–(iv), II (i)–(iii), and III. In general, a set R is called a **commutative ring** if it has addition and multiplication defined so as to satisfy the properties I (i)–(iv), II (i)–(iii), and III in Theorem A.1 . We denote the zero element and the identity by 0 and 1, respectively.

EXAMPLE A.2 – THE COMMUTATIVE RING $\mathbf{Z}[x]$. The set $\mathbf{Z}[x]$ of all polynomials with integral coefficients

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m, \qquad m \geq 0,\ a_j \in \mathbf{Z},$$

becomes a commutative ring relative to ordinary sum and product. The zero element is 0, namely, the polynomial with all coefficients equal to 0, and the identity element is 1, namely, the polynomial for which $a_0 = 1$ and all other coefficients are equal to 0. The additive inverse of $f(x)$ is given by

$$-a_0 + (-a_1)x + (-a_2)x^2 + \cdots + (-a_m)x^m, \qquad m \geq 0,\ a_j \in \mathbf{Z}.$$

For two elements a, b of a commutative ring R the difference $a - b$ is defined as

$$a - b = a + (-b),$$

where $-b$ is the additive inverse of b. We can check that this has the property of the usual difference as follows. Set

$$a - b = c.$$

Then

$$\begin{aligned}c + b &= \{a + (-b)\} + b = a + \{(-b) + b\} \\ &= a + 0 = a.\end{aligned}$$

In particular, if $R = \mathbf{Z}/(n)$, then

$$\begin{aligned}\bar{a} - \bar{b} &= \bar{a} + (-\bar{b}) = \bar{a} + \overline{-b} \\ &= \overline{a + (-b)} = \overline{a - b};\end{aligned}$$

hence we can define the difference $\bar{a} - \bar{b}$ as the last term $\overline{a - b}$.

Let us consider $\mathbf{Z}/(n)$ again. For an element $\bar{a} \neq \bar{0}$, does there exist an element $\bar{b}$ such that

$$\bar{a} \cdot \bar{b} = \bar{1}?$$

We need the following lemma,

LEMMA A.3. *If* m *and* n *are relatively prime integers, that is, if their greatest common divisor is 1, then there exist integers* a, b *such that*

$$am + bn = 1.$$

PROOF. We can write a simple proof by using Euclid's algorithm. Here, however, we state a proof by using the notion of ideal which we discuss later (see § A.2). We define a subset (m, n) of $\mathbf{Z}$ as

$$(m, n) = \{\alpha m + \beta n | \alpha, \beta \in \mathbf{Z}\}.$$

If integers d, e belong to (m, n), then

$$d + e \in (m, n).$$

Furthermore, for any integer r, we have

$$rd \in (m, n).$$

In particular, if $d \in (m, n)$, then $-d \in (m, n)$. Now let d_0 be the smallest positive number contained in (m, n). We show that (m, n) coincides with the set of all multiples of d_0. Since $rd_0 \in (m, n)$ for any integer r, the set (m, n) contains all the multiples of d_0. Conversely, let δ be an arbitrary element of (m, n). We show that δ is a multiple of d_0. If $\delta = 0$, the assertion is clear. Hence we assume that $\delta \neq 0$. We may assume that δ is positive by replacing it by $-\delta$, if necessary. Divide δ by d_0 and let ϵ be the remainder, that is,

$$\delta = \ell d_0 + \epsilon, \qquad 0 \leq \epsilon < d_0.$$

Since $\delta \in (m, n)$ and $-\ell d_0 \in (m, n)$, we have

$$\epsilon = \delta + (-\ell d_0) \in (m, n),$$

If $\epsilon > 0$, this contradicts the assumption that d_0 is the smallest positive integer contained in (m, n). It follows that $\epsilon = 0$, that is, δ is a multiple of d_0. We have thus shown that

$$(m, n) = \{rd_0 | r \in \mathbf{Z}\}.$$

From the definition of d_0, there exist integers a, b such that

$$d_0 = am + bn.$$

We show that $d_0 = 1$. Since m, n are elements of (m, n), they are multiples of d_0. Hence if $d_0 \geq 2$, d_0 must be a common divisor of m and n. But we assumed that m and n are relatively prime — a contradiction. Thus we conclude that $d_0 = 1$.

We may derive the following important result from this lemma.

COROLLARY A.1. *If integers m and n are relatively prime, then in $\mathbf{Z}/(n)$ there is a unique residue class $\bar{a}$ such that $\bar{a} \cdot \bar{m} = \bar{1}$.*

PROOF. By Lemma A.3 there exist integers a, b such that

$$am + bn = 1$$

and hence

$$am \equiv 1 \pmod{n},$$

which implies

$$\bar{a} \cdot \bar{m} = \bar{1}.$$

If we have furthermore

$$\bar{b} \cdot \bar{m} = 1,$$

we get

$$(\bar{a} - \bar{b}) \cdot \bar{m} = \bar{0}.$$

This means

$$(a - b)m \equiv 0 \pmod{n}.$$

Since m and n are relatively prime, it follows that $a - b$ is a multiple of n, that is, $\bar{a} = \bar{b}$.

THEOREM A.2. *We have the following assertions concerning $\mathbf{Z}/(n)$.*

(i) *If n is a prime number p, then $\mathbf{Z}(p)$ satisfies* I (i)-(iv), II (i)-(iii), III (i) *in Theorem A.1 and has furthermore the following properties:*
II (iv) *(existence of a multiplicative inverse) For any element $\bar{a} \neq \bar{0}$, there exists a unique $\bar{b}$ such that $\bar{a} \cdot \bar{b} = \bar{1}$. (We denote this element by $\bar{a}^{-1}$.)*

(ii) *If n is a composite number, there exist $\bar{a} \neq \bar{0}, \bar{b} \neq \bar{0}$ in $\mathbf{Z}/(n)$ such that $\bar{a} \cdot \bar{b} = \bar{0}$. (Such elements $\bar{a}, \bar{b}$ are called* **zero divisors***).*

PROOF.

(i) If $\bar{a} \neq \bar{0}$, then a and p are relatively prime, since p is a prime. By Corollary A.1, there is a unique $\bar{b}$ such that $\bar{a}\bar{b} = \bar{1}$.

(ii) Since n is a composite number, we can write

$$n = m\ell, \qquad m \geq 2,\ \ell \geq 2.$$

In $\mathbf{Z}/(n)$, we have

$$\bar{m} \neq \bar{0},\ \bar{\ell} \neq \bar{0},$$

but

$$\bar{m} \cdot \bar{\ell} = \bar{n} = \bar{0}.$$

By virtue of Theorem A.2, we see that we can add, subtract, multiply and divide in $\mathbf{Z}/(p)$. If $\bar{a} \neq \bar{0}$, then, given $\bar{c}$, there exists an element $\bar{d}$ such that $\bar{d} \cdot \bar{a} = \bar{c}$. In fact, take $\bar{d} = \bar{c} \cdot \bar{a}^{-1}$. Then multiplying both sides by $\bar{a}$ we obtain

$$\begin{aligned} \bar{d} \cdot \bar{a} &= (\bar{c} \cdot \bar{a}^{-1}) \cdot \bar{a} = \bar{c} \cdot (\bar{a}^{-1} \cdot \bar{a}) \\ &= \bar{c} \cdot (\bar{a} \cdot \bar{a}^{-1}) = \bar{c} \cdot \bar{1} = \bar{c}. \end{aligned}$$

This justifies the role of $\bar{d}$ as $\bar{c} \div \bar{a}$.

A system satisfying the properties I (i)-(iv), II (i)-(iv), and III in Theorems A.1 and Theorem A.2 is called a **field**, more precisely, a **commutative field**. In a field, we can do addition, subtraction, multiplication and division, as we have demonstrated above.

§ A.2. The polynomial ring $\mathbf{Q}[x]$

In this section, we consider the set $\mathbf{Q}[x]$ of all polynomials with rational coefficients. A polynomial is said to be **irreducible** if it cannot be decomposed as a product of two non-constant polynomials of lower degrees. Whether a polynomial is irreducible or not depends on where we take the coefficients. For example, $x^2 + 1$ cannot be factored as a product of polynomials in $\mathbf{Q}[x]$, but if we allow complex coefficients, then we can factor it in the form

$$x^2 + 1 = (x + i)(x - i).$$

In this section, we consider polynomials with rational coefficients, namely, elements in $\mathbf{Q}[x]$, unless otherwise noted.

A polynomial $f(x)$ with rational coefficients can be factored into a product of irreducible polynomials

$$f(x) = cf_1(x)^{n_1} f_2(x)^{n_2} \dots f_\ell(x)^{n_\ell},$$

where c is a nonzero rational number and the n_j's are integers ≥ 1. Such a factorization is essentially unique, that is, if

$$f(x) = dg_1(x)^{m_1} g_2(x)^{m_2} \dots g_k(x)^{m_k},$$

then $\ell = k$ and, by re-indexing the $g_j(x)$'s, we get

$$f_j(x) = a_j g_(x), \qquad a_j \neq 0,\ a_j \in \mathbf{Q}, 1 \leq j \leq \ell.$$

To prove this fact, we use Euclid's algorithm, which is valid in $\mathbf{Q}[x]$. The fundamental fact is the following well-known lemma.

LEMMA A.4. *Let $f(x)$ and $g(x)$ be two polynomials of degree m and n, respectively. Then there exist unique polynomials $a(x), g(x) \in \mathbf{Q}[x]$ such that*

$$f(x) = a(x)g(x) + r(x), \quad \deg r(x) < n,$$

where $\deg r(x)$ denotes the degree of $r(x)$.

The polynomial ring $\mathbf{Q}[x]$ has properties similar to those of the ring of integers $\mathbf{Z}$. We can make arguments similar to those in § A.1. Here, however, we provide somewhat different arguments, starting with the notion of ideal.

DEFINITION A.2. A subset I of $\mathbf{Q}[x]$ is called an **ideal** if it has the following properties.

(i) For two elements $f(x)$ and $g(x) \in I$, we have $f(x) + g(x) \in I$;

(ii) For any element $a(x) \in \mathbf{Q}[x]$ and for any element $f(x) \in I$, we have $a(x)f(x) \in I$.

It follows from (i) and (ii) that for any $a(x)$ and $b(x) \in \mathbf{Q}[x]$ and for any $f(x)$ and $g(x) \in I$ we have

$$a(x)f(x) + b(x)g(x) \in I.$$

Since $\mathbf{Q} \subset \mathbf{Q}[x]$, we have

$$\alpha f(x) + \beta g(x) \in I$$

for any $\alpha, \beta \in \mathbf{Q}$ and for any $f(x), g(x) \in I$. In fact, ideals in $\mathbf{Q}[x]$ have a simple form.

LEMMA A.5. *An ideal I of $\mathbf{Q}[x]$ coincides with the set $(h(x))$ of all multiples of a certain polynomial $h(x)$, namely,*

$$(h(x)) = \{a(x)h(x) | a(x) \in \mathbf{Q}[x]\}.$$

PROOF. Pick a nonzero polynomial $h(x)$ of lowest degree that is contained in I. If the degree of $h(x)$ is 0, that is, if $h(x)$ is a constant $c \neq 0$, then $1 = \frac{1}{c} \cdot c \in I$. In this case, for any polynomial $f(x) \in \mathbf{Q}[x]$ we have

$$f(x) = f(x) \cdot 1 \in I,$$

which means that $I = \mathbf{Q}[x]$. On the other hand, $(c) = \mathbf{Q}[x]$, so that the conclusion of Lemma A.5 is valid.

