Moscow Lectures 9

Takashi Takebe

Elliptic Integrals and Elliptic Functions

NATIONAL RESEARCH UNIVERSITY

Skoltech
Skolkovo Institute of Science and Technology

Springer

Moscow Lectures

Volume 9

Moscow Lectures is a new book series with textbooks arising from all mathematics institutions in Moscow.

Takashi Takebe

Elliptic Integrals and Elliptic Functions

Springer

Takashi Takebe
Toshima-ku, Tokyo, Japan

ISSN 2522-0314 ISSN 2522-0322 (electronic)
Moscow Lectures
ISBN 978-3-031-30264-0 ISBN 978-3-031-30265-7 (eBook)
https://doi.org/10.1007/978-3-031-30265-7

Mathematics Subject Classification (2020): 33E05, 14H55, 14H52

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The purpose of this book is to present various aspects of the theory of elliptic integrals and elliptic functions. Today elliptic functions are often treated as nice examples in textbooks on complex analysis. However, such a rich set of objects is much more than just a collection of nice examples but a driving force of mathematics in the eighteenth and the nineteenth centuries. Many celebrated mathematicians, including notably Euler, Legendre, Abel, Jacobi, Gauss, Riemann, Liouville and Weierstrass, contributed to the development of the theory of elliptic integrals and elliptic functions.

Most mathematicians (and certainly this includes me) probably agree with Richard Bellmann, who wrote 'The theory of elliptic functions is the fairyland of mathematics. The mathematician who once gazes upon this enchanting and wondrous domain crowded with the most beautiful relations and concepts is forever captivated'[1].

Moreover, elliptic integrals and elliptic functions are also important in applications. They appear in various domains of physics and mathematics in unexpected ways. I first encountered elliptic functions in the field of physics and was captivated indeed.

In this monograph I have tried to give clear motivations ('Why do we consider such objects?') and to develop the theory following a natural way of thinking (e.g., 'If there are such and such things, then you would surely expect this one'). As a result, the story roughly traces the historical development, which means that those definitions and theorems which typically enter first on the stage in a usual 'theory of elliptic functions', will appear later on. Nevertheless, I am confident that most (though, of course, not all) important parts of the theory are covered. Those who are already familiar with elliptic functions may enjoy this book as a 'side reader'.

The contents of this book are mainly based on a series of lectures at the Faculty of Mathematics, National Research University Higher School of Economics (Moscow, Russia), which I gave in the spring semesters of 2014, 2016, 2018, 2020 and in the autumn semester of 2021. Following these lectures (in 2014 and 2016) and

[1] In the foreword of [B].

adding further material (e.g., conformal mappings between the upper half plane and a rectangle), I wrote a series of articles in Japanese for the 'Sugaku Seminar' (Math Seminar) magazine in Japan from April 2017 till September 2018 on the same subject intended mainly for undergraduate students and lovers of mathematics with a minimum knowledge of college calculus. When I published these articles in book form ([Tkb]), I added yet more materials (e.g., addition formulae for general elliptic functions, infinite product expansions for theta functions). In this English version I omit several reviews on elementary calculus and, besides minor changes, add, for example, the classification of real elliptic integrals, the Weierstrass zeta and sigma functions, the inversion problem from the modulus $k(\tau)$ to τ and the relation of the arithmetic-geometric mean and theta zero-values.

Prerequisites of this book are:

- elementary calculus, including general topology of $\mathbb{R}^n$. (Several facts about integrals are reviewed in Section A.1.)
- (after Chapter 6) complex analysis up to the argument principle. (Several facts are reviewed in Section A.2.)
- some knowledge of elementary homology theory in topology is welcome, but the very minimum which is necessary is explained in Section 6.2.2.

Since my cultural background is Japanese, some references are written in Japanese. I have tried to add counterparts to such references written in Western languages but could not do this in all cases. I apologise for this incompleteness.

Moscow,
October 2022 *Takashi Takebe*

Acknowledgements

The author cordially thanks those who gave me various comments and suggestions for the present manuscript or the Japanese version [Tkb], especially, Atsushi Nakayashiki, Hiroyuki Ochiai, Tadashi Tokieda and Hideshi Yamane among others. The author is grateful to the language editor Barnaby Sheppard, whose suggestions were quite helpful.

Special thanks are due to Ian Marshall, who read the manuscript and gave me not only mathematical but linguistic suggestions.

Contents

Part II Complex Part

Chapter 0
Introduction

In this chapter, apart from establishing rigorous definitions and logic, we will survey the various themes in the main part of the book to get an overview of the theory. We also pick up several topics which we shall not deal with later, in order to show the breadth and depth of the theory of elliptic functions.

0.1 What is an Elliptic Function in one Phrase?

Today's short answer to this question would be

'a meromorphic function on $\mathbb{C}$ with two periods independent over $\mathbb{R}$'.

"Is that all?" Yes, that's it!

Details of this definition will be examined later. Here we only explain what 'two periods independent over $\mathbb{R}$' means. For example, the trigonometric function $\sin x$ satisfies the equation $\sin(x+2\pi) = \sin x$, so we say that $\sin x$ has period 2π. Likewise, if a function $f(z)$ ($z \in \mathbb{C}$) satisfies

$$f(z+\omega_1) = f(z+\omega_2) = f(z)$$

for two non-zero complex numbers ω_1 and ω_2, we say '$f(z)$ has two periods ω_1 and ω_2'. However, if, say, $\omega_2 = 2\omega_1$, the equation $f(z+\omega_2) = f(z)$ follows trivially from $f(z+\omega_1) = f(z)$. In general, if $\omega_2 = r\omega_1$ ($r \in \mathbb{Q}$), two periods reduce to one. If $\omega_2 = \alpha\omega_1$ (where α is an irrational number) and f is an analytic function, it can be shown that f is a constant function. Otherwise, namely, if the complex numbers ω_1 and ω_2 are linearly independent over $\mathbb{R}$ as vectors in a plane $\mathbb{C} \cong \mathbb{R}^2$ and span a parallelogram (Fig. 0.1), we say that periods ω_1 and ω_2 are independent over $\mathbb{R}$.

Perhaps you might find the above definition of an elliptic function to be too simple. Indeed, elliptic functions are often treated just as 'nice examples' in textbooks of complex analysis.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_1

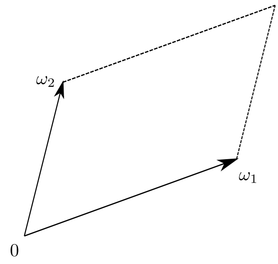

Fig. 0.1 Two independent periods.

In fact, this is because (at least some aspects of) complex analysis itself was developed so that it could explain properties of elliptic functions. In the nineteenth century the theory of elliptic functions was a cutting-edge research theme, which was the cradle and the prototype of various parts of modern mathematics such as algebraic geometry.

But the above definition does not include the word ‘ellipse’. To answer the natural question “Why does such an object have the name ‘*elliptic*’ function?”, we need to go back to the eighteenth century, when *elliptic integrals*, which is the first half of the title of this book, were studied. An elliptic integral is an integral of the form,

$$\int R(x, \sqrt{\varphi(x)})\, dx,$$

where $R(x,s)$ is a rational function (= polynomial/polynomial), and $\varphi(x)$ is a polynomial of degree three or four. It is called ‘an elliptic integral’ as the formula of arc length of an ellipse is expressed by an integral of this form, which we shall discuss in Chapter 1. Legendre (Legendre, Adrien-Marie, 1752–1833) studied integrals of this kind in detail and published a voluminous book on the subject.

Let us take the example of the fundamental rational function $\dfrac{1}{s}$ for $R(x,s)$, and regard the above integral as a function of the upper limit u, after fixing the lower limit u_0:

$$z(u) = \int_{u_0}^{u} \frac{dx}{\sqrt{\varphi(x)}}.$$

Abel (Abel, Niels Henrik, 1802–1829), Jacobi (Jacobi, Carl Gustav Jacob, 1804–1851) and Gauss (Gauß, Johann Carl Friedrich, 1777–1855) found that the inverse function $u(z)$ to $z(u)$ is a nice function — that is, a well-defined function which has good properties like trigonometric functions — and has two periods as a complex function on $\mathbb{C}$. This nice function is the elliptic function mentioned before.[1]

[1] The name ‘elliptic function’ is given by Jacobi. The word “elliptic function” was used by Legendre earlier, but Legendre’s “elliptic function” means an elliptic integral in today’s usage.

Just now we viewed $z(u)$ as being a function on the complex plane $\mathbb{C}$. But upon careful consideration it becomes apparent that it is not so simple. We encounter a problem when we treat the integral as a complex (contour) integral: which sign of $\pm\sqrt{\varphi(x)}$ should we take? Indeed, the complex root is a multi-valued function! We have to fix one of two signs or 'uniformise' the multi-valued function somehow. We shall introduce the notion of Riemann surface later, but suffice it to say for now that we split the domain of definition into two, as there are two values of the function. We shall construct a surface by cutting and pasting domains of definition and then redefine the function $\sqrt{\varphi(x)}$ on it.

Let us return to elliptic integrals. We can add points at infinity to the Riemann surface on which the integral $z(u)$ is defined. The idea is similar to the construction of the Riemann sphere from the plane $\mathbb{C}$ in complex analysis. The resulting surface has the shape of a torus (Fig. 0.2) and is called an 'elliptic curve'.

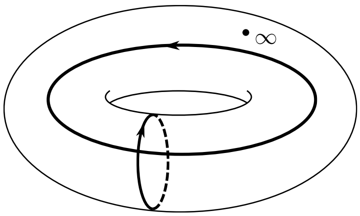

Fig. 0.2 A torus and closed curves on it.

"Hang on! Isn't it a surface? Are you mistaken?" Sorry, it *is* a 'curve', because it is *one-dimensional* 'over the complex field $\mathbb{C}$'. And, please don't confuse it with an 'ellipse'![2]

Cutting along two curves on the torus in Fig. 0.2, we obtain a quadrilateral, which corresponds to the parallelogram in Fig. 0.1. There is a correspondence between closed curves on the elliptic curve and periods of the elliptic function:

$$\text{a period of the elliptic function } u(z) \quad = \quad \int_{\text{a closed curve on the elliptic curve}} \frac{dz}{\sqrt{\varphi(z)}}.$$

Depending on which curve we integrate along, we obtain ω_1, ω_2 or other periods (= a linear combination of ω_1 and ω_2 with integral coefficients). Through this identification of the parallelogram and the cut-open elliptic curve we can define an elliptic function as 'a meromorphic function on the elliptic curve'.

[2] What makes the situation complicated is that there is yet another object called an 'elliptic *surface*', which is two-dimensional over $\mathbb{C}$, namely, four-dimensional over $\mathbb{R}$.

0.2 What Properties Do Elliptic Functions Have?

We have defined the notion of an elliptic function. But that does not mean we know 'what it is', unless we know its properties. A standard approach to determining the properties of something is to compare it with other well-known things. From our high school days we know functions like polynomials, rational functions and trigonometric functions very well. Let us compare elliptic functions with them.

Both polynomials and rational functions are considered on $\mathbb{C}$ to begin with. But in complex analysis we often add infinity to $\mathbb{C}$ to form the Riemann sphere and then regard polynomials and rational functions as functions on it. According to Liouville's theorem in complex analysis, they can be defined as 'meromorphic functions on the Riemann sphere'. Replacing 'the Riemann sphere' in this definition by 'the elliptic curve', we obtain the definition of elliptic functions at the end of the previous section.

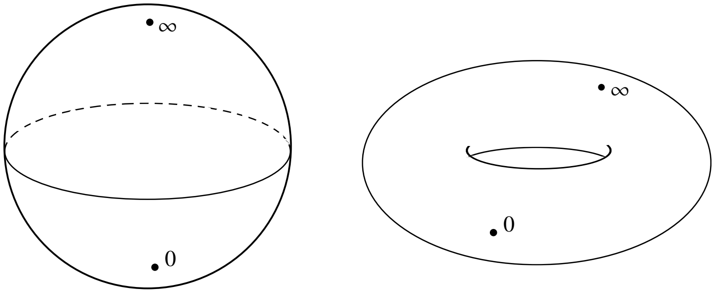

Fig. 0.3 The Riemann sphere and an elliptic curve.

In this sense we can say that rational functions (including polynomials) and elliptic functions are of the same species, but just living in different places.

A rational function $f(z)$ has the form

$$f(z) = C\frac{(z-\alpha_1)\cdots(z-\alpha_N)}{(z-\beta_1)\cdots(z-\beta_M)},$$

while an elliptic function is represented by means of so-called theta functions as, for example,

$$f(z) = C\frac{\theta_{11}(z-\alpha_1)\cdots\theta_{11}(z-\alpha_N)}{\theta_{11}(z-\beta_1)\cdots\theta_{11}(z-\beta_N)}.$$

Now let us turn to the analogy with trigonometric functions. An important property of a trigonometric function is its periodicity. For instance, $\sin z$ and $\cos z$ have period 2π: $\sin(z+2\pi) = \sin z$, $\cos(z+2\pi) = \cos z$. Namely, when we regard them as functions on $\mathbb{C}$, they return to their former values when they are shifted by 2π.

Therefore we may consider these standard trigonometric functions as living on an infinitely long cylinder of circumference 2π (Fig. 0.4).

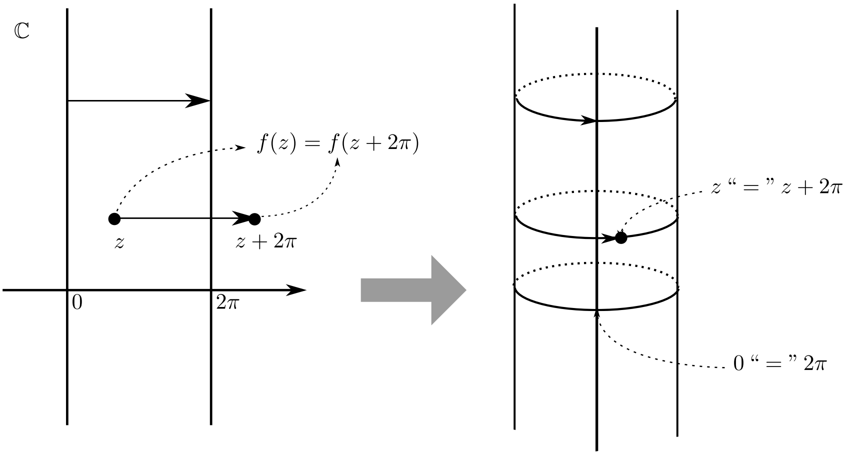

Fig. 0.4 Plane $\mod 2\pi \cong$ cylinder.

Imagine that the elliptic curve, on which an elliptic function lives, is made of rubber. Stretch it, it might snap, but don't stop, just stretch. If you have stretched it infinitely far, you will get a cylinder. Accordingly elliptic functions, inhabitants of the elliptic curve, become trigonometric functions. Speaking more precisely, elliptic functions become trigonometric functions in the limit when the ratio of the two periods ω_1 and ω_2 tends to $\sqrt{-1} \times \infty$. Therefore trigonometric functions and elliptic functions are close relatives and have much in common.

As an example let us look at addition formulae. The sine function has the addition formula,

$$\sin(x+y) = \sin x \cos y + \sin y \cos x,$$

which expresses the value of the function at $x+y$ as a polynomial of values of trigonometric functions at x and y. (For general trigonometric functions we need rational functions instead of polynomials.) Elliptic functions also have addition formulae. For example, Jacobi's sn function, a cousin of the sine function, satisfies the following formula:

$$\operatorname{sn}(x+y) = \frac{\operatorname{sn} x \operatorname{cn} y \operatorname{dn} y + \operatorname{sn} y \operatorname{cn} x \operatorname{dn} x}{1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y}.$$

Here, cn and dn are also elliptic functions (cn is a cousin of the cosine function) and k is a parameter related to the periods.

Another point of contact between elliptic functions and trigonometric functions is differential equations. As we know, the derivative of $\sin x$ is $\cos x$ and $\cos^2 x =$

$1-\sin^2 x$, from which it follows that the function $\sin x$ satisfies a differential equation

$$((\sin x)')^2 = 1 - \sin^2 x.$$

Similarly, $\operatorname{sn} x$ satisfies

$$((\operatorname{sn} x)')^2 = (1 - \operatorname{sn}^2 x)(1 - k^2 \operatorname{sn}^2 x).$$

The right-hand side is a polynomial of degree 4 in sn, so the differential equation for $\operatorname{sn} x$ is more complicated than that for $\sin x$. Indeed the factor $1 - \operatorname{sn}^2 x$ corresponds to the right-hand side of the differential equation for $\sin x$. Another elliptic function $\wp(x)$, called the Weierstrass $\wp$-function,[3] satisfies the differential equation

$$((\wp(z))')^2 = 4(\wp(z))^3 - g_2\wp(z) - g_3,$$

whose right-hand side is a polynomial of degree 3. Here g_2 and g_3 are parameters related to the periods.

0.3 What Use Are Elliptic Functions?

We have seen that elliptic functions have various good properties. Because of these properties they appear in various places in physics as well as in mathematics.

First let us have look at applications in physics.

Pendulum

Perhaps the most well-known application of elliptic functions is in the description of the motion of a pendulum.

As we learned in school, a pendulum has an important property called *isochronism*: the period of swing of a pendulum depends neither on the mass of the bob nor on the amplitude of the swing. This makes it possible to use a pendulum as a clock.

Isochronism can be shown by the differential equation, *the equation of motion*:

$$ml\,\frac{d^2}{dt^2}\,\varphi(t) = -mg\,\varphi(t).$$

Here m is the mass of the bob, l is the length of the pendulum, $\varphi(t)$ is angle at the time t (the line $\varphi = 0$ is vertical) and g is the acceleration of gravity (cf. Fig. 0.5). A solution of this equation has the form $\varphi(t) = A\sin(\omega t + \alpha)$ $\left(\omega = \sqrt{\dfrac{g}{l}}\right)$. The period

[3] $\wp$ is the lower case script letter ‘p’, which is probably used only for this function.

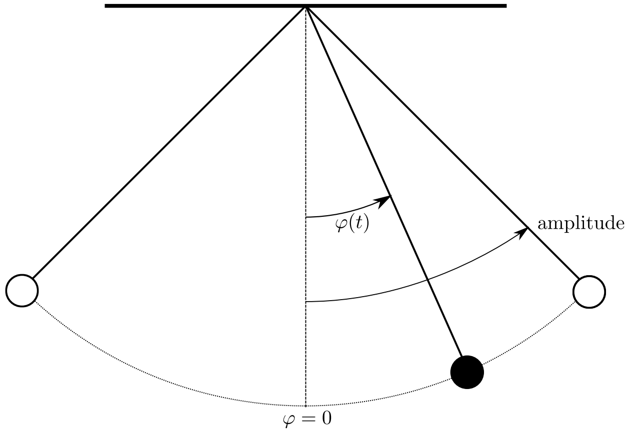

Fig. 0.5 A pendulum.

of the pendulum is $T = \dfrac{2\pi}{\omega} = 2\pi\sqrt{\dfrac{l}{g}}$, which does not involve m or the amplitude A of the swing. This is the reason for the isochronism of a pendulum. (The constant α in the solution is related to the initial position of the bob.)

However, if we are being completely rigorous, then the truth is that isochronism *does not hold*! True 'isochronism' can be stated as follows: the period of swing of a pendulum does not depend on the mass of the bob, and, *if the amplitude is small*, it *practically* does not depend on the amplitude, either. We shall explain how to derive the above equation in the main text, but, roughly speaking, we approximate $\sin\varphi(t)$ by $\varphi(t)$ when the amplitude is small. Thanks to this approximation, the differential equation becomes simpler (linear) and can be solved by a trigonometric function. When the amplitude is large, we cannot use approximation and have to solve a *non-linear* differential equation

$$m\frac{d^2\varphi}{dt^2} = -m\omega^2 \sin\varphi,$$

which cannot be solved simply by trigonometric functions. Instead, elliptic functions appear in the solution. The amplitude determines the periods (ω_1, ω_2) of the elliptic functions, which, in turn, determine the period of the pendulum. Therefore the period of swing *does depend on the amplitude*.

Skipping rope

Another interesting application of elliptic functions in physics is to the description of a skipping rope. The shape taken by a skipping rope in motion is the graph of an elliptic function (Jacobi's sn function). We shall show this in Section 5.2.

Soliton equations

The field of research called 'integrable systems' in mathematical physics deals with exactly solvable systems, whose solutions can be written down without approximation. In this area elliptic functions play an important role. So-called 'soliton equations', among which the KdV equation and the Toda lattice equation are well known, have exact solutions expressed by elliptic functions.

Indeed, the Toda lattice equation, which occupies an important position among various soliton equations, was discovered by Toda[4] (Toda, Morikazu, 1917–2010), who looked for 'a non-linear lattice with solutions described by elliptic functions'.

Solvable lattice models

Another different-looking application in physics can be found in statistical mechanics. Solvable lattice models (which provide a kind of mathematical model of crystals, integrable in the above sense) are sometimes defined by matrices known as R-matrices. There are models whose R-matrices are expressed by elliptic functions. The trigonometric limit, which we discussed in Section 0.2, of those elliptic R-matrices are also R-matrices for other solvable lattice models. Such trigonometric R-matrices are at the origin of the algebraic structures called 'quantum groups' discovered in 1985. The elliptic R-matrices also define algebraic structures. In fact one of the elliptic algebras, the Sklyanin algebra, was discovered even earlier than quantum groups by Sklyanin in 1982 ([Sk1]) and its representation theory was studied in 1983 ([Sk2]). Such elliptic algebras are much more complicated than quantum groups and are still under active research. (In fact, I first got acquainted with elliptic functions when I was studying elliptic solvable lattice models.)

Of course, not only in physics but also in mathematics, elliptic integrals and elliptic functions have a number of applications and related topics. (We shall mention topics related to elliptic curves in the next section.)

[4] He describes how he discovered the Toda lattice in [To1].

Arithmetic-geometric mean

One of the origins of the study of elliptic integrals is the arithmetic-geometric mean. There are several 'means' of two real numbers a and b (we assume $a \geq b \geq 0$). The two most well-known means are the arithmetic mean $\dfrac{a+b}{2}$ and the geometric mean $\sqrt{ab}$. If we apply these two operations iteratively and take the limit, we obtain the *arithmetic-geometric mean*. More precisely, we define two sequences $\{a_n\}_{n=0,1,2,\dots}$, $\{b_n\}_{n=0,1,2,\dots}$ by the initial values $a_0 = a$, $b_0 = b$ and the recurrence relations,

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}.$$

We can show that both sequences have the same limit: $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$. This is the definition of the arithmetic-geometric mean. We shall explain in detail how the arithmetic-geometric mean is expressed by an elliptic integral. Gauss conjectured this relation after intense numerical calculation and then proved it. He also correctly predicted that it would 'open a completely new area of analysis'.

Formula for solving quintic equations

In the history of mathematics there is the celebrated episode concerning Abel and Galois (Galois, Évariste, 1811–1832), who proved independently that polynomial equations of degree five or higher with general coefficients are "not solvable". More precisely, any polynomial equation in one variable possesses a solution in $\mathbb{C}$, but a general solution formula in radicals exists only for equations of degree four or less. (There are special equations of higher degree whose solutions can be expressed by means of arithmetic operations and radicals.)

This Abel–Ruffini theorem is well known partly because of the tragic lives of Abel and Galois and partly because it features the paradoxical appearance of a "proof of impossibility" (although it is by no means a paradox). In fact Abel's work on quintic equations is closely related to his study of elliptic integrals.[5] Moreover, if elliptic integrals and theta functions are permitted besides arithmetic operations and radicals, we *can* write down a solution formula for a quintic equation.[6]

0.4 A Small Digression on Elliptic Curves

As we do not touch upon the algebraic theory of elliptic curves in the main text, let us mention some related topics here.

[5] For example, [Tks].

[6] For example, [U] Ch. 6, [PS] Ch. 7.

Using the differential equations satisfied by elliptic functions (cf. Section 0.2), we can describe elliptic curves by algebraic equations. For example, the differential equation for the $\wp$-function becomes an algebraic equation

$$y^2 = 4x^3 - g_2 x - g_3$$

through the substitution $x = \wp(z)$, $y = \wp'(z)$. Adding one infinity ∞ ($x = \infty$, $y = \infty$) to the set of pairs (x, y) of complex numbers satisfying the above equation, we obtain 'the same thing' as the elliptic curve.

In algebraic geometry, this definition is more usual. What is good about this algebraic definition is that we do not have to consider the coefficients and coordinates x and y as complex numbers. We can take any field (e.g., a finite field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$) or even a ring R as the set of values of coordinates. Such considerations are fruitful in number theory, too. Elliptic curves used for 'elliptic curve cryptography' in cryptology and elliptic curves used in the proof of Fermat's last theorem are of this kind.

Let us return to the definition of a elliptic curve as the torus obtained by gluing parallel edges of a parallelogram spanned by $(\omega_1, \omega_2) \in \mathbb{C}^2$. It is easy to see that another pair (ω_1', ω_2') may define the same elliptic curve as (ω_1, ω_2). The condition that two such pairs define one and the same elliptic curve is as follows: there exist integers a, b, c, d such that $ad - bc = \pm 1$ and

$$\frac{\omega_2'}{\omega_1'} = \frac{a\omega_2 + b\omega_1}{c\omega_2 + d\omega_1}.$$

When we define an elliptic curve by a cubic equation $y^2 = 4x^3 - g_2 x - g_3$, there is also the possibility that two pairs of coefficients (g_2, g_3) and (g_2', g_3') lead to the same elliptic curve. A necessary and sufficient condition for the coincidence of the two elliptic curves is that the function

$$j(g_2, g_3) = \frac{(12g_2)^3}{g_2^3 - 27g_3^2}$$

takes the same value for both curves: $j(g_2, g_3) = j(g_2', g_3')$.

If we further pursued this direction, we would develop the theory of moduli spaces of elliptic curves or the theory of modular functions and modular forms in number theory.

In fact, Gauss studied such theories, but his results remained unpublished during his lifetime. After his death the results in his notes were published, although nobody understood them at the time of publication. It took mankind almost seventy years to understand what Gauss meant.[7] We shall touch on this topic at the end of this book. (See Remark 17.12.)

---

[7] For example, [Kl] Kap. I, or [INU] part II, Ch. 3, §2.

0.5 Structure of this book

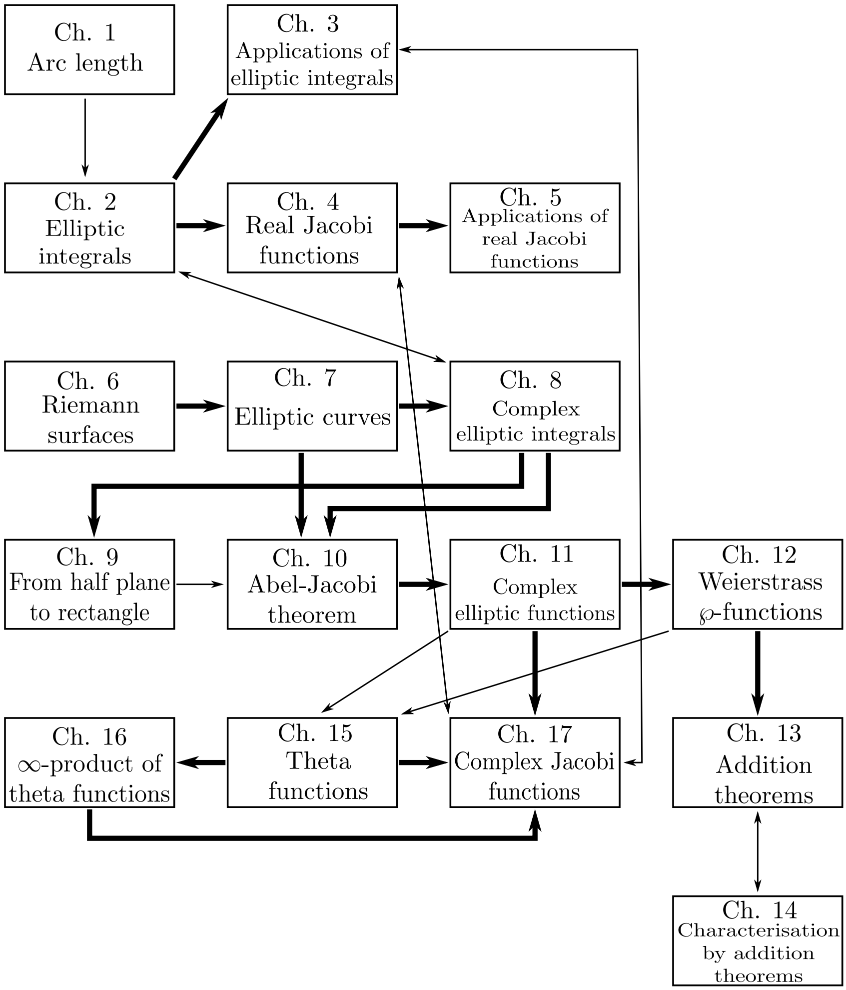

Fig. 0.6 Relation among chapters.

In principle it is assumed that this book will be read in order of the chapters, but there are chapters which can be skipped. In the above figure, thicker arrows show strong logical connections of the chapters and thinner arrows mean that there are connections or correspondences of materials in those chapters.

Part I
Real Part

In the first half of this book we study elliptic integrals and elliptic functions on $\mathbb{R}$. The exception is the classification theorem of elliptic integrals, which is simpler for complex-valued elliptic integrals than for real-valued ones.