Next we assume $\deg h(x) \geq 1$. For any $g(x) \in I$ we have unique $a(x), r(x) \in \mathbf{Q}[x]$ such that

$$g(x) = a(x)h(x) + r(x), \quad \deg r < \deg h.$$

Here $g(x), h(x) \in I$ implies

$$r(x) = g(x) - a(x)h(x) = g(x) + (-a(x))h(x) \in I.$$

By definition of $h(x)$, we conclude that $r(x) = 0$. Therefore $g(x) = a(x)h(x)$, showing that $I \subset (h(x))$. On the other hand, $h(x) \in I$ implies $(h(x)) \subset I$, so that $I = (h(x))$.

COROLLARY A.2. *If $f(x)$ and $g(x)$ are relatively prime polynomials in $\mathbf{Q}[x]$, then there exist polynomials $a(x), b(x) \in \mathbf{Q}[x]$ such that*

$$a(x)f(x) + b(x)g(x) = 1.$$

PROOF. Set

$$I = (f(x), g(x)) = \{\alpha(x)f(x) + \beta(x)g(x) | \alpha(x), \beta(x) \in \mathbf{Q}[x]\}.$$

It is easily verified that $(f(x), g(x))$ is an ideal of $\mathbf{Q}[x]$. By Lemma A.5, there exists $h(x) \in \mathbf{Q}[x]$ such that

$$(f(x), g(x)) = (h(x)).$$

Since $f(x), g(x) \in I$, we get $f(x) \in (h(x)), \quad g(x) \in (h(x))$. Hence

$$f(x) = p(x)h(x), \quad g(x) = q(x)h(x)$$

for some polynomials $p(x), q(x)$. Thus $h(x)$ is a common divisor of $f(x)$ and $g(x)$. But $f(x)$ and $g(x)$ being relatively prime, it follows that $h(x)$ must be a constant $c \neq 0$. Thus $I = \mathbf{Q}[x]$, and there exist $a(x), b(x) \in \mathbf{Q}[x]$ such that

$$a(x)f(x) + b(x)g(x) = 1.$$

With these preparations we shall now generalize the notion of congruence and define the ring $\mathbf{Q}[x]/I$ — an analogue of $\mathbf{Z}/(n)$ in the preceding section. For any polynomial $f(x) \in \mathbf{Q}[x]$ we define a subset of $\mathbf{Q}[x]$,

$$\overline{f(x)} = \{g(x) \in \mathbf{Q}[x] | f(x) - g(x) \in I\},$$

and call it the **residue class** for $f(x)$ modulo I. If

$$f(x) - h(x) \in I,$$

then for $g(x) \in \overline{f(x)}$ we have

$$h(x) - g(x) = (h(x) - f(x)) + (f(x) - g(x)) \in I,$$

which shows that $g(x) \in \overline{h(x)}$. Conversely, if $g(x) \in \overline{h(x)}$, then

$$f(x) - g(x) = (f(x) - h(x)) + (h(x) - g(x)) \in I,$$

which shows that $g(x) \in \overline{f(x)}$. We have thus shown that if $f(x) - h(x) \in I$, then the residue classes mod I for $f(x)$ and $h(x)$ coincide. that is,

$$\overline{f(x)} = \overline{h(x)}.$$

It also follows that if $g(x) \in \overline{f(x)}$, then

$$\overline{f(x)} = \overline{g(x)},$$

because $f(x) - g(x) \in I$ by definition of $\overline{f(x)}$. Furthermore, if two residue classes $\overline{f_1(x)}$ and $\overline{f_2(x)}$ have a non-empty intersection, then in fact

$$\overline{f_1(x)} = \overline{f_2(x)}.$$

That is, we have the analogue of Lemma A.2 of the preceding section. We shall thus denote by $\mathbf{Q}[x]/I$ the set of all distinct residue classes modulo I. The definition is somewhat too abstract, but what we have is an analogue of $\mathbf{Z}/(n)$ in the preceding section.

Recall that for the ring $\mathbf{Z}/(n)$

$$a \equiv b \,(\mathrm{mod}\ n)$$

implies $\bar{a} = \bar{b}$. If we denote by (n) the set of all multiples of n, then it is an ideal of $\mathbf{Z}$. (The definition of ideal in $\mathbf{Z}$ is obvious by replacing $\mathbf{Z}[x]$ by $\mathbf{Z}$ in Definition A.2.) Then

$$a \equiv b \,(\mathrm{mod}\ n)$$

is equivalent to

$$a - b \in (n),$$

and the residue class $\bar{a}$ mod n may be rewritten in the form

$$\bar{a} = \{b \in \mathbf{Z} | a - b \in (n)\}.$$

The analogy between $\mathbf{Q}[x]/I$ and $\mathbf{Z}/(n)$ should be obvious.

Before we show that $\mathbf{Q}[x]/I$ has the structure of a commutative ring, we examine a few examples.

EXAMPLE A.3. $f(x) = x^2+1$ is an irreducible polynomial in $\mathbf{Q}[x]$. If we divide $g(x) \in \mathbf{Q}[x]$ by $f(x)$, we may write

$$g(x) = a(x)(x^2+1) + \alpha x + \beta.$$

Hence

$$g(x) - (\alpha x + \beta) \in (f(x)).$$

In $\mathbf{Q}[x]/(f(x))$ we have

$$\overline{g(x)} = \overline{\alpha x + \beta}.$$

In other words, a residue class modulo the ideal (x^2+1) coincides with the residue class of a linear polynomial or a constant. We also see that a necessary and sufficient condition for

$$\overline{\alpha x + \beta} = \overline{\alpha' x + \beta'}$$

is

$$\alpha x + \beta - (\alpha' x + \beta') \in (x^2+1),$$

which is equivalent to

$$\alpha = \alpha', \quad \beta = \beta'.$$

Thus we see that

$$\mathbf{Q}[x]/(x^2+1) = \{\overline{\alpha x + \beta} \mid \alpha, \beta \in \mathbf{Q}\}.$$

EXAMPLE A.4. We may factor

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

in $\mathbf{Z}[x]$. As in Example A.3, any polynomial $g(x)$ can be written in the form

$$g(x) = b(x)(x^2-1) + \alpha x + \beta.$$

By a similar discussion, we have also

$$\mathbf{Q}[x]/(x^2-1) = \{\overline{\alpha x + \beta} | \alpha, \beta \in \mathbf{Q}\}.$$

The two examples above, $\mathbf{Q}[x]/(x^2+1)$ and $\mathbf{Q}[x]/(x^2-1)$, look the same as sets; the difference appears when these sets are considered with the structures of rings. We define sum and product in $\mathbf{Q}[x]/I$ just as we did for $\mathbf{Z}/(n)$.

DEFINITION A.3. In $\mathbf{Q}[x]/I$ we define

$$\overline{f(x)} + \overline{g(x)} = \overline{f(x) + g(x)}$$
$$\overline{f(x)} \cdot \overline{g(x)} = \overline{f(x)g(x)}.$$

As in the case of $\mathbf{Z}/(n)$, it is easy to show that

$$\overline{f_1(x)} = \overline{f_2(x)}, \quad \overline{g_1(x)} = \overline{g_2(x)}$$

imply

$$\overline{f_1(x)} + \overline{g_1(x)} = \overline{f_2(x)} + \overline{g_2(x)}$$
$$\overline{f_1(x)} \cdot \overline{g_1(x)} = \overline{f_2(x)} \cdot \overline{g_2(x)}.$$

Therefore sum and product are defined for residue classes.

EXAMPLE A.5 (DIFFERENCE BETWEEN $R_1 = \mathbf{Q}[x]/(x^2+1)$ AND $R_2 = \mathbf{Q}[x]/(x^2-1)$). In $R_2 = \mathbf{Q}[x]/(x^2-1)$, we have

$$\overline{x+1} \neq \bar{0}, \quad \overline{x-1} \neq \bar{0},$$

but

$$\overline{x+1} \cdot \overline{x-1} = \overline{x^2-1} = \bar{0}.$$

Elements whose products with non-zero elements are 0 are called **zero divisors**.

In $R_1 = \mathbf{Q}[x]/(x^2+1)$, there are no zero divisors. Suppose $\overline{g(x)} \cdot \overline{h(x)} = \bar{0}$. Then $g(x)h(x)$ must be a multiple of x^2+1, and x^2+1 must divide $g(x)$ or $h(x)$ and hence $\overline{g(x)} = \bar{0}$ or $\overline{h(x)} = \bar{0}$.

We may generalize the discussions above and prove the following theorem.

THEOREM A.3. *For an ideal I of $\mathbf{Q}[x]$, the following assertions hold.*

(i) *$\mathbf{Q}[x]/I$ becomes a commutative ring relative to sum and product given in Definition A.3. (This is called the* **residue ring** *of $\mathbf{Q}[x]$ by I.) That is, it has properties* I (i)-(iv), II (i)-(iii), *and* III *in Theorem A.1.*

(ii) *The ideal I is of the form $(f(x))$. If $f(x)$ is an irreducible polynomial, then for any $\overline{g(x)} \neq \bar{0}$, there is a unique $\overline{h(x)} \in \mathbf{Q}[x]$ such that*

$$\overline{g(x)} \cdot \overline{h(x)} = \bar{1}.$$

In other words, the commutative ring $\mathbf{Q}[x]$ has property II (iv) *and is a field. On the other hand, if $f(x)$ is reducible, then $\mathbf{Q}[x]/I$ has zero divisors, that is, there exist $\overline{g(x)} \neq \bar{0}, \overline{h(x)} \neq \bar{0}$ such that*

$$\overline{g(x)} \cdot \overline{h(x)} = \overline{f(x)} = \bar{0}.$$

PROOF. The proof of (i) is obvious from Definition A.3.

Proof of (ii). If $f(x)$ is an irreducible polynomial, then $\overline{g(x)} \neq \bar{0}$ implies that $f(x)$ and $g(x)$ are relatively prime. By Corollary A.2, there exist $a(x), h(x) \in \mathbf{Q}[x]$ such that

$$a(x)f(x) + h(x)g(x) = 1.$$

Hence

$$\overline{h(x)g(x)} = \bar{1}.$$

By Definition A.3, we get

$$\overline{h(x)} \cdot \overline{g(x)} = \bar{1}.$$

On the other hand, if $f(x)$ is reducible, then we can factor

$$f(x) = g(x)h(x), \qquad \deg g(x) \geq 1, \ \deg h(x) \geq 1$$

and get

$$\overline{g(x)} \neq \bar{0}, \ \overline{h(x)} \neq \bar{0}.$$

Then we have

$$\overline{g(x)} \cdot \overline{h(x)} = \overline{g(x)h(x)} = \overline{f(x)} = \bar{0},$$

showing that $\mathbf{Q}[x]/I$ has zero divisors.

An ideal I is called a **prime ideal** if $\mathbf{Q}[x]/I$ has no zero divisor. From the theorem above we get

COROLLARY A.3. *For an ideal $I = (f(x))$ of $\mathbf{Q}[x]$ to be a prime ideal, it is necessary and sufficient that $f(x)$ is an irreducible polynomial in $\mathbf{Q}[x]$. In this case, $\mathbf{Q}[x]/I$ is a field*

EXAMPLE A.6. The polynomial $f(x) = x^2 + 1$ is irreducible in $\mathbf{Q}[x]$, and $I = (f(x))$ is a prime ideal. Therefore $\mathbf{Q}[x]/(f(x))$ is a field. In this field we have

$$\bar{x}^2 = -\bar{1}.$$

From Example A.3, every element in $\mathbf{Q}[x]/(f(x))$ can be written in the form

$$\bar{\alpha}\bar{x} + \bar{\beta}, \quad \alpha, \beta \in \mathbf{Q}.$$

On the other hand, we set

$$\mathbf{Q}(i) = \{a + bi \mid a, b \in \mathbf{Q}\},$$

where i is the imaginary unit. We find that it is a field as follows. We have

$$(a+bi)(c+di) = (ac-bd) + (ad+bc)i$$
$$\frac{1}{a+bi} = \frac{a-bi}{a^2+b^2},$$

and the properties I (i)-(iv), II (i)-(iv), III in Theorems A.1 and A.2. Now we define a mapping

$$\phi : a + bi \in \mathbf{Q}(i) \longmapsto \bar{a} + \bar{b}\bar{x} \in \mathbf{Q}[x]/(f(x)),$$

and easily find that it is injective and surjective and that

$$\phi((a+bi)(c+di)) = \phi(a+bi)\phi(c+di).$$

Thus we may say that $\mathbf{Q}(i)$ and $\mathbf{Q}[x]/(f(x))$ are isomorphic as fields (that is, they have the same structure relative to addition, subtraction, multiplication, and division).

This example shows that the field $\mathbf{Q}[x]/(x^2+1)$ may be regarded as the field $\mathbf{Q}(i)$ obtained by adjoining i to $\mathbf{Q}$. More generally, as we shall show later, if $f(x)$ is an irreducible polynomial, then $\mathbf{Q}[x]/(f(x))$ can be regarded as a field obtained by adjoining a root of $f(x)$ to $\mathbf{Q}$.

EXAMPLE A.7. The polynomial $f(x) = x^2 - 2$ is irreducible in $\mathbf{Q}[x]$, because $\sqrt{2}$ is an irrational number. Hence $\mathbf{Q}[x]/(x^2-2)$ is a field. By Example A.3, we can write an arbitrary element in this field in the form

$$\bar{\alpha} + \bar{\beta} \cdot \bar{x}, \quad \alpha, \beta \in \mathbf{Q}.$$

The inverse of $\bar{\alpha} + \bar{\beta}\bar{x} \neq \bar{0}$ can be found by solving

$$(\bar{\alpha} + \bar{\beta} \cdot \bar{x}) \cdot (\bar{\gamma} + \bar{\delta} \cdot \bar{x}) = \bar{1}. \tag{A.1}$$

By using $\bar{x}^2 = \bar{2}$ and

$$\begin{aligned}(\bar{\alpha} + \bar{\beta}\bar{x})(\bar{\alpha} - \bar{\beta} \cdot \bar{x}) &= \bar{\alpha}^2 - \bar{\beta}^2 \cdot \bar{x}^2 \\ &= \bar{\alpha}^2 - \bar{2} \cdot \bar{\beta}^2 \\ &= \overline{\alpha^2 - 2\beta^2},\end{aligned}$$

and by multiplying both sides of (A.1) by $\alpha - \beta \cdot \bar{x}$, we obtain

$$\overline{\alpha^2 - 2\beta^2} \cdot (\bar{\gamma} + \bar{\delta}\bar{x}) = \bar{\alpha} - \bar{\beta} \cdot \bar{x}.$$

Since $\alpha^2 - 2\beta^2$ is a non-zero rational number, we have

$$\bar{\gamma} + \bar{\delta} \cdot \bar{x} = \overline{\left(\frac{\alpha}{\alpha^2 - 2\beta^2}\right)} - \overline{\left(\frac{\beta}{\alpha^2 - 2\beta^2}\right)} \cdot \bar{x}.$$

Incidentally, this is essentially the same operation of rationalizing denominators, that is,

$$\frac{1}{\alpha + \beta\sqrt{2}} = \frac{\alpha - \beta\sqrt{2}}{(\alpha + \beta\sqrt{2})(\alpha - \beta\sqrt{2})} = \frac{\alpha - \beta\sqrt{2}}{\alpha^2 - 2\beta^2}.$$

On the other hand, it is easy to directly verify that

$$\mathbf{Q}(\sqrt{2}) = \{a + b\sqrt{2} | a, b \in \mathbf{Q}\}$$

is a field and that the mapping

$$\phi : a + b\sqrt{2} \in \mathbf{Q}(\sqrt{2}) \longmapsto \bar{a} + \bar{b} \cdot \bar{x} \in \mathbf{Q}[x]$$

is an isomorphism of the two fields.

§ A.3. Commutative rings and fields

Our discussions so far may have given the reader some rough ideas about commutative rings and fields. In this section we provide the general theory of rings and fields.

Given a set R, suppose we can assign to any pair of elements a and b a certain element c in R; we say that we have a binary operation in R. We express an operation by a certain symbol; for instance, to say that there is an operation $\cdot$ means that the element $c \in R$ corresponding to the pair (a, b) will be denoted by $a \cdot b$. Note that generally $a \cdot b$ and $b \cdot a$ are distinct.

With these preparations we define commutative fields and rings.

DEFINITION A.4. A set K provided with two binary operations of sum and product expressed by $+$ and $\cdot$ (or more precisely, $(K, +, \cdot)$ is called a **commutative field** if the following properties hold.

I Properties of addition.

(i) (commutative law) $a + b = b + a$ for any $a, b \in K$.

(ii) (associative law) $a + (b + c) = (a + b) + c$ for $a, b, c \in K$.

(iii) (existence of a zero) $a + 0 = a$ for any $a \in K$.

(iv)(existence of additive inverse) For any $a \in K$, there is a $b \in K$ such that $a + b = 0$ (we denote b by $-a$ in the following).

II Properties of multiplication.

(i) (commutative law) $a \cdot b = b \cdot a$ for any $a, b \in K$.

(ii) (associative law) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for any $a, b, c \in K$.

(iii) (existence of an identity) $a \cdot 1 = a$ for any $a \in K$.

(iv) (existence of a multiplicative inverse) For any $a \neq 0$ in K, there is a $c \in K$ such that $a \cdot c = 1$ (c is called the inverse and denoted by a^{-1}).

III (distributive law) $a \cdot (b + c) = a \cdot b + a \cdot c$.

The uniqueness of $-a$, 1, and a^{-1} in I (iv), II (iii), and II (iv), respectively, will be shown in a moment. We normally assume that $0 \neq 1$ because it is convenient not to think of $\{0\}$ as a field.

If the condition II (iv) is not assumed in Definition A.4, we call K a **commutative ring**. When the condition II (i) is not assumed, we get the notions of a

non-commutative field and a non-commutative ring. Though they are important, we shall consider only commutative fields and rings in the following.

We shall now show that there is only one zero. Suppose 0_1 and 0_2 have the property that

$$a + 0_1 = a, \quad a + 0_2 = a \quad \text{for every } a \in K.$$

By setting $a = 0_2$ in the first equation, we get $0_2 + 0_1 = 0_2$, By setting $a = 0_1$ in the second equation, we get $0_1 + 0_2 = 0_1$.

By I (i), we have $0_2 + 0_1 = 0_1 + 0_2$ and thus $0_1 = 0_2$.

We may similarly show that an identity element is unique. Suppose 1 and $1'$ have the property that

$$a \cdot 1 = a, \quad a \cdot 1' = a \text{ for every } a \in K.$$

By setting $a = 1'$ in the first equation, we get

$$1' \cdot 1 = 1'.$$

By setting $a = 1$ in the second equation we have

$$1 \cdot 1' = 1.$$

Using II (i) we get

$$1 = 1 \cdot 1' = 1' \cdot 1 = 1'.$$

As for the additive inverse, suppose

$$a + b = 0 \quad a + b' = 0 \text{ for some } a \in K.$$

By adding b' to both sides of the first equation we have

$$b' + (a + b) = b'.$$

Using I (ii) and I (i) we have

$$\begin{aligned} b' + (a + b) &= (b' + a) + b = (a + b') + b \\ &= a + (b' + b) = a + (b + b') \end{aligned}$$

and hence

$$b' = a + (b + b').$$

Similarly, from the second equation we have

$$b + (a + b') = b$$

and, using I (i), (ii), we get

$$b = a + (b + b'),$$

and thus $b = b'$. We leave it as an exercise for the reader to write a similar proof for the uniqueness of the multiplicative inverse.

EXAMPLE A.8. The set $\mathbf{Q}$ of all rational numbers, the set $\mathbf{R}$ of all real numbers, and the set $\mathbf{C}$ of all complex numbers — each forms a commutative field relative to ordinary sum and product. For a positive integer n which is not a square, we set

$$\mathbf{Q}(\sqrt{n}) = \{a + b\sqrt{n} | a, b \in \mathbf{Q}\}.$$

If $n = -m, m \geq 1$, then we agree that

$$\sqrt{n} = \sqrt{m}\, i.$$

Now for any non-square integer n, $\mathbf{Q}(\sqrt{n})$, which is obviously a subset of $\mathbf{C}$, is closed under ordinary addition and multiplication and is a field. The inverse of $a + b\sqrt{n}$ is given by

$$\frac{1}{a + b\sqrt{n}} = \frac{a - b\sqrt{n}}{a^2 - b^2 n} = \frac{a}{a^2 - b^2 n} + \frac{-b}{a^2 - b^2 n}\sqrt{n}.$$

EXAMPLE A.9. Let α be a root of an irreducible polynomial of degree n with rational coefficients, and set

$$\mathbf{Q}(\alpha) = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1} | a_0, a_1, \ldots, a_{n-1} \in \mathbf{Q}\}.$$

This subset of $\mathbf{C}$ is closed under the usual addition and multiplication. For addition this is obvious. To show that it is closed under multiplication, we write

$$f(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n, \quad b_n \neq 0.$$

Since $f(\alpha) = 0$, we have

$$\alpha^n = -\frac{1}{b_n}(b_0 + b_1\alpha + b_2\alpha^2 + \cdots + b_{n-1}\alpha^{n-1}) \in \mathbf{Q}(\alpha).$$

Using this relation repeatedly, we find that $\alpha^m \in \mathbf{Q}(\alpha)$ for any positive integer m. It follows that $\mathbf{Q}(\alpha)$ contains all the numbers of the form

$$\sum_{j=0}^{m} c_j \alpha^j, \quad c_j \in \mathbf{Q}.$$

Therefore $\mathbf{Q}(\alpha)$ is closed under multiplication. To show that $\mathbf{Q}(\alpha)$ is a field, it suffices to show that for any non-zero element of $\mathbf{Q}(\alpha)$

$$\beta = c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_{n-1}\alpha^{n-1}$$

the inverse exists. By setting

$$g(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1}x^{n-1}$$

we see that $f(x)$ and $g(x)$ are relatively prime, since $f(x)$ is irreducible. By Corollary A.2, there exist polynomials $a(x), b(x) \in \mathbf{Q}(x)$ such that

$$a(x)f(x) + b(x)g(x) = 1.$$

Substituting α for x in this equation, we have

$$b(\alpha) \cdot \beta = 1$$

because $\beta = g(\alpha)$. Since $b(\alpha) \in \mathbf{Q}(\alpha)$, it follows that the inverse of β is equal to $b(\alpha)$.