Chapter 1
The Arc Length of Curves

1.1 The Arc Length of the Ellipse

To begin with, let us consider the following assertion, familiar to everyone:

The circumference of a circle of radius a is $2\pi a$.

As a matter of fact, this is nothing more than a paraphrase of the definition of π: 'The number π is the ratio of a circle's circumference to its diameter'. However, if you pursue logical rigour, there are many gaps to be filled. "Why is the ratio of circumference to diameter the same for any circle?" (This is *not* true for creatures living on a sphere. In fact, according to the general theory of relativity this is not true in our universe either because of gravity.) "What is the *length* of a curve, after all?" and so on. To those who want answers to these questions: I know what you are worrying about, but filling all those gaps would take us far from the main theme. So let us compromise and be content with 'confirming' that 'the circumference of a circle $= 2\pi a$' by using the formula for the length of a curve usually given in a calculus course.

Recall that a smooth curve on the plane $\mathbb{R}^2$ has a parametrisation,

$$\gamma : [a, b] \ni t \mapsto \gamma(t) = (x(t), y(t)) \in \mathbb{R}^2,$$

where $x(t)$ and $y(t)$ are smooth functions, which means that $x(t)$ and $y(t)$ are differentiable and their derivatives $x'(t)$ and $y'(t)$ are continuous[1] (Fig. 1.1).

The arc length of γ (the length from $\gamma(a)$ to $\gamma(b)$ along γ) is equal to the integral

$$\int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt. \tag{1.1}$$

[1] Exactly speaking, smoothness of a curve requires non-vanishing of the vector $(x'(t), y'(t))$, which is not important here.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_2

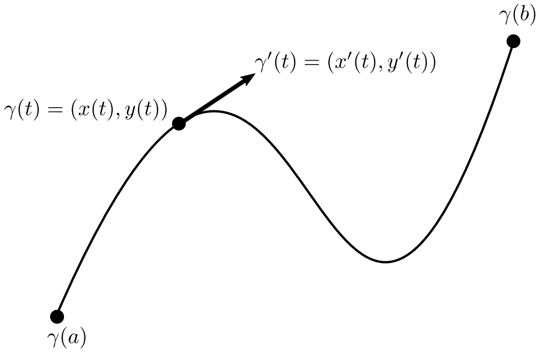

Fig. 1.1 A smooth curve.

We have to go back to the definition of 'the length of a curve' to prove this formula, for which we refer to any textbook of calculus. Here we only explain the meaning of the formula. If we regard the variable t as 'time', $\gamma(t) = (x(t), y(t))$ are the coordinates of a moving point. The derivative $\gamma'(t) = (x'(t), y'(t))$ is its velocity vector and the integrand in (1.1) is its length '=' the speed of the point. Namely, the above formula is a generalisation, from the case of a point moving with constant velocity to the case of a point moving with variable velocity, of the trivial formula 'the distance travelled is equal to the speed multiplied by the time'.

Now, we apply the formula (1.1) to a circle. We parametrise the arc in Fig. 1.2 (the bold curve) as

$$(x(\varphi), y(\varphi)) = (a\cos\varphi, a\sin\varphi), \qquad \varphi \in [0, \theta]. \tag{1.2}$$

The length of this arc is

$$\begin{aligned} &\int_0^\theta \sqrt{\left(\frac{d}{d\varphi} a\sin\varphi\right)^2 + \left(\frac{d}{d\varphi} a\cos\varphi\right)^2}\, d\varphi \\ &= \int_0^\theta \sqrt{a^2\cos^2\varphi + a^2\sin^2\varphi}\, d\varphi = \int_0^\theta a\, d\varphi = a\theta, \end{aligned} \tag{1.3}$$

as is well known. If the arc is the whole circle, θ is equal to 2π, and thus we confirmed that the arc length of the circle is $2\pi a$. Fine, everything went well.

Remark 1.1 "Wait, we defined π by the length of a circle and then we used it in the computation of the length of a circle. Isn't that circular reasoning?" You are very careful. To make the argument rigorous, for example, we define π using a period of trigonometric functions and then define a circle by (1.2). Trigonometric functions should be defined, for example, by power series, not using figures, because the notion

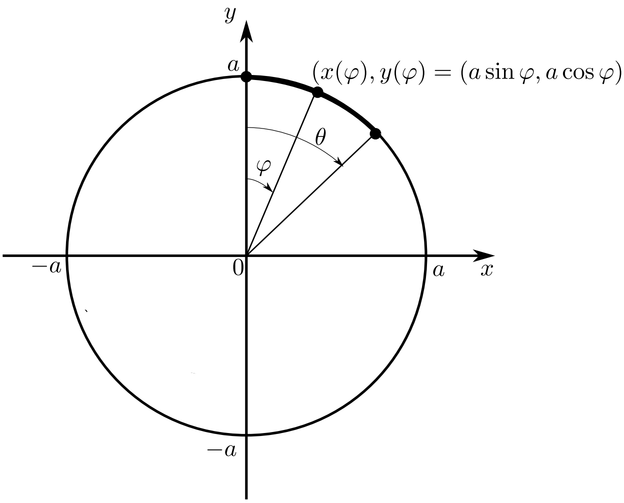

Fig. 1.2 An arc of a circle (the bold part corresponds to $\varphi \in [0, \theta]$).

of 'angle' is defined by the arc length, using the formula proved above in the opposite direction.[2] We do not go into details here.

If we restrict θ to $\left[0, \dfrac{\pi}{2}\right]$ — namely, if we consider an arc in the first quadrant — we can parametrise the curve by means of the x-coordinate:

$$(x, y(x)) = (x, \sqrt{a^2 - x^2}), \qquad x \in [0, x_0], \tag{1.4}$$

where $x_0 = a\sin\theta$. Using this parametrisation, we may compute the arc length as

$$\begin{aligned}
&\int_0^{x_0} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_0^{x_0} \sqrt{1 + \frac{x^2}{a^2 - x^2}}\, dx \\
&= a\int_0^{x_0} \frac{dx}{\sqrt{a^2 - x^2}} = a\int_0^{x_0/a} \frac{dz}{\sqrt{1 - z^2}} \qquad (z = x/a) \\
&= a \arcsin z \Big|_{z=0}^{z=x_0/a} = a \arcsin\left(\frac{x_0}{a}\right),
\end{aligned} \tag{1.5}$$

which agrees with (1.3), because $x_0 = a\sin\theta$.

Now, let us consider a cousin of the circle, the ellipse. We learned the formula of the arc length of a circle in elementary school, but no such formula for an ellipse was taught in school. We must discover for ourselves what we are not taught. First

[2] This is a toy analogue of the relation of an elliptic integral and an elliptic function, which we shall discuss in Chapter 4.

we parametrise an arc of an ellipse $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$ as

$$(x(\varphi), y(\varphi)) = (a\sin\varphi, b\cos\varphi). \tag{1.6}$$

Here we assume $a > b > 0$. As in the case of a circle, φ varies in the interval $[0, \theta]$ (Fig. 1.3). Note that φ is *not* the angle between the coordinate axis and the line connecting the origin with a point on the ellipse. It is the angle between the y-axis and the line connecting the origin with a point on the auxiliary circle of radius a.

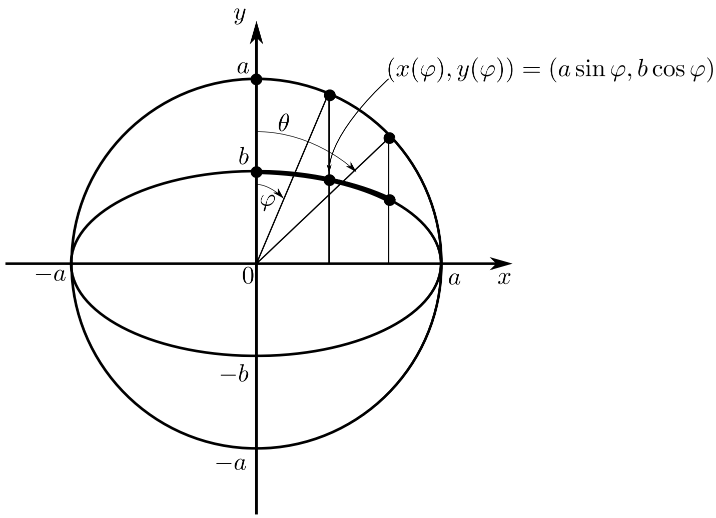

Fig. 1.3 An arc of an ellipse (the bold part corresponds to $\varphi \in [0, \theta]$).

The arc length should be computed by the formula automatically. Let us try.

$$\begin{aligned}
&\text{The arc length of the bold curve in Fig. 1.3}\\
&= \int_0^\theta \sqrt{\left(\frac{d}{d\varphi}a\sin\varphi\right)^2 + \left(\frac{d}{d\varphi}b\cos\varphi\right)^2}\,d\varphi\\
&= \int_0^\theta \sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}\,d\varphi\\
&= a\int_0^\theta \sqrt{1-\frac{a^2-b^2}{a^2}\sin^2\varphi}\,d\varphi = a\int_0^\theta \sqrt{1-k^2\sin^2\varphi}\,d\varphi.
\end{aligned} \tag{1.7}$$

Here we put $k := \sqrt{\dfrac{a^2-b^2}{a^2}}$, which is called the *eccentricity* of the ellipse.

The computation (1.3) of the arc length of a circle corresponds to the case $k = 0$. When $k \neq 0$, namely, if the ellipse is not a circle, it is known that the integral (1.7) which expresses the arc length of an ellipse is not an elementary function.[3] (So it is not taught in school.) Therefore we give it a special name. The integral

$$E(k,\theta) := \int_0^{\theta} \sqrt{1-k^2\sin^2\varphi}\, d\varphi \tag{1.8}$$

depending on k ($0 \le k < 1$) and θ is called the *incomplete elliptic integral of the second kind*. When $\theta = \dfrac{\pi}{2}$,

$$E(k) := E\left(k, \frac{\pi}{2}\right) = \int_0^{\frac{\pi}{2}} \sqrt{1-k^2\sin^2\varphi}\, d\varphi \tag{1.9}$$

is called the *complete elliptic integral of the second kind*. When it is clear which one (incomplete or complete) is meant, we just say 'elliptic integral of the second kind'. The parameter k is called the *modulus* in this context.

Using these notations, we have

$$\begin{aligned} \text{The arc length of the arc } 0 \le \varphi \le \theta &= a\,E\left(\sqrt{\frac{a^2-b^2}{a^2}}, \theta\right), \\ \text{The circumference of an ellipse } &= 4a\,E\left(\sqrt{\frac{a^2-b^2}{a^2}}\right). \end{aligned} \tag{1.10}$$

As is the case of a circle (1.4), in the first quadrant ($0 \le \theta \le \dfrac{\pi}{2}$) we can use a Cartesian parametrisation of the ellipse,

$$(x, y(x)) = \left(x, b\sqrt{1-\frac{x^2}{a^2}}\right), \qquad x \in [0, x_0], \tag{1.11}$$

where $x_0 = a\sin\theta$. Computation of the arc length by this parametrisation may be carried out as follows:

$$\begin{aligned} \text{Arc length} &= a\,E(k,\theta) \\ &= \int_0^{x_0} \sqrt{1+\left(\frac{dy}{dx}\right)^2}\, dx \\ &= \int_0^{x_0} \sqrt{1+\frac{b^2}{a^2}\frac{(x/a)^2}{1-(x/a)^2}}\, dx \\ &= a\int_0^{\sin\theta} \sqrt{\frac{1-k^2z^2}{1-z^2}}\, dz, \qquad (z = x/a), \end{aligned} \tag{1.12}$$

[3] An elementary function is a function obtained by composing finitely many arithmetic operations, radicals, exponential functions, trigonometric functions and their inverse functions. We shall discuss this class of functions a little in Section 2.1.

which is an algebraic expression in the sense that it does not involve trigonometric functions in the integrand. Similarly elliptic integrals of the second kind can be expressed as

$$
\begin{aligned}
E(k,\theta) &= \int_0^{\sin\theta} \sqrt{\frac{1-k^2z^2}{1-z^2}}\,dz,\\
E(k) &= \int_0^1 \sqrt{\frac{1-k^2z^2}{1-z^2}}\,dz.
\end{aligned}
\tag{1.13}
$$

(They can be obtained just by making the change of variables $z = \sin\varphi$ in (1.8) and (1.9).)

There are other curves, the arc lengths of which are expressed by the elliptic integral of the second kind. We leave the details to the reader as exercises.

Exercise 1.2 Find an expression for the arc length of the graph of $y = b\sin\frac{x}{a}$ $(a,b > 0)$ between the origin $(0,0)$ and the point $\left(x_0, b\sin\frac{x_0}{a}\right)$ $(x_0 > 0)$. To which arc corresponds the complete elliptic integral of the second kind?

Exercise 1.3 How about conics (quadratic curves) other than ellipses? The answer is that the arc length of a hyperbola is expressed by the elliptic integral of the second kind and the arc length of a parabola becomes an elementary function. Actually, the expression for the hyperbola is rather messy, so here we show a 'complex' expression, which is easily derived.

(i) A hyperbola $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$: Let us parametrise it by hyperbolic functions as $(x,y) = (a\cosh t, b\sinh t)$. Show that the arc length between $t = 0$ and $t = t_0 > 0$ is (formally) expressed as $-ib\,E\left(\frac{\sqrt{a^2+b^2}}{b}, it_0\right)$ $(i = \sqrt{-1})$.

(Hint: Hyperbolic functions are defined by $\sinh t = \frac{e^t - e^{-t}}{2}$, $\cosh t = \frac{e^t + e^{-t}}{2}$, from which follow fundamental identities $\cosh^2 t - \sinh^2 t = 1$, $\frac{d}{dt}\sinh t = \cosh t$, $\frac{d}{dt}\cosh t = \sinh t$. They are related to trigonometric functions by Euler's formula, $e^{it} = \cos t + i\sin t$: $\cosh t = \cos it$, $i\sinh t = \sin it$.)

(ii) A parabola $y = ax^2$ $(a > 0)$: Find its arc length between $(0,0)$ and (x_0, ax_0^2) $(x_0 > 0)$.

(Hint: Messy calculation of an indefinite integral. The answer is an elementary function, but the expression is somewhat simplified by the use of a hyperbolic function $\sinh t$ and its inverse function $\operatorname{arsinh} u$.)

Your natural question would be "You talked about elliptic integrals of the *second* kind. Are there elliptic integrals of the *first* kind?" This question will be answered in the next section.

1.2 The Lemniscate and its Arc Length

We have shown that arc lengths of quadratic curves are expressed in terms of elliptic integrals or elementary functions. (For an ellipse, (1.10), for a hyperbola and a parabola, Exercise 1.3.) Quadratic curves are called 'quadratic' because they are defined by quadratic equations, but these curves also have a geometric characterisation.

For example, an ellipse is defined as the orbit of a point P such that the sum of distances $\overline{PA}$ and $\overline{PB}$ from two fixed points A and B is constant: $\overline{PA}+\overline{PB}=l$. Similarly, a hyperbola is defined as the orbit of a point P such that the difference of distances $\overline{PA}$ and $\overline{PB}$ from two fixed points A and B is constant: $|\overline{PA}-\overline{PB}|=l$. See Fig. 1.4.

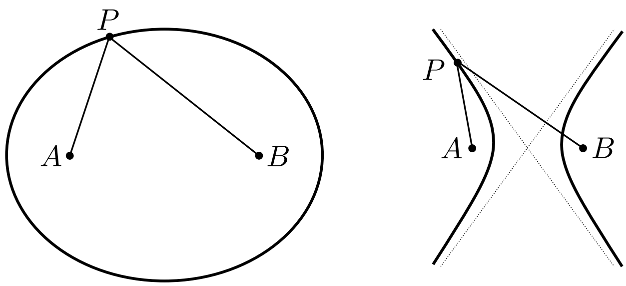

Fig. 1.4 An ellipse ($\overline{PA}+\overline{PB}=l$) and a hyperbola ($|\overline{PA}-\overline{PB}|=l$).

After hearing such definitions of an ellipse and a hyperbola, haven't you ever had a question like this?: "Why should we restrict ourselves to the sum and difference? Can we not take the *product* instead?" Let us investigate this question.[4]

Fix two points $A=(-a,0)$ and $B=(a,0)$ on the x-axis ($a>0$). The distances from a point $P=(x,y)$ to A and B are

$$\overline{PA}=\sqrt{(x+a)^2+y^2}, \qquad \overline{PB}=\sqrt{(x-a)^2+y^2}.$$

Their product is

[4] The answer to the question "How about the *ratio*?" is simpler. It leads to the circles of Apollonius.

$$\begin{aligned}\overline{PA}\cdot\overline{PB} &= \sqrt{(x+a)^2+y^2}\sqrt{(x-a)^2+y^2}\\ &= \sqrt{(x^2-a^2)^2+2y^2(x^2+a^2)+y^4}\\ &= \sqrt{(x^2+y^2)^2-2a^2(x^2-y^2)+a^4}.\end{aligned}$$

Hence, the curve, on which the product of distances from two fixed points, A and B, is constant l^2, is defined by the equation

$$(x^2+y^2)^2-2a^2(x^2-y^2)+a^4=l^4. \tag{1.14}$$

Such curves are called *Cassini ovals*, and, in particular, when $l=a$, the *lemniscate*[5], which has the equation

$$(x^2+y^2)^2=2a^2(x^2-y^2). \tag{1.15}$$

The Cassini oval and the lemniscate are defined by quartic polynomials in the coordinates x and y. So they are more complicated than the quadratic curves.

For later use we rewrite the equations (1.14) and (1.15) in terms of polar coordinates (r,φ), $x=r\cos\varphi$, $y=r\sin\varphi$. As $x^2+y^2=r^2$ and $x^2-y^2=2r^2\cos 2\varphi$, the Cassini oval (1.14) is defined by

$$r^4-4a^2r^2\cos 2\varphi+a^4=l^4, \tag{1.16}$$

and the lemniscate (1.15) is defined by

$$r^4=2a^2r^2\cos 2\varphi,\ \text{namely } r^2=2a^2\cos 2\varphi. \tag{1.17}$$

The Cassini oval is a convex closed smooth simple curve when $l\geq\sqrt{2}a$, a cocoon-shaped closed smooth simple curve when $a<l<\sqrt{2}a$ and two convex closed smooth simple curves when $0<l<a$, but the lemniscate ($l=a$) is a curve with a crossing at the origin. See Fig. 1.5. We do not prove these facts, as our only object of interest is the lemniscate.

In Fig. 1.6 the lemniscate in Fig. 1.5 is picked up and some data are added. Since the left hand sides in (1.17) are non-negative, $\cos 2\varphi$ cannot be negative. For this reason there is no part of the curve in the regions $\frac{\pi}{4}<\varphi<\frac{3\pi}{4}$ and $\frac{5\pi}{4}<\varphi<\frac{7\pi}{4}$.

The expression (1.17) is a parametrisation of the lemniscate by the angle φ. To compute the arc length we introduce a more convenient parameter. If we look only at the first quadrant, the argument φ moves only in the interval $\left[0,\frac{\pi}{4}\right]$. The function $\cos 2\varphi$ monotonically decreases from 1 to 0 on this interval. Therefore there is a unique ψ in $\left[0,\frac{\pi}{2}\right]$ satisfying

$$\cos 2\varphi=(\cos\psi)^2. \tag{1.18}$$

[5] The lemniscate was introduced and studied by Jacob Bernoulli (1654–1705) and Johann Bernoulli (1667–1748) in connection with a problem in physics (the problem of isochrones).

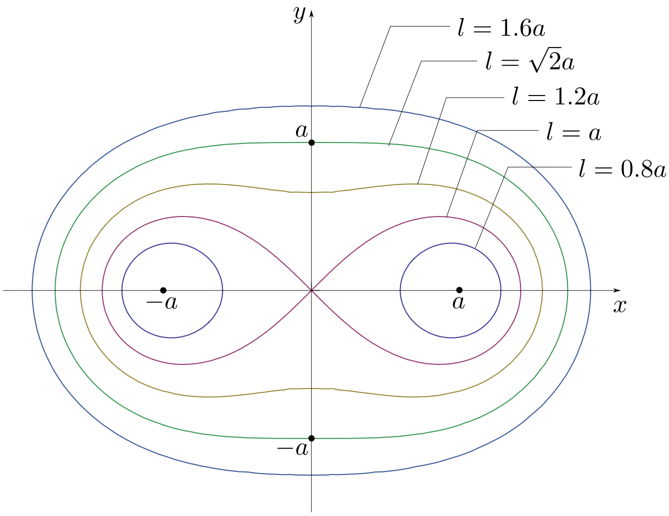

Fig. 1.5 Cassini ovals ($l = 1.6a$, $\sqrt{2}a$, $1.2a$, $0.8a$) and a lemniscate ($l = a$).

In terms of this variable, the expression (1.17) is rewritten as

$$r = \sqrt{2}a \cos\psi. \tag{1.19}$$

The x- and y-coordinates of a point on the lemniscate are expressed by ψ as follows: Squaring the equation (1.19), we have

$$x^2 + y^2 = 2a^2 \cos^2\psi.$$

Squaring this once more, we obtain $(x^2 + y^2)^2 = r^4 = 4a^4 \cos^4\psi$. Thus (1.15) leads to

$$x^2 - y^2 = 2a^2 \cos^4\psi.$$

From these expressions of the sum and the difference of x^2 and y^2 follow

$$\begin{aligned} x^2 &= a^2 \cos^2\psi\,(1 + \cos^2\psi), \\ y^2 &= a^2 \cos^2\psi\,(1 - \cos^2\psi). \end{aligned}$$

Namely, for $\psi \in \left[0, \dfrac{\pi}{2}\right]$,

$$\begin{aligned} x &= \sqrt{2}a \cos\psi \sqrt{1 - \frac{1}{2}\sin^2\psi}, \\ y &= a \cos\psi \sin\psi = \frac{a}{2}\sin 2\psi. \end{aligned} \tag{1.20}$$

The geometric meaning of this parametrisation is illustrated in Fig. 1.7.

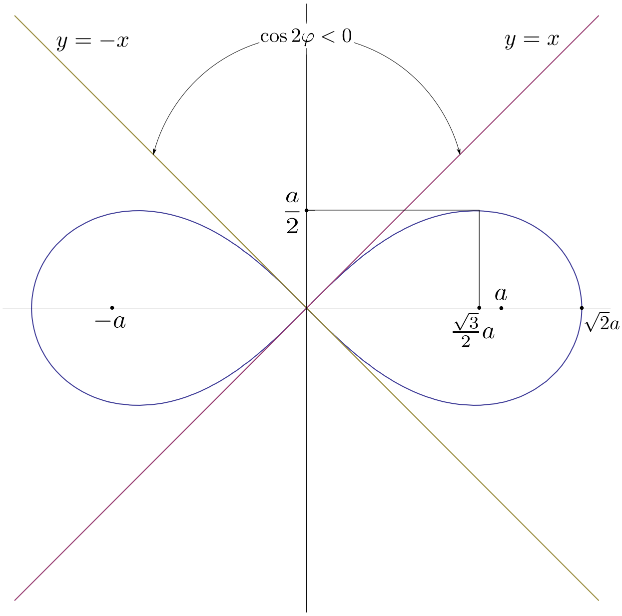

Fig. 1.6 Lemniscate.

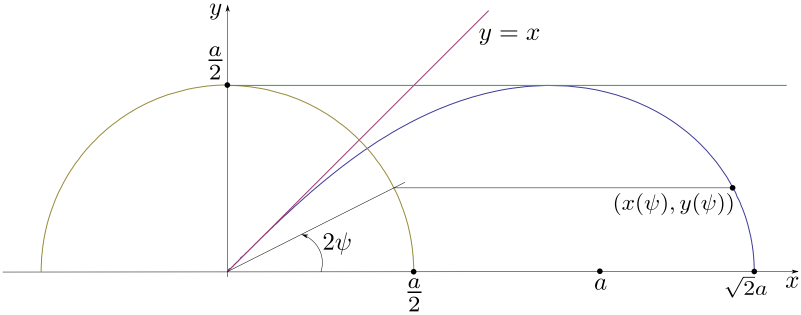

Fig. 1.7 Parametrisation of a lemniscate.

We use the parameter ψ to compute the arc length of the lemniscate.[6] The derivatives of $x = x(\psi)$ and $y = y(\psi)$ are

[6] The following computations in this section are quite complicated, so we omit details. When you are checking them, be not discouraged!

$$\frac{dx}{d\psi} = \sqrt{2}a \frac{\sin\psi}{\sqrt{1-\frac{1}{2}\sin^2\psi}} \left(-\frac{3}{2} + \sin^2\psi\right),$$
$$\frac{dy}{d\psi} = a(1 - 2\sin^2\psi),$$

from which follows

$$\left(\frac{dx}{d\psi}\right)^2 + \left(\frac{dy}{d\psi}\right)^2 = \frac{a^2}{1-\frac{1}{2}\sin^2\psi}.$$

Substituting this into (1.1) with t replaced by ψ, we obtain the arc length of the portion of the lemniscate corresponding to $\psi \in [0,\theta]$ to be

$$a\int_0^\theta \frac{d\psi}{\sqrt{1-\frac{1}{2}\sin^2\psi}}.$$

The arc in the first quadrant is $\left\{(x(\psi), y(\psi)) \mid \psi \in \left[0, \frac{\pi}{2}\right]\right\}$, so its length is

$$a\int_0^{\frac{\pi}{2}} \frac{d\psi}{\sqrt{1-\frac{1}{2}\sin^2\psi}},$$

and the total length of the lemniscate is the quadruple of this.

In general, the integral

$$F(k,\theta) := \int_0^\theta \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}} \tag{1.21}$$

is called the *incomplete elliptic integral of the first kind*. It is called the *complete elliptic integral of the first kind* when $\theta = \dfrac{\pi}{2}$, and then is often denoted as

$$K(k) := F\left(k, \frac{\pi}{2}\right) = \int_0^{\frac{\pi}{2}} \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}. \tag{1.22}$$

The parameter k is called the *modulus*.

In these notations,

$$\begin{aligned} \text{The arc length of the arc } 0 \le \psi \le \theta &= a\,F\left(\frac{1}{\sqrt{2}}, \theta\right), \\ \text{The arc length of the whole lemniscate} &= 4a\,K\left(\frac{1}{\sqrt{2}}\right). \end{aligned} \tag{1.23}$$

The elliptic integrals of the first kind also have an expression without trigonometric functions in the integrand — as in (1.13) for the elliptic integral of the second kind

— which can be obtained by changing the integration variable from ψ to $z = \sin\psi$ in (1.21). The inside of the radical symbol is $1 - k^2 z^2$ and, since

$$\frac{dz}{d\psi} = \cos\psi = \sqrt{1 - z^2},$$

we can rewrite

$$d\psi = \frac{dz}{\sqrt{1 - z^2}}.$$

Thus the algebraic expressions of the elliptic integrals of the first kind are

$$\begin{aligned} F(k,\theta) &= \int_0^{\sin\theta} \frac{dz}{\sqrt{(1 - z^2)(1 - k^2 z^2)}}, \\ K(k) &= \int_0^1 \frac{dz}{\sqrt{(1 - z^2)(1 - k^2 z^2)}}. \end{aligned} \tag{1.24}$$

Exercise 1.4 Using the change of variable of integration $\eta^2 = 1 - y^2$, express the following integrals in terms of the elliptic integrals of the first kind $F(k,\varphi)$ $(x = \cos\varphi)$ and $K(k)$ $(k \in \mathbb{R})$:

$$\begin{aligned} f(x) &= \int_0^x \frac{dy}{\sqrt{1 - y^4}}, \\ L &= \int_0^1 \frac{dy}{\sqrt{1 - y^4}}. \end{aligned}$$

Remark 1.5 The origin of the study of the elliptic integrals goes back to the integral $\int \frac{dy}{\sqrt{1 - y^4}}$ in this exercise. An 18th century Italian mathematician Fagnano (Giulio Carlo de' Toschi di Fagnano, 1682–1766) discovered a formula for this integral, which corresponds to an addition formula for elliptic functions. During the examination for Fagnano's membership in the Berlin Academy, Euler (Euler, Leonhard, 1707–1783) learned of this result and extended it.[7]

In this chapter we encountered elliptic integrals of the *first kind* and the *second kind*. It is natural for you to ask following questions:

- What on earth are those *elliptic integrals*?
- Are there elliptic integrals of the *third kind*, of the *fourth kind* and so on?

We shall answer these questions in the next chapter. (A short answer to the first question was given in the introduction, but we shall explain it in detail.)

[7] See [Ay] or Ch. 2–Ch. 4 of [Tks] for details.

Chapter 2
Classification of Elliptic Integrals

2.1 What is an Elliptic Integral?

We saw in the last chapter that the arc lengths of the ellipse and the lemniscate are expressed in terms of elliptic integrals of the second and the first kind, respectively:

$$\int \sqrt{1-k^2\sin^2\varphi}\,d\varphi = \int \sqrt{\frac{1-k^2x^2}{1-x^2}}\,dx,$$
$$\int \frac{d\psi}{\sqrt{1-k^2\sin^2\psi}} = \int \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

It is natural to call the former an *elliptic* integral, but why call the latter 'elliptic', even though the curve is not an ellipse? In fact, today the word 'elliptic integral' is a general term used in the following sense.

Definition 2.1 Let $R(x,s)$ be a rational function in two variables x and s depending on s non-trivially, $\dfrac{\partial R}{\partial s} \neq 0$, and $\varphi(x)$ be a square-free polynomial of degree three or four in x; that is, $\varphi(x)$ has three or four distinct roots. Then, an integral of the form

$$\int R(x, \sqrt{\varphi(x)})\,dx \tag{2.1}$$

is called an *elliptic integral*.