EXAMPLE A.10 (POLYNOMIAL RINGS OVER A FIELD K). We denote by $K[x]$ the set of all polynomials with coefficients in a field K,

$$\{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n | n \geq 0, a_j \in K\},$$

and call it the **polynomial ring** of one variable over K. Relative to the ordinary sum and product it becomes a commutative ring. The zero element 0 and the identity element 1 of K give the zero element 0 and the identity element 1 of $K[x]$, respectively, when viewed as constant polynomials. $K[x]$ has properties analogous to those of $\mathbf{Q}[x]$ in the preceding section. In particular, we have factorization in $K[x]$.

By increasing the number of variables, we can define $K[x_1, x_2, \ldots, x_n]$ as the set of all polynomials in n variables with coefficients in K,

$$\{\sum a_{i_1 i_2 \ldots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} | i_1, i_2, \ldots, i_n \geq 0, a_{i_1 i_2 \ldots i_n} \in K\}.$$

It is easy to see that $K[x]$ is a commutative ring with the zero element 0 and the identity element 1. The polynomial rings play a fundamental role in the theory of commutative rings.

Now let us state a few facts we can easily derive from the definitions of a field or a commutative ring K. For any element $a \in K$ we have from III and I (iii)

$$a \cdot 0 + a \cdot 0 = a \cdot (0 + 0) = a \cdot 0$$

and adding $-(a \cdot 0)$ by I (iv) we get

$$a \cdot 0 = 0.$$

Using the additive inverse -1 of the identity element 1, we have

$$1 + (-1) = 0.$$

Multiplying both sides by a we have

$$a \cdot \{1 + (-1)\} = 0.$$

By the distributive law the left-hand side is

$$a \cdot 1 + a \cdot (-1) = 0.$$

Since $a \cdot 1 = a$, we get

$$-a = a \cdot (-1) = (-1) \cdot a.$$

For any elements $a, b \in K$ we define $a - b$ by

$$a - b = a + (-b).$$

Then we find that

$$a - b = c$$

and

$$a = b + c$$

are equivalent, and $a - b$ has properties analogous to ordinary subtraction. Furthermore if K is a field and $b \neq 0$, then $a \div b$ or a/b is defined by

$$a/b = a \cdot b^{-1}.$$

It is easy to see that

$$a/b = c$$

and

$$a = bc$$

are equivalent, and a/b has properties analogous to those of ordinary division. When we work with a field K in general, we normally write $a \cdot b^{-1}$ instead of $a \div b$ or a/b. In the following we shall omit the symbol for product unless it is necessary.

DEFINITION A.5. A subset I of a commutative ring R is called an **ideal** of R if it has the following properties:

(i) For any elements $\alpha, \beta \in I$, we have $\alpha + \beta \in I$.

(ii) For any element $a \in R$ and for any element $\alpha \in I$, we have $a\alpha \in I$.

This is the same form as the definition of an ideal of $\mathbf{Q}[x]$ in the preceding section. Since

$$-\beta = (-1)\beta$$

as we stated before, we see that $\beta \in I$ implies $-\beta \in I$ by (ii), and hence $\alpha - \beta = \alpha + (-\beta) \in I$ by (i).

Now for $\alpha_1, \alpha_2, \ldots, \alpha_\ell$ in a commutative ring R, we set

$$(\alpha_1, \alpha_2, \ldots, \alpha_\ell) = \{a_1\alpha_1 + a_2\alpha_2 + \cdots + a_\ell\alpha_\ell\},$$

which has the properties (i) and (ii) in Definition A.5, as is easy to see. It is called the **ideal generated by** $(\alpha_1, \alpha_2, \ldots, \alpha_\ell)$. In particular, $(0) = \{0\}$ and $(1) = R$.

LEMMA A.6. *The ideals in a field K are (0) or $K = (1)$.*

PROOF. Consider an ideal $I \neq (0)$ of a field K. There is an element $a \neq 0$ in I. Since K is a field, there is $\alpha^{-1} \in K$. Then by (ii) we have $1 = \alpha^{-1} \cdot \alpha \in I$. Hence for any $a \in K$, we have $a = a \cdot 1 \in I$, that is $I = K$.

DEFINITION A.6. A mapping ϕ from a commutative ring R into a commutative ring S is called a **homomorphism** if it has the following two properties

(i) For any $a, b \in R$, we have

$$\phi(a + b) = \phi(a) + \phi(b)$$
$$\phi(ab) = \phi(a)\phi(b).$$

(ii) If we denote the identity of R (resp. S) by 1_R (resp. 1_S), then

$$\phi(1_R) = 1_S.$$

LEMMA A.7. *For a given homomorphism $\phi : R \to S$ from a commutative ring R to a commutative ring S, the subset of R*

$$\ker \phi = \{\alpha \in R | \phi(\alpha) = 0\},$$

where 0 is the zero element of S, is an ideal of R, called the **kernel** *of ϕ.*

PROOF. We show that $\ker\phi$ has properties (i) and (ii) of Definition A.5. If $\alpha, \beta \in \ker\phi$, then by Definition A.6 we get

$$\phi(\alpha+\beta) = \phi(\alpha)+\phi(\beta) = 0+0 = 0,$$

which shows $\alpha+\beta \in \ker\phi$. Furthermore, for any element $a \in R$ we have, again by Definition A.6,

$$\phi(a\alpha) = \phi(a)\phi(\alpha) = \phi(a)\cdot 0 = 0,$$

which shows $a\alpha \in \ker\phi$. Thus we have shown that $\ker\phi$ is an ideal of R.

If $\ker\phi$=$\{0\}$, we say that ϕ is an **injective homomorphism to** S. If furthermore $\phi(R) = S$, then ϕ is called an **isomorphism**. In this case, we may think of R and S as having the same structure as commutative rings. In particular, if R is a field, then it has only (0) and R as ideals; thus a homomorphism $\phi : R \to S$ has $\ker\phi = (0)$ (since $\phi(1_R) = 1_S$ for a homomorphism ϕ). Therefore we see that ϕ is an injective homomorphism. Given an ideal I of a commutative ring R, we can form a new commutative ring R/I, called the **residue ring** of R by I. As in the preceding section, for any $a \in R$, define a subset $\bar{a}$ of R,

$$\bar{a} = \{\alpha \in R | a-\alpha \in I\}.$$

It is called the **residue class** containing a modulo I. It is easy to see that if $\bar{a}\cap\bar{b} \neq \emptyset$, then $\bar{a} = \bar{b}$. We denote by R/I the set of all residue classes modulo I. We define sum and product in R/I by

$$\bar{a}+\bar{b} = \overline{a+b}$$
$$\bar{a}\cdot\bar{b} = \overline{ab}.$$

We find that $\bar{0}$ and $\bar{1}$ are the zero element and the identity element of R/I. It now follows that R/I is a commutative ring. The natural mapping

$$\phi : a \in R \longmapsto \bar{a} \in R/I$$

is a homomorphism, as is easily verified from the definitions of sum and product in R/I.

EXAMPLE A.11. The mapping $\phi : \mathbf{Q}[x] \to \mathbf{C}$ defined by

$$\phi : f(x) \in \mathbf{Q}[x] \longmapsto f(i) \in \mathbf{C}$$

is a homomorphism, and

$$\ker\phi = \{g(x) \in \mathbf{Q}[x] | g(i) = 0\}.$$

Since $f(x) \in \mathbf{Q}[x]$ has rational coefficients, we have

$$\overline{f(i)} = f(-i),$$

where the bar denotes the conjugate (this is true as long as the coefficients are real numbers). Hence $g(i) = 0$ implies $g(-i) = 0$, and thus $g(x)$ is divisible by $(x-i)(x+i) = x^2+1$. It follows that $\ker\phi$ is the ideal generated by x^2+1, that is,

$$\ker\phi = (x^2+1).$$

As we saw in Example A.6 of the preceding section, we can identify

$$\mathbf{Q}[x]/\ker\phi = \mathbf{Q}[x]/(x^2+1)$$

with $\mathbf{Q}(i)$. Now

$$\mathrm{im}\phi = \{\phi(a) | a \in \mathbf{Q}[x]\}$$

is called the **image** of ϕ, and we can show that $\mathrm{im}\phi = \mathbf{Q}(i)$. We may replace $\mathbf{Q}[x]$ by the ring $\mathbf{R}[x]$ of all polynomials with real coefficients and define a homomorphism by

$$\phi : f(x) \in \mathbf{R}[x] \longmapsto f(i) \in \mathbf{C}.$$

By a similar argument as above we find that

$$\ker\phi = (x^2+1),$$

and $\mathbf{R}[x]/(x^2+1)$ can be identified with $\mathbf{C}$.

EXAMPLE A.12. We consider a mapping ψ from the ring of polynomials in two variables $\mathbf{C}[x,y]$ into the ring of polynomials in one variable $\mathbf{C}[t]$ given by

$$\psi : f(x,y) \in \mathbf{C}[x,y] \longmapsto f(t^2,t^3) \in \mathbf{C}[t].$$

It is easy to see that ψ is a homomorphism. Suppose

$$g(x,y) = \sum a_{mn}x^m y^n$$

belongs to $\ker\psi$. Then

$$g(t^2,t^3) = \sum a_{mn}t^{2m+3n} = 0. \tag{A.2}$$

Suppose

$$t^{2m+3n} = t^{2m'+3n'}, \text{ that is, } 2m+3n = 2m'+3n'$$

for some pairs $(m,n) \neq (m',n')$. Then we get from $(m,n) \neq (m',n')$

$$2(m-m') = 3(n'-n), \text{ hence } m-m' = 3\ell, \quad n'-n = 2\ell, \quad \ell \geq 1.$$

From this we can show that $g(x,y)$ is divisible by x^3-y^2; the details are left as an exercise for the reader. Therefore

$$\ker\psi = (x^3-y^2).$$

How about

$$\mathrm{im}\psi = \{f(t^2,t^3) | \mathbf{C}[x,y]\}?$$

By noting that $\psi(f(x,y)) = t^{2m+3n}$, we see that there is no polynomial $f(x,y)$ such that $\psi(f(x,y)) = t$. On the other hand, we have

$$\psi(1) = t, \quad \psi(x) = t^2, \quad \psi(y) = t^3,$$

$$\psi(x^2) = t^4, \quad \psi(xy) = t^5, \quad \psi(y^2) = t^6, \ldots.$$

We find that $\mathrm{im}\psi$ is the set of all polynomials that do not contain first-order terms in t, namely,

$$A = \{a_0 + \sum_{j=2}^{m} a_j t^j | m \geq 2, a_j \in \mathbf{C}\}.$$

The subset $A \subset \mathbf{C}(t)$ is closed under sum and product in $\mathbf{C}(t)$ and forms a commutative ring. The fact that $A \neq \mathbf{C}(t)$ is related to the fact that the figure determined by x^3-y^2 in $\mathbf{C}^2$ has a singular point at the origin $(0,0)$. See § 2.3 (a).

Given a homomorphism $\phi : R \to S$ of commutative rings, we may define a mapping $\bar{\phi}$ of the residue ring $R/\ker\phi$ into S by

$$\bar{\phi} : \bar{a} \in R/\ker\phi \longmapsto \phi(a) \in S.$$

Then $\bar{\phi}$ is an injective homomorphism. The proof of this assertion is left as an exercise for the reader.

DEFINITION A.7. Let R be a commutative ring.

(i) If two elements $a \neq 0, b \neq 0$ satisfy $ab = 0$, we say that a is a **zero divisor** (and so is b). A commutative ring is called an **integral domain** if it has no zero divisor.

(ii) An ideal I of R is called a **prime ideal** if R/I is an integral domain.

We have already seen examples of zero divisors in (ii) of Theorem A.2 in Example A.5, and in (ii) of Theorem A.3. It is easy to show that a field has no zero divisors by using the existence of a multiplicative inverse.

A commutative ring is an integral domain if and only if $ab = 0$ implies $a = 0$ or $b = 0$. We may also paraphrase the definition of a prime ideal as follows. An ideal I is a prime ideal if and only if, in the residue ring R/I, $\bar{a} \cdot \bar{b} = \bar{0}$ implies $\bar{a} = 0$ or $\bar{b} = 0$. Since $\bar{a} \cdot \bar{b} = \bar{0}$ means $ab \in I$, we can say that I is a prime ideal if and only if $ab \in I$ implies $a \in I$ or $b \in I$. Or taking the contrapositive, we may say that I is a prime ideal if and only if $a \notin I$ and $b \notin b$ imply $ab \notin I$. In fact, we used this definition of a prime ideal in § 2.4(b).

EXAMPLE A.13. In the polynomial ring $R = \mathbf{C}[x_0, x_1, x_2, x_3]$, consider homogeneous polynomials

$$F = x_0x_3 - x_1x_2, \quad G = x_1^2 - x_0x_2, \quad H = x_2^2 - x_1x_3.$$

Then the ideal $I = (F, G, H)$ generated by F, G, and H is a prime ideal (see Example 2.18 and Exercise 2.8). The ideal $J = (F, G)$ generated by F and G is not a prime ideal. We may show that $H \notin J$ by showing that

$$H = AF + BG$$

leads to a contradiction by comparison of the coefficients on the two sides. Therefore we see that in R/J we have $\bar{H} \neq \bar{0}$. We have also $\bar{x}_0 \neq \bar{0}$. On the other hand, from

$$x_0(x_2^2 - x_1x_3) = -x_1F - x_2G$$

we get

$$\bar{x}_0 \cdot \bar{H} = \bar{0}.$$

Thus $\bar{x}_0, \bar{H}$ are zero divisors and J is not a prime ideal. (Furthermore, Example 2.19 gives a geometric reason why J is not a prime ideal.)

An ideal I of a commutative ring R is called a **maximal ideal** if $I \neq R$ and if an ideal J of R containing I must coincide with R or I.

THEOREM A.4. *An ideal I of a commutative ring R is a maximal ideal if and only if R/I is a field. In particular, a maximal ideal is a prime ideal.*

PROOF. Suppose I is a maximal ideal. Pick an element $a \in R, a \notin I$, and set

$$J = \{\alpha a + b | \alpha \in R, b \in I\}.$$

Then J is an ideal of R, because $\alpha a + b, \beta a + c \in J$ imply

$$(\alpha a + b) + (\beta a + c) = (\alpha + \beta)a + (b + c) \in J$$

as well as

$$r(\alpha a + b) = (r\alpha)a + rb \in J$$

for every $r \in R$. Since $a \notin I$, we have $I \neq J$. Since I is a maximal ideal, we must have $J = R$; in particular, $1 \in J$. This means that there exist $\alpha \in R, b \in I$ such that

$$1 = \alpha a + b.$$

In R/I, this means

$$\bar{1} = \bar{\alpha} \cdot \bar{a},$$

that is, the existence of a multiplicative inverse of $\bar{a} \neq \bar{0}$. Hence R/I is a field.

Conversely, assume that R/I is a field. Take an ideal $J \neq I$ such that $I \subset J$. If we take an element $a \in J, a \notin I$, then $\bar{a} \neq \bar{0}$ in R/I. Hence there is $\bar{b} \in R/I$ such that $\bar{a} \cdot \bar{b} = \bar{1}$. This means

$$ab - 1 \in I.$$

By setting $c = ab - 1$, we obtain

$$1 = ab - c, \quad c \in I.$$

Since $a, c \in J$, we get $1 \in J$, which implies $J = R$. Hence I is a maximal ideal.

EXAMPLE A.14. The set of all polynomials $\mathbf{Z}[x]$ with integral coefficients is a commutative ring relative to ordinary sum and product. The ideal (p) of $\mathbf{Z}[x]$ generated by a prime number p is a prime ideal of $\mathbf{Z}[x]$.The residue ring $\mathbf{Z}[x]/(p)$ has the same commutative ring structure as the polynomial ring $\mathbf{Z}/(p)[x]$, namely, the set of all polynomials in x over the field $Z/(p)$. Since $\mathbf{Z}/(p)[x]$ is an integral domain, it follows that (p) is a prime ideal. Next, pick a polynomial $f(x) \in \mathbf{Z}[x]$ and consider the ideal $(p, f(x))$ generated by p and $f(x)$. If we write

$$f(x) = \sum_{j=0}^{n} a_j x^j, \quad a_j \in \mathbf{Z},$$

and set

$$\bar{f}(x) = \sum_{j=0}^{n} \bar{a}_j x^j,$$

where $\bar{a}_j$ is the element in $\mathbf{Z}/(p)$ determined by a_j, then $\bar{f}(x) \in \mathbf{Z}/(p)[x]$. Suppose $\bar{f}[x]$ is reducible in $\mathbf{Z}/(p)[x]$, that is,

$$\bar{f}(x) = \bar{g}(x)\bar{h}(x), \quad \deg\bar{g}(x) \geq 1, \deg\bar{h}(x) \geq 1,$$

where $\bar{g}(x), \bar{h}(x)$ denote the polynomials over $\mathbf{Z}/(p)$ obtained from $g(x), h(x)$ by considering their coefficients in $\mathbf{Z}/(p)$, just like $\bar{f}(x)$ corresponding to $f(x)$. Then the coefficients of $f(x) - g(x)h(x)$ are divisible by p, and hence

$$g(x)h(x) \in (p, f(x)).$$

On the other hand, $g(x) \notin (p, f(x))$ and $h(x) \notin (p, f(x))$. This means that $(p, f(x))$ is not a prime ideal.

If $\bar{f}(x)$ is irreducible in $\mathbf{Z}/(p)[x]$, then $\mathbf{Z}/(p)[x]/(\bar{f}(x))$ is a field, by arguments similar to those in Lemma A.5 and Corollary A.2 of the preceding section. On the other hand, we can show that $\mathbf{Z}[x]/(p, f(x))$ and $\mathbf{Z}/(p)[x]/(\bar{f}(x))$ have the same commutative ring structure (that is, they are isomorphic). It now follows that $(p, f(x))$ is a maximal ideal of $\mathbf{Z}[x]$.

The next theorem is fundamental for proving the zero point theorem of Hilbert. The proof requires some more preparations and is hence omitted here.

THEOREM A.5. *Every maximal ideal of the polynomial ring* $\mathbf{C}[x_1, x_2, \ldots, x_n]$ *over* $\mathbf{C}$ *is of the form* $(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$, *where* $a_j \in \mathbf{C}$.

The ideal $(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$ coincides with the kernel, $\ker \phi$, of the homomorphism

$$\phi : f(x_1, x_2, \ldots, x_n) \in \mathbf{C}[x_1, x_2, \ldots, x_n] \longmapsto f(a_1, a_2, \ldots, a_n) \in \mathbf{C}.$$

We note that this theorem is based on the fact that a polynomial in one variable over $\mathbf{C}$ has roots in $\mathbf{C}$. For example, in $\mathbf{R}[x_1, x_2]$, the ideal $(x_1 - a, x_2^2 + 1), a \in \mathbf{R}$, is maximal.

A field K is said to be **algebraically closed** if every polynomial $f(x)$ with coefficients in K has roots in K. The complex number field $\mathbf{C}$ is algebraically closed (this is called the **fundamental theorem of algebra**). Theorem A.5 is valid when $\mathbf{C}$ is replaced by any algebraically closed field K. Incidentally, it is known that for every field k there is an algebraically closed field $\bar{k}$ that contains k.

§ A.4. Finite fields

In the preceding section we gave a general definition of a field. Let us consider the result of adding the identity element 1 of a field K n times:

$$\underbrace{1 + 1 + \cdots + 1}_{n}.$$

If this is not 0 for any positive integer n, we say that the **characteristic** of K is zero. The fields $\mathbf{Q}, \mathbf{R}$, and $\mathbf{C}$ are examples. On the other hand, it is possible that this sum is 0 for some n. For example, in the field $\mathbf{Z}/(p)$ considered in § A.1, we have

$$\underbrace{\bar{1} + \bar{1} + \cdots + \bar{1}}_{p} = \overline{\underbrace{1 + 1 + \cdots + 1}_{p}} = \bar{p} = \bar{0}.$$

When the characteristic of K is not 0, the smallest positive integer n for which

$$\underbrace{1 + 1 + \cdots + 1}_{n} = 0$$

is called the **characteristic** of K.

LEMMA A.8. *The characteristic of a field is* 0 *or a prime number.*

PROOF. Suppose the characteristic of a field K is n. If n can be factored as

$$n = m\ell, \quad m \geq 2, \ell \geq 2,$$

we set

$$a = \underbrace{1 + 1 + \cdots + 1}_{m}$$

$$b = \underbrace{1 + 1 + \cdots + 1}_{\ell}.$$

By definition of the characteristic, we have $a \neq 0, b \neq 0$. by using the distributive law, we see that

$$\begin{aligned} ab &= a\underbrace{(1 + 1 + \cdots + 1)}_{\ell} \\ &= \underbrace{a + a + \cdots + a}_{\ell} \\ &= \underbrace{\underbrace{1 + 1 + \cdots + 1}_{m} + \underbrace{1 + 1 + \cdots + 1}_{m} + \underbrace{1 + 1 + \cdots + 1}_{m}}_{\ell} \\ &= \underbrace{1 + 1 + \cdots + 1}_{n} = 0. \end{aligned}$$

Thus a and b are zero divisors, a contradiction because K is a field. This proves that n is a prime number.

Now let us consider a field K of characteristic $p \geq 2$. For any integer m such that $1 \leq m < p$, we denote by $\bar{m}$ the sum

$$\underbrace{1 + 1 + \cdots + 1}_{m}.$$

Consider the subset of K

$$\mathbf{F}_p = \{0, 1, \bar{2}, \ldots, \overline{p-1}\}.$$

It turns out that $\mathbf{F}_p$ is a field relative to sum and product in K, because we easily see that

$$\bar{m} + \bar{n} = \overline{m + n}$$
$$\bar{m} \cdot \bar{n} = \overline{mn}.$$

Furthermore, we leave the existence of an additive inverse and of a multiplicative inverse as an exercise; they can be done similarly to (1) of Theorem A.2 in § A.1. As a matter of fact, beyond similarity of notation we can simply show that $\mathbf{F}_p$ and $Z/(p)$ have the same structure. We identify the two in the sequel. We call $\mathbf{F}_p$ the **prime field** of characteristic p. All fields K of characteristic p always contain the prime field $\mathbf{F}_p$ and can be constructed from it.