The elliptic integral of the first kind meets this definition with $R(x,s) = \dfrac{1}{s}$, $\varphi(x) = (1-x^2)(1-k^2x^2)$. The elliptic integral of the second kind has a different form from (2.1) at first glance, but by rewriting it as

$$\int \sqrt{\frac{1-k^2x^2}{1-x^2}}\,dx = \int \frac{1-k^2x^2}{\sqrt{(1-x^2)(1-k^2x^2)}}\,dx,$$

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_3

it also satisfies the conditions with $R(x,s) = \dfrac{1-k^2x^2}{s}$, $\varphi(x) = (1-x^2)(1-k^2x^2)$.

In this chapter we consider elliptic integrals as indefinite integrals. As we have learned in elementary calculus courses, indefinite integrals or primitive functions are related to definite integrals by the fundamental theorem of calculus: if $F(x)$ satisfies $F'(x) = f(x)$, then

$$F(x) = \int_a^x f(y)\,dy + \text{const.}$$

(We assume all 'good' properties of $f(x)$ like continuity.) In this chapter, we forget about definite integrals and regard integrals naively as *antiderivatives*. As we shall see, in this context (indefinite) integrals are algebraic objects.

Before going into elliptic integrals, let us recall more elementary integrals. Integrals $\displaystyle\int R(x)\,dx$ of a rational function $\left(= \dfrac{\text{polynomial}}{\text{polynomial}}\right)$ $R(x)$ with real coefficients can be expressed as a combination of rational functions, logarithms and the inverse tangent functions. Here we recall only the idea (or steps) of the proof of this fact and for details refer to any textbook of calculus.[1]

To begin with, we find the partial fraction decomposition of the integrand $R(x)$:

$$R(x) = P(x) + \sum_{j=1}^{l_1}\sum_{m=1}^{m_j} \frac{c_{jm}}{(x-a_j)^m} + \sum_{j=l_1+1}^{l_2}\sum_{m=1}^{m_j} \frac{d_{jm}x + e_{jm}}{((x-a_j)^2+b_j^2)^m}. \tag{2.2}$$

Here $P(x)$ is a polynomial, a_j, b_j, c_{jm}, d_{jm}, e_{jm} are real numbers and l_1, l_2, m_j are positive integers. This expression has many terms, but the point is that we decompose the integrand so that the primitive function of each part can be easily found.

The polynomial $P(x)$ is integrated by the formula

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} \qquad (n \in \mathbb{N})$$

and the result is a polynomial.

Among the remaining parts, the primitive functions of the terms with $m = 1$ are found by the following fundamental formulae[2] with changes of variables.

$$\int \frac{dx}{x-a} = \log(x-a), \quad \int \frac{dx}{x^2+1} = \arctan x,$$
$$\int \frac{2x\,dx}{x^2+1} = \log(x^2+1).$$

[1] Hereafter we omit constants of integration.

[2] You might say, "The symbols $|\cdot|$ are missing in the logarithm." Don't worry. Since $\log(-x) = \log x + \pi i$, the difference is nothing more than a constant of integration. Moreover, in complex analysis, one *should not* take $\log|x-a|$ as the primitive function of $\dfrac{1}{x-a}$. Of course, in real analysis (and in applications in, e.g., physics), one often has to take $\log|x-a|$. Here, it is not an essential problem, so *never mind*.

When m is larger than 1, we reduce the problem to the case $m = 1$, using integration by parts.

So we know how to integrate rational functions. (The actual computation is usually terrible, though.) As a next step, let us extend our interest to encompass slightly more complicated integrals of the form

$$\int R(x, \sqrt{\varphi(x)})\, dx,$$

having the same form as the elliptic integral (2.1), but for $\varphi(x)$ we take a linear function (or an affine function) $ax+b$. In this case, the substitution $y = \sqrt{ax+b}$ gives

$$x = \frac{y^2 - b}{a}, \quad dx = \frac{2y\, dy}{a},$$

which turn the above integral into an integral of a rational function in y. Therefore our integrals are a combination of rational functions of $y = \sqrt{\varphi(x)}$, log and arctan.

Let us proceed further to the case when $\varphi(x)$ is a quadratic polynomial ax^2+bx+c. This case is reduced to an integral of a rational function by changing the variable by trigonometric functions, which we discussed before. Here we compute two simple examples, again referring textbooks of calculus for futher details.

$$\begin{aligned}
\int \sqrt{1+x^2}\, dx &= \int \sqrt{1+\tan^2\theta}\, \frac{d\theta}{\cos^2\theta} \qquad (x = \tan\theta)\\
&= \int \frac{d\theta}{\cos^3\theta}\\
&= 2\int \frac{(1+t^2)^2}{(1-t^2)^3}\, dt \qquad \left(t = \tan\frac{\theta}{2}\right)\\
&= \cdots\\
&= \frac{1}{2}\left(x\sqrt{1+x^2} + \log(x+\sqrt{1+x^2})\right).
\end{aligned}$$

(Any integral of a rational function of $\sin\theta$ and $\cos\theta$ is converted to an integral of a rational function by means of the change of variables $t = \tan\dfrac{\theta}{2}$.)

We shall meet the following integral later in this book:

$$\begin{aligned}
\int \frac{dx}{\sqrt{1-x^2}} &= \int \frac{d\sin\theta}{\cos\theta} \qquad (x = \sin\theta)\\
&= \int \frac{\cos\theta\, d\theta}{\cos\theta} = \int d\theta = \theta = \arcsin x.
\end{aligned}$$

Many of you probably remember this integral as a formula. Here we recalled the steps of its computation as an example.

In general, if the integrand is a rational function of a quadratic irrational function $\sqrt{ax^2+bx+c}$, we rewrite it as $\sqrt{y^2+1}$, $\sqrt{1-y^2}$ or $\sqrt{y^2-1}$ by the change of variable $x \mapsto x - \dfrac{b}{2a}$ and rescaling. These expressions turn into rational functions of trigonometric functions by the changes of variables $y = \tan\theta$, $y = \sin\theta$ and $y = \dfrac{1}{\cos\theta}$ respectively. Then, applying $t = \tan\dfrac{\theta}{2}$, the problem reduces to the integral of a rational function. Sometimes one can use hyperbolic functions $\sinh\theta$ or $\cosh\theta$, but we omit the details. (For example, it is simpler to use $x = \sinh\theta$ for the first example $\int \sqrt{1+x^2}\,dx$.)

Summarising, if $\varphi(x)$ is a polynomial of degree not greater than two, an integral of the form $\int R(x, \sqrt{\varphi(x)})\,dx$ is expressed by a combination of rational functions, logarithms, inverse trigonometric functions and radicals. Let us call such a function *elementary*.

However, when $\deg\varphi(x)$ is greater than two, it can be proved that such an integral is not an elementary function in general.[3] In other words, an elliptic integral gives a 'new' function.

In Definition 2.1 $\varphi(x)$ was required to be square-free, because, if, for example, $\varphi(x)$ has a square factor like $\varphi(x) = (x-a)(x-b)^2$, then $\sqrt{\varphi(x)} = \sqrt{x-a} \times (x-b)$ and the integral $\int R(x, \sqrt{\varphi(x)})\,dx$ is an elementary function, as we saw above.

If the polynomial $\varphi(x)$ is of degree five or higher, such integrals are called *hyperelliptic integrals*, which are more complicated than elliptic ones. (There is no such curve as a "hyperellipse".) The complexity of elliptic and hyperelliptic integrals comes from the complexity of the associated geometric objects (Riemann surfaces), as we shall see later in the study of elliptic integrals in $\mathbb{C}$.

2.2 Classification of Elliptic Integrals

In the previous chapter we introduced elliptic integrals of the 'first kind' and the 'second kind'. Why are they special? Is there anything like "an elliptic integral of the n-th kind"?

Actually, according to the theorem which we are going to discuss now, any elliptic integral is expressed as a combination of elliptic integrals of the first kind, the second kind and the third kind to be defined in the theorem below.

Here a remark is in order. Although we are restricting ourselves to the real world, that is, we are talking about elliptic integrals and elliptic functions over $\mathbb{R}$ until we reach Chapter 6, we use *complex numbers* in this classification theorem. Thus, in this section a function may be complex-valued: for example coefficients in the rational function $R(x,s)$ and the polynomial $\varphi(x)$ are allowed to be complex numbers; a

[3] A detailed proof is in [Hi] Appendix A-5 or in [C]. The idea is to use Liouville's theory on integration. (For Liouville's theory, see, for example, [Ro].)

rational function like $R(x,s) = \dfrac{1}{(x-i)s}$ and a polynomial like $\varphi(x) = x^3 + i$ $(i = \sqrt{-1})$ can appear.

Why does this extension to complex coefficients make the classification simpler? As an example, let us recall how we integrated $\dfrac{1}{1+x^2}$. We know that $\displaystyle\int \frac{dx}{1+x^2} =$ $\arctan x$, but if we integrate this function as a complex-valued rational function, we can compute it like this:

$$\int \frac{dx}{1+x^2} = \frac{1}{2i} \int \left(\frac{1}{x-i} - \frac{1}{x+i} \right) dx = \frac{\log(x-i) - \log(x+i)}{2i}.$$

So, any integral of a rational function can be expressed in terms of log and we do not have to use arctan. (We have only to replace arctan by using the above formula.)

The situation is similar with elliptic integrals. The classification of *real* elliptic integrals is more complicated. We shall consider it briefly in the next section.

Theorem 2.2 (Legendre–Jacobi standard form)

Any elliptic integral

$$\int R(x, \sqrt{\varphi(x)})\, dx$$

is a linear combination of the following functions:

- *integrals of rational functions and quadratic irrational functions in x (expressed by elementary functions (rational functions, logarithms, radicals and their combinations);*
- *rational functions multiplied by* $\sqrt{\varphi(x)}$*;*
- *the elliptic integral of the first kind,* $\displaystyle\int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}$*;*
- *the elliptic integral of the second kind,*

$$\int \sqrt{\frac{1-k^2z^2}{1-z^2}}\, dz = \int \frac{1-k^2z^2}{\sqrt{(1-z^2)(1-k^2z^2)}}\, dz;$$

- *the elliptic integral of the third kind,*

$$\int \frac{dz}{(z^2-\alpha^2)\sqrt{(1-z^2)(1-k^2z^2)}}.$$

Here k is a constant determined by $\varphi(x)$ (the modulus *of the elliptic integral) and α is a constant determined by $R(x,s)$ and $\varphi(x)$.*

We need some remarks here.

- The word ‘combination’ in the definition of elementary functions means a finite number of arithmetic operations and compositions. As we shall see in the proof of the theorem, there appear only simple compositions.

- The variable z in the elliptic integrals of the first, the second and the third kind is obtained by changes of variables, and in general is different from the original variable x.

Proof (of Theorem* 2.2*) The proof of the classification theorem is roughly divided into two steps:

(I) cast the polynomial $\varphi(x)$ in standard form,

$$\varphi_k(z) = (1-z^2)(1-k^2z^2);$$

(II) bring the elliptic integral to standard form, using recurrence relations. □

(I) Standardising $\varphi(x)$

The definition of an elliptic integral requires $\varphi(x)$ to be a square-free polynomial of degree three or four. So, the classification theorem claims that even a polynomial of degree three can be turned into a polynomial $\varphi_k(x)$ of degree four. We explain this trick later. First, let us assume $\deg \varphi(x) = 4$.

Since we are considering polynomials with complex coefficients, we can factorise $\varphi(x)$ as

$$\varphi(x) = a(x-\alpha_0)(x-\alpha_1)(x-\alpha_2)(x-\alpha_3).$$

The constant a is non-zero and, because $\varphi(x)$ is square-free, $\alpha_0, \dots, \alpha_3$ are distinct complex numbers. Let us bring this polynomial to the form $\varphi_k(x)$.

The main tool is a fractional linear transformation (a Möbius transformation). Fractional linear transformations are usually covered as a topic in courses on complex analysis, but the definition is simple and can be given here without any technical machinery:

$$z \mapsto T(z) := \frac{Az+B}{Cz+D},$$

where $AD - BC \neq 0$. This map gives a bijection from the *Riemann sphere*[4] to itself. (If $AD - BC = 0$, namely, if $A : B = C : D$, the image of this map is one point.) The meaning of the point at infinity ∞ can be found in any course on complex analysis, but here you can just regard it as 'a fraction with denominator 0'. Thus $Cz + D = 0 \Rightarrow T(z) = \infty$. Similarly, $T(\infty)$ can be regarded as the limit when z tends to ∞: $T(\infty) = \lim\limits_{z\to\infty} \dfrac{Az+B}{Cz+D} = \dfrac{A}{C}$.

Lemma 2.3 *We can choose A, B, C, D and k, such that*

$$\begin{aligned} T(1) &= \alpha_0, & T(-1) &= \alpha_1, \\ T(k^{-1}) &= \alpha_2, & T(-k^{-1}) &= \alpha_3. \end{aligned} \tag{2.3}$$

---

[4] As a set the Riemann sphere is the union $\mathbb{C} \cup \{\infty\}$ of the set of complex numbers and the single element set consisting of the point at infinity. It is also called the *projective line* over $\mathbb{C}$. I prefer to denote it by $\mathbb{P}^1(\mathbb{C})$ or simply $\mathbb{P}^1$, but there are various notations, $\mathbb{C}P^1$, $\hat{\mathbb{C}}$, $\bar{\mathbb{C}}$ and so on.

This lemma follows from a property of fractional linear transformations. We leave the proof to the reader as an exercise (Exercise 2.4), and proceed further, admitting the lemma.

Let us rewrite the elliptic integral (2.1), changing the variable by $x = T(z)$ in Lemma 2.3. The square root of $\varphi(x)$ becomes the square root of the product of factors of the form $T(z) - \alpha_j$. For example, the factor corresponding to $j = 2$ has the form

$$\begin{aligned} T(z) - \alpha_2 &= T(z) - T(k^{-1}) \\ &= \frac{Az+B}{Cz+D} - \frac{Ak^{-1}+B}{Ck^{-1}+D} = (\text{const.}) \times \frac{z-k^{-1}}{Cz+D}. \end{aligned}$$

You can check it directly, but the following trick might help: It is easy to see that the denominator is $Cz + D$, while the numerator is a polynomial in z of degree one. When $z = k^{-1}$, $T(z) - \alpha_2$ should be 0, from which it follows that the numerator is a constant multiple of $z - k^{-1}$.

By multiplying all such factors for $j = 0, 1, 2, 3$ we obtain

$$\begin{aligned} \sqrt{\varphi(x)} &= \sqrt{a \prod_{j=0}^{3} (T(z) - \alpha_j)} \\ &= (\text{const.}) \sqrt{\frac{(z-1)(z+1)(z-k^{-1})(z+k^{-1})}{(Cz+D)^4}} \qquad (2.4) \\ &= (\text{const.}) \frac{\sqrt{\varphi_k(z)}}{(Cz+D)^2}. \end{aligned}$$

Hence the integrand of the elliptic integral becomes

$$R(x, \sqrt{\varphi(x)}) = R\left(\frac{Az+B}{Cz+D}, (\text{const.}) \frac{\sqrt{\varphi_k(z)}}{(Cz+D)^2} \right).$$

It looks complicated, but the important point is that the square root appears only in the form $\sqrt{\varphi_k(z)}$. As $R(x,s)$ is a rational function and we substitute a rational function of z in place of x, the integrand is expressed as

$$R(x, \sqrt{\varphi(x)}) = \tilde{R}(z, \sqrt{\varphi_k(z)})$$

for a certain rational function $\tilde{R}(z,t)$. It remains to rewrite 'dx' in the elliptic integral. By the formula for the change of variables,

$$dx = \frac{dx}{dz} dz = \frac{AD - BC}{(Cz+D)^2} dz,$$

which is nothing more than a rational function multiplied by dz. Thus the elliptic integral is rewritten as

$$\int R(x,\varphi(x))\,dx = \int \tilde{\tilde{R}}(z,\sqrt{\varphi_k(z)})\,dz$$

with an appropriate rational function $\tilde{\tilde{R}}(z,t)$. We have standardised $\varphi(x)$ to $\varphi_k(z)$ in this way.

Up to now we have been assuming that the degree of $\varphi(x)$ is four. The procedure for a polynomial of degree three is almost the same. The main difference comes from the factorisation,

$$\varphi(x) = a(x-\alpha_1)(x-\alpha_2)(x-\alpha_3).$$

There are only three roots of $\varphi(x)$, so we take $\alpha_0 = \infty$ additionally, and take k and $T(z)$ satisfying (2.3). (Recall that both the domain and the image of T are the Riemann sphere, which contains ∞.) In this case, as $T(1) = \infty$, the denominator of $T(z) = \dfrac{Az+B}{Cz+D}$ becomes 0 when $z = 1$. That is, $Cz+D$ is a constant multiple of $z-1$ and, instead of (2.4), we have

$$\begin{aligned}
\sqrt{\varphi(x)} &= \sqrt{a\prod_{j=1}^{3}(T(z)-\alpha_j)} \\
&= (\text{const.})\sqrt{\frac{(z+1)(z-k^{-1})(z+k^{-1})}{(z-1)^3}} \\
&= (\text{const.})\sqrt{\frac{(z-1)(z+1)(z-k^{-1})(z+k^{-1})}{(z-1)^4}} \\
&= (\text{const.})\frac{\sqrt{\varphi_k(z)}}{(z-1)^2}.
\end{aligned}$$

The rest of the standardising procedure of $\varphi(x)$ is the same as that for the case $\deg\varphi = 4$.

Exercise 2.4 (The proof of Lemma 2.3): (i) Check that a fractional linear transformation preserves the *anharmonic ratio* (*cross ratio*) of four points α_0, α_1, α_2, α_3, $\lambda := \dfrac{(\alpha_0-\alpha_2)(\alpha_1-\alpha_3)}{(\alpha_0-\alpha_3)(\alpha_1-\alpha_2)}$:

$$\frac{(\alpha_0-\alpha_2)(\alpha_1-\alpha_3)}{(\alpha_0-\alpha_3)(\alpha_1-\alpha_2)} = \frac{(T(\alpha_0)-T(\alpha_2))(T(\alpha_1)-T(\alpha_3))}{(T(\alpha_0)-T(\alpha_3))(T(\alpha_1)-T(\alpha_2))}. \tag{2.5}$$

Then, using (2.5) and assuming the existence of T satisfying (2.3), show that

$$k = \frac{1-\sqrt{\lambda}}{1+\sqrt{\lambda}} \text{ or } k = \frac{1+\sqrt{\lambda}}{1-\sqrt{\lambda}}. \tag{2.6}$$

(Note that T^{-1} is also a fractional linear transformation.)

(ii) Show that $k^2 \neq 0, 1$. (Hint: First show that the anharmonic ratio of four distinct complex numbers cannot be 0, 1 or ∞.)

(iii) Assume that k is what we have found in (i). Show that the fractional linear transformation $x = T(z)$ satisfying (2.3) is obtained by solving the equation

$$\frac{(x-\alpha_2)(\alpha_1-\alpha_3)}{(x-\alpha_3)(\alpha_1-\alpha_2)} = \frac{(z-k^{-1})(-1+k^{-1})}{(z+k^{-1})(-1-k^{-1})} \tag{2.7}$$

in z. (Hint: It is possible to check this directly by solving the equation, but it is not necessary. First, note that z is expressed as a fractional linear transformation of x, because the inverse of a fractional linear transformation and the composition of two fractional linear transformations are both fractional linear transformations. If x in the left-hand side of the above equation takes one of the values α_j $(j = 2, 1, 3, 0)$, then the values of both sides are 0, 1, ∞, λ, from which it follows that the corresponding values of z are what we need.)

Why do we need k, which we found in (i)? □

Remark 2.5 The assignment (2.3) of α_i $(i = 0, \dots, 3)$ to $T(\pm 1)$ and $T(\pm k^{-1})$ is one particular choice, but we may take any assignment. The expression of k in terms of λ changes according to this choice. For example, if α_1 and α_2 are interchanged and T is defined by

$$\begin{aligned} T(1) &= \alpha_0, & T(k^{-1}) &= \alpha_1, \\ T(-1) &= \alpha_2, & T(-k^{-1}) &= \alpha_3, \end{aligned} \tag{2.8}$$

the corresponding anharmonic ratio becomes $1-\lambda$ and therefore

$$k = \frac{1 \mp \sqrt{1-\lambda}}{1 \pm \sqrt{1-\lambda}} = \frac{2-\lambda \mp 2\sqrt{1-\lambda}}{\lambda} \tag{2.9}$$

instead of (2.6).

By permutations of $(\alpha_0, \dots, \alpha_3)$, the anharmonic ratio takes one of the values:

$$\lambda,\ \frac{1}{\lambda},\ 1-\lambda,\ \frac{1}{1-\lambda},\ \frac{\lambda}{1-\lambda},\ \frac{\lambda-1}{\lambda},$$

and k varies correspondingly. This is related to the modular transformation of theta functions (cf. Section 15.5), but we do not go into details. □

Exercise 2.6 Let $R(x,s)$ be a rational function of (x,s) and let $\varphi(x)$ be a polynomial of degree four without multiple roots. Show that the elliptic integral $\displaystyle\int R(x, \sqrt{\varphi(x)})\,dx$ can be rewritten in the form $\displaystyle\int \tilde{R}(y, \sqrt{\psi(y)})\,dy$ for some rational function $\tilde{R}(y, s')$ of (y, s'), where $\psi(y)$ is a polynomial of degree three. (Hint: use a fractional linear transformation which brings one of roots of $\varphi(x)$ to infinity.) □

(II) Standardising the elliptic integral

Now we are ready to prove the classification theorem. Although we know that we may assume that the square roots appear in the form $\sqrt{\varphi_k(z)}$, we use this fact only at the last step. What is important until then is not the explicit form of $\varphi_k(z)$ but the fact that $\varphi(z) = \varphi_k(z)$ is a polynomial of degree four without multiple roots.

First, notice the trivial fact that the square of $s = \sqrt{\varphi(x)}$ is a polynomial: $s^2 = \varphi(x)$. Consequently, we can replace even powers of s, i.e., $s^2, s^4, \dots$, by polynomials, $\varphi(x), \varphi(x)^2, \dots$, and odd powers of s, i.e., $s, s^3, s^5, \dots$ by $s \times$ (polynomial in x), $s, s\varphi(x), s\varphi(x)^2, \dots$. Therefore, we may assume that the rational function $R(x,s)$ in the elliptic integral is of the form $\dfrac{P(x,s)}{Q(x,s)}$, where $P(x,s)$ and $Q(x,s)$ are polynomials in (x,s) of degree one in s:

$$\begin{aligned} P(x,s) &= P_0(x) + P_1(x)s, \\ Q(x,s) &= Q_0(x) + Q_1(x)s. \end{aligned}$$

Moreover, we can rewrite $R(x, \sqrt{\varphi(x)}) = \dfrac{P(x, \sqrt{\varphi(x)})}{Q(x, \sqrt{\varphi(x)})}$ as follows.

$$\begin{aligned} R(x, \sqrt{\varphi(x)}) &= \frac{P_0(x) + P_1(x)s}{Q_0(x) + Q_1(x)s} \\ &= \frac{(P_0(x) + P_1(x)s)(Q_0(x) - Q_1(x)s)}{(Q_0(x) + Q_1(x)s)(Q_0(x) - Q_1(x)s)} \\ &= \frac{\tilde{P}_0(x) + \tilde{P}_1(x)s}{Q_0(x)^2 - Q_1(x)^2\varphi(x)} \\ &= R_0(x) + \tilde{R}_1(x)s = R_0(x) + \frac{R_1(x)}{s}. \end{aligned}$$

Here, $\tilde{P}_0$, $\tilde{P}_1$ are polynomials, R_0, $\tilde{R}_1$ are rational functions and $R_1(x) = \tilde{R}_1(x)\,\varphi(x)$.

Perhaps at first the classification might seem quite difficult, as we need to consider all rational functions in x and $s = \sqrt{\varphi(x)}$. But, in fact, we have reduced the problem to the classification of integrals of two types: $R_0(x)$ and $\dfrac{R_1(x)}{\sqrt{\varphi(x)}}$. Since we know how to integrate a rational function, the remaining problem is the integration of $\dfrac{R_1(x)}{\sqrt{\varphi(x)}}$.

Next, we find the partial fraction decomposition of $R_1(x)$. When only real coefficients are allowed, it is of the form (2.2), in which fractions with denominators of the form $((x-a)^2 + b^2)^m$ appear, but the partial fraction decomposition with complex coefficients is simpler:

$$R_1(x) = \tilde{P}(x) + \sum_{j=1}^{l} \sum_{m=1}^{m_j} \frac{c_{jm}}{(x - a_j)^m}. \tag{2.10}$$

As before, $\tilde{P}(x)$ is a polynomial, a_j, c_{jm} are complex numbers and l, m_j are integers. Thus we have shown that the elliptic integral $\displaystyle\int \frac{R_1(x)}{\sqrt{\varphi(x)}}dx$ is a linear combination of integrals of the following forms ($s = \sqrt{\varphi(x)}$):

- $\displaystyle I_n := \int \frac{x^n}{s}\,dx$ $(n = 0, 1, 2, \ldots)$,
- $\displaystyle J_n(\alpha) := \int \frac{dx}{(x-\alpha)^n s}$ $(n = 0, 1, 2, \ldots)$.

They are not independent but satisfy certain recurrence relations.

The recurrence relation for the integrals I_n is derived from the relation

$$\begin{aligned}\frac{d}{dx}x^n s &= nx^{n-1}s + \frac{x^n}{2s}\frac{d\varphi}{dx} \\ &= \frac{nx^{n-1}\varphi(x)}{s} + \frac{x^n \times (\text{polynomial of third degree})}{s}.\end{aligned}$$

Integrating it, we obtain relations for the integrals I_n:

$$x^n s = \begin{cases} c_{n,n+3}I_{n+3} + \cdots + c_{n,n-1}I_{n-1}, & (n \neq 0), \\ c_{0,3}I_3 + \cdots + c_{0,0}I_0, & (n = 0). \end{cases}$$

(The coefficients $c_{n,i}$ are determined by $\varphi(x)$.) Successive application of this relation reduces all I_n to a linear combination of (polynomial) $\times\, s$, I_2, I_1 and I_0.

The recurrence relations among the integrals $J_n(\alpha)$ comes from the following relation:

$$\begin{aligned}\frac{d}{dx}\frac{s}{(x-\alpha)^n} &= \frac{-ns}{(x-\alpha)^{n+1}} + \frac{1}{2(x-\alpha)^n s}\frac{d\varphi}{dx} \\ &= \frac{1}{(x-\alpha)^{n+1}s}\left(-n\varphi(x) + \frac{x-\alpha}{2}\frac{d\varphi}{dx}\right) \\ &= \frac{1}{(x-\alpha)^{n+1}s}\sum_{i=0}^{4} d_{n,i}(x-\alpha)^i.\end{aligned} \tag{2.11}$$

Here the coefficients $d_{n,i}$ are determined by $\varphi(x)$ from

$$-n\varphi(x) + \frac{x-\alpha}{2}\frac{d\varphi}{dx} = \sum_{i=0}^{4} d_{n,i}(x-\alpha)^i. \tag{2.12}$$

Explicit forms of two of them, $d_{n,0}$ and $d_{n,1}$, will be necessary later:

$$d_{n,0} = -n\varphi(\alpha),\ d_{n,1} = \left(\frac{1}{2} - n\right)\varphi'(\alpha). \tag{2.13}$$

They are obtained by the use of Taylor's formula for coefficients of the Taylor expansion.

Integrating (2.11), we obtain relations between the integrals $J_n(\alpha)$:

$$\frac{s}{(x-\alpha)^n} = d_{n,0}\,J_{n+1} + \cdots + d_{n,4}\,J_{n-3}. \tag{2.14}$$

(We omit '(α)' in $J_n(\alpha)$ for a while.)

There are two cases depending on the value of α. If $\varphi(\alpha) \neq 0$, $d_{n,0} \neq 0$ because of (2.13), which means that J_{n+1} is a linear combination of $J_n, \dots, J_{n-3}$ and $\dfrac{s}{(x-\alpha)^n}$ by (2.14). Recursively applying this procedure, we can express J_n ($n \geq 2$) as a linear combination of J_1, J_0, J_{-1}, J_{-2} and (a rational function) $\times s$.

In the other case ($\varphi(\alpha) = 0$) the formulae (2.13) lead to $d_{n,0} = 0$ and $d_{n,1} \neq 0$ because $\varphi(x)$ is square free. Hence the recurrence relation (2.14) means that J_n is a linear combination of $J_{n-1}, \dots, J_{n-3}$ and $\dfrac{s}{(x-\alpha)^n}$. Recursive application of this procedure reduces J_n ($n \geq 1$) to a linear combination of J_0, J_{-1}, J_{-2} and (a rational function)$\times s$.

Here, from the definition, J_0, J_{-1} and J_{-2} are

$$J_0 = \int \frac{dx}{s}, \quad J_{-1} = \int \frac{x-\alpha}{s}\,dx, \quad J_{-2} = \int \frac{(x-\alpha)^2}{s}\,dx,$$

each of which is a linear combination of I_0, I_1, I_2.