Now $\mathbf{F}_p$ has p elements. A field with a finite number of elements is called a **finite field** or **Galois field**. In a finite field K, if we keep adding the identity

element, the results cannot be all distinct. Thus there exist integers $m > \ell \geq 1$ such that

$$\underbrace{1+1+\cdots+1}_{m} = \underbrace{1+1+\cdots+1}_{\ell},$$

which implies

$$\underbrace{1+1+\cdots+1}_{m-\ell} = 0.$$

Therefore the characteristic of a finite field must be a prime number $p \geq 2$.

How can we construct finite fields other than the prime fields ? We start with

EXAMPLE A.15. The prime field $\mathbf{F}_2$ of characteristic 2 consists of the zero element 0 and the identity element 1 and is the simplest finite field, $\mathbf{F}_2 = \{0, 1\}$. Now consider the polynomial ring $\mathbf{F}_2[x]$ of all polynomials with coefficients in $\mathbf{F}_2$. In $\mathbf{F}_2$, the polynomial $x^2 + 1$ is reducible. In F_2, we have $2 = 0$ (more precisely, in the notation above we should write $\bar{2} = 0$, but for simplicity we shall skip bars in the following). Therefore we have

$$(x+1)^2 = x^2 + 2x + 1 = x^2 + 1.$$

On the other hand, $\mathbf{F}_2[x]$ has only x and $x+1$ as linear polynomials, and x^2+x+1 cannot be expressed as a product of these linear polynomials. Thus $x^2 + x + 1$ is irreducible. Now we can show, by a method similar to that for Theorem A.3 (2), that the commutative ring $K = \mathbf{F}_2[x]/(x^2+x+1)$ is a field. If we denote by α the residue class $\bar{x}$ of x, then we have

$$\alpha^2 + \alpha + 1 = 0,$$

and every element in K can be expressed in the form

$$a\alpha + b, \quad a, b \in \mathbf{F}_2.$$

The 4 elements of K are

$$0, 1, \alpha, \alpha + 1.$$

Since $1 + 1 = 0$, we have $-1 = 1$ and hence

$$\alpha^2 = -\alpha - 1 = \alpha + 1.$$

Furthermore, we get

$$\begin{aligned} \alpha^3 &= \alpha \cdot \alpha^2 = \alpha(\alpha + 1) \\ &= \alpha^2 + \alpha = -1 = 1, \end{aligned}$$

and hence we can write

$$K = \{0, \alpha, \alpha^2, \alpha^3\}.$$

It also follows from $\alpha^3 = 1$ that α^{-1} is equal to $\alpha^2 + \alpha + 1$.

EXAMPLE A.16. The polynomial $x^3 + x + 1$ is irreducible in $\mathbf{F}_2$, and hence

$$K = \mathbf{F}_2[x]/(x^3 + x + 1)$$

is a field of characteristic 2. Denoting by β the residue class $\bar{x}$ of x, we have

$$\beta^3 + \beta + 1 = 0.$$

Since every element of K is the residue class of a certain polynomial of degree at most 2, it is expressible in the form

$$a + b\beta + c\beta^2, \quad a, b, c \in \mathbf{F}_2.$$

It follows that K has $2^3 = 8$ elements. Using $\beta^3 + \beta + 1 = 0$, we get

$$\begin{aligned} \beta^3 &= -\beta - 1 = \beta + 1 \\ \beta^4 &= \beta \cdot \beta^3 = \beta(\beta + 1) = \beta^2 + \beta \\ \beta^5 &= \beta(\beta^2 + \beta) = \beta^3 + \beta^2 = 1 + \beta + \beta^2 \\ \beta^6 &= \beta(1 + \beta + \beta^2) = \beta + \beta^2 + \beta^3 = \beta + \beta^2 + \beta + 1 = 1 + \beta^2 \\ \beta^7 &= \beta(1 + \beta^2) = \beta + \beta^3 = \beta + \beta + 1 = 1, \end{aligned}$$

and conclude that

$$K = \{0, \beta, \beta^2, \beta^3, \ldots, \beta^7\}.$$

EXAMPLE A.17. In the polynomial ring $\mathbf{F}_3[x]$ over the prime field $\mathbf{F}_3 = \{0, 1, 2\}$ of characteristic 3, the polynomial $x^2 + x + 2$ is irreducible. ($\mathbf{F}_3[x]$ has only 6 linear polynomials $\pm x + 1, \pm x - 1, \pm x$; and note that $2 = -1$.) The field

$$K = \mathbf{F}_3[x]/(x^2 + x + 2)$$

has characteristic 3. Denoting by γ the residue class $\bar{x}$ of x, we have

$$\gamma^2 + \gamma + 2 = 0.$$

Every element of K can be expressed in the form

$$a + b\gamma, \quad a, b \in \mathbf{F}_3,$$

and

$$K = \{0, \gamma, \gamma^2, \gamma^3, \ldots, \gamma^8\}.$$

This is left for the reader to verify as an exercise.

The example above turns out to be a special case of the following theorem.

THEOREM A.6. *Consider an irreducible polynomial $f(x)$ of degree n in the polynomial ring $F[x]$ over a field F. Then*

$$K = F[x]/(f(x))$$

is a field. Denoting by α the residue class $\bar{x}$ of x, we have

$$K = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n_1}\alpha^{n-1} | a_j \in F\},$$

which has dimension n as a vector space over F. If F is a finite field with q elements, then K has q^n elements.

PROOF. We have built up a long explanation toward this theorem since § A.1; by now the proof should be obvious. First of all, note that $F[x]$ has factorization and that the analogues of Lemma A.5 of § A.2 and Corollary A.2 are valid. Hence we can use the proof of Theorem A.3 here as it is. Division by $f(x)$ in $F[x]$ has residues of degree $\leq n-1$. Regarding elements of F as elements of K, we see that every element of K can be put in the form

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1}.$$

From Theorem A.6 we see that α is a root of the equation $f(x) = 0$, that is, $f(\alpha) = 0$. Unlike the case of complex coefficients, we do not know in advance the field in which we solve $f(x) = 0$; we could say that we are creating a field containing a root of $f(x) = 0$ by forming $K = F[x]/(f(x))$. If we are given a root α of $f(x) = 0$ in some way, and if $f(x)$ is irreducible in $F[x]$, then we can show, as in Example A.9, that

$$F(\alpha) = \{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1} | a_j \in F\}.$$

In fact, we can show the the mapping

$$\phi : \overline{g(x)} \in F[x]/(f(x)) \longmapsto g(\alpha) \in F(\alpha)$$

is an isomorphism between the two fields K and $F(\alpha)$. We say that $F(\alpha)$ is the field obtained by adjoining α to F, and that it is an extension of of degree n of F. We shall identify $K = F[x]/(f(x))$ and $F(\alpha)$ by ϕ.

We shall summarize the fundamental properties of finite fields as follows.

THEOREM A.7.

(i) *The number of elements in a finite field of characteristic p is p^m, $m \geq 1$.*

(ii) *If a finite field K of characteristic p has $q = p^n$ elements, then for any element a in $K^* = K - \{0\}$ we have*

$$a^{q-1} = 1.$$

Furthermore, we have

$$K^* = \{\alpha, \alpha^2, \alpha^3, \ldots, \alpha^{q-1}\}$$

for an appropriate choice of $\alpha \in K^*$. (In group theory terminology, K^* is a cyclic group or order $q-1$. An element α as above is called a **primitive element**).

(iii) *If two finite fields K_1 and K_2 have the same number of elements, they are isomorphic as fields. In other words, a field with q elements is essentially unique. This field is denoted by $GF(q)$.*

We have to omit the proof except for (i), which follows easily from the fact that a finite field K of characteristic p contains the prime field $\mathbf{F}_p$ and is a vector space over $\mathbf{F}_p$. As for (ii) the reader should refer to Examples A.15, A.16, and A.17. For (iii), see, for example, p. 412 of G. Birkhoff and S. MacLane: *A Survey of Modern Algebra*, Third Edition, MacMillan, New York, 1965

In conclusion, we state phenomena peculiar to fields of characteristic p.

LEMMA A.9. *For polynomials over a field K of characteristic p we have*

$$(x+y)^{p^\nu} = x^{p^\nu} + y^{p^\nu}, \quad \nu = 1, 2, 3, \ldots.$$

PROOF. We use induction on ν. For $\nu = 1$, the binomial formula says that

$$(x+y)^p = \sum_{j=0}^{p} \binom{p}{j} x^{p-j} y^j.$$

If $j \neq 0, \neq p$, then $\binom{p}{j}$ is a multiple of p, which means that it is 0 in a field of characteristic p. Therefore

$$(x+y)^p = x^p + y^p.$$

Now assume that the formula is valid up to ν. Then

$$\begin{aligned}(x+y)^{p^{\nu+1}} &= \{(x+y)^{p^\nu}\}^p = (x^{p^\nu} + y^{p^\nu})^p \\ &= (x^{p^\nu})^p + (y^{p^\nu})^p = x^{p^{\nu+1}} y^{p^{\nu+1}},\end{aligned}$$

showing that the formula is valid for $\nu + 1$.

COROLLARY A.4. *Suppose a polynomial $f(x)$ over a finite field K of characteristic p has a root α. If K has $q = p^m$ elements. then α^q is also a root of $f(x)$.*

PROOF. Let

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n, \quad a_j \in K.$$

By the lemma above, we have

$$f(x)^q = a_0^q + a_1^q x^q + a_2^q x^{2q} + \cdots a_n^q x^{nq}.$$

By Theorem A.7 (ii), we get

$$a_j^q = a_j,$$

and hence

$$f(x)^q = a_0^q + a_1^q \alpha^q + a_2(\alpha^q)^2 + \cdots a_n^q(\alpha^q)^n = f(x^q).$$

Therefore we get

$$0 = f(\alpha)^q = a_0 + a_1\alpha^q + a_2(\alpha^q)^2 + \cdots + a_n(\alpha^q)^n = f(\alpha^q),$$

which shows that α^q is a root of $f(x)$.

§ A.5. Localization and local rings

A commutative ring R is called a **local ring** if it has only one maximal ideal. Local rings play an important role in algebraic geometry. Hence we shall explain the method of getting a local ring from a commutative ring by localization. For simplicity, we shall assume that R is an integral domain.

A subset S of a commutative ring R is called a **multiplicatively closed set** if it satisfies the following conditions:

(i) If $a, b \in S$, then $ab \in S$;

(ii) $0 \notin S$, $1 \in S$.

Given R and a multiplicatively closed subset S, we shall define a new commutative ring $S^{-1}R$ as follows. (The definition below might be considered a generalization of a fraction.)

In the set of all expressions $\frac{a}{s}$, where $a \in R, s \in S$, we define an equivalence relation by

$$\frac{a}{s} \equiv \frac{a'}{s'} \text{ if and only if } as' = a's. \tag{A.3}$$

This is an analogue to the equality of two fractions.