So far, we have shown that any elliptic integral $\displaystyle\int R(x, \sqrt{\varphi(x)})\,dx$ is a linear combination of an integral of a rational function, (a rational function)$\times\sqrt{\varphi(x)}$, and integrals of the following four types:

$$I_0 = \int \frac{dx}{\sqrt{\varphi(x)}}, \quad I_1 = \int \frac{x\,dx}{\sqrt{\varphi(x)}}, \quad I_2 = \int \frac{x^2\,dx}{\sqrt{\varphi(x)}},$$
$$J_1(\alpha) = \int \frac{dx}{(x-\alpha)\sqrt{\varphi(x)}}.$$

As the final step, we use the explicit form of $\varphi(x) = \varphi_k(x) = (1-x^2)(1-k^2x^2)$. The first integral I_0 is nothing but the elliptic integral of the first kind. The next one, I_1, is an elementary function, since by means of the change of variable $x = t^2$ it can be rewritten as

$$I_1 = \frac{1}{2}\int \frac{dt}{\sqrt{(1-t)(1-kt)}}.$$

The integral I_2 can be rewritten as

$$I_2 = \int \frac{\frac{1}{k^2}(1-(1-k^2x^2))}{\sqrt{(1-x^2)(1-k^2x^2)}}\,dx = \frac{1}{k^2}I_0 - \frac{1}{k^2}\int \sqrt{\frac{1-k^2x^2}{1-x^2}}\,dx,$$

which is a linear combination of elliptic integrals of the first and the second kinds. Lastly,

$$\begin{aligned} J_1(\alpha) &= \int \frac{(x+\alpha)\,dx}{(x^2-\alpha^2)\sqrt{(1-x^2)(1-k^2x^2)}} \\ &= \frac{1}{2}\int \frac{dt}{(t-\alpha^2)\sqrt{(1-t)(1-k^2t)}} + \alpha \int \frac{dx}{(x^2-\alpha^2)\sqrt{(1-x^2)(1-k^2x^2)}} \end{aligned}$$

is a linear combination of an elementary function (an integral of a quadratic irrational function) and the elliptic integral of the third kind.

The proof of the classification theorem is completed. □

Exercise 2.7 Reduce the elliptic integral $\displaystyle\int \frac{x^4\,dx}{\sqrt{(1-x^2)(1-2x^2)}}$ to standard form (a linear combination of an elementary function, the elliptic integrals of the first/second/third kinds).

In the classification theorem, we used a fourth-degree polynomial $\varphi_k(x) = (1-x^2)(1-k^2x^2)$. We can also use other polynomials to make such standard forms. For example, if $\varphi(x) = x(1-x)(1-\lambda x)$, we can take the following integrals as the standard forms:

$$\int \frac{dx}{\sqrt{x(1-x)(1-\lambda x)}}, \quad \int \frac{x\,dx}{\sqrt{x(1-x)(1-\lambda x)}}, \quad \int \frac{dx}{(x-\alpha)\sqrt{x(1-x)(1-\lambda x)}}.$$

They are called *Riemann's standard forms*. The proof is completely parallel to the proof of the above classification theorem, Theorem 2.2.

2.3 Real Elliptic Integrals.

Let us recall that an integral of a *real* rational function can always be expressed as a combination of *real* functions, as we discussed in Section 2.1.

In contrast to integrals of rational functions, in the case of elliptic integrals, even when all the coefficients of $\varphi(x)$ and $R(x,s)$ are real numbers, the proof of Theorem 2.2 does not guarantee that an integral $\displaystyle\int R(x,\sqrt{\varphi(x)})dx$ is a linear combination of *real* elliptic integrals of the first, the second and the third kinds and integrals of *real* rational functions. This is because there were two points in that proof where the use of complex numbers was essential:

1. When we standardised $\varphi(x)$, we factorised it by means of the fundamental theorem of algebra.

2. In the decomposition (2.10) of $R_1(x)$ powers of quadratic polynomials did not appear in the denominators, in contrast to the real case (2.2).

However, examining the details of the factorisation of $\varphi(x)$ in the first point, we can show that we can always take the real modulus of the elliptic integrals appearing in Theorem 2.2.

As is well known, if a real polynomial has a complex (non-real) root, its complex conjugate is also a root. So, if $\deg\varphi(x) = 4$, there are the following three cases for the roots $\alpha_0, \dots, \alpha_3$:

(i) All roots are real: $\alpha_i \in \mathbb{R}$ $(i = 0, \dots, 3)$.
(ii) Two roots are real and two roots are complex conjugate to each other: $\alpha_0, \alpha_1 \in \mathbb{R}$, $\alpha_3 = \bar{\alpha}_2 \notin \mathbb{R}$.
(iii) All roots are not real: $\alpha_1 = \bar{\alpha}_0 \notin \mathbb{R}$, $\alpha_3 = \bar{\alpha}_2 \notin \mathbb{R}$.

(i) The case $\alpha_i \in \mathbb{R}$ $(i = 0, \dots, 3)$: We may order the roots as $\alpha_2 < \alpha_0 < \alpha_1 < \alpha_3$. Then their anharmonic ratio satisfies the inequalities

$$0 < \lambda = \frac{(\alpha_0 - \alpha_2)(\alpha_1 - \alpha_3)}{(\alpha_0 - \alpha_3)(\alpha_1 - \alpha_2)} = \frac{\alpha_0 - \alpha_2}{\alpha_1 - \alpha_2} \cdot \frac{\alpha_3 - \alpha_1}{\alpha_3 - \alpha_0} < 1,$$

and hence $0 < k < 1$ by taking the first expression in (2.6). The coefficients of the fractional linear transformation $x = T(z) = \dfrac{Az + B}{Cz + D}$ are also real numbers (cf. Exercise 2.4 (iii)).

(ii) The case $\alpha_0, \alpha_1 \in \mathbb{R}$ and $\alpha_3 = \bar{\alpha}_2$: In this case the absolute value of the anharmonic ratio is 1:

$$\begin{aligned} |\lambda| &= \left| \frac{(\alpha_0 - \alpha_2)(\alpha_1 - \alpha_3)}{(\alpha_0 - \alpha_3)(\alpha_1 - \alpha_2)} \right| = \left| \frac{\alpha_0 - \alpha_2}{\alpha_1 - \alpha_2} \right| \left| \frac{\alpha_1 - \alpha_3}{\alpha_0 - \alpha_3} \right| \\ &= \left| \frac{\alpha_0 - \alpha_2}{\alpha_1 - \alpha_2} \right| \left| \overline{\left(\frac{\alpha_1 - \alpha_2}{\alpha_0 - \alpha_2} \right)} \right| = 1. \end{aligned}$$

Therefore there exists a real number θ such that $\sqrt{\lambda} = e^{i\theta}$. Substituting this into the first expression of (2.6),

$$k = \frac{1 - e^{i\theta}}{1 + e^{i\theta}} = \frac{e^{-i\theta/2} - e^{i\theta/2}}{e^{-i\theta/2} + e^{i\theta/2}} = -i \tan\frac{\theta}{2}.$$

So k is purely imaginary and $k^2 < 0$.

If x is real, z defined by (2.7) is also real. This can be proved by taking the complex conjugate of (2.7) and inverting both sides. (Recall that $\alpha_3 = \bar{\alpha}_2$ and $\overline{k^{-1}} = -k^{-1}$.) Using another change of variable

$$\zeta := \sqrt{1 - k^2 z^2},$$

which sends a real number z to a real number ζ, we can rewrite, for example, the elliptic integral of the first kind with modulus k as

$$\int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = -\frac{1}{\sqrt{k^2-1}} \int \frac{d\zeta}{\sqrt{(1-\zeta^2)(1-h^2\zeta^2)}}, \tag{2.15}$$

where $h = \dfrac{1}{\sqrt{1-k^2}}$, $0 < h < 1$. Thus, although the coefficient is complex, we have reduced the integral to the real elliptic integral of the first kind with modulus h, $0 < h < 1$.

Exercise 2.8 Check (2.15) and show that the real elliptic integrals of the second and the third kinds are also reduced to the elliptic integrals of the same kind with modulus h by the same argument as above.

(iii) The case $\alpha_1 = \bar{\alpha}_0$ and $\alpha_3 = \bar{\alpha}_2$: In this case the anharmonic ratio is a positive real number, since

$$\begin{aligned}\lambda &= \frac{(\alpha_0-\alpha_2)(\alpha_1-\alpha_3)}{(\alpha_0-\alpha_3)(\alpha_1-\alpha_2)} = \frac{(\alpha_0-\alpha_2)\overline{(\alpha_0-\alpha_2)}}{(\alpha_0-\alpha_3)\overline{(\alpha_0-\alpha_3)}} \\ &= \frac{|\alpha_0-\alpha_2|^2}{|\alpha_0-\alpha_3|^2}.\end{aligned} \tag{2.16}$$

We may assume that $\operatorname{Im}\alpha_0 > 0$ and $\operatorname{Im}\alpha_2 > 0$. This means α_0 and α_2 are on the same side of the real axis, while α_0 and $\alpha_3 = \bar{\alpha}_2$ are on the opposite side, from which the inequality $\lambda < 1$ follows. So, as in the case (i), we have $0 < k < 1$.

In contrast to the case (i), the fractional linear transformation $x = T(z)$ gives a purely imaginary number z. Note that $\left|\dfrac{\alpha_1-\alpha_3}{\alpha_1-\alpha_2}\right| = \left|\dfrac{-1+k^{-1}}{-1-k^{-1}}\right|$.

Exercise 2.9 Show this equality. (Hint: Show that the left-hand side is equal to the square root of λ by the same computation as (2.16). The right-hand side is equal to $\sqrt{\lambda}$ because of (2.6).)

Hence, by taking the absolute values of both sides of (2.7), we have

$$\left|\frac{x-\alpha_2}{x-\alpha_3}\right| = \left|\frac{z-k^{-1}}{z+k^{-1}}\right|.$$

If x is a real number, the left-hand side is 1 because $\alpha_3 = \bar{\alpha}_2$, from which it follows that $|z-k^{-1}| = |z+k^{-1}|$. Since z is of the same distance from k^{-1} and from $-k^{-1}$ on the real axis, it is on the imaginary axis.

Therefore, if we define another variable ζ by

$$\zeta := \frac{iz}{\sqrt{1-z^2}},$$

it is real for $z \in i\mathbb{R}$ and the elliptic integral of the first kind is rewritten as follows:

$$\int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = -i \int \frac{d\zeta}{\sqrt{(1-\zeta^2)(1-h^2\zeta^2)}}, \tag{2.17}$$

where $h = \sqrt{1-k^2}$ $(0 < h < 1)$.

Exercise 2.10 Check (2.17) and show that the real elliptic integrals of the second and the third kinds are also reduced to the elliptic integrals of the same kind with modulus h by the same argument as above.

Thus we have a good understanding of what elliptic integrals are. Or, at least, let us assume so. Now, it is time to *use* them, which is the theme of the next chapter.

Chapter 3
Applications of Elliptic Integrals

In this chapter, we are going to see how elliptic integrals are applied in mathematics and physics. Good mathematical objects appear in many situations.

3.1 The Arithmetic-Geometric Mean

We begin with the *arithmetic-geometric mean*, which was one of the origins of the study of elliptic integrals and elliptic functions, but at the beginning elliptic integrals are not in sight. Please be patient.

The inequality between the arithmetic mean and the geometric mean of two positive real numbers a and b is well known:

$$\sqrt{ab} \leq \frac{a+b}{2}. \tag{G<A}$$

For the moment, we assume that a is not less than b ($a \geq b > 0$). The 'mean' should be between a and b, and we have an inequality (G<A), so there is the order

$$b \leq \sqrt{ab} \leq \frac{a+b}{2} \leq a.$$

Let us repeat this procedure: Put $a_0 := a$, $b_0 := b$ and define sequences $\{a_n\}_{n=0,1,2,\dots}$ and $\{b_n\}_{n=0,1,2,\dots}$ by recurrence relations,

$$a_{n+1} := \frac{a_n + b_n}{2}, \qquad b_{n+1} := \sqrt{a_n b_n}. \tag{3.1}$$

Then, the above inequalities imply

$$b_n \leq b_{n+1} \leq a_{n+1} \leq a_n,$$

and the two sequences shrink inward as

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_4

$$b = b_0 \le b_1 \le \cdots \le b_n \le \cdots \le a_n \le \cdots \le a_1 \le a_0 = a. \tag{3.2}$$

If $a = b$, all a_n's and b_n's are equal. If $a < b$, both sequences $\{a_n\}_{n=0,1,2,\ldots}$ and $\{b_n\}_{n=0,1,2,\ldots}$ are bounded (they are between a and b), $\{a_n\}$ is monotonically decreasing and $\{b_n\}$ is monotonically increasing, which means those sequences have limits:

$$\alpha := \lim_{n\to\infty} a_n, \qquad \beta := \lim_{n\to\infty} b_n.$$

Moreover these limits coincide: $\alpha = \beta$. The proof is simple. Let us prove that the difference $c_n := a_n - b_n$ of the two sequences converges to zero. From the inequality $b_{n-1} \le b_n$ and the definition (3.1) of a_n it follows that

$$0 \le c_n = a_n - b_n \le a_n - b_{n-1} = \frac{a_{n-1} + b_{n-1}}{2} - b_{n-1} = \frac{c_{n-1}}{2}.$$

Hence,

$$0 \le c_n \le \frac{c_{n-1}}{2} \le \cdots \le \frac{c_0}{2^n}, \tag{3.3}$$

which leads to $c_n \to 0$ ($n \to \infty$). Of course the order of a and b is irrelevant to the conclusion that $\{a_n\}$ and $\{b_n\}$ have the same limit. So, we do not impose a condition on the order of a and b below.

This common limit is called the *arithmetic-geometric mean* (AGM) of a and b. We denote it by $M(a,b)$:

$$M(a,b) := \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n. \tag{3.4}$$

Example 3.1 Gauss computed several examples of the arithmetic-geometric mean ([G] Bd.III, pp. 363–364). We compute two of them with a computer to the 21-st decimal place. Observe how the sequences converge.

$$\begin{aligned}
a &= 1, & b &= 0.2,\\
a_1 &= 0.6, & b_1 &= 0.447213595499957939282\ldots,\\
a_2 &= 0.523606797749978969641\ldots, & b_2 &= 0.518004012822270290537\ldots,\\
a_3 &= 0.520805405286124630089\ldots, & b_3 &= 0.520797870939876244136\ldots,\\
a_4 &= 0.520801638113000437112\ldots, & b_4 &= 0.520801638099375678366\ldots,\\
a_5 &= 0.520801638106188057739\ldots, & b_5 &= 0.520801638106188057739\ldots.
\end{aligned}$$

The inequality (3.3) claims that the difference $a_n - b_n$ is less than half of the preceding difference $a_{n-1} - b_{n-1}$ at each step, but, as is shown above, the sequences converge much faster. We do not go into details on this point.

Gauss computed this example to the 21-st decimal place for $a_1, \ldots, a_4$ and $b_1, \ldots, b_3$ and to the 15-th decimal place for a_5, b_4 and b_5. His calculation was incorrect[1] for b_2. (The error was about 2×10^{-15}.) Therefore his final results (a_5 and b_5) differed from the correct results by one in the 15-th decimal place.

The next example will be important later. (See Example 3.4.)

$$\begin{aligned}
a &= \sqrt{2} = 1.41421356237309504880 2\ldots, & b &= 1, \\
a_1 &= 1.20710678118654752440 1\ldots, & b_1 &= 1.18920711500272106671 8\ldots, \\
a_2 &= 1.19815694809463429555 9\ldots, & b_2 &= 1.19812352149312012260 7\ldots, \\
a_3 &= 1.19814023479387720908 3\ldots, & b_3 &= 1.19814023467730720579 8\ldots, \\
a_4 &= 1.19814023473559220744 1\ldots, & b_4 &= 1.19814023473559220743 9\ldots.
\end{aligned}$$

Gauss computed this correctly to the 21-st decimal place.

Today we can compute such examples instantly using computers, but in the eighteenth century, when computers did not exist, Gauss computed them *by hand*. (These examples comprise less than one third of Gauss's results on the arithmetic-geometric mean.) I had thought that Gauss was a completely "theoretical" scientist, but, seeing such tremendous computations, my impression of him changed.

Now we return to the general theory of the arithmetic-geometric mean and see what properties it has.

(i) *Symmetry* : $M(a,b) = M(b,a)$ by definition.
(ii) *Homogeneity* : for a positive real number $\lambda > 0$, $M(\lambda a, \lambda b) = \lambda M(a,b)$. (Since the definitions (3.1) of the arithmetic and the geometric means are homogeneous of degree one, if we replace a and b with λa and λb, the sequences $\{a_n\}$ and $\{b_n\}$ are replaced with $\{\lambda a_n\}$ and $\{\lambda b_n\}$ and, as a result, the arithmetic-geometric mean is multiplied by λ.)
(iii) $M(a,a) = a$.
(iv) $M(a,b) = M(a_n, b_n)$ $(n = 0, 1, 2, \ldots)$. (The limit does not change by throwing away a finite number of terms.)

By the homogeneity $M(a,b) = b\,M\left(\frac{a}{b}, 1\right) = a\,M\left(1, \frac{b}{a}\right)$. Therefore, in order to study properties of $M(a,b)$, it is enough to study $M(1,b)$ or $M(a,1)$. In addition, thanks to the symmetry $M(a,1) = M(1,a)$, we have only to examine $M(1,b)$.

Thank you for your patience, here comes the elliptic integral!

Theorem 3.2 *Assume* $0 < k < 1$. *The arithmetic-geometric mean is expressed by the complete elliptic integral of the first kind as follows:*

$$M(1,k) = \frac{\pi}{2} \frac{1}{K(k')}. \tag{3.5}$$

Here k' *is the* complementary modulus *defined by* $k' := \sqrt{1-k^2}$.

[1] This error was pointed out by Hiroyuki Ochiai, who read the draft of [Tkb].

For convenience's sake we recall the definition of the complete elliptic integral of the first kind ((1.22) and (1.24)) here.

$$K(k') = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k'^2\sin^2\theta}} = \int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k'^2z^2)}}. \tag{3.6}$$

Proof It was Gauss, who first proved Theorem 3.2. He gave two proofs, one with theta functions and one by a change of the integration variable. Here we show a proof ([Ne]) by a simpler change of the integration variable found in the twentieth century. At the end of this book (Section 17.4) we give another proof, which follows the strategy of Gauss's first proof.

The following is the key lemma.

Lemma 3.3 *Let a and b be positive real numbers and define $I(a,b)$ by*

$$I(a,b) := \int_0^{\pi/2} \frac{d\theta}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}}. \tag{3.7}$$

Then it is expressed by the arithmetic-geometric mean as

$$I(a,b) = \frac{\pi}{2}\frac{1}{M(a,b)}. \tag{3.8}$$

The theorem follows immediately from this lemma. Setting $a = 1$ and $b = k$, we have

$$I(1,k) = \frac{\pi}{2}\frac{1}{M(1,k)}.$$

On the other hand, from (3.7) follows

$$\begin{aligned} I(1,k) &= \int_0^{\pi/2} \frac{d\theta}{\sqrt{\cos^2\theta + k^2\sin^2\theta}} \\ &= \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-(1-k^2)\sin^2\theta}} = K(k'). \end{aligned}$$

Thus we obtain (3.5). This is the end of the proof of Theorem 3.2. □

Proof (of Lemma* 3.3.*) We can prove the identity

$$I(a,b) = I\left(\frac{a+b}{2}, \sqrt{ab}\right), \tag{3.9}$$

which means that $I(a,b)$ does not change by replacing a and b with their arithmetic and geometric means. Recursive application of this identity shows

$$I(a,b) = I(a_1,b_1) = \cdots = I(a_n,b_n) = \cdots \longrightarrow I\big(M(a,b),M(a,b)\big). \tag{3.10}$$

(Here we changed the order of the limit $n \to \infty$ and the integral which defines $I(a_n, b_n)$. We have only to check the uniform convergence of the integrand.)

For any constant c the integral $I(c,c)$ is easily computed:

$$I(c,c) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{c^2\cos^2\theta + c^2\sin^2\theta}}$$
$$= \int_0^{\frac{\pi}{2}} \frac{d\theta}{c\sqrt{\cos^2\theta + \sin^2\theta}} = \int_0^{\frac{\pi}{2}} \frac{d\theta}{c} = \frac{\pi}{2c}.$$

Substituting $c = M(a,b)$ into this equation and using (3.10), we obtain $I(a,b) = \dfrac{\pi}{2M(a,b)}$.

It remains to prove (3.9). We change the integration variable twice. The integral $I(a,b)$ is equal to

$$I(a,b) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}}$$
$$= \int_0^{\frac{\pi}{2}} \frac{d\theta}{\cos\theta\sqrt{a^2 + b^2\tan^2\theta}}.$$

If we change the variable from θ to $t = b\tan\theta$, the inside of the radical sign becomes $a^2 + t^2$. On the other hand, we have

$$\frac{dt}{d\theta} = \frac{b}{\cos^2\theta}, \quad b^2 + t^2 = \frac{b^2}{\cos^2\theta},$$

which implies

$$\frac{d\theta}{\cos\theta} = \frac{\cos\theta}{b} dt = \frac{dt}{\sqrt{b^2 + t^2}}.$$

Summarising, we have obtained

$$I(a,b) = \int_0^{\infty} \frac{dt}{\sqrt{(a^2 + t^2)(b^2 + t^2)}}. \tag{3.11}$$

(Do not forget to change the integration range.)

The next change of the variable is $u = \dfrac{1}{2}\left(t - \dfrac{ab}{t}\right)$. (See Fig. 3.1.)

It is easy to check the identities

$$\frac{du}{dt} = \frac{1}{2}\left(1 + \frac{ab}{t^2}\right) = \frac{\sqrt{ab + u^2}}{t},$$
$$(a^2 + t^2)(b^2 + t^2) = t^2((a+b)^2 + 4u^2),$$

which lead to

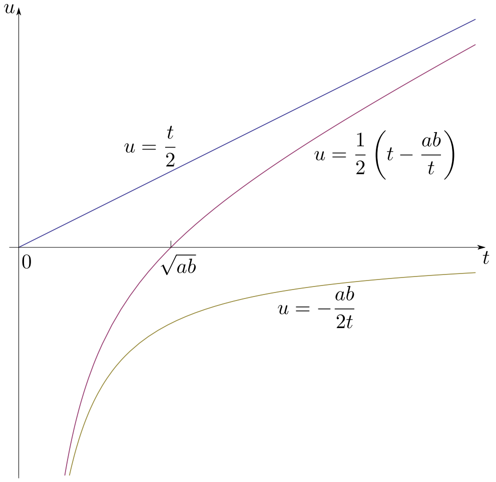

Fig. 3.1 The change of the variable $u(t) = \dfrac{1}{2}\left(t - \dfrac{ab}{t}\right)$.

$$
\begin{aligned}
I(a,b) &= \int_{-\infty}^{\infty} \frac{du}{\sqrt{((a+b)^2+4u^2)(ab+u^2)}} \\
&= 2\int_{0}^{\infty} \frac{du}{\sqrt{((a+b)^2+4u^2)(ab+u^2)}} \\
&= \int_{0}^{\infty} \frac{du}{\sqrt{\left(\left(\frac{a+b}{2}\right)^2+u^2\right)((\sqrt{ab})^2+u^2)}}.
\end{aligned}
$$

Comparing this expression with (3.11), we conclude

$$
I(a,b) = I\left(\frac{a+b}{2}, \sqrt{ab}\right),
$$

which is nothing but (3.9). □

Up to now, we have assumed that $0 < k < 1$. It is easy to see that the equation

$$
M(k,1) = \frac{\pi}{2}\frac{1}{K(k')}, \qquad k' := \sqrt{1-k^2}
$$

holds for $k > 1$, too. In this case, since $1-k^2 < 0$, k' is not a real number. However, the number k' appears in the definition of $K(k')$, (3.6), only in the form $k'^2 \in \mathbb{R}$ and there is no imaginary number in the proof of Theorem 3.2 for $k > 1$.

Example 3.4 For $k = \sqrt{2}$ ($k' = \sqrt{-1}$), $K(\sqrt{-1}) = \dfrac{\pi}{2M(\sqrt{2}, 1)}$, which means

$$\int_0^1 \frac{dx}{\sqrt{1-x^4}} = \frac{\pi}{2M(\sqrt{2}, 1)}.$$

In fact, Gauss conjectured this equation before proving Theorem 3.2. He had computed an elliptic integral $\dfrac{\varpi}{2} = \displaystyle\int_0^1 \frac{dx}{\sqrt{1-x^4}}$ by hand(!) and obtained $\dfrac{\varpi}{\pi} = 0.834626841674030\ldots$ ([G] Bd.X Abt1, p. 169). By comparing this result with the computation which we saw in Example 3.1, Gauss observed

$$M(\sqrt{2}, 1) = \frac{\pi}{\varpi}.$$

(This is one of the proofs that Gauss had an abnormal sense of numbers. How can an ordinary person notice that the product of

$$0.834626841674030\ldots \text{ and } 1.1981402347355922074\ldots$$

is 1?) Twenty-two year old Gauss wrote in his diary (30th May 1799; [G] Bd.X Abt1, p. 542),

> '*Terminum medium arithmetico-geometricum inter* 1 *et* $\sqrt{2}$ *esse* $= \dfrac{\pi}{\varpi}$ *usque ad figuram undecimam comprobavimus, qua re demonstrata prorsus novus campus in analysi certo aperietur.* (We verified to the eleventh decimal place that the arithmetic-geometric mean between 1 and $\sqrt{2}$ is equal to $\dfrac{\pi}{\varpi}$. If this will be proved, a completely new field of analysis is surely opened.)'

Gauss proved it later ([G] Bd.III, p. 352– (1818) and Bd.X Abt1, p. 194– (1799))[2] and further developed his own theory of elliptic functions but never published while he was alive ([Kl] Erstes Kapitel, [Tkg] Ch. 9).

With this historical background, Theorem 3.2 claims that the complete elliptic integral of the first kind and the arithmetic-geometric mean are essentially the same thing. Using this fact, we can easily derive several properties of the complete elliptic integral of the first kind from the properties of arithmetic-geometric mean. Please try to show the following identity in this way.

Exercise 3.5 Prove

$$K(k) = \frac{1}{1+k} K\left(\frac{2\sqrt{k}}{1+k}\right) = \frac{2}{1+k'} K\left(\frac{1-k'}{1+k'}\right),$$

by using the properties of the arithmetic-geometric mean. (Hint: Use homogeneity (Property (ii) of the arithmetic-geometric mean) and $M(a, b) = M(a_n, b_n)$ (Property (iv)).)

[2] According to [Ka] p. 103, Landen (J. Landen, 1719–1790) had already proved (3.5) before Gauss.

3.2 Motion of a Simple Pendulum

Now let us turn to a problem in physics: a simple pendulum. This is the most well-known application of elliptic integrals.

A *simple pendulum* is a really simple system. It is a weight (bob; mass m) suspended from a fixed point (pivot) by a cord or a rod like Fig. 3.2. The cord (or the rod) is so light that we can neglect its mass and it is so stiff that it does not bend and its length l does not change during motion. We also assume that there is no friction nor air drag. (See Fig. 3.2.)

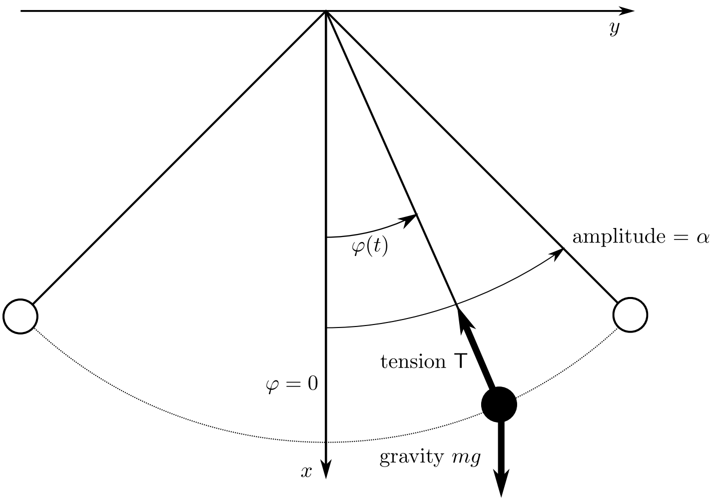

Fig. 3.2 A simple pendulum.

The gravity mg (g is the gravitational acceleration) pulls the bob down and the tension T pulls the bob to the direction of the cord. If we take the x- and y-coordinates as in Fig. 3.2 and denote the coordinates of the bob as $(x(t), y(t)) = (l\cos\varphi(t), l\sin\varphi(t))$, the acceleration of the bob is

$$\frac{d^2}{dt^2}\begin{pmatrix} l\cos\varphi(t) \\ l\sin\varphi(t) \end{pmatrix} = \frac{d}{dt}\begin{pmatrix} -l\dot{\varphi}\sin\varphi \\ l\dot{\varphi}\cos\varphi \end{pmatrix} = \begin{pmatrix} -l\ddot{\varphi}\sin\varphi - l\dot{\varphi}^2\cos\varphi \\ l\ddot{\varphi}\cos\varphi - l\dot{\varphi}^2\sin\varphi \end{pmatrix}.$$

(Here we denote the derivatives of the function $\varphi(t)$ with respect to t by $\dfrac{d\varphi}{dt} = \dot{\varphi}$ and $\dfrac{d^2\varphi}{dt^2} = \ddot{\varphi}$.) Therefore Newton's equation of motion for this system is

$$ml\begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}\ddot{\varphi} - ml\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}\dot{\varphi}^2 = \begin{pmatrix} -\mathsf{T}\cos\varphi + mg \\ -\mathsf{T}\sin\varphi \end{pmatrix}. \tag{3.12}$$

Multiplying the x- and y-components by $-\sin\varphi$ and $\cos\varphi$ respectively and summing up, we obtain $ml\ddot{\varphi} = -mg\sin\varphi$, namely, a differential equation

$$\frac{d^2\varphi}{dt^2} = -\omega^2 \sin\varphi, \qquad \omega := \sqrt{\frac{g}{l}}. \tag{3.13}$$

This is a *non-linear* differential equation, whose right-hand side contains a non-linear function $\sin\varphi$ of the unknown function $\varphi(t)$.

If the amplitude (= the maximum value of the angle φ) is small, the function $\varphi(t)$ is always small and we can approximate $\sin\varphi(t)$ by $\varphi(t)$: $\sin\varphi \approx \varphi$. Then the equation (3.13) is approximated by

$$\frac{d^2\varphi}{dt^2} = -\omega^2\varphi,$$

which is a *linear* ordinary differential equation of the second order. The set of all solutions is a linear space and the general solution is expressed as

$$\varphi(t) = c_1 \cos\omega t + c_2 \sin\omega t \tag{3.14}$$

with two constants $c_1, c_2 \in \mathbb{R}$. (In other words, we can take $\cos\omega t$ and $\sin\omega t$ as the basis vectors of the solution space.) The solution $\varphi(t)$ has period

$$T_0 = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{l}{g}} \tag{3.15}$$

due to the periodicity of the trigonometric functions. The period T_0 does not depend on the mass m of the bob or on the amplitude α, which proves the so-called *isochronism*.

In introductory courses on physics, the analysis of a pendulum usually stops here. Let us go further and consider the case when 'the amplitude is large', which means solving the differential equation (3.13) without approximation. Since it is not a linear differential equation, there is no general theory of solution. However we can solve the equation (3.13) as follows.

First, we multiply the equation of motion by $\dot{\varphi}$ and rewrite it as

$$\frac{d^2\varphi}{dt^2}\frac{d\varphi}{dt} = -\omega^2 \sin\varphi \frac{d\varphi}{dt},$$
$$\frac{1}{2}\frac{d}{dt}\left(\frac{d\varphi}{dt}\right)^2 = \frac{d}{dt}(\omega^2\cos\varphi).$$

Integrating the last equation in t, we obtain

$$\frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2 = \omega^2\cos\varphi + (\text{const.}). \tag{3.16}$$

In other words,

$$\tilde{E} := \frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2 - \omega^2 \cos\varphi \tag{3.17}$$

is a constant independent of the time variable t: $\dfrac{d\tilde{E}}{dt} = 0$. In fact, $\tilde{E}$ is proportional to the mechanical energy E of the pendulum: $E = ml^2\tilde{E}$. (The kinetic energy of the pendulum is $\dfrac{ml^2}{2}\left(\dfrac{d\varphi}{dt}\right)^2$, the potential energy is $-mgl\cos\varphi$ and the total mechanical energy is their sum $= ml^2\tilde{E}$.) The statement '$\tilde{E}$ does not depend on t' is the *conservation of mechanical energy*: 'the mechanical energy E does not depend on t'. We do not go into the physical details here.

If the pendulum is 'swinging'[3], the angle $\varphi(t)$ has a maximum, the *amplitude*, which we denote by α as in Fig. 3.2. At some time $t = t_0$ the function $\varphi(t)$ attains this maximum value $\varphi(t_0) = \alpha$, and its derivative vanishes there: $\dot{\varphi}(t_0) = 0$. So, if we put $t = t_0$ in (3.17), the first term in the right-hand side is zero and $\tilde{E} = -\omega^2\cos\alpha$. Substituting this into (const.) in (3.16), we have

$$\begin{aligned}\frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2 &= \omega^2(\cos\varphi - \cos\alpha)\\ &= 2\omega^2\left(\sin^2\frac{\alpha}{2} - \sin^2\frac{\varphi}{2}\right).\end{aligned}$$

Thus we obtain

$$\frac{d\varphi}{dt} = 2\omega\sqrt{\sin^2\frac{\alpha}{2} - \sin^2\frac{\varphi}{2}}. \tag{3.18}$$

Since $|\varphi| \le \alpha < \pi$ (cf. Fig. 3.2), we have $\left|\sin\dfrac{\varphi}{2}\right| < \sin\dfrac{\alpha}{2}$. In particular, the inside of the radical symbol in (3.18) is non-negative.

Using this fact, we introduce another variable θ as follows. By setting $k := \sin\dfrac{\alpha}{2}$, $-1 \le k^{-1}\sin\dfrac{\varphi}{2} \le 1$. Hence, we can define a variable θ by $k^{-1}\sin\dfrac{\varphi}{2} = \sin\theta$, or,

$$\begin{aligned}\theta &= \theta(\varphi) := \arcsin\left(k^{-1}\sin\frac{\varphi}{2}\right),\\ \varphi &= \varphi(\theta) := 2\arcsin(k\sin\theta).\end{aligned}$$

The right-hand side of (3.18) becomes

$$2\omega\sqrt{k^2 - k^2\sin^2\theta} = 2k\omega\cos\theta,$$

[3] This is not a trivial assumption. We shall discuss other states of a pendulum in Exercise 3.6.

and the left-hand side is obtained by the chain rule, $\dfrac{d\varphi}{dt} = \dfrac{d\varphi}{d\theta}\dfrac{d\theta}{dt}$. ($\theta(\varphi(t))$ is abbreviated as θ.) To find the derivative $\dfrac{d\varphi}{d\theta}$, we differentiate the definition of θ, $k\sin\theta = \sin\dfrac{\varphi}{2}$, by θ:

$$k\frac{d}{d\theta}\sin\theta = \frac{d}{d\theta}\left(\sin\frac{\varphi}{2}\right),$$

which leads to

$$k\cos\theta = \frac{1}{2}\cos\frac{\varphi}{2}\frac{d\varphi}{d\theta} = \frac{1}{2}\sqrt{1-k^2\sin^2\theta}\frac{d\varphi}{d\theta}.$$

Summarising, the equation (3.18) becomes

$$\frac{2k\cos\theta}{\sqrt{1-k^2\sin^2\theta}}\frac{d\theta}{dt} = 2k\omega\cos\theta,$$

or, cancelling $2k\cos\theta$,

$$\frac{1}{\sqrt{1-k^2\sin^2\theta}}\frac{d\theta}{dt} = \omega.$$

Integrating from $t=0$ to t, we obtain

$$\int_{\theta(\varphi(0))}^{\theta(\varphi(t))}\frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} = \int_0^t \omega\, dt,$$

by changing the integration variable from t to θ in the left-hand side. We may assume $\varphi(0)=0$ by shifting the time variable t for simplicity. Since $\theta(0)=0$, the lower end of the integral in the left-hand side becomes $\theta(\varphi(0))=0$:

$$\int_0^{\theta(\varphi(t))}\frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} = \omega t. \tag{3.19}$$

This is the relation between the time t and the angle φ.

It is time to recall our main theme. Yes, it *is* the elliptic integral. More precisely, the left-hand side of (3.19) is nothing but the incomplete elliptic integral of the first kind $F(k,\theta)$. The time $t(\varphi)$, at which the angle takes the value φ, is expressed as

$$t(\varphi) = \frac{1}{\omega}F(k,\theta(\varphi)) = \sqrt{\frac{l}{g}}F\left(\sin\frac{\alpha}{2},\theta(\varphi)\right). \tag{3.20}$$

The period of the pendulum is the time taken for the bob to complete the motion: 'the lowest position $\longrightarrow$ the rightmost position $\longrightarrow$ the lowest position $\longrightarrow$ the leftmost position $\longrightarrow$ the lowest position'. This is equal to four times the time taken for the bob to move from its lowest to hightest position. When the angle φ attains its maximum α, the variable θ is equal to $\dfrac{\pi}{2}$. Therefore,

$$\begin{aligned} \text{period } T &= 4 \times (\text{time from } \theta = 0 \text{ till } \theta = \frac{\pi}{2}) \\ &= 4\sqrt{\frac{l}{g}} F\left(\sin\frac{\alpha}{2}, \frac{\pi}{2}\right) = 4\sqrt{\frac{l}{g}} K\left(\sin\frac{\alpha}{2}\right). \end{aligned} \tag{3.21}$$

The complete elliptic integral of the first kind $K\left(\sin\frac{\alpha}{2}\right)$ depends on the amplitude α and, consequently, the period of the pendulum depends on the amplitude. Alas! The isochronism does not hold rigorously.

Returning to the theme of the first half of this chapter, let us express the period of the pendulum in terms of the arithmetic-geometric mean. We know that the arithmetic-geometric mean and the complete elliptic integral of the first kind are connected by (3.5). As we want to apply it to $K\left(\sin\frac{\alpha}{2}\right)$ in (3.21), we need the complementary modulus k' corresponding to $k = \sin\frac{\alpha}{2}$. Since k and k' satisfy $k^2 + k'^2 = 1$, we can take $k' = \cos\frac{\alpha}{2}$:

$$K\left(\sin\frac{\alpha}{2}\right) = \frac{\pi}{2}\frac{1}{M\left(1, \cos\frac{\alpha}{2}\right)},$$

from which follows

$$\text{Period } T = 2\pi\sqrt{\frac{l}{g}}\frac{1}{M\left(1, \cos\frac{\alpha}{2}\right)}.$$

The first part of the right-hand side, $2\pi\sqrt{\frac{l}{g}}$, is the period T_0 in (3.15) for a small amplitude. In general, the period of the pendulum can be computed as T_0 with correction by $\dfrac{1}{M\left(1, \cos\frac{\alpha}{2}\right)}$:

$$T = \frac{T_0}{M\left(1, \cos\frac{\alpha}{2}\right)}. \tag{3.22}$$

If you are very good at experiments and can measure the angle α, the period T, the length l and the gravitational acceleration g extremely accurately, you can 'measure' the arithmetic-geometric mean by the following formula!

$$M\left(1, \cos\frac{\alpha}{2}\right) = \frac{T_0}{T}.$$

Well, I do not recommend this method to find the arithmetic-geometric mean, though, as, apart from the problem of measurement error, the original definition (3.4) converges quite fast, as was observed in Example 3.1, and is quite practical.

Anyway, we can conclude the following from (3.22):

- When the amplitude α is close to 0, $\cos\frac{\alpha}{2}$ is almost 1 and the arithmetic-geometric mean $M\left(1, \cos\frac{\alpha}{2}\right)$ is approximated by 1, which means that we have approximate isochronism.
  Put differently, when α tends to 0, the period T converges to T_0, which means $K\left(\sin\frac{\alpha}{2}\right)$ converges to $\frac{\pi}{2}$.
- When $0 < \alpha < \pi$, $0 < \cos\frac{\alpha}{2} < 1$ and the arithmetic-geometric mean of 1 and $\cos\frac{\alpha}{2}$ is less than 1. Hence the period T is larger than T_0.
- When the amplitude gets larger and the pendulum almost stands upside down,[4] α tends to π and $K\left(\sin\frac{\alpha}{2}\right)$ diverges to ∞, which means that the period becomes longer and longer. We shall discuss such behaviour of the elliptic integral in the next chapter.

The expression (3.20) gives a correspondence between the time and the angle, but we have not yet described the angle as a function of the time as we did in (3.14). This is another topic in the next chapter, which leads to the *elliptic function*.

Exercise 3.6 If we swing the pendulum strongly enough, it begins to rotate rather than swing. In this case the angle φ is a monotonically increasing (or decreasing) function of t and the 'period' T is defined as the time taken for the bob to rotate once around the pivot: If $\varphi(t_1) = \varphi_1$, then $\varphi(t_1 + T) = \varphi_1 + 2\pi$ (or $\varphi_1 - 2\pi$).

Express this period using $\tilde{E}$ (equivalently, using the energy of the pendulum $E = ml^2\tilde{E}$) and the elliptic integral. (Hint: Use the modulus $k_0 := \sqrt{\frac{2\omega^2}{\omega^2 + \tilde{E}}}$. In the main text above, the modulus k of the elliptic integral was k_0^{-1}.)

[4] For such an experiment, you have to use a rod, not a cord.

Chapter 4
Jacobi's Elliptic Functions on $\mathbb{R}$

In this chapter we introduce elliptic functions as inverse functions of elliptic integrals.

We use several convergence theorems in real analysis, which we cite in Appendix A.

4.1 Jacobi's Elliptic Functions

In the previous chapter we showed that the motion of a simple pendulum is described by means of the incomplete elliptic integral of the first kind as

$$
\begin{aligned}
t(\varphi) &= \sqrt{\frac{l}{g}} F\left(\sin\frac{\alpha}{2}, \theta(\varphi)\right) \\
&= \sqrt{\frac{l}{g}} \int_0^{\theta(\varphi)} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}.
\end{aligned}
\tag{4.1}
$$

Here t is the time, α is the amplitude, $k = \sin\dfrac{\alpha}{2}$ and the angle φ is related to the variable θ by $\sin\theta = k^{-1}\sin\dfrac{\varphi}{2}$. This formula expresses the time t in terms of the angle φ, which makes it easy to compute the period. But it is not natural as a 'description of motion'.

A more natural description would have the form 'the angle φ = a function of the time t'. Namely, we should consider the inverse function of the above formula.

So, we put aside the motion of a pendulum and consider the inverse function of the incomplete elliptic integral of the first kind here. We assume that the modulus k of the elliptic integrals satisfies the inequality $0 \le k < 1$. For later use we mainly adopt the algebraic expression of elliptic integrals, which does not include the sine, as in the following definition.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_5

Definition 4.1 *Jacobi's elliptic function* $\mathrm{sn}(u) = \mathrm{sn}(u,k)$ is the inverse function of the incomplete elliptic integral of the first kind,

$$u(x) = u(x,k) = \int_0^x \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}, \tag{4.2}$$

which is regarded as a function of x ($x \in [-1,1]$):

$$\mathrm{sn}(u(x,k),k) = x, \qquad u(\mathrm{sn}(u,k),k) = u. \tag{4.3}$$

Since the integrand of $u(x)$, $\dfrac{1}{\sqrt{(1-z^2)(1-k^2z^2)}}$, is always positive on the interval $z \in (-1,1)$ and an even function, the integral $u(x)$ is a continuous monotonically increasing odd function of x. Moreover $u(1) = K(k)$, where $K(k)$ is the complete elliptic integral of the first kind

$$K(k) = \int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} \tag{4.4}$$

and $u(-1) = -K(k)$. (See Fig. 4.1.)

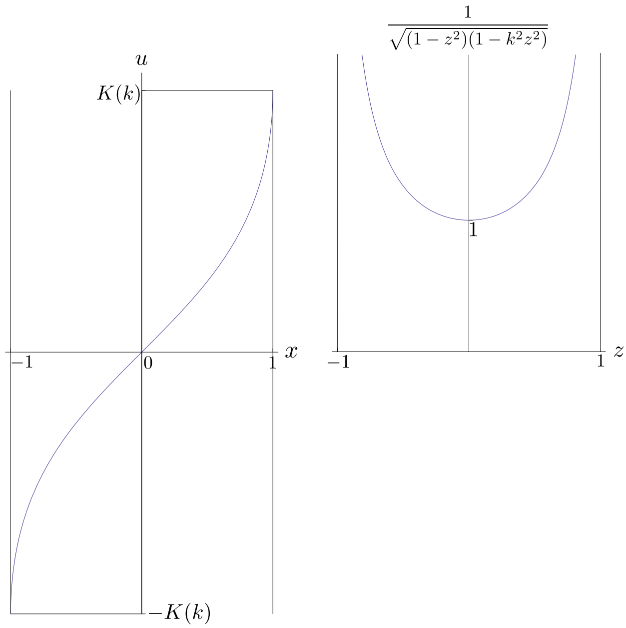

Fig. 4.1 The incomplete elliptic integral of the first kind $u(x)$ and its integrand ($k = 0.8$, $K = K(k)$).

Therefore the inverse function $\operatorname{sn}(u)$ of $u(x)$ is a monotonically increasing odd continuous function defined on $[-K(k), K(k)]$ and takes values

$$\operatorname{sn}(0) = 0, \qquad \operatorname{sn}(\pm K(k)) = \pm 1. \tag{4.5}$$

The graph is as in Fig. 4.2, which is the mirror image of the left graph in Fig. 4.1 with respect to the straight line $u = x$.

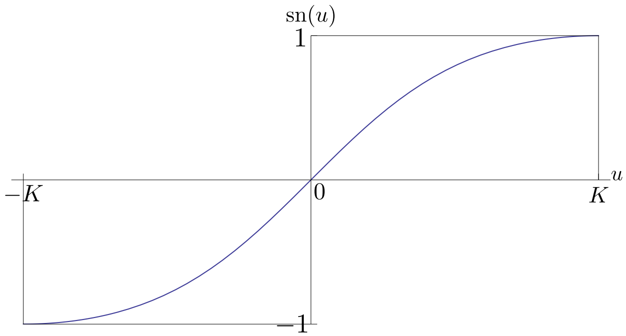

Fig. 4.2 $\operatorname{sn}(u)$ on $[-K(k), K(k)]$ ($k = 0.8$, $K = K(k)$).

Remark 4.2 (On convergence of improper integrals) We first encountered the complete elliptic integral of the first kind in Section 1.2 as the length of the lemniscate. Since it is natural that such a nice curve has a length, we did not emphasise it there, but, as a matter of fact, the integral (4.4) is an improper integral, as the integrand $\dfrac{1}{\sqrt{(1-z^2)(1-k^2z^2)}}$ diverges at the endpoint $z = 1$ of the interval $[0, 1]$ of integration. We can confirm the convergence of this improper integral by the comparison test as follows: first, let us rewrite the integrand of $K(k)$ as

$$\frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}} = \frac{1}{\sqrt{(1+z)(1-k^2z^2)}} \times \frac{1}{\sqrt{1-z}}. \tag{4.6}$$

The only part which diverges in the right-hand side is $\dfrac{1}{\sqrt{1-z}}$. (Recall that we are assuming that $k < 1$.) The rest is estimated by a constant on $[0, 1]$ as

$$\frac{1}{\sqrt{(1+z)(1-k^2z^2)}} \le C_k := \frac{1}{\sqrt{1-k^2}}. \tag{4.7}$$

Applying the inequality (4.7) to (4.6), we have an estimate of the integrand of $K(k)$ from above,

$$\frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}} \leq \frac{C_k}{\sqrt{1-z}}. \tag{4.8}$$

On the other hand, the improper integral

$$\int_0^1 \frac{C_k}{\sqrt{1-z}}\,dz = -2C_k\sqrt{1-z}\Big|_{z=0}^{z=1} = 2C_k$$

is convergent. Therefore the complete elliptic integral of the first kind $K(k) = \int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}$, which is smaller than the above integral, also converges and takes a finite value.

We gave this proof as an example of a proof of convergence of improper integrals. Actually, we can prove the convergence of $K(k)$ more easily, recalling the expression of $K(k)$ which we saw first: by the change of integration variable $z = \sin\varphi$ we have an expression of $K(k)$ with the sine function as (1.22). The integrand of this expression is bounded on $\left[0, \frac{\pi}{2}\right]$, which means it is not an improper integral, and hence trivially converges.

The complete elliptic integral of the second kind is also an improper integral in the expression (1.13), whose convergence can be shown in a similar way.

There is no definite reading of the notation 'sn'. (I usually read it just as 'es-en'.) At first glance this notation might seem stupid, as it is like a misspelled "sin". In fact, as similarity of notations shows, $\mathrm{sn}(u) = \mathrm{sn}(u,k)$ is a cousin of $\sin u$ and, when $k = 0$, they really coincide. Precisely speaking, setting k to 0 in the above definition, we have

$$\begin{aligned}\mathrm{sn}(u,0) &= \text{the inverse function of} \left(u(x) = \int_0^x \frac{dz}{\sqrt{1-z^2}} = \arcsin x\right)\\ &= \sin u.\end{aligned}$$

(Recall that the integral $\int \frac{dz}{\sqrt{1-z^2}}$ appeared in Section 2.1.) Correspondingly, the complete elliptic integral $K(k)$ becomes

$$K(0) = \int_0^1 \frac{dz}{\sqrt{1-z^2}} = \arcsin(1) = \frac{\pi}{2}.$$

In other words, Definition 4.1 for $k = 0$ is equivalent to defining the function $\sin u$ on the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (cf. Fig. 4.3).

Since the integrand $\frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}}$ of $u(x,k)$ is continuous (and, moreover, differentiable) with respect to k, $K(k)$ and $u(x,k)$ are continuous functions of k and, consequently, $\mathrm{sn}(u,k)$ is continuous in k. For example, $\lim_{k\to 0} K(k) = \frac{\pi}{2}$.

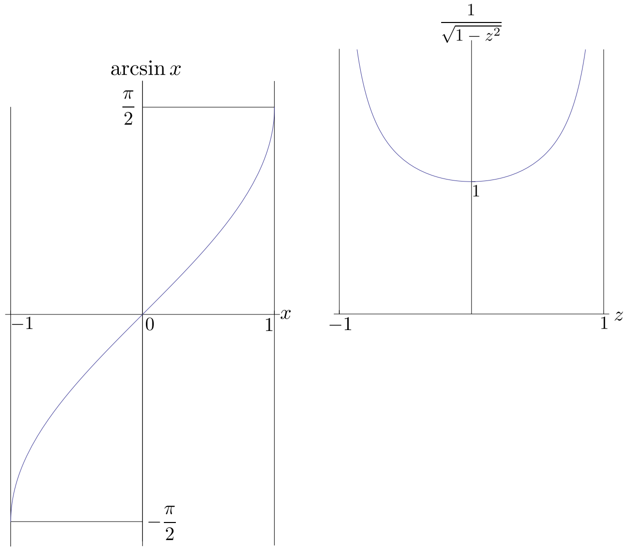

Fig. 4.3 $\arcsin(x)$ and its integrand.

Remark 4.3 (on continuity) In (4.2) the function $u(x,k)$ is defined as a usual definite Riemann integral when $|x|<1$, so it is not difficult to show the continuity of $u(x,k)$ in k. Exactly speaking, it comes from the equality of type 'limit of integral = integral of limit',

$$\lim_{k\to k_0}\int_0^x \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = \int_0^x \frac{dz}{\sqrt{(1-z^2)(1-k_0^2z^2)}},$$

and we need to show the uniform convergence of the integrand. If $0<|z|<|x|<1$, then we have a concrete estimate,

$$\begin{aligned}
&\left|\frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}} - \frac{1}{\sqrt{(1-z^2)(1-k_0^2z^2)}}\right| \\
&= \frac{\left|\sqrt{1-k_0^2z^2}-\sqrt{1-k^2z^2}\right|}{\sqrt{(1-z^2)(1-k^2z^2)(1-k_0^2z^2)}} \\
&= \frac{|k^2-k_0^2||z|^2}{\sqrt{(1-z^2)(1-k^2z^2)(1-k_0^2z^2)}\left(\sqrt{1-k^2z^2}+\sqrt{1-k_0^2z^2}\right)}
\end{aligned}$$

$$< \frac{|k^2 - k_0^2||x|^2}{\sqrt{(1-x^2)(1-k^2x^2)(1-k_0^2x^2)}\left(\sqrt{1-k^2x^2}+\sqrt{1-k_0^2x^2}\right)},$$

from which follows the desired uniform convergence of the integrand when k tends to k_0 by the Weierstrass M-test. Thus we can prove that $u(x,k)$ is continuous in k by Proposition A.1.

In fact, in order to show the continuity of $u(x,k)$ in k, it is simpler to use a theorem of multivariable calculus (Proposition A.2), which states that a definite Riemann integral of a two-variable function in one of the variables gives a continuous function in the remaining variable. (The above estimate shows that the integrand is a continuous function in (z,k).)

As for the continuity of the complete elliptic integral $K(k)$ in k, it is slightly more difficult, because it is defined by an improper integral, as was mentioned in Remark 4.2.

The integrand $\dfrac{1}{\sqrt{(1-z^2)(1-k^2z^2)}}$ is a continuous function of (z,k) on $[0,1)\times[0,1)$ and satisfies the estimate (4.8). If k belongs to a closed interval $[c,d]\subset[0,1)$, the constant C_k is always not less than C_d, $C_k \le C_d$ by the definition of C_k, (4.7). Therefore the integrand of $K(k)$ satisfies the following estimate:

$$\frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}} \le M(z) := \frac{C_d}{\sqrt{1-z}}.$$

Since the improper integral $\displaystyle\int_0^1 M(z)\,dz$ converges, $K(k)$ converges uniformly on $[c,d]$ by Weierstrass's M-test, Proposition A.5. The conditions for Proposition A.3 are fulfilled and $K(k)$ is continuous in k. □

As the function $u(x)$ and the incomplete elliptic integral of the first kind $F(k,\theta)$ are connected by $F(k,\theta)=u(\sin\theta)$, we have

(4.9) $$\operatorname{sn}(F(k,\theta),k)=\sin\theta, \quad F(k,\arcsin(\operatorname{sn}(u,k)))=u.$$

The inverse function $\operatorname{am}(u)=\operatorname{am}(u,k)$ of $F(k,\theta)$ in θ is called *Jacobi's amplitude function*:

$$\operatorname{am}(F(k,\theta),k)=\theta, \qquad F(k,\operatorname{am}(u,k))=u.$$

Using it, we can rewrite the relations (4.9) as

$$\operatorname{sn}u=\sin(\operatorname{am}u), \qquad \operatorname{am}u=\arcsin(\operatorname{sn}u).$$

The graph of $\operatorname{am}(u)$ is shown in Fig. 4.4.

We have defined the elliptic function $\operatorname{sn}u$ only on the interval $[-K(k),K(k)]$. As we mentioned above, $\operatorname{sn}u=\operatorname{sn}(u,k)$ becomes $\sin u$ when $k=0$, and $\sin u$ is defined everywhere on $\mathbb{R}$. So the definition of $\operatorname{sn}u$ should also be extended to $\mathbb{R}$.

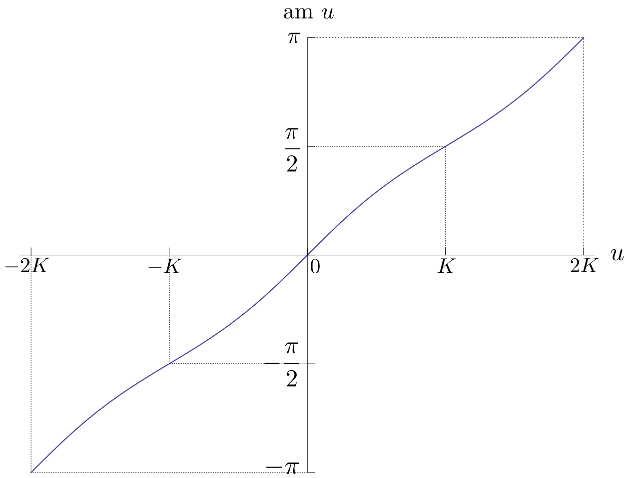

Fig. 4.4 Jacobi's amplitude function $\mathrm{am}(u)$ ($k = 0.8$, $K = K(k)$).

Recall that an important property of the trigonometric function $\sin u$ is its (anti-)periodicity:

$$\sin(u+\pi) = -\sin u, \qquad \sin(u+2\pi) = \sin u.$$

Following this property of $\sin u$, let us extend $\operatorname{sn} u$ to $\mathbb{R}$ by

$$\begin{aligned} \operatorname{sn}(u+2K(k),k) &= -\operatorname{sn}(u,k), \\ \operatorname{sn}(u+4K(k),k) &= \operatorname{sn}(u,k). \end{aligned} \tag{4.10}$$

We shall explain why this is the 'correct' extension later by using the addition theorem (Exercise 4.8 (ii)), and give another explanation after extending the elliptic integrals and the elliptic functions to the complex plane. For now, please be convinced of correctness of the above definition by the graph Fig. 4.5, which is a natural extension of Fig. 4.2.

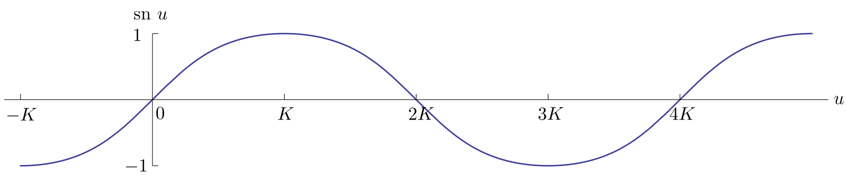

Fig. 4.5 $\operatorname{sn}(u)$ on $\mathbb{R}$ ($k = 0.8$, $K = K(k)$).

4.2 Properties of Jacobi's Elliptic Functions

4.2.1 Troika of Jacobi's elliptic functions

The trigonometric function sine has a partner, cosine. Jacobi's elliptic function sn has two partners, cn (a cousin of cos) and dn. Using sn, cn and dn together, we can formulate various properties of them as in the case of the trigonometric functions.

First let us define cn and dn on the interval $-K(k) \le u \le K(k)$ by

$$\begin{aligned} \operatorname{cn}(u) = \operatorname{cn}(u,k) &:= \sqrt{1-\operatorname{sn}^2(u,k)}, \\ \operatorname{dn}(u) = \operatorname{dn}(u,k) &:= \sqrt{1-k^2\operatorname{sn}^2(u,k)}. \end{aligned} \tag{4.11}$$

Here the signs of roots are taken so that $\operatorname{cn}(u) \ge 0$ and $\operatorname{dn}(u) \ge 0$ in this region. In particular $\operatorname{cn}(0) = \operatorname{dn}(0) = 1$. On $\mathbb{R}$ we extend them by periodicity as in the case of $\operatorname{sn} u$ $(K = K(k))$:

$$\operatorname{cn}(u+2K) = -\operatorname{cn}(u), \qquad \operatorname{cn}(u+4K) = \operatorname{cn}(u), \tag{4.12}$$

$$\operatorname{dn}(u+2K) = \operatorname{dn}(u). \tag{4.13}$$

They are also called *Jacobi's elliptic functions*. Their graphs are shown in Fig. 4.6.