Note: When R is not an integral domain, the condition is replaced by the existence of $s'' \in S$ such that $s''(as' - a's) = 0$.

With this identification, we consider the set $S^{-1}R$ of the equivalence classes of all $\frac{a}{s}$, $a \in R, s \in S$. We define sum and product in $S^{-1}R$ by

$$\frac{a}{s} + \frac{b}{t} = \frac{at + bs}{st}$$
$$\frac{a}{s} \cdot \frac{b}{t} = \frac{ab}{st}.$$

Since

$$\frac{a}{s} \equiv \frac{at}{st}, \quad \frac{b}{t} \equiv \frac{bs}{st},$$

where $s, t \in S$, is obvious from (A.3), the definitions of sum and product make sense. It follows that $S^{-1}R$ is a commutative ring with $\frac{0}{1}$ as the zero element and $\frac{1}{1}$ as the identity element. The mapping

$$\psi : a \in R \longmapsto \frac{a}{1} \in S^{-1}R$$

is an injective homomorphism of R into $S^{-1}R$; we may identify a with $\frac{a}{1}$. Thus we may write 0 and 1 for the zero element and the identity of $S^{-1}R$.

Note: When R is not an integral domain, we have

$$\ker \psi = \{a \in R | \text{ there is an } s \in S \text{ such that } sa = 0\},$$

so that ψ is not necessarily injective.

EXAMPLE A.18. For an integral domain R, $S = R - \{0\}$ is multiplicatively closed. In this case, we see that $S^{-1}R$ is a field. In fact, each element of $S^{-1}R$ is of the form $\frac{b}{a}$, where $a, b \in R, a \neq 0$, and $\frac{b}{a} = 0$ only when $b = 0$ by (A.3). Hence if $\frac{b}{a} \neq 0$, then $\frac{a}{b} \in S^{-1}R$. By definition of product, we get

$$\frac{b}{a} \cdot \frac{a}{b} = 1.$$

This means that $S^{-1}R$ is a field. It is called the **field of quotients** of the integral domain R. As we already stated, we can identify a with $\frac{a}{1}$ and consider $R \subset S^{-1}R$. For $R = \mathbf{Z}$, its field of quotients is $\mathbf{Q}$ – the passage from $\mathbf{Z}$ to $\mathbf{Q}$ is nothing but the construction of fractions. When R is the ring of polynomials $K[x]$ over a field K, its field of quotients is the field of rational functions $K(x)$, that is,

$$K(x) = \left\{ \frac{g(x)}{f(x)} \mid f(x), g(x) \in K[x], f(x) \neq 0 \right\}.$$

For the ring $K[x_1, x_2, \ldots, x_n]$ of polynomials in n variables, the field of quotients is the field of rational functions in n variables,

$$K(x_1, x_2, \ldots, x_n) = \left\{ \frac{g(x_1, x_2, \ldots, x_n)}{f(x_1, x_2, \ldots, x_n)} \mid f \neq 0, f, g \in K[x_1, x_2, \ldots, x_n], \right\}.$$

EXAMPLE A.19. For a prime ideal $\mathfrak{p}$ of an integral domain R, set

$$S = R - \mathfrak{p} = \{a \in R | a \notin \mathfrak{p}\}.$$

Since $0 \in \mathfrak{p}$, we have $0 \notin S$. Furthermore $1 \notin \mathfrak{p}$ implies $1 \in S$. For $a, b \in S$, we have $a, b \notin \mathfrak{p}$, which implies $ab \notin \mathfrak{p}$, since $\mathfrak{p}$ is a prime ideal. Hence

$$ab \in S.$$

This proves that S is multiplicatively closed. Denote $S^{-1}R$ by $R_{\mathfrak{p}}$. We may think of R as a subset of $R_{\mathfrak{p}}$. If we denote by $\mathfrak{p}R_{\mathfrak{p}}$ the ideal of $R_{\mathfrak{p}}$ generated by $\mathfrak{p}$, then we see that

$$\mathfrak{p}R_{\mathfrak{p}} = \left\{ \frac{s}{a} | s \in \mathfrak{p}, a \notin \mathfrak{p} \right\}.$$

It is also easy to see that

$$\mathfrak{p}R_{\mathfrak{p}} \cap R = \mathfrak{p}.$$

Using these facts, we can naturally define a mapping

$$\phi : R/\mathfrak{p} \longrightarrow R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$$

by associating to the residue class of $a \in R$ modulo $\mathfrak{p}$ the residue class of a modulo $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$. If $\bar{a} \in R/\mathfrak{p}$ is not the zero element, then $a \notin \mathfrak{p}$. Therefore $\frac{1}{a} \in R_{\mathfrak{p}}, a \notin \mathfrak{p}$, and $\frac{1}{a} \notin \mathfrak{p}R_{\mathfrak{p}}$. If we denote the residue class of a modulo $\mathfrak{p}R_{\mathfrak{p}}$ by $\bar{a}$ and the residue class of $\frac{1}{a}$ by $\overline{(\frac{1}{a})}$, then

$$\bar{a} \cdot \overline{(\frac{1}{a})} = \bar{1}$$

in $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$. In general, if $\frac{b}{a} \in R_{\mathfrak{p}}$ and $\frac{b}{a} \notin \mathfrak{p}R_{\mathfrak{p}}$, then $a \notin \mathfrak{p}$, $b \notin \mathfrak{p}$, and $\frac{a}{b} \in R_{\mathfrak{p}}$. Therefore

$$\overline{(\frac{b}{a})} \cdot \overline{(\frac{a}{b})} = 1$$

in $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$. Thus we find that $\mathfrak{p}R_{\mathfrak{p}}$ is a maximal ideal of $R_{\mathfrak{p}}$. Since we have

$$\overline{(\frac{b}{a})} = \bar{b} \cdot \overline{(\frac{1}{a})},$$

it follows that $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$ is the field of quotients of the integral domain $R/\mathfrak{p}$.

Now we see that the only maximal ideal of $R_{\mathfrak{p}}$ is $\mathfrak{p}R_{\mathfrak{p}}$, and hence $R_{\mathfrak{p}}$ is a local ring. Let $J \neq R_{\mathfrak{p}}$ be an ideal of $R_{\mathfrak{p}}$. If J is not contained in $\mathfrak{p}R_{\mathfrak{p}}$, then there exists an element $\frac{b}{a}$ such that

$$\frac{b}{a} \in J, \quad \frac{b}{a} \notin \mathfrak{p}R_{\mathfrak{p}}.$$

From $\frac{b}{a} \notin \mathfrak{p}R_{\mathfrak{p}}$ we find

$$a \notin \mathfrak{p}, \quad b \notin \mathfrak{p},$$

and hence

$$\frac{a}{b} \in R_{\mathfrak{p}}.$$

Since J is an ideal of $R_{\mathfrak{p}}$, we get

$$\frac{a}{b} \cdot \frac{b}{a} = 1 \in J,$$

which implies $J = R_{\mathfrak{p}}$ — a contradiction. Hence $J \subset \mathfrak{p}R_{\mathfrak{p}}$, from which we see that $\mathfrak{p}R_{\mathfrak{p}}$ is the only maximal ideal of $R_{\mathfrak{p}}$.

If $R = K[x]$ and $\mathfrak{p} = (x - a)$, then we find that

$$R_{\mathfrak{p}} = \left\{ \frac{g(x)}{f(x)} | f(a) \neq 0, f(x), g(x) \in K[x] \right\}$$

and

$$\mathfrak{p}R_{\mathfrak{p}} = \left\{ \frac{(x-a)h(x)}{f(x)} | f(a) \neq 0, f(x)h(x) \in K[x] \right\}.$$

If $R = K[x, y]$ and $\mathfrak{p} = (y^2 - x^3)$, then we have

$$R_{\mathfrak{p}} = \left\{ \frac{g(x,y)}{f(x,y)} | f(x,y) \text{ is not divisible by } y^2 - x^3, f, g \in K[x,y] \right\}.$$

If we define a mapping ϕ from $R_{\mathfrak{p}}$ into $K(t)$ by

$$\phi : \frac{g(x,y)}{f(x,y)} \in R_{\mathfrak{p}} \longmapsto \frac{g(t^2,t^3)}{f(t^2,t^3)} \in K(t),$$

we find that it is a homomorphism and that

$$\ker \phi = \mathfrak{p}R_{\mathfrak{p}}.$$

It now follows that $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$ and $K(t)$ are isomorphic as fields. Finally, we mention one important local ring.

EXAMPLE A.20 (THE RING OF FORMAL POWER SERIES). Denote by $K[[x]]$ the set of all **formal power series** with coefficients in K

$$f(x) = \sum_{j=0}^{\infty} a_j x^j, \quad a_j \in K.$$

We can define sum and product of formal power series as natural extensions of those operations for polynomials. Then $K[[x]]$ is a commutative ring. If $a_0 \neq 0$ in the power series $f(x)$, then there exists a formal power series $g(x)$ such that

$$f(x)g(x) = 1.$$

By setting

$$g(x) = \sum_{j=0}^{\infty} b_j x^j$$

we see that $f(x)g(x) = 1$ is expressed by the series of conditions

$$\begin{aligned}
&a_0 b_0 = 1 \\
&a_0 b_1 + a_1 b_0 = 0 \\
&a_0 b_2 + a_1 b_1 + a_2 b_0 = 0 \\
&\cdots\cdots\cdots \\
&a_0 b_k + a_1 b_{k-1} + a_2 b_{k-2} + \cdots a_k b_0 = 0 \\
&\cdots\cdots\cdots .
\end{aligned}$$

Since $a_0 \neq 0$, we get $b_0 = 1/a_0, \quad b_1 = -a_1/a_0^2, \ldots$, finding the coefficients b_k and hence $g(x)$ uniquely. We can now show that $K[[x]]$ has only one maximal ideal, (x), generated by x, that is,

$$(x) = \left\{ \sum_{j=1}^{\infty} a_j x^j \in K[[x]] \right\}.$$

Every power series $h(x)$ can be expressed in the form

$$h(x) = x^m \sum_{j=0}^{\infty} c_j x^j, \quad c_0 \neq 0,$$

and

$$\sum_{j=0}^{\infty} c_j x^j$$

has an inverse in $k[[x]]$. Thus in the field of quotients for $K[[x]]$, we can write

$$\frac{1}{h(x)} = \frac{1}{x^m} \left\{ \frac{1}{c_0} - \frac{c_1}{c_0^2} x + \cdots \right\}.$$

It follows that the field of quotients of $K[[x]]$ is the set of all formal power series with a finite number of negative power terms

$$K((x)) = \left\{ \sum_{j=-m}^{\infty} a_j x^j | a_j \in K \right\}.$$

Each element of $K((x))$ is called a **formal Laurent series**. In the same fashion, the ring of formal power series in several variables $K[[x_1, x_2, \ldots, x_n]]$ is a local ring.

References

[1] M. Reid, Undergraduate Algebraic Geometry, London Math. Soc. Student Texts 12, Cambridge University Press, 1988.

This is a good introduction for those who have just finished reading our book.

[2] K. Iwasawa, Algebraic Functions, Trans. Math. Monographs, vol. 118, American Mathematical Society, Providence, 1993 (translation of the revised Japanese edition, Iwanami Shoten, 1973).