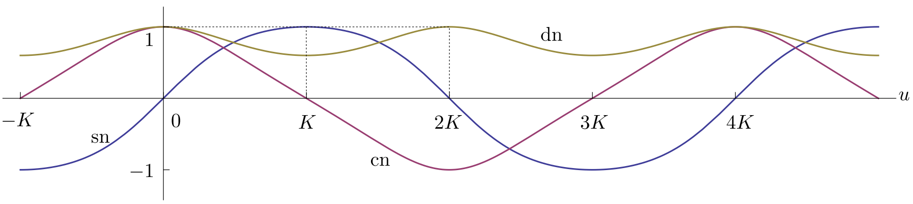

Fig. 4.6 $\operatorname{sn}(u)$, $\operatorname{cn}(u)$ and $\operatorname{dn}(u)$ on $\mathbb{R}$ $(k = 0.8,\ K = K(k))$.

Since sn takes values as in (4.5), cn and dn take values as follows, correspondingly:

$$\begin{aligned} \operatorname{cn}(0) &= 1, \quad \operatorname{cn}(\pm K) = 0, \\ \operatorname{dn}(0) &= 1, \quad \operatorname{dn}(\pm K) = \sqrt{1-k^2}. \end{aligned} \tag{4.14}$$

The function sn being odd, cn and dn defined by (4.11) are even functions.

At $k = 0$ $\operatorname{sn}(u,k)$ becomes $\sin u$. Correspondingly $\operatorname{cn}(u,k)$ becomes $\cos u$ and $\operatorname{dn}(u,k)$ becomes the constant function 1, which immediately follows from the definitions (4.11).

Exercise 4.4 When the modulus k tends to 1, the limits are expressed in terms of hyperbolic functions:

$$\mathrm{sn}(u,k) \to \tanh u := \frac{\sinh u}{\cosh u},$$
$$\mathrm{cn}(u,k), \mathrm{dn}(u,k) \to \frac{1}{\cosh u}.$$

Show these formulae. (Hint: Show that the limit $k \to 1$ of the incomplete elliptic integral of the first kind $u(x)$ defined by (4.2) is artanh, the inverse function to tanh.[1] One has to change the order of the limit and integrals, the proof of which may be omitted.)

The graphs of sn, cn and dn for $k = 0, 0.5, 0.8$ and 1 are shown in Fig. 4.7, Fig. 4.8 and Fig. 4.9.

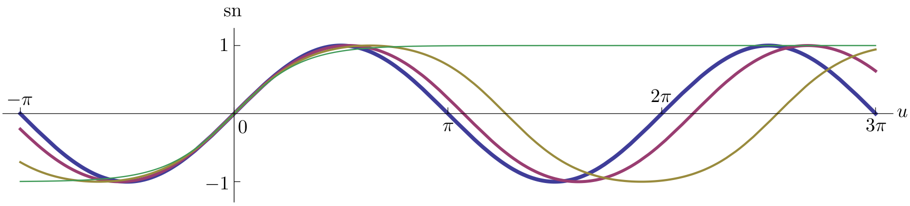

Fig. 4.7 $\mathrm{sn}(u,k)$ for $k = 0$ ($\sin u$; the thickest curve), $k = 0.5$, $k = 0.8$ and $k = 1$ ($\tanh u$; the thinnest curve).

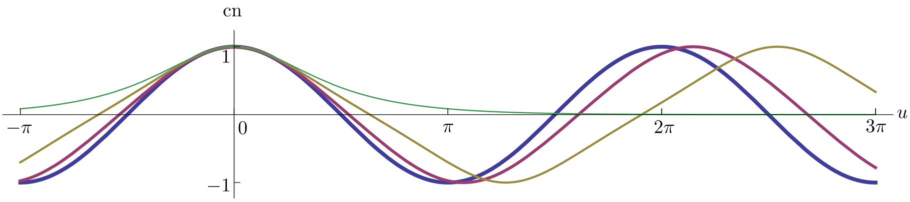

Fig. 4.8 $\mathrm{cn}(u,k)$ for $k = 0$ ($\cos u$; the thickest curve), $k = 0.5$, $k = 0.8$ and $k = 1$ ($\frac{1}{\cosh u}$; the thinnest curve).

Remark 4.5 (on the limit of $K(k)$ when $k \to 1$) As is seen from Exercise 4.4 and Fig. 4.7, when k tends to 1, the period of sn gets larger and finally diverges:

[1] The symbol 'artanh' might scare you, but don't be afraid. It is an elementary function, since, by solving the equation $x = \tanh u = \dfrac{e^u - e^{-u}}{e^u + e^{-u}}$ in u, we have $u = \operatorname{artanh} x = \dfrac{1}{2} \log\left(\dfrac{1+x}{1-x}\right)$.

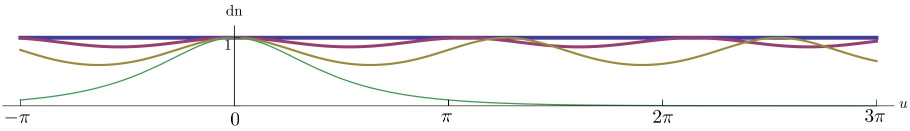

Fig. 4.9 $\mathrm{dn}(u,k)$ for $k=0$ (1; the thickest line), $k=0.5$, $k=0.8$ and $k=1$ ($\dfrac{1}{\cosh u}$; the thinnest curve).

$\lim_{k\to 1} K(k) = \infty$. This can be shown by putting $\lim_{k\to 1}$ inside the integral in the definition of $K(k)$. For example, using the improper integral (4.4), we can show the divergence as follows.

We have an inequality

$$K(k) > \int_0^{1-\delta'} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} \tag{4.15}$$

for any $\delta' \in (0,1)$, because the integrand of the integral (4.4) is positive. The right-hand side is not an improper integral and thus when k approaches to 1, the integral approaches $\displaystyle\int_0^{1-\delta'} \frac{dz}{1-z^2} = \operatorname{artanh}(1-\delta')$. Since $\operatorname{artanh}(1)=\infty$, the above inequality means that $K(k)$ grows unboundedly, when k tends to 1. Thus we have $\lim_{k\to 1} K(k) = \infty$.

If you want to use an ε-δ type argument, you need to show that for any $M>0$ there exists $\delta>0$ such that for any k satisfying $1-\delta<k<1$, $K(k)$ is larger than M: $K(k)>M$.

Since artanh is an increasing function and $\lim_{x\to 1} \operatorname{artanh} x = \infty$, there exists a $\delta'>0$ such that $\operatorname{artanh}(1-\delta') > M+1$. As is shown above,

$$\lim_{k\to 1} \int_0^{1-\delta'} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = \operatorname{artanh}(1-\delta'),$$

from which it follows that there exists a $\delta>0$ such that for any $k\in(1-\delta,1)$

$$\left| \int_0^{1-\delta'} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} - \operatorname{artanh}(1-\delta') \right| < 1.$$

Combining this estimate with the inequality $\operatorname{artanh}(1-\delta') > M+1$, we obtain

$$\int_0^{1-\delta'} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} > M.$$

This together with (4.15) shows $K(k)>M$. □

4.2.2 Derivatives

As sn is defined by the elliptic *integral*, it is easy to find its derivative. Recall that the inverse function of $\operatorname{sn} u$ is

$$u(x) = \int_0^x \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}.$$

Hence, by the fundamental theorem of calculus, in other words, because integration and derivation are inverse to each other, we have

$$\begin{aligned}
\frac{d}{du}\operatorname{sn} u &= \frac{1}{\left.\dfrac{d}{dx}u(x)\right|_{x=\operatorname{sn} u}} = \left(\left.\frac{1}{\sqrt{(1-x^2)(1-k^2x^2)}}\right|_{x=\operatorname{sn} u} \right)^{-1} \\
&= \sqrt{1-\operatorname{sn}^2 u}\sqrt{1-k^2\operatorname{sn}^2 u} \\
&= \operatorname{cn} u \operatorname{dn} u.
\end{aligned}$$

From this it follows directly by the chain rule that

$$\begin{aligned}
\frac{d}{du}\operatorname{cn} u &= \frac{d}{du}\sqrt{1-\operatorname{sn}^2 u} = \frac{-2\operatorname{sn} u \dfrac{d\operatorname{sn} u}{du}}{2\sqrt{1-\operatorname{sn}^2 u}} \\
&= -\operatorname{sn} u \operatorname{dn} u, \\
\frac{d}{du}\operatorname{dn} u &= \frac{d}{du}\sqrt{1-k^2\operatorname{sn}^2 u} = \frac{-2k^2\operatorname{sn} u \dfrac{d\operatorname{sn} u}{du}}{2\sqrt{1-k^2\operatorname{sn}^2 u}} \\
&= -k^2 \operatorname{sn} u \operatorname{cn} u.
\end{aligned}$$

(Exactly speaking, the above computation is true on the open interval $(-K(k), K(k))$, and we have to extend this result to $\mathbb{R}$ by using periodicity (4.10), (4.12), (4.13).)

Summarising, we have

$$\begin{aligned}
\frac{d\operatorname{sn} u}{du} &= \operatorname{cn} u \operatorname{dn} u, \\
\frac{d\operatorname{cn} u}{du} &= -\operatorname{sn} u \operatorname{dn} u, \\
\frac{d\operatorname{dn} u}{du} &= -k^2\operatorname{sn} u \operatorname{cn} u,
\end{aligned} \tag{4.16}$$

which is consistent with the limits $k \longrightarrow 0$:

$$\frac{d\sin u}{du} = \cos u, \quad \frac{d\cos u}{du} = -\sin u.$$

4.2.3 Addition formulae

One of the important properties of trigonometric functions is the addition formula. For example, the addition formula for the sine function is

$$\sin(u+v) = \sin u \cos v + \cos u \sin v. \tag{4.17}$$

The hyperbolic function also has addition formulae. Recall that hyperbolic functions are defined by the exponential function as

$$\sinh u = \frac{e^u - e^{-u}}{2}, \qquad \cosh u = \frac{e^u + e^{-u}}{2}.$$

It is easy to show by these definitions the addition formulae of sinh and cosh,

$$\begin{aligned} \sinh(u+v) &= \sinh u \cosh v + \sinh v \cosh u, \\ \cosh(u+v) &= \cosh u \cosh v + \sinh v \sinh u, \end{aligned}$$

from which the addition formula of $\tanh u = \dfrac{\sinh u}{\cosh u}$ follows:

$$\tanh(u+v) = \frac{\tanh u + \tanh v}{1 + \tanh u \tanh v}. \tag{4.18}$$

As we have mentioned in Section 4.1, when $k = 0$, sn becomes sin, while sn converges to tanh, when k tends to 1. Therefore, it is natural to guess, "Doesn't $\mathrm{sn}(u,k)$ have an addition formula like $\sin u$ and $\tanh u$?" Let us consider how we can interpolate the two addition formulae (4.17) and (4.18).

These two addition formulae have quite different expressions, so it seems difficult to 'interpolate' them, but by rewriting them as follows[2] we can find a natural interpolation. First, let us rewrite the addition formula (4.17) of sine without using cosine:

$$\sin(u+v) = \sin u \frac{d \sin v}{dv} + \frac{d \sin u}{du} \sin v. \tag{4.19}$$

If we substitute the sine functions in the right-hand side by tanh, we have

$$\tanh u \frac{d \tanh v}{dv} + \frac{d \tanh u}{du} \tanh v = (\tanh u + \tanh v)(1 - \tanh u \tanh v)$$

as $\dfrac{d \tanh u}{du} = 1 - \tanh^2 u$. By this formula, we can rewrite the addition formula (4.18) of tanh as

[2] We follow the discussion in Ch. 8 of [To2].

$$\tanh(u+v) = \frac{\tanh u \dfrac{d\tanh v}{dv} + \dfrac{d\tanh u}{du}\tanh v}{1-\tanh^2 u \tanh^2 v}. \tag{4.20}$$

Comparing two formulae (4.19) and (4.20), we can infer that the equation

$$\begin{aligned}\operatorname{sn}(u+v) &= \frac{\operatorname{sn} u \dfrac{d\operatorname{sn} v}{dv} + \dfrac{d\operatorname{sn} u}{du}\operatorname{sn} v}{1-k^2\operatorname{sn}^2 u\operatorname{sn}^2 v} \\ &= \frac{\operatorname{sn} u\operatorname{cn} v\operatorname{dn} v + \operatorname{sn} v\operatorname{cn} u\operatorname{dn} u}{1-k^2\operatorname{sn}^2 u\operatorname{sn}^2 v},\end{aligned} \tag{4.21}$$

which interpolates (4.19) and (4.20), would be the addition formula of sn. *Indeed it is.*[3]

How can we prove this? Let us substitute $u+v$ in (4.21) by a constant c, replace v by $c-u$ and define a function of u by the right-hand side of (4.21):

$$F(u) := \frac{\operatorname{sn} u\operatorname{cn}(c-u)\operatorname{dn}(c-u) + \operatorname{sn}(c-u)\operatorname{cn} u\operatorname{dn} u}{1-k^2\operatorname{sn}^2 u\operatorname{sn}^2(c-u)}. \tag{4.22}$$

Claim: $F(u)$ is constant as a function of u: $\dfrac{dF}{du} = 0$.

According to this claim, $F(u) = F(0)$ for any u. Since $\operatorname{sn} 0 = 0$ and $\operatorname{cn} 0 = \operatorname{dn} 0 = 1$, it is easy to see that $F(0) = \operatorname{sn} c$, that is to say,

$$\operatorname{sn} c = \frac{\operatorname{sn} u\operatorname{cn}(c-u)\operatorname{dn}(c-u) + \operatorname{sn}(c-u)\operatorname{cn} u\operatorname{dn} u}{1-k^2\operatorname{sn}^2 u\operatorname{sn}^2(c-u)}.$$

By resetting $c = u+v$ we obtain the addition formula (4.21).

The proof of the claim $\dfrac{dF}{du} = 0$ goes as follows. Let us denote the numerator and the denominator in the definition of the function $F(u)$ by N and D respectively:

$$\begin{aligned} N &:= \text{the numerator of } F(u) \\ &= \operatorname{sn} u\operatorname{cn}(c-u)\operatorname{dn}(c-u) + \operatorname{sn}(c-u)\operatorname{cn} u\operatorname{dn} u \\ D &:= \text{the denominator of } F(u) \\ &= 1-k^2\operatorname{sn}^2 u\operatorname{sn}^2(c-u). \end{aligned}$$

As

[3] Of course, there are many other possible interpolations of (4.19) and (4.20). (For example, expressions with k and k^3 instead of k^2 in the denominator also interpolate them.) The candidate (4.21) above is proposed because we know the correct answer. To discover it without any background knowledge, you would have to make many trials and errors.

$$\frac{dF}{du} = \frac{\dfrac{dN}{du} D - N \dfrac{dD}{du}}{D^2},$$

it is sufficient to show that the denominator in the right-hand side vanishes. So the goal is

$$\frac{dN}{du} D = N \frac{dD}{du}. \tag{4.23}$$

This is an easy (but lengthy) computation with the help of formulae (4.16). Please try it yourself.

Exercise 4.6 Prove (4.23).

The addition formulae of cn and dn,

$$\operatorname{cn}(u+v) = \frac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{sn} v \operatorname{dn} u \operatorname{dn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}, \tag{4.24}$$

$$\operatorname{dn}(u+v) = \frac{\operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} u \operatorname{cn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v} \tag{4.25}$$

follow from the addition formula (4.21) of sn and Definition 4.11. Their proofs are also simple computations, so we leave them to the reader as an exercise. Notice that, when $k = 0$, the addition formula (4.24) of cn reduces to that of the cosine.

Exercise 4.7 Prove the addition formulae (4.24) and (4.25). (Hint: The computation in the proof of (4.24) would be simplified by rewriting the denominator in the right-hand side as $\operatorname{cn}^2 u + \operatorname{sn}^2 u \operatorname{dn}^2 v$ or $\operatorname{cn}^2 v + \operatorname{sn}^2 v \operatorname{dn}^2 u$. In the case of dn, use (the denominator) $= \operatorname{dn}^2 u + k^2 \operatorname{sn}^2 u \operatorname{cn}^2 v = \operatorname{dn}^2 v + k^2 \operatorname{sn}^2 v \operatorname{cn}^2 u$ instead.)

Exercise 4.8 (i) Using the addition formulae and (4.5), (4.14), prove

$$\begin{gathered} \operatorname{sn}(u+K) = \frac{\operatorname{cn} u}{\operatorname{dn} u}, \quad \operatorname{cn}(u+K) = -k' \frac{\operatorname{sn} u}{\operatorname{dn} u}, \\ \operatorname{dn}(u+K) = \frac{k'}{\operatorname{dn} u}. \end{gathered} \tag{4.26}$$

($K = K(k)$, $k' := \sqrt{1-k^2}$.) Further prove

$$\operatorname{sn}(2K) = 0, \quad \operatorname{cn}(2K) = -1, \quad \operatorname{dn}(2K) = 1, \tag{4.27}$$

by the formulae (4.26).

(ii) Show that the periodicities (4.10), (4.12) and (4.13) are consistent with the results of (i) and the addition formulae.

The results of Exercise 4.8 (i) for sn and cn correspond to the following formulae for the trigonometric functions ($k = 0$).

$$\sin\left(u+\frac{\pi}{2}\right)=\cos u, \qquad \cos\left(u+\frac{\pi}{2}\right)=-\sin u,$$
$$\sin\pi=0, \qquad \cos\pi=-1.$$

In the next chapter, we apply Jacobi's elliptic functions introduced in this chapter to problems in physics. One of the problems is the description of motion of a pendulum, which we mentioned at the beginning of this chapter. The other topic in the next chapter shows that we are already familiar with the graph of sn from our childhood; it is the shape taken by a skipping rope in motion.

Chapter 5
Applications of Jacobi's Elliptic Functions

In the previous chapter we defined Jacobi's elliptic function sn as the inverse function of the incomplete elliptic integral of the first kind, introduced cn and dn and studied their properties. These Jacobi's elliptic functions appear in various problems, from which we pick up two applications to physics in this chapter.

5.1 Motion of a Simple Pendulum

In Section 3.2 we solved the equation of motion of a simple pendulum,

$$\frac{d^2\varphi}{dt^2} = -\omega^2 \sin\varphi, \qquad \omega := \sqrt{\frac{g}{l}}. \tag{5.1}$$

Here $\varphi = \varphi(t)$ is the angle from vertical at time t, l is the length of the pendulum and g is the gravitational acceleration.

If the amplitude (the maximal angle) is equal to α (Fig. 5.1), the time $t = t(\varphi)$ when the angle becomes φ is expressed in terms of the elliptic integral of the first kind,

$$\omega\, t(\varphi) = \int_0^{k^{-1}\sin\frac{\varphi}{2}} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}, \tag{5.2}$$

where the modulus k is equal to $\sin\dfrac{\alpha}{2}$. In Section 3.2 we calculated the period of the pendulum by this formula.

However, when we describe the motion of the pendulum, it is natural to express the angle φ as a function of the time t. For that purpose we can use the inverse function of the elliptic integral, Jacobi's elliptic function sn. The formula (5.2) means

$$\mathrm{sn}(\omega t, k) = \frac{1}{k}\sin\frac{\varphi}{2},$$

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_6

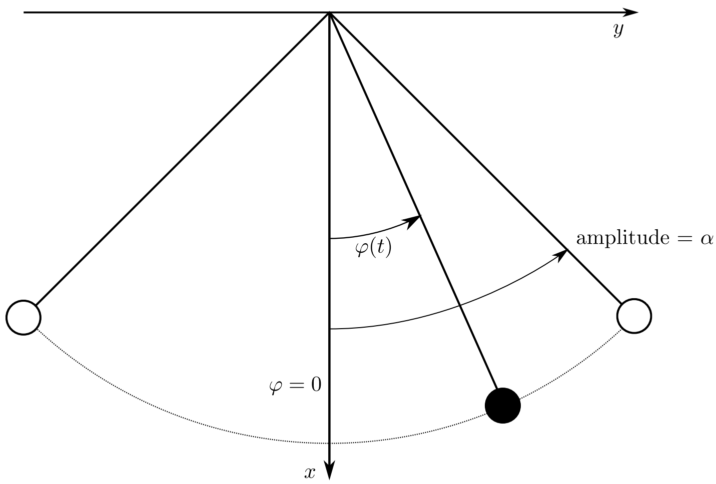

Fig. 5.1 A simple pendulum.

from which it follows that the angle at time t is expressed as

$$\varphi(t) = 2\arcsin\bigl(k\,\mathrm{sn}(\omega t, k)\bigr). \tag{5.3}$$

The verification that the function (5.3) is a solution of the equation (5.1) is a good exercise in computing derivatives of the elliptic functions. The first derivative of $\varphi(t)$ is

$$\frac{d}{dt}\varphi(t) = \frac{2}{\sqrt{1-k^2\,\mathrm{sn}^2(\omega t)}}\,\frac{d}{dt}k\,\mathrm{sn}(\omega t),$$

because of the formula $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$ and the chain rule. (Here we omit k in $\mathrm{sn}(u,k)$.) The denominator in the right-hand side is equal to $\mathrm{dn}(\omega t)$ by the definition of dn, $\mathrm{dn}(u) = \sqrt{1-k^2\,\mathrm{sn}^2(u)}$. Using the formula $(\mathrm{sn}\,u)' = \mathrm{cn}\,u\;\mathrm{dn}\,u$, we have

$$\begin{aligned}\frac{d}{dt}\varphi(t) &= \frac{2}{\mathrm{dn}(\omega t)}\,k\omega\;\mathrm{cn}(\omega t)\;\mathrm{dn}(\omega t)\\ &= 2k\omega\;\mathrm{cn}(\omega t).\end{aligned} \tag{5.4}$$

Differentiating (5.4) once more and using $(\mathrm{cn}\,u)' = -\,\mathrm{sn}\,u\;\mathrm{dn}\,u$, we have

$$\frac{d^2}{dt^2}\varphi(t) = -2k\omega^2\;\mathrm{sn}(\omega t)\;\mathrm{dn}(\omega t). \tag{5.5}$$

On the other hand, $k\;\mathrm{sn}(\omega t) = \sin\dfrac{\varphi}{2}$ because of (5.3). Therefore,

$$\operatorname{dn}(\omega t) = \sqrt{1 - k^2 \operatorname{sn}^2(\omega t)}$$
$$= \sqrt{1 - \sin^2 \frac{\varphi}{2}} = \cos \frac{\varphi}{2}.$$

Substituting this into (5.5), we obtain

$$\frac{d^2}{dt^2}\varphi(t) = -2\omega^2 \sin \frac{\varphi}{2} \cos \frac{\varphi}{2} = -\omega^2 \sin \varphi,$$

which confirms the equation of motion, (5.1).

When the amplitude α is small, the modulus $k = \sin \frac{\alpha}{2}$ is close to 0. As was explained in Section 4.1, $\operatorname{sn}(u,k)$ is approximated by $\sin u$ when the modulus k is close to 0. The modulus $k = \sin \frac{\alpha}{2}$ itself is approximated by $\frac{\alpha}{2}$. On the other hand, $\arcsin x$ is approximated by x, when x is small.[1] Thus we obtain

$$\varphi(t) \approx \alpha \sin(\omega t)$$

from (5.3), which is nothing but the formula for the angle of a pendulum with small amplitude. We learned it in high school physics or first year college physics.

On the contrary, when the amplitude gets larger, the period becomes longer, because the period of sn becomes larger as the modulus k grows. As we saw in Section 4.1, the complete elliptic integral $K(k)$ diverges in the limit $k \to 1$. Correspondingly the period of sn ($= 4K(k)$) diverges and sn becomes the hyperbolic tangent tanh. The modulus $k = \sin \frac{\alpha}{2}$ approaches 1, when $\frac{\alpha}{2}$ approaches $\frac{\pi}{2} = 90°$, namely, α comes close to $\pi = 180°$, which means that the pendulum is almost upside-down. (For this experiment we should use a rod like a pendulum for a clock instead of a cord.)

In the extreme case $\alpha = 180°$ the bob stays still (though unstable) at the top of the circle drawn by itself. We can interpret this situation as the case when the period is infinity.

If α becomes even larger, the pendulum does not swing but rotates. This is the situation which you considered in Exercise 3.6.

5.2 The Shape of a Skipping Rope

Let us recall the shape of the graph of function $\operatorname{sn}(u,k)$, Fig. 5.2.

The goal of this section is to show that, under certain reasonable conditions, the shape of a skipping rope is the arc between 0 and $2K$ in this graph. First we derive an ordinary differential equation for the shape of the skipping rope and then solve

[1] This is a consequence of the approximation $\sin y \approx y$.

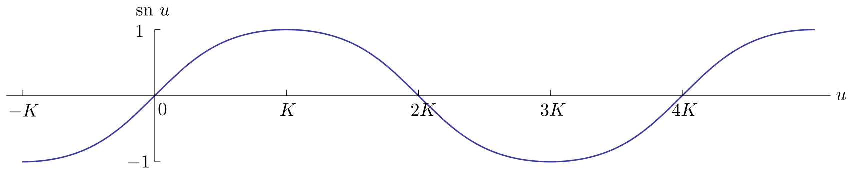

Fig. 5.2 Jacobi's sn ($k = 0.8$, $K = K(k)$).

it. Later we add another derivation of the differential equation by the variational method.

5.2.1 Derivation of the differential equation

We assume the following geometric conditions:

- The ends of the rope are fixed at points $(X,Y,Z) = (0,0,0)$ and $(2a,0,0)$ ($a > 0$) in the space $\mathbb{R}^3$.
- The whole rope always lies in a plane containing the X-axis, which rotates around the X-axis with angular velocity ω. (You can rotate a long rope for group skipping to make a spiral, but here we do not consider such shapes.[2])

We shall impose two more physical conditions later, but for the moment we assume the above two conditions only. When the rope is in the XZ-plane, we have a curve like in Fig. 5.3. We want to determine the shape of this curve.

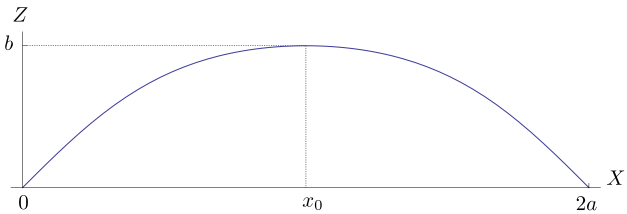

Fig. 5.3 A skipping rope.

As we are assuming that the rope is in one plane, the curve is expressed as the graph of a function $y = y(x)$. Here we should note that this system of xy-coordinates is rotating together with the rope; the x-axis is the same as the X-axis, but the y-axis is rotating around the x-axis. If we add the z-axis perpendicular to the xy-plane, the situation is shown in Fig. 5.4.

[2] See [To2], Chapter 4 for general cases.

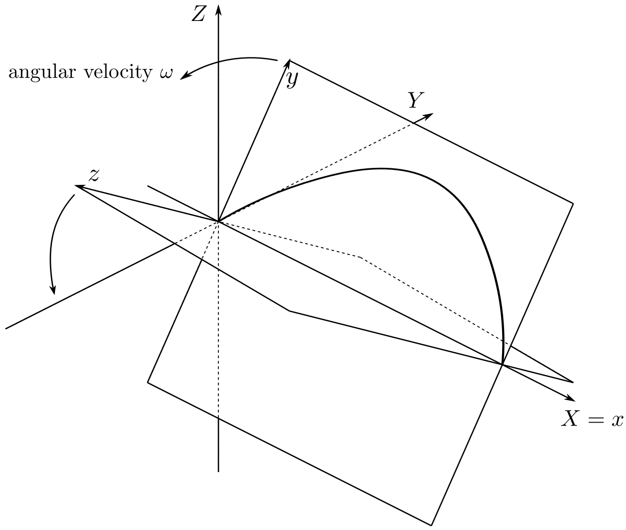

Fig. 5.4 Rotating coordinate system xyz.

The yz-coordinates are obtained by rotating the YZ-coordinates by angle ωt. Hence, the relation between the two coordinate systems is as follows:

$$\begin{aligned} X &= x, \\ Y &= y\cos\omega t - z\sin\omega t, \\ Z &= y\sin\omega t + z\cos\omega t. \end{aligned} \tag{5.6}$$

It is important that the Y- and the Z-coordinates of a point fixed to the xyz-coordinate system depend on time t.

Newton's well-known second law of motion '(mass) × (acceleration) = (force)' holds for special coordinate systems called the *inertial frames of reference* or the *inertial systems*. In a coordinate system moving with acceleration (= moving with *non-constant* velocity) with respect to an inertial system the second law of motion needs modification.

In our case the XYZ-coordinates are fixed to the ground and can be regarded as an inertial system.[3] In this coordinate system the equation of motion of a point mass is

$$m\frac{d^2}{dt^2}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} F_X \\ F_Y \\ F_Z \end{pmatrix}, \tag{5.7}$$

[3] When we consider the motion of a skipping rope, the acceleration by motion of the Earth (revolution around its own axis and around the sun) is negligible. If the acceleration of the Earth is not negligible (for example, the Foucault pendulum), even the Earth-bound system of coordinates cannot be regarded as an inertial system.

where m is the mass, $(X,Y,Z) = (X(t),Y(t),Z(t))$ is the position of the point mass at time t, and F_X, F_Y and F_Z are the X-, Y- and Z-components of the force.

We consider the skipping rope to consist of very small parts and that each part is moving according to the above equation of motion as a point mass. The force $\begin{pmatrix} F_X \\ F_Y \\ F_Z \end{pmatrix}$ working on that part is a composite of gravitation and tension from neighbouring parts.

The equation (5.7) is simple, but we want to find the function $y = y(x)$, not $X(t)$, $Y(t)$, $Z(t)$. Therefore we have to rewrite the equation in the xyz-coordinate system, using the transformation rules (5.6). The x-coordinate is simply the same as the X-coordinate:

$$m\frac{d^2x}{dt^2} = F_X.$$

For the y- and the z-coordinates we need some computation. By differentiating (5.6) twice, we have

$$\begin{aligned}
\frac{d^2Y}{dt^2} &= \frac{d^2y}{dt^2}\cos\omega t - 2\omega\frac{dy}{dt}\sin\omega t - \omega^2 y\cos\omega t \\
&\quad - \frac{d^2z}{dt^2}\sin\omega t - 2\omega\frac{dz}{dt}\cos\omega t + \omega^2 z\sin\omega t, \\
\frac{d^2Z}{dt^2} &= \frac{d^2y}{dt^2}\sin\omega t + 2\omega\frac{dy}{dt}\cos\omega t - \omega^2 y\sin\omega t \\
&\quad + \frac{d^2z}{dt^2}\cos\omega t - 2\omega\frac{dz}{dt}\sin\omega t - \omega^2 z\cos\omega t.
\end{aligned} \tag{5.8}$$

These expressions are large, but we obtain a simpler expression by multiplying $\frac{d^2Y}{dt^2}$ by $m\cos\omega t$, $\frac{d^2Z}{dt^2}$ by $m\sin\omega t$, summing them up and substituting (5.7):

$$F_Y\cos\omega t + F_Z\sin\omega t = m\frac{d^2y}{dt^2} - m\omega^2 y - 2m\omega\frac{dz}{dt}. \tag{5.9}$$

Similarly, multiplying $\frac{d^2Y}{dt^2}$ in (5.8) by $-m\sin\omega t$, $\frac{d^2Z}{dt^2}$ by $m\cos\omega t$, summing them up and substituting (5.7), we obtain

$$-F_Y\sin\omega t + F_Z\cos\omega t = m\frac{d^2z}{dt^2} - m\omega^2 z + 2m\omega\frac{dy}{dt}. \tag{5.10}$$

We shall denote the left-hand side of (5.9) by F_y and the left-hand side of (5.10) by F_z. These are composite forces of gravitation and tension expressed in the xyz-coordinates. Thus we obtain the equations of motion in the xyz-coordinates as follows:

$$m\frac{d^2y}{dt^2} = F_y + m\omega^2 y + 2m\omega\frac{dz}{dt}, \tag{5.11}$$

$$m\frac{d^2z}{dt^2} = F_z + m\omega^2 z - 2m\omega\frac{dy}{dt}. \tag{5.12}$$

We know that F_y and F_z are composite forces of gravitation and tension, but there are additional terms, $m\omega^2 y + 2m\omega\frac{dz}{dt}$ and $m\omega^2 z - 2m\omega\frac{dy}{dt}$, in the right-hand sides, which were absent in the XYZ-coordinates. They do not have a corresponding force like F_y or F_z, but, if a person who knows only the inertial system were to enter into this xyz-coordinate system, then this person would feel an "(apparent) force" corresponding to these additional terms. The part

$$\boldsymbol{F}\mathrm{c} = \begin{pmatrix} F_{\mathrm{c},y} \\ F_{\mathrm{c},z} \end{pmatrix} := \begin{pmatrix} m\omega^2 y \\ m\omega^2 z \end{pmatrix} \tag{5.13}$$

is called the *centrifugal force* and the part

$$\boldsymbol{F}_{\mathrm{Cor}} := \begin{pmatrix} 2m\omega\dfrac{dz}{dt} \\ -2m\omega\dfrac{dy}{dt} \end{pmatrix} \tag{5.14}$$

is called the *Coriolis force*.[4]

We use the xyz-coordinate system, in which the skipping rope stands still. Hence $\frac{dz}{dt}$ and $\frac{dy}{dt}$ vanish and the Coriolis force does not appear in our problem.

Now we add physical conditions:

- The linear density ρ (= mass per unit length) is constant, the rope does not expand or contract and has length l.
- The rope is rotating so fast that the gravitation is negligible compared to the centrifugal force.

Under these assumptions, we consider the equation of motion of the part of the rope on the interval $[x, x+\Delta x]$ of the x-axis (Fig. 5.5).

The length Δx of the interval is so small that we can approximate this part by a point mass. The z-coordinate being always zero, the only equation which we have to solve is equation (5.11) for the y-coordinate.

The ends of the part of the rope are at (x, y), $(x+\Delta x, y(x+\Delta x))$. By the definition of the derivative we have

$$\Delta y := y(x+\Delta x) - y(x) = y'(x)\,\Delta x + o(\Delta x), \qquad y' = \frac{dy}{dx}. \tag{5.15}$$

[4] Apparent force is also generated in a coordinate system in general acceleration motion, not limited to rotation. Such a "force" is called an *inertial force* in general.

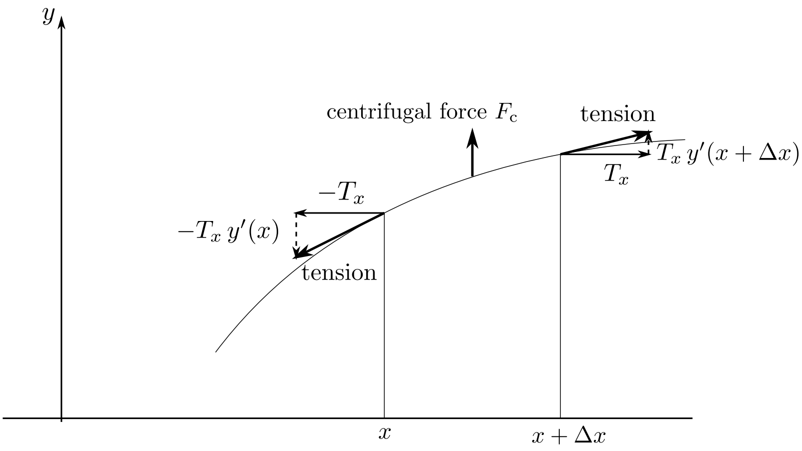

Fig. 5.5 Forces acting on a small part of the rope.

(For those who are not familiar with the symbol o: $o(\Delta x)$ stands for any quantity which tends to 0 faster than Δx when $\Delta x \to 0$: $h(x) = o(\Delta x) \Leftrightarrow \lim_{\Delta x \to 0} \dfrac{h(x)}{\Delta x} = 0$. The above equation is equivalent to the usual definition of the derivative $y'(x) = \lim_{\Delta x \to 0} \dfrac{y(x+\Delta x) - y(x)}{\Delta x}$.)

The part of the rope can be regarded as a linear segment, which has length

$$\sqrt{\Delta x^2 + \Delta y^2} = \sqrt{1 + y'^2}\,\Delta x + o(\Delta x). \tag{5.16}$$

Hence its mass is

$$m = \rho \sqrt{1 + y'^2}\,\Delta x + o(\Delta x). \tag{5.17}$$

The centrifugal force acting on this part can be found easily. By the formula (5.13) the centrifugal force is $m\omega^2 y$ in the y-direction,[5] which gives

$$F_c(x) = \rho \omega^2 y \sqrt{1 + y'^2}\,\Delta x + o(\Delta x) \tag{5.18}$$

together with the formula (5.17).

Next, we compute the tension. Each point of the rope is pulled from both sides to the tangential directions. Let us decompose this force into the x- and the y-components as in Fig. 5.5.

[5] Exactly speaking, the y-coordinate of the segment depends on the position and, therefore, the centrifugal force depends on the position. But the difference is irrelevant to the final result, as it vanishes when $\Delta x \to 0$.

Recall that the rope does not move in the xyz-coordinate system. Hence, in particular, the rightward and the leftward tension in the x-direction should balance: if the rightward tension is T_x, the leftward tension is $-T_x$. Moreover because of Newton's third law (the action-reaction law) T_x does not depend on the position and is a constant function of x.

Then, what is the tension in the y-direction? At the left end $(x, y(x))$ of the segment of the rope the tension vector points along a straight line with slope $y'(x)$ in the negative direction. The x-component of this vector is $-T_x$, as we showed above. Thus the y-component of the tension vector at this point is $-T_x\, y'(x)$. Similarly at the right end $(x+\Delta x, y(x+\Delta x))$, the slope of the tension vector is $y'(x+\Delta x)$ and its x-component is T_x, which shows that its y-component is

$$T_x\, y'(x+\Delta x) = T_x\, y'(x) + T_x\, y''(x)\Delta x + o(\Delta x).$$

Summing up the results here, we obtain the y-component of the tension acting on this segment.

$$F_y = T_x y''(x)\Delta x + o(\Delta x). \tag{5.19}$$

This F_y is what we denoted by the same symbol in (5.11). Indeed F_y is the y-component of the 'true force' (= not apparent like centrifugal force), which is equal to the tension under our assumption that the gravitation is negligible.

Let us repeat once more that the rope stands still in the xyz-coordinate system, which means that the left-hand side of the equation of motion (5.11) and the Coriolis force proportional to $\dfrac{dz}{dt}$ are zero. Thus we obtain $F_y + m\omega^2 y = 0$. By (5.18) and (5.19) this means

$$\rho\omega^2\, y\sqrt{1+y'^2}\Delta x + T_x y''\Delta x + o(\Delta x) = 0.$$

Dividing both sides by Δx and taking the limit $\Delta x \to 0$, we obtain an ordinary differential equation

$$\rho\omega^2 y\sqrt{1+y'^2} + T_x y'' = 0 \tag{5.20}$$

for the function $y(x)$.

5.2.2 Solution of the differential equation and an elliptic function

If we solve the equation (5.20), we get the shape of the skipping rope. But this equation is a complicated combination of the unknown function y and its first and second derivatives. You might have doubts, "Is it really solvable?" Fortunately (or, miraculously) *it is solvable*. Let us simplify it as follows: multiplying (5.20) by

$\dfrac{2y'}{\sqrt{1+y'^2}}$, we have

$$\rho\omega^2(y^2)' + T_x \frac{(y'^2)'}{\sqrt{1+y'^2}} = 0. \tag{5.21}$$

The integral of the first term in the left-hand side is $\rho\omega^2 y^2$. The second term can also be integrated by the change of the integration variable, $\tilde{y} = y'^2$:

$$\begin{aligned}\int T_x \frac{(y'^2)'}{\sqrt{1+y'^2}}\, dx &= T_x \int \frac{d\tilde{y}}{\sqrt{1+\tilde{y}}} \\ &= 2T_x\sqrt{1+\tilde{y}} + (\text{integration constant}).\end{aligned}$$

(We use the fact that T_x does not depend on x.) Thus the integral of the equation (5.21) is

$$\rho\omega^2 y^2 + 2T_x\sqrt{1+y'^2} = (\text{constant}).$$

To determine the constant on the right-hand side, let us fix x to x_0, where $y'(x_0) = 0$, and denote the value $y(x_0)$ by b. (If the shape of the skipping rope is as in Fig. 5.3, b is the maximum value of the function $y(x)$.) Then the above constant is equal to $\rho\omega^2 b^2 + 2T_x$:

$$\rho\omega^2 y^2 + 2T_x\sqrt{1+y'^2} = \rho\omega^2 b^2 + 2T_x.$$

By solving this expression in y'^2 we obtain a differential equation of the first order,

$$y'^2 = \frac{\rho\omega^2}{T_x}(b^2 - y^2)\left(1 + \frac{\rho\omega^2}{4T_x}(b^2 - y^2)\right). \tag{5.22}$$

As there are no second derivatives and no square roots, it is really simpler than (5.20).

You might have already felt at this stage that there is an elliptic integral or an elliptic function behind this. You have really good sense. The differential equation (5.22) means that 'the square of the derivative of y is a polynomial of degree four in y', in other words, '$\dfrac{dy}{dx}$ is the square root of a polynomial of degree four in y'. Roughly speaking,

$$\frac{dy}{dx} = \sqrt{\text{quartic polynomial of } y}.$$

By dividing it by the right-hand side and then integrating it by x we have

$$\int \frac{dy}{\sqrt{\text{quartic polynomial of } y}} = \int dx = x.$$

The left-hand side is an elliptic integral of y and the right-hand side is x, from which it follows that y is the inverse function of an elliptic integral. *It is nothing but an elliptic function*!

Let us make this argument precise. To simplify the notations, we introduce the following functions and constants:

(5.23) $$\eta(x) := \frac{y(x)}{b},$$

(5.24) $$k^2 := \frac{\rho\omega^2 b^2/4T_x}{1+\rho\omega^2 b^2/4T_x},$$

(5.25) $$c^2 := \frac{\rho\omega^2}{T_x}\left(1+\frac{\rho\omega^2 b^2}{4T_x}\right) = \frac{4k^2}{(1-k^2)^2 b^2}.$$

With these symbols the equation (5.22) is rewritten as

$$\begin{aligned} b^2\eta'^2 &= \frac{\rho\omega^2 b^2}{T_x}(1-\eta^2)\left(1+\frac{\rho\omega^2 b^2}{4T_x}(1-\eta^2)\right) \\ &= b^2c^2(1-\eta^2)(1-k^2\eta^2). \end{aligned}$$

Dividing both sides by b^2 and taking square roots, we have

(5.26) $$\frac{d\eta}{dx} = c\sqrt{(1-\eta^2)(1-k^2\eta^2)},$$

or

$$\frac{1}{\sqrt{(1-\eta^2)(1-k^2\eta^2)}}\frac{d\eta}{dx} = c.$$

By integrating this equation by x from 0 to x, the left-hand side becomes an integral in η and we have

$$\int_{\eta(0)}^{\eta(x)} \frac{d\eta}{\sqrt{(1-\eta^2)(1-k^2\eta^2)}} = cx.$$

By our assumption, $y(0) = 0$, thus $\eta(0) = 0$. Therefore the left-hand side is the incomplete elliptic integral of the first kind. Taking the inverse function, we obtain

$$\eta(x) = \operatorname{sn}(cx, k), \text{ i.e., } y(x) = b\operatorname{sn}(cx, k),$$

which proves that the shape of the skipping rope is a graph of Jacobi's elliptic function sn.

Since we assume that the rope intersects with the x-axis at $x = 0$ and $x = 2a$, $y(2a) = b\operatorname{sn}(ca, k)$ should be 0. As we have seen in the previous chapter, the function $\operatorname{sn}(u, k)$ vanishes when u is an integer multiple of $2K(k)$. Therefore the rope takes the shape in Fig. 5.3 when $2ca = 2K(k)$. So, $c = \dfrac{K(k)}{a}$ and

$$y(x) = b \operatorname{sn}\left(\frac{K(k)x}{a}, k\right). \tag{5.27}$$

Thus we have found the shape of the skipping rope, but let us examine 'how' it is determined. When we want to draw the graph of the function (5.27), we need values of the constants a, k and b. The constant a is given from the beginning by the position of an end of the rope, $x = 2a$. The constants b and k are related by (5.24), in which the linear density ρ, the angular velocity ω and the x-component of the tension T_x are parameters. On the other hand, the relation $c = \dfrac{K(k)}{a}$ and (5.25) lead to

$$b = \frac{2k}{k'^2 c} = \frac{2ak}{k'^2 K(k)}, \quad (k' = \sqrt{1-k^2}). \tag{5.28}$$

Hence the constants b and k are determined by the system of equations (5.24) and (5.28) with parameters a, ρ, ω and T_x. An unpleasant thing here is that the tension T_x appears. The parameters a, ρ and ω can be easily measured and are controllable, while T_x is determined indirectly and is hard to measure. Fortunately, there is another parameter which replaces it; the length of the rope l.

The length of the rope is the length of the curve $(x, y(x)) = (x, b\operatorname{sn}(cx))$ $(x \in [0, 2a])$. By the formula (1.1) of the length of a curve it is expressed as follows:

$$l = \int_0^{2a} \sqrt{\left(\frac{dx}{dx}\right)^2 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_0^{2a} \sqrt{1 + y'^2}\, dx. \tag{5.29}$$

The expression inside the square root is equal to

$$\begin{aligned} 1 + y'^2 &= 1 + b^2 c^2 \operatorname{cn}^2(cx) \operatorname{dn}^2(cx) \\ &= 1 + \frac{4k^2}{k'^4} \operatorname{cn}^2(cx) \operatorname{dn}^2(cx) \end{aligned}$$

by the derivation formula (4.16) of sn and (5.28). Using the slightly technical formula

$$\operatorname{cn}^2(cx) = 1 - \operatorname{sn}^2(cx) = 1 - \frac{1}{k^2}(1 - \operatorname{dn}^2(cx)),$$

we have

$$\begin{aligned} 1 + y'^2 &= 1 + \frac{4k^2}{k'^4}\left(-\frac{k'^2}{k^2} + \frac{1}{k^2}\operatorname{dn}^2(cx)\right)\operatorname{dn}^2(cx) \\ &= \left(1 - \frac{2}{k'^2}\operatorname{dn}^2(cx)\right)^2. \end{aligned}$$

Therefore the integrand in (5.29) is $1 - \dfrac{2}{k'^2}\operatorname{dn}^2(cx)$ up to sign. To determine the sign, recall that dn is defined by $\operatorname{dn} u = \sqrt{1 - k^2 \operatorname{sn}^2 u}$ and that $\operatorname{sn} u$ satisfies $-1 \le \operatorname{sn} u \le 1$, from which the inequalities

$$k' = \sqrt{1-k^2} \le \mathrm{dn}\, u \le 1$$

follow (Fig. 5.6). Thus we have $\dfrac{2}{k'^2}\,\mathrm{dn}^2(cx) \ge 2$.

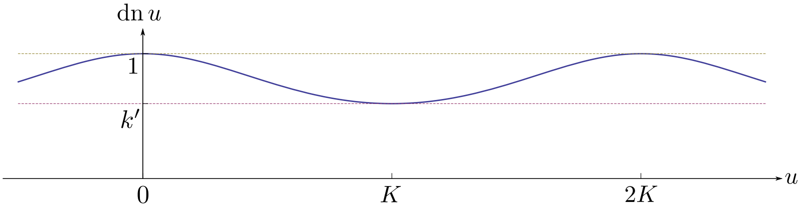

Fig. 5.6 $k' \le \mathrm{dn}\, u \le 1$ $(K = K(k))$.

Hence $\sqrt{1+y'^2} = \dfrac{2}{k'^2}\,\mathrm{dn}^2(cx) - 1$ and we can rewrite (5.29) as

$$\begin{aligned} l &= \int_0^{2a} \left(\frac{2}{k'^2}\,\mathrm{dn}^2(cx) - 1 \right) dx = \frac{2}{k'^2} \int_0^{2a} \mathrm{dn}^2(cx)\, dx - 2a \\ &= \frac{2a}{k'^2 K(k)} \int_0^{2K(k)} \mathrm{dn}^2(u)\, du - 2a, \end{aligned} \tag{5.30}$$

changing the integration variable from x to $u = cx = \dfrac{K(k)x}{a}$. The above expression is essentially an integral of dn^2, which might seem terribly difficult, but, in fact it has already appeared before. Since $\mathrm{dn}\, u$ is symmetric around the axis $u = K(k)$ (Fig. 5.6),

$$\int_0^{2K(k)} \mathrm{dn}^2(u)\, du = 2 \int_0^{K(k)} \mathrm{dn}^2(u)\, du.$$

Then, by a change $\sin\phi = \mathrm{sn}\, u$, or, equivalently, $\phi = \arcsin(\mathrm{sn}\, u) = \mathrm{am}\, u$ ($\mathrm{am}\, u$ is Jacobi's amplitude function in Section 4.1) of the integration variable, we can rewrite the above integral as

$$\int_0^{K(k)} \mathrm{dn}^2(u)\, du = \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2\phi}\, d\phi, \tag{5.31}$$

which is the complete elliptic integral of the second kind $E(k)$, (1.9). The proof of (5.31) is a very good exercise in computing Jacobi's elliptic functions, so we leave it to the reader.

Exercise 5.1 Show (5.31).

Thus (5.30) is equivalent to

$$l = \frac{4a\,E(k)}{(1-k^2)K(k)} - 2a, \text{ or, } \frac{l}{4a} + \frac{1}{2} = \frac{E(k)}{(1-k^2)K(k)}, \tag{5.32}$$

which give a relation between given parameters (a and l) and the modulus k.

Summarising, the constants in (5.27) are determined, in principle, as follows:

1. First, solving the equation (5.32) with the parameters l (the length of the rope) and a ($x = 2a$ is the position of an end of the rope),[6] find the modulus k.
2. Next, determine b by (5.28).

We defined k in (5.24) by the linear density ρ, the angular velocity ω and the tension T_x, but, as we have shown here, both k and b are determined only from a and l. The curve stays unchanged even when ρ and ω are changed.[7] The relation (5.24) is interpreted as a formula determining the tension from these quantities.

5.2.3 The variational method

We have derived and solved the equation of the shape of the skipping rope in the previous section. In this section we rederive the equation by the *variational method* using a potential. The explanation of the physics will be rather sketchy.

Here by 'potential' $V = V(x,y,z)$ we mean a sort of potential energy, which satisfies

$$(\text{force}) = -\operatorname{grad} V = \left(-\frac{\partial V}{\partial x}, -\frac{\partial V}{\partial y}, -\frac{\partial V}{\partial z}\right).$$

In our case we consider a potential $V(y)$ for the y-component $m\omega^2 y$ of the centrifugal force:

$$V(y) = -\frac{m}{2}\omega^2 y^2.$$

We consider a potential for each small part of the rope and sum up all of them to obtain a potential for the centrifugal force of the rope.

A small part of the rope above the interval $[x, x+\Delta x]$ of the x-axis has mass $\rho\sqrt{1+y'^2}\Delta x$ as shown in (5.17). The corresponding potential is $-\dfrac{\rho\omega^2}{2} y^2 \sqrt{1+(y')^2}\Delta x$. (We omit terms of order $o(\Delta x)$.) The limit $\Delta x \to 0$ of the sum of all of them (from $x = 0$ to $x = 2a$) is the potential for the whole rope and is expressed by an integral,

$$U[y] = -\frac{\rho\omega^2}{2}\int_0^{2a} y^2\sqrt{1+(y')^2}\,dx. \tag{5.33}$$

[6] As the right-hand side of (5.32) is a complicated function including elliptic integrals, we cannot solve it explicitly. This is why the phrase 'in principle' is added.

[7] Of course, under the condition that ω is so large that gravitation is negligible compared to the centrifugal force.

The mapping U assigns a real number to each function (curve) $y = y(x)$. That is to say, it is a 'function of functions' and sometimes called a *functional*. Physics tells us that the function y which makes $U[y]$ minimal gives the solution. Finding such a minimising function is called the *variational problem*.

In order to make $U[y]$ lower, we 'have only to' make the integrand larger. Can we make the integrand as large as we want? No, that is impossible, because there are constraints to y:

- The ends of the rope are fixed: $y(0) = y(2a) = 0$.
- The length of the rope is fixed: $\displaystyle\int_0^{2a} \sqrt{1 + (y'(x))^2}\, dx = l$. In other words, y satisfies a functional equation,

$$G[y] := \int_0^{2a} \sqrt{1 + (y'(x))^2}\, dx - l = 0. \tag{5.34}$$

We have to find an extremal function $\mathring{y}(x)$ under these conditions.

You have probably already encountered 'constrained extremal problems' in an analysis course. Of course I do not mean an extremal problem of functionals, but an extremal problem of a usual function like: 'Find the extrema of a function $f(y_1, y_2) = y_1 + y_2$ under the constraint $g(y_1, y_2) = \sqrt{y_1^2 + y_2^2} - l = 0$'. If you have encountered such problems before, then you should have heard about a standard method to solve this kind of problem, the *method of Lagrange multipliers*. If the constraint $g(y_1, y_2) = 0$ is absent, we can find a candidate $(\mathring{y}_1, \mathring{y}_2)$ for the extremal points by the equation $\operatorname{grad} f(\mathring{y}_1, \mathring{y}_2) = \left(\dfrac{\partial f}{\partial y_1}(\mathring{y}_1, \mathring{y}_2), \dfrac{\partial f}{\partial y_2}(\mathring{y}_1, \mathring{y}_2)\right) = (0, 0)$. When the condition $g(y_1, y_2) = 0$ is imposed, a candidate for the extremal points satisfies the following:

- There exists a $\lambda \in \mathbb{R}$ such that

$$\operatorname{grad} f(\mathring{y}_1, \mathring{y}_2) - \lambda \operatorname{grad} g(\mathring{y}_1, \mathring{y}_2) = (0, 0).$$

- $g(\mathring{y}_1, \mathring{y}_2) = 0$.

The constant λ is called the *Lagrange multiplier*. The statement can be reformulated by using a function $\mathcal{L}(y_1, y_2, \lambda) := f(y_1, y_2) - \lambda g(y_1, y_2)$ as follows:

$$\operatorname{grad} \mathcal{L}(\mathring{y}_1, \mathring{y}_2, \lambda) = \left(\frac{\partial \mathcal{L}}{\partial y_1}, \frac{\partial \mathcal{L}}{\partial y_2}, \frac{\partial \mathcal{L}}{\partial \lambda}\right)(\mathring{y}_1, \mathring{y}_2, \lambda) = (0, 0, 0), \tag{5.35}$$

which means that, although the variables y_1 and y_2 are constrained by $g(y_1, y_2) = 0$, by introducing a new variable λ we can look for an extremal point as if the variables of $\mathcal{L}(y_1, y_2, \lambda)$ are free (non-constrained).

This method can be applied to our extremal problem of the functional $U[y]$ defined by (5.33) with a constraint (5.34), $G[y] = 0$. Let us fix two small functions

$\alpha(x)$ and $\beta(x)$ on the interval $[0,2a]$ satisfying $\alpha(0)=\alpha(2a)=0$, $\beta(0)=\beta(2a)=0$, and define two functions of extra variables $(\varepsilon_1,\varepsilon_2)$:

$$f(\varepsilon_1,\varepsilon_2) := U[\mathring{y}+\varepsilon_1\alpha+\varepsilon_2\beta], \tag{5.36}$$

$$g(\varepsilon_1,\varepsilon_2) := G[\mathring{y}+\varepsilon_1\alpha+\varepsilon_2\beta]. \tag{5.37}$$

If $\mathring{y}(x)$ is an extremal function of the functional $U[y]$ under the condition $G[y]=0$, the point $(\varepsilon_1,\varepsilon_2)=(0,0)$ is an extremal point of $f(\varepsilon_1,\varepsilon_2)$ under the condition $g(\varepsilon_1,\varepsilon_2)=0$. Note that the boundary conditions are satisfied because $\alpha(0)=\alpha(2a)=0$ and $\beta(0)=\beta(2a)=0$.

By the method of Lagrange multipliers, there exists a constant λ such that function

$$h(\varepsilon_1,\varepsilon_2,\lambda) := f(\varepsilon_1,\varepsilon_2)-\lambda g(\varepsilon_1,\varepsilon_2)$$

satisfies

$$\frac{\partial h}{\partial\varepsilon_1}(0,0)=\frac{\partial h}{\partial\varepsilon_2}(0,0)=\frac{\partial h}{\partial\lambda}(0,0)=0. \tag{5.38}$$

The last equality $\dfrac{\partial h}{\partial\lambda}(0,0)=0$ means that $g(0,0)=0$, namely $G[\mathring{y}]=0$. What do the first two equalities $\dfrac{\partial h}{\partial\varepsilon_1}(0,0)=\dfrac{\partial h}{\partial\varepsilon_2}(0,0)=0$ mean? Let us define

$$\begin{aligned}\mathcal{L}[y,\lambda] &:= U[y]-\lambda G[y]\\ &= -\frac{\rho\omega^2}{2}\int_0^{2a} y^2\sqrt{1+(y')^2}\,dx-\lambda\left(\int_0^{2a}\sqrt{1+(y')^2}\,dx-l\right),\end{aligned} \tag{5.39}$$

$$L(y_1,y_2,\lambda) := -\frac{\rho\omega^2}{2}y_1^2\sqrt{1+y_2^2}-\lambda\sqrt{1+y_2^2}. \tag{5.40}$$

Then,

$$h(\varepsilon_1,\varepsilon_2,\lambda)=\mathcal{L}[\mathring{y}+\varepsilon_1\alpha+\varepsilon_2\beta,\lambda], \tag{5.41}$$

$$\mathcal{L}[y,\lambda]=\int_0^{2a} L\big(y(x),y'(x)\big)\,dx+\lambda l. \tag{5.42}$$

Using these notations, we have

$$\begin{aligned}\frac{\partial h}{\partial\varepsilon_1}(0,0) &= \frac{\partial}{\partial\varepsilon_1}\mathcal{L}[\mathring{y}+\varepsilon_1\alpha+\varepsilon_2\beta]\bigg|_{(\varepsilon_1,\varepsilon_2)=(0,0)}\\ &= \int_0^{2a}\left(\frac{\partial L}{\partial y_1}\alpha+\frac{\partial L}{\partial y_2}\alpha'\right)dx\\ &= \int_0^{2a}\left(\frac{\partial L}{\partial y_1}\alpha-\frac{d}{dx}\frac{\partial L}{\partial y_2}\alpha\right)dx.\end{aligned}$$

In the third line we integrated the second term by parts and turned α' into α. You might ask "Haven't you forgetten the boundary term $\left.\dfrac{\partial L}{\partial y_2}\alpha\right|_{x=0}^{x=2a}$ of the integration by parts?" No. Here the boundary condition $\alpha(0) = \alpha(2a) = 0$ plays the role. The boundary term is equal to

$$\frac{\partial L}{\partial y_2}\big(\mathring{y}(x), \mathring{y}'(x)\big)\,\alpha(x)\Big|_{x=0}^{x=2a} = \left.\frac{\partial L}{\partial y_2}\right|_{x=2a}\alpha(2a) - \left.\frac{\partial L}{\partial y_2}\right|_{x=0}\alpha(0),$$

and vanishes because $\alpha(0) = \alpha(2a) = 0$.