This celebrated book deals with the algebraic theory of algebraic functions based on the theory of valuations and the analytic theory of closed Riemann surfaces. It is recommended for all students of mathematics, regardless of their special interests.

[3] C.L. Siegel, Topics in Complex Function Theory, I, II, III, Wiley-Interscience, 1969, 1971, 1973.

This authoritative treatise gives a lucid treatment, from elliptic functions to the theory of Jacobian varieties.

[4] H. Matsumura, Commutative Algebra, 2nd ed. Benjamin, 1980.

This is a very readable book on the theory of commutative rings.

[5] D. Mumford, The Red Book of Varieties and Schemes, Lecture Notes in Math. No. 1358, Springer, 1988.

A well-known book giving an introduction to the theory of schemes.

[6] A. Grothendieck and J.A. Dieudonné, Eléments de Géométrie Algébrique, Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques, vol. 4, 8, 11, 17, 20, 24, 28, 32, 1960-67. Revised edition of EGA I, Springer, 1971.

If you have the prerequisites on commutative rings and homological algebra, you might want to read this voluminous unfinished work. It is easier than some other introductory books.

[7] I.R. Schafarevitch, Basic Algebraic Geometry I, II, Springer, 1994.

[8] J.P. Serre, Faisceaux algébriques cohérents, Ann. of Math. 61(1955), 197-278 (reprinted in his Complete Works, Vol. 1, pp. 301-339).

This is a pioneering work on algebraic geometry that uses sheaf theory. You may learn sheaf cohomology theory from this book as a supplement to your reading of [5].

[9] J.P. Serre, Géométrie algébrique et géométrie analytique, Ann. Inst. Fourier 6(1956), 1-42 (reprinted in his Complete Works, Vol. 1, pp. 402-443).

This paper is generally cited as "GAGA".

[10] J.W.S. Cassels, Lectures on Elliptic Curves, London Math. Soc. Student Texts 24, Cambridge University Press, 1991.

This book and the next are unique introductions to elliplic curves.

[11] J. H. Silverman and J. Tate, Rational Points on Elliptic Curves, Springer, 1992.

Recommended for those who want to study elliptic curves, including computer-aided experimental aspect of the theory of elliptic curves.

[12] J. H. Silverman, The Arithmetic of Elliptic Curves, Springer, 1986.

[13] D.A. Cox and D. O'Shea, Ideals, Varieties, and Algorithms, Springer, 1992; 2nd ed., 1996.

A good book dealing with ideal theory in polynomial rings and algebraic varieties. You will profit by using software such as Maple or Mathematica.

[14] C. Moreno, Algebraic Curves over Finite Fields, Cambridge University Press, 1991.

This book and the next deal with applications of algebraic geometry to coding theory.

[15] J.H. van Lint and G. van der Geer, Introduction to Coding Theory and Algebraic Geometry, Birkhäuser, 1989.

[16] D.M. Bressoud, Factorization and Primality Testing, Springer, 1989.

The last chapter deals with applications of elliptic curves to primality testing, which is also included in [10].

[17] D. Mumford, Algebraic Geometry I: Complex Projective Varieties, Springer, 1976.

This contains important results in algebraic geometry over the complex numbers.

[18] P. Griffiths and J. Harris, Principles of Algebraic Geometry, J. Wiley, 1978.

We recommend Chapter 2 of the book for complex curves and Riemann surfaces.

[19] C.H. Clemens, A Scrap Book of Complex Curve Theory, Plenum Press, 1980.

The book contains interesting information on algebraic curves and Riemann surfaces.

[20] W. Fulton, Algebraic Curves, Benjamin, 1969.

The book contains an algebraic treatment of algebraic curves.

Index

Index for Definitions, Theorems, etc.

(The page numbers are given in parentheses)

Selected Titles in This Series

(*Continued from the front of this publication*)

128 **V. P. Orevkov,** Complexity of proofs and their transformations in axiomatic theories, 1993
127 **F. L. Zak,** Tangents and secants of algebraic varieties, 1993
126 **M. L. Agranovskiĭ,** Invariant function spaces on homogeneous manifolds of Lie groups and applications, 1993
125 **Masayoshi Nagata,** Theory of commutative fields, 1993
124 **Masahisa Adachi,** Embeddings and immersions, 1993
123 **M. A. Akivis and B. A. Rosenfeld,** Élie Cartan (1869–1951), 1993
122 **Zhang Guan-Hou,** Theory of entire and meromorphic functions: deficient and asymptotic values and singular directions, 1993
121 **I. B. Fesenko and S. V. Vostokov,** Local fields and their extensions: A constructive approach, 1993
120 **Takeyuki Hida and Masuyuki Hitsuda,** Gaussian processes, 1993
119 **M. V. Karasev and V. P. Maslov,** Nonlinear Poisson brackets. Geometry and quantization, 1993
118 **Kenkichi Iwasawa,** Algebraic functions, 1993
117 **Boris Zilber,** Uncountably categorical theories, 1993
116 **G. M. Fel'dman,** Arithmetic of probability distributions, and characterization problems on abelian groups, 1993
115 **Nikolai V. Ivanov,** Subgroups of Teichmüller modular groups, 1992
114 **Seizô Itô,** Diffusion equations, 1992
113 **Michail Zhitomirskiĭ,** Typical singularities of differential 1-forms and Pfaffian equations, 1992
112 **S. A. Lomov,** Introduction to the general theory of singular perturbations, 1992
111 **Simon Gindikin,** Tube domains and the Cauchy problem, 1992
110 **B. V. Shabat,** Introduction to complex analysis Part II. Functions of several variables, 1992
109 **Isao Miyadera,** Nonlinear semigroups, 1992
108 **Takeo Yokonuma,** Tensor spaces and exterior algebra, 1992
107 **B. M. Makarov, M. G. Goluzina, A. A. Lodkin, and A. N. Podkorytov,** Selected problems in real analysis, 1992
106 **G.-C. Wen,** Conformal mappings and boundary value problems, 1992
105 **D. R. Yafaev,** Mathematical scattering theory: General theory, 1992
104 **R. L. Dobrushin, R. Kotecký, and S. Shlosman,** Wulff construction: A global shape from local interaction, 1992
103 **A. K. Tsikh,** Multidimensional residues and their applications, 1992
102 **A. M. Il'in,** Matching of asymptotic expansions of solutions of boundary value problems, 1992
101 **Zhang Zhi-fen, Ding Tong-ren, Huang Wen-zao, and Dong Zhen-xi,** Qualitative theory of differential equations, 1992
100 **V. L. Popov,** Groups, generators, syzygies, and orbits in invariant theory, 1992
99 **Norio Shimakura,** Partial differential operators of elliptic type, 1992
98 **V. A. Vassiliev,** Complements of discriminants of smooth maps: Topology and applications, 1992 (revised edition, 1994)
97 **Itiro Tamura,** Topology of foliations: An introduction, 1992
96 **A. I. Markushevich,** Introduction to the classical theory of Abelian functions, 1992
95 **Guangchang Dong,** Nonlinear partial differential equations of second order, 1991
94 **Yu. S. Il'yashenko,** Finiteness theorems for limit cycles, 1991
93 **A. T. Fomenko and A. A. Tuzhilin,** Elements of the geometry and topology of minimal surfaces in three-dimensional space, 1991
92 **E. M. Nikishin and V. N. Sorokin,** Rational approximations and orthogonality, 1991

(See the AMS catalog for earlier titles)

Selected Titles in This Series

166 **Kenji Ueno,** An introduction to algebraic geometry, 1997

165 **V. V. Ishkhanov, B. B. Lur′e, and D. K. Faddeev,** The embedding problem in Galois theory, 1997

164 **E. I. Gordon,** Nonstandard methods in commutative harmonic analysis, 1997

163 **A. Ya. Dorogovtsev, D. S. Silvestrov, A. V. Skorokhod, and M. I. Yadrenko,** Probability theory: Collection of problems, 1997

162 **M. V. Boldin, G. I. Simonova, and Yu. N. Tyurin,** Sign-based methods in linear statistical models, 1997

161 **Michael Blank,** Discreteness and continuity in problems of chaotic dynamics, 1997

160 **V. G. Osmolovskiĭ,** Linear and nonlinear perturbations of the operator div, 1997

159 **S. Ya. Khavinson,** Best approximation by linear superpositions (approximate nomography), 1997

158 **Hideki Omori,** Infinite-dimensional Lie groups, 1997

157 **V. B. Kolmanovskiĭ and L. E. Shaĭkhet,** Control of systems with aftereffect, 1996

156 **V. N. Shevchenko,** Qualitative topics in integer linear programming, 1997

155 **Yu. Safarov and D. Vassiliev,** The asymptotic distribution of eigenvalues of partial differential operators, 1997

154 **V. V. Prasolov and A. B. Sossinsky,** Knots, links, braids and 3-manifolds. An introduction to the new invariants in low-dimensional topology, 1997

153 **S. Kh. Aranson, G. R. Belitsky, and E. V. Zhuzhoma,** Introduction to the qualitative theory of dynamical systems on surfaces, 1996

152 **R. S. Ismagilov,** Representations of infinite-dimensional groups, 1996

151 **S. Yu. Slavyanov,** Asymptotic solutions of the one-dimensional Schrödinger equation, 1996

150 **B. Ya. Levin,** Lectures on entire functions, 1996

149 **Takashi Sakai,** Riemannian geometry, 1996

148 **Vladimir I. Piterbarg,** Asymptotic methods in the theory of Gaussian processes and fields, 1996

147 **S. G. Gindikin and L. R. Volevich,** Mixed problem for partial differential equations with quasihomogeneous principal part, 1996

146 **L. Ya. Adrianova,** Introduction to linear systems of differential equations, 1995

145 **A. N. Andrianov and V. G. Zhuravlev,** Modular forms and Hecke operators, 1995

144 **O. V. Troshkin,** Nontraditional methods in mathematical hydrodynamics, 1995

143 **V. A. Malyshev and R. A. Minlos,** Linear infinite-particle operators, 1995

142 **N. V. Krylov,** Introduction to the theory of diffusion processes, 1995

141 **A. A. Davydov,** Qualitative theory of control systems, 1994

140 **Aizik I. Volpert, Vitaly A. Volpert, and Vladimir A. Volpert,** Traveling wave solutions of parabolic systems, 1994

139 **I. V. Skrypnik,** Methods for analysis of nonlinear elliptic boundary value problems, 1994

138 **Yu. P. Razmyslov,** Identities of algebras and their representations, 1994

137 **F. I. Karpelevich and A. Ya. Kreinin,** Heavy traffic limits for multiphase queues, 1994

136 **Masayoshi Miyanishi,** Algebraic geometry, 1994

135 **Masaru Takeuchi,** Modern spherical functions, 1994

134 **V. V. Prasolov,** Problems and theorems in linear algebra, 1994

133 **P. I. Naumkin and I. A. Shishmarev,** Nonlinear nonlocal equations in the theory of waves, 1994

132 **Hajime Urakawa,** Calculus of variations and harmonic maps, 1993

131 **V. V. Sharko,** Functions on manifolds: Algebraic and topological aspects, 1993

130 **V. V. Vershinin,** Cobordisms and spectral sequences, 1993

129 **Mitsuo Morimoto,** An introduction to Sato's hyperfunctions, 1993

(*Continued in the back of this publication*)