Thus we have

$$\frac{\partial h}{\partial \varepsilon_1}(0,0) = \int_0^{2a}\left(\frac{\partial L}{\partial y_1}\big(\mathring{y}(x), \mathring{y}'(x)\big) - \frac{d}{dx}\frac{\partial L}{\partial y_2}\big(\mathring{y}(x), \mathring{y}'(x)\big)\right)\alpha(x)\,dx. \tag{5.43}$$

Similarly,

$$\frac{\partial h}{\partial \varepsilon_2}(0,0) = \int_0^{2a}\left(\frac{\partial L}{\partial y_1}\big(\mathring{y}(x), \mathring{y}'(x)\big) - \frac{d}{dx}\frac{\partial L}{\partial y_2}\big(\mathring{y}(x), \mathring{y}'(x)\big)\right)\beta(x)\,dx. \tag{5.44}$$

Recall that $\alpha(x)$ and $\beta(x)$ were arbitrarily chosen. Hence from $\dfrac{\partial h}{\partial \varepsilon_1}(0,0) = \dfrac{\partial h}{\partial \varepsilon_2}(0,0) = 0$ it follows that the expression inside the parentheses of the integrands of (5.43) and (5.44) should be zero. This leads to the *Euler–Lagrange equation*, $\dfrac{\partial L}{\partial y_1} - \dfrac{d}{dx}\dfrac{\partial L}{\partial y_2} = 0$, which is usually denoted as

$$\frac{\partial L}{\partial y}(y, y', \lambda) - \frac{d}{dx}\frac{\partial L}{\partial y'}(y, y', \lambda) = 0. \tag{5.45}$$

Remark 5.2 The Euler–Lagrange equation (5.45) can be regarded as a consequence of the 'infinite-dimensional' method of Lagrange multipliers as follows.

In the example of the extremal points of $f(y_1, y_2) = 0$ under conditions $g(y_1, y_2) = 0$, we have two independent variables y_i ($i \in \{1, 2\}$). Instead of just two variables, let us take infinitely many variables $y(x)$ ($x \in [0, 2a]$). We regard the variable 'x' as an index. Then, replacing 'function $f(y_1, y_2)$' by 'functional $U[y(x)]$' and 'condition $g(y_1, y_2) = 0$' by 'condition $G[y] = 0$', the method of Lagrange multipliers is formulated as follows: There exists a constant λ such that the functional $\mathcal{L}[y, \lambda] = U[y] - \lambda G[y]$ satisfies

$$\left(\frac{\delta \mathcal{L}}{\delta y}, \frac{\partial \mathcal{L}}{\partial \lambda}\right)[\mathring{y}, \lambda] = (0, 0), \tag{5.46}$$

which is a functional version of (5.35).

What is 'a functional version of a partial derivative' $\dfrac{\delta \mathcal{L}}{\delta y}$? To answer this question, recall what a derivative of a (usual) function is. Derivatives of a (usual) function $\mathcal{L}(y_1, y_2)$ are coefficients in the equation

$$\begin{aligned}\Delta\mathcal{L} &:= \mathcal{L}(y_1+\Delta y_1, y_2+\Delta y_2) - \mathcal{L}(y_1, y_2)\\ &= \frac{\partial \mathcal{L}}{\partial y_1}(y_1, y_2)\Delta y_1 + \frac{\partial \mathcal{L}}{\partial y_2}(y_1, y_2)\Delta y_2 + o(|(\Delta y_1, \Delta y_2)|),\end{aligned} \tag{5.47}$$

which expresses a small deviation of the function $\mathcal{L}(y_1, y_2)$ corresponding to a shift by a small vector $(\Delta y_1, \Delta y_2)$. ($|(\Delta y_1, \Delta y_2)| = \sqrt{\Delta y_1^2 + \Delta y_2^2}$ is the length of the vector $(\Delta y_1, \Delta y_2)$.)

In the case of a functional $\mathcal{L}[y, \lambda]$, we use a small function $\delta y(x)$ instead of a small vector $(\Delta x, \Delta y)$. The deviation $\delta\mathcal{L}$ of the functional corresponding to the shift by δy is

$$\begin{aligned}\delta\mathcal{L} &:= \mathcal{L}[y+\delta y, \lambda] - \mathcal{L}[y, \lambda]\\ &= \int_0^{2a} \big(L(y(x)+\delta y(x), y'(x)+\delta y'(x), \lambda) - L(y, y', \lambda)\big)\, dx.\end{aligned}$$

A computation similar to (5.43) leads to

$$\delta\mathcal{L} = \int_0^{2a} \left(\frac{\partial L}{\partial y_1} - \frac{d}{dx}\frac{\partial L}{\partial y_2}\right)\delta y\, dx + o(|(\delta y, \delta y')|). \tag{5.48}$$

In the finite-dimensional case, $\dfrac{\partial \mathcal{L}}{\partial y_i}$ is the coefficient of Δy_i. So, we can interpret the coefficient of δy in the expression (5.48), $\dfrac{\partial L}{\partial y_1} - \dfrac{d}{dx}\dfrac{\partial L}{\partial y_2}$ as $\dfrac{\delta \mathcal{L}}{\delta y}$. In this sense, the Euler–Lagrange equation is equivalent to $\dfrac{\delta \mathcal{L}}{\delta y} = 0$. □

Using the explicit expression (5.40) of $L(y_1, y_2, \lambda)$, we can compute the Euler–Lagrange equation (5.45) as follows. Since

$$\begin{aligned}\frac{\partial L}{\partial y} &= 2y\sqrt{1+y'^2},\\ \frac{d}{dx}\frac{\partial L}{\partial y'} &= 2y\frac{y'^2}{\sqrt{1+y'^2}} + \frac{y''(y^2-\lambda)}{(1+y'^2)^{3/2}},\end{aligned}$$

the equation (5.45) multiplied by $\sqrt{1+y'^2}$ is equal to

$$2y - (y^2-\lambda)\frac{y''}{1+y'^2} = 0, \text{ or, } \frac{y''}{1+y'^2} = \frac{2y}{y^2-\lambda}, \tag{5.49}$$

which is a second-order ordinary differential equation in y. The integrals of both sides of the last equation are easily computed:

$$\frac{1}{2}\log\left(1+y'^2\right) = \log(\lambda - y^2) + (\text{constant}).$$

(Please check by differentiation.[8]) Exponentiating this equation, we obtain

$$1 + y'^2 = C(\lambda - y^2)^2. \tag{5.50}$$

($C = \exp(\text{constant})$.) This is a first-order ordinary differential equation and, as it has the form 'square of y' is a quartic polynomial of y', the solution has to be expressed in terms of Jacobi's elliptic function sn, as was the case with (5.22). The details are left to the reader, but let us comment on several important points.

As in the previous section, we look for a solution $y(x)$ of (5.50) which has maximum value b at $x = x_0$ as in Fig. 5.3. Substituting $y'(x_0) = 0$ in (5.50), we have $C = (\lambda - b^2)^{-2}$. Therefore equation (5.50) is rewritten as

$$\begin{aligned}\left(\frac{dy}{dx}\right)^2 &= \left(\frac{\lambda - y^2}{\lambda - b^2}\right)^2 - 1 \\ &= \frac{(b^2 - y^2)(2\lambda - b^2 - y^2)}{(\lambda - b^2)^2}.\end{aligned}$$

The function $\eta(x) := \dfrac{y(x)}{b}$ should satisfy an equation,

$$\frac{d\eta}{dx} = c\sqrt{(1-\eta^2)(1-k^2\eta^2)}. \tag{5.51}$$

Here the constants c and k are defined by the equations

$$c^2 = \frac{2\lambda - b^2}{(\lambda - b^2)^2}, \quad k^2 = \frac{b^2}{2\lambda - b^2}. \tag{5.52}$$

The differential equation (5.51) is nothing but the equation (5.26) for $\eta(x)$.

The procedure to determine the constants b and c is almost the same as before. The positions of the zeros of sn determine the constant c as $c = \dfrac{K(k)}{a}$. The constant b is determined from the condition 'the length of the rope $= l$', as in the previous section, where we used (5.25), $c^2 = \dfrac{4k^2}{(1-k^2)^2 b^2}$, which can be easily derived from (5.52).

[8] We assume that the rope has a shape as in Fig. 5.3, therefore $y > 0$, $y'' < 0$. In this case from (5.49) it follows that $y^2 - \lambda = 2y\dfrac{1+y'^2}{y''} < 0$ and the argument of log in the right-hand side is positive.

Let us now stop our discussion of elliptic integrals and elliptic functions of *real numbers*. We shall extend them to *complex numbers*, and for that purpose we are going to prepare their stages, *Riemann surfaces* and *elliptic curves*, from the next chapter.

Part II
Complex Part

In the second half of this book we extend elliptic integrals and elliptic functions to $\mathbb{C}$. Actually they live not on $\mathbb{C}$ but on so-called Riemann surfaces or elliptic curves, which are introduced in Chapter 6 and Chapter 7 respectively.

Chapter 6
Riemann Surfaces of Algebraic Functions

Up to the previous chapter we were discussing elliptic integrals and elliptic functions in the 'real world'. We are now going to take a step toward the 'imaginary direction'. There are many things in mathematics which are invisible in the realm of real numbers but become clearly visible through complex numbers. Many complicated theorems over the real field become simpler over the complex field. The classification theorem of elliptic integrals in Chapter 2 is one such example. We shall see that the true nature of elliptic functions is revealed when they are defined on the complex plane.

In fact, it is more natural to consider elliptic integrals and elliptic functions to 'live' not on the complex 'plane' but on a 'surface consisting of complex numbers' called a Riemann surface. In this chapter we explain the general facts about Riemann surfaces which will be needed later. In the next chapter (Chapter 7), we shall construct a Riemann surface, on which elliptic integrals live. The complex elliptic integrals shall be discussed from Chapter 8.

We freely use elementary facts of complex functions, which can be found in any textbooks on complex analysis, for example, [Ah], [Lv] or [Na]. Several important theorems are gathered in Section A.2 for the readers's convenience.

6.1 Riemann Surfaces of Algebraic Functions

We are aiming at constructing elliptic integrals $\displaystyle\int R(x, \sqrt{\varphi(x)})\,dx$ over $\mathbb{C}$, but in this section we first examine the definition of square roots. The integrals will be discussed in the next section (Section 6.2).

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_7

6.1.1 What is the problem?

Is there still anything more to discuss about the square root, even though all of us learned it in our schooldays? Yes. The problem is its *multi-valuedness*.

When we consider square roots of real numbers as a real function, we take the following definition: 'The square root $\sqrt{a}$ of a is a number whose square is equal to a: $(\sqrt{a})^2 = a$. Only *non-negative* numbers, $a \geq 0$, have square roots. The square roots are also *non-negative*: $\sqrt{a} \geq 0$. We do not consider the square root of a negative number.'

In the complex world any complex number z, not only non-negative real numbers, has a square root, $\sqrt{z}$. The square roots of -1 are $\pm i$, the square roots of i are $\pm\dfrac{1+i}{\sqrt{2}}$. Well, this is a good news, but a little *too* good, and it causes a problem: Which sign should we take?

If a complex number w satisfies $w^2 = z$, of course $-w$ also satisfies the same equation, $(-w)^2 = z$. When z is a positive real number, as we mentioned above, we can avoid the problem of signs by defining $\sqrt{z}$ to be a positive number. However, when z is a complex number, there is *no natural choice* of signs.

You may have the following idea: let us express z in polar form as $z = r\, e^{i\varphi}$ ($r = |z|$, $\varphi = \arg z$; Fig. 6.1). The square root $\sqrt{r}$ is defined naturally, as r is a positive real number. Using it, can't we define $\sqrt{z}$ by "$\sqrt{z} := \sqrt{r} e^{i\varphi/2}$"?

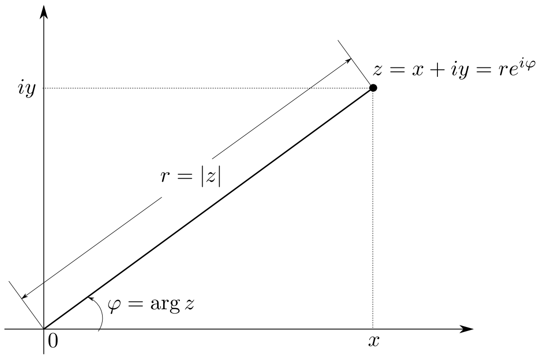

Fig. 6.1 Complex number $z = x + iy = r\, e^{i\varphi}$.

Good point, but, unfortunately, even this method does not define $\sqrt{z}$ uniquely. Why? Because *the argument* $\varphi = \arg z$ *is not uniquely defined*. If one moves a complex number z around the origin 0 once, its argument increases by 2π. In general the argument has an ambiguity of $2\pi\mathbb{Z} = \{2\pi n \mid n \in \mathbb{Z}\}$:

$$z = re^{i\varphi} = re^{i(\varphi\pm 2\pi)} = re^{i(\varphi\pm 4\pi)} = \cdots = re^{i(\varphi+2n\pi)}.$$

Correspondingly, the above "definition" gives

$$\sqrt{z} = \sqrt{r}e^{i(\varphi+2n\pi)/2} = \sqrt{r}e^{i\varphi/2+in\pi} = (-1)^n\sqrt{r}e^{i\varphi/2},$$

and there is no natural way to determine the sign $(-1)^n$.

Since the value $\sqrt{z}$ is not defined uniquely, the function $z \mapsto \sqrt{z}$ is called a *multi-valued* function. (A 'usual' function which assigns one and only one value to each z is called a *single-valued* function.)

6.1.2 Then, what should we do?

How can we solve this problem of signs? Roughly speaking, there are two ways out.

1. Restrict the range of the argument.
2. Split the domain of definition into two.

The first method is simple: 'If it is not uniquely determined, determine it uniquely for yourself.' If the range of the argument is restricted to, for example, $-\pi < \arg z \leq \pi$, it is uniquely determined and, consequently, $\sqrt{z}$ is defined as a single-valued function. This method suffices for many situations, but not always.

If z goes around the origin as in Fig. 6.2, the argument of z tends to π but suddenly jumps to $-\pi$, when z crosses the negative part of the real axis, which means that $\arg z$ and, as a consequence, $\sqrt{z}$ are discontinuous functions. It is not convenient for a fundamental function like $\sqrt{z}$ to be discontinuous, as we want to differentiate such functions.

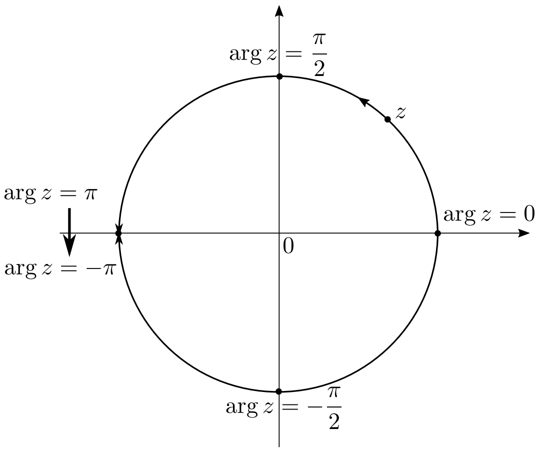

Fig. 6.2 If $-\pi \leq \arg z < \pi$, $\arg z$ is discontinuous.

Moreover, as you see from the phrase 'for example' above, the range of the argument can be chosen arbitrarily. Instead of $-\pi < \arg z \leq \pi$, the interval $0 \leq \arg z < 2\pi$ is also used quite often, and, if necessary, we can take an interval like $-\dfrac{7\pi}{3} <$

$\arg z \leq -\dfrac{\pi}{3}$. Thus, the discontinuity of $\sqrt{z}$ changes according to the choice of the range of the argument. Later we shall consider complex integrals like $\displaystyle\int R(z, \sqrt{\varphi(z)})\,dz$ along various curves. Therefore we want to avoid such ambiguous discontinuity of the square root.

Thus the second solution to the problem of signs becomes important. This is an idea due to Riemann (Riemann, Georg Friedrich Bernhard, 1826–1866),[1] who proposed to 'split the number z into two, as there are two choices of square roots', more precisely, to consider that there are two points $(z,+)$ and $(z,-)$ behind one complex number z $(\neq 0)$. Correspondingly, a 'small' domain $D \subset \mathbb{C}$ is split into two domains D_+ and D_- (Fig. 6.3).

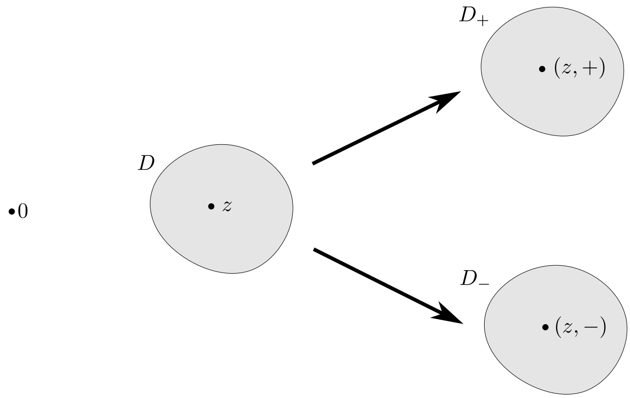

Fig. 6.3 Splitting a domain D.

Here we mean by the adjective 'small' that the argument $\arg z$ of a point in the domain D can be determined as a single-valued continuous function. For example, we can define $\arg z$ uniquely as a continuous function on any domain D contained in a sector $\{z \mid \theta_0 < \arg z < \theta_1\}$ $(0 < \theta_1 - \theta_0 < 2\pi)$ (Fig. 6.4, left). On the other hand, on the annulus $A_{r_1,r_2} := \{z \mid r_1 < |z| < r_2\}$ (Fig. 6.4, right) the same problem as in the previous subsection occurs and we cannot define $\arg z$ as a single-valued continuous function, so we exclude such a domain from the following considerations.

Under these assumptions, using the prescribed argument $\arg z$ on D, we define the square root of $z = r\,e^{i\varphi}$ ($r = |z|$, $\varphi = \arg z$) by

$$\begin{cases} \sqrt{(z,+)} &= +\sqrt{r}e^{i\varphi/2}, \\ \sqrt{(z,-)} &= -\sqrt{r}e^{i\varphi/2}. \end{cases} \tag{6.1}$$

The right-hand sides are two square roots of z, and we take each of them as the square root of different points.

[1] The original articles are I (1851) and VI (1857) in [Ri].

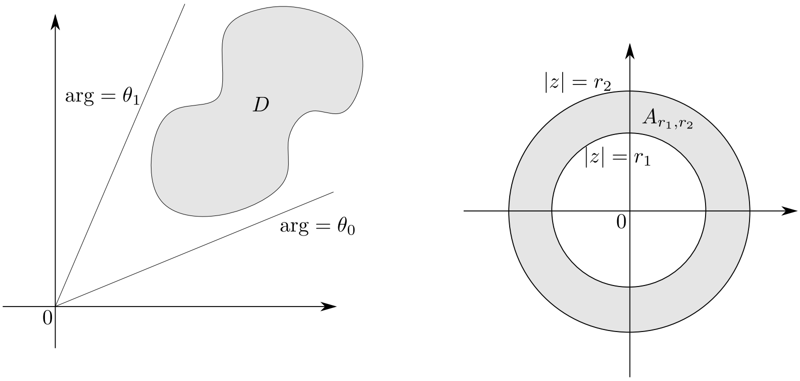

Fig. 6.4 A domain in a sector and an annulus.

However the point $z = 0$ is an exception. Since there is no complex number w other than 0 which satisfies $w^2 = 0$, the square root of 0 is uniquely determined, $\sqrt{0} = 0$, and there is no need to split it.

Thus we have defined the function $\sqrt{z}$ on 'small domains' and at $z = 0$. How can we extend this function 'nicely' to all complex numbers?

Answer: Glue two complex planes (= two copies of $\mathbb{C}$, which we call *sheets* and denote them by $\mathbb{C}_+$ and $\mathbb{C}_-$) in the following way (Fig. 6.5):

1. Cut the negative sides $(-\infty, 0)$ of the real axes of both sheets.
2. Identify the upper side (the boundary of the upper half plane) of the cut of $\mathbb{C}_+$ with the lower side (the boundary of the lower half plane) of the cut of $\mathbb{C}_-$.
3. Identify the lower side of the cut of $\mathbb{C}_+$ with the upper side of the cut of $\mathbb{C}_-$.
4. Identify 0 of $\mathbb{C}_\pm$ and regard them as one point.

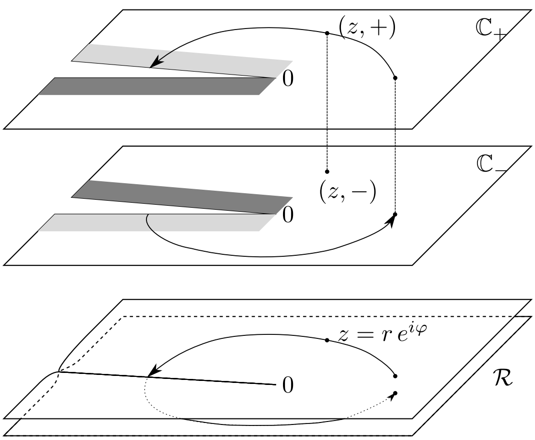

Fig. 6.5 Construction of the Riemann surface of $\sqrt{z}$ by gluing two $\mathbb{C}$'s.

In Fig. 6.5 it looks as if the two sheets cross at the negative part of the real axes to form an intersection. This is nothing more than an apparent crossing caused by the restriction that we have to draw the picture in three-dimensional space. The true intersection of the two sheets is only the origin, $\{0\}$. It is difficult to understand from the picture and the above construction, but we shall construct the same Riemann surface by another method in the next subsection, which should convince you that the intersection of two sheets is just one point.

We assume that the argument $\varphi = \arg z$ takes values in $(-\pi, \pi]$ on both sheets $\mathbb{C}_\pm$ and define the square root by (6.1). Thus the square root $\sqrt{z}$ of $z = r\,e^{i\varphi}$ is equal to $\sqrt{r}e^{i\varphi/2}$ on $\mathbb{C}_+$ and to $-\sqrt{r}e^{i\varphi/2}$ on $\mathbb{C}_-$. When we restricted the values of the argument, the square root $\sqrt{z}$ was not continuous. How about the redefined square root? Let us take $z = r\,e^{i\varphi}$ $(r > 0)$ in $\mathbb{C}_+$ and trace the value of $\sqrt{z}$, moving its argument φ in the interval $[0, 2\pi]$ from 0 to 2π.

(i) At the starting point, $\sqrt{z} = \sqrt{r}e^{i0/2} = \sqrt{r}$.
(ii) $0 \le \varphi \le \pi$: z moves on the upper sheet $\mathbb{C}_+$ as $(z, +)$. By the definition (6.1) $\sqrt{z} = \sqrt{r}e^{i\varphi/2}$.
(iii) $\pi < \varphi \le 2\pi$: When φ exceeds π, z transfers to another sheet and moves on $\mathbb{C}_-$ as $(z, -)$. Since the argument is always in $(-\pi, \pi]$, $\arg z = \varphi - 2\pi$. Thus we have

$$\sqrt{z} = \sqrt{(z,-)} = -\sqrt{r}e^{i(\varphi-2\pi)/2} = \sqrt{r}e^{i\varphi/2}$$

by (6.1). Therefore $\sqrt{z}$ is expressed by $\sqrt{r}e^{i\varphi/2}$ as in (i), which means that it is a continuous function of r and φ, and, consequently, of z.
(iv) Even when φ reaches 2π, z remains in the sheet $\mathbb{C}_-$ and does not return to the starting point in $\mathbb{C}_+$. The value of $\sqrt{z}$ becomes $-\sqrt{r}e^{i(2\pi-2\pi)/2} = -\sqrt{r}$.

If the argument φ goes further from 2π to 4π, $z = r\,e^{i\varphi}$ goes around the origin once more, transfers from the sheet $\mathbb{C}_-$ to $\mathbb{C}_+$ at $\varphi = 3\pi$ and returns to the starting point when $\varphi = 4\pi$. The sign of $\sqrt{z}$ changes once more by the same procedure as above and $\sqrt{z}$ gets back to $\sqrt{r}$ as expected.

Summarising, by this definition the function $\sqrt{z}$ is neither discontinuous nor multi-valued, which suggests that it should be defined on the 'stage' constructed as[2]

$$\mathcal{R} := (\mathbb{C}_+ \smallsetminus \{0\}) \cup \{0\} \cup (\mathbb{C}_- \smallsetminus \{0\}),$$

which is shown in Fig. 6.5. The set $\mathcal{R}$ is called the *Riemann surface of* $\sqrt{z}$. The point 0 is called the *branch point* of $\mathcal{R}$, because $\mathcal{R}$ branches at 0 as in Fig. 6.5.

6.1.3 Another construction

In the previous section we constructed the Riemann surface of $\sqrt{z}$ in a quite handmade way by 'cutting and gluing', which we described not very rigorously. (It is not difficult

[2] The symbol '$\smallsetminus$' denotes the difference of sets: $A \smallsetminus B := \{x \in A \mid x \notin B\}$.

to make it rigorous, but the description becomes much more complicated.) Here in this section we introduce a more systematic construction, from which, however, the shape of the Riemann surface is not easy to imagine.

In the previous section we split a point z ($\neq 0$) into two points $(z, \pm)$. The purpose of attaching '$\pm$' is merely to distinguish them. So, let us attach the value of $\sqrt{z}$ itself instead of just signs and consider pairs of complex numbers $(z, w = \pm\sqrt{r}\, e^{i\varphi/2})$ as points of $\mathcal{R}$. In other words,

$$\mathcal{R} := \{(z,w) \mid w^2 - z = 0\} \subset \mathbb{C}^2. \tag{6.2}$$

Is this really the same as what we constructed before? Yes. Each sheet $\mathbb{C}_\pm$ is embedded into $\mathcal{R}$ as follows:

$$\mathbb{C}_+ \ni z = re^{i\varphi} \mapsto (z, +\sqrt{r}e^{i\varphi/2}) \in \mathcal{R},$$
$$\mathbb{C}_- \ni z = re^{i\varphi} \mapsto (z, -\sqrt{r}e^{i\varphi/2}) \in \mathcal{R},$$

where $-\pi < \varphi \leq \pi$. Let us check that they are glued together in $\mathcal{R}$ as in Fig. 6.5. We have only to examine the negative parts of the real axes of $\mathbb{C}_\pm$:

(i) When a point $z_+ = re^{i\varphi_+}$ in $\mathbb{C}_+$ approaches the negative part of the real axis from the upper half plane, the argument φ_+ grows to π and, correspondingly, z approaches $re^{i\pi} = -r$. Since $e^{i\varphi_+/2}$ tends to $e^{i\pi/2} = i$, $(z, +\sqrt{r}\, e^{i\varphi_+/2})$ approaches $(-r, i\sqrt{r})$.

(ii) On the other hand, when a point $z_- = re^{i\varphi_-}$ in $\mathbb{C}_-$ approaches the negative part of the real axis from the lower half plane, the argument φ_- decreases to $-\pi$ and z approaches $re^{-i\pi} = -r$. Since $e^{i\varphi_-/2}$ tends to $e^{-i\pi/2} = -i$, $(z, -\sqrt{r}\, e^{i\varphi_-/2})$ approaches $(-r, -(-i)\sqrt{r}) = (-r, i\sqrt{r})$.

Thus, when a point approaches a negative real number $-r$ either from above in $\mathbb{C}_+$ or from below in $\mathbb{C}_-$, it tends to one and the same point $(-r, i\sqrt{r})$ in $\mathcal{R}$ defined by (6.2), which means that the upper side of the cut of $\mathbb{C}_+$ and the lower side of the cut of $\mathbb{C}_-$ glue together as we described in the previous section.[3] We can show similarly that the lower side of the cut of $\mathbb{C}_+$ and the upper side of the cut of $\mathbb{C}_-$ glue together.

Although it is hard to see the concrete 'shape' of $\mathcal{R}$ from the construction (6.2), it has the following advantages:

- Two points correspond to one $z(\neq 0)$ naturally. The negative part of the real axis does not require a special treatment.
- It is obvious that one point $(0,0)$ corresponds to 0.
- A local coordinate system is defined at each point. As a consequence $\mathcal{R}$ becomes a *one-dimensional complex manifold*.

[3] If you feel that I am tediously repeating the same thing, it proves that you have perfectly understood the arguments in the previous section. Indeed, the goal is 'making $\sqrt{z}$ a single-valued function' and we are repeatedly confirming that this is realised on $\mathcal{R}$.

Some readers may be unfamiliar with the word 'manifold'. A manifold is a geometric object, around each point of which a system of coordinates is defined *locally*, which makes calculus possible.[4]

Since the general theory of manifolds is not necessary for us, we do not quote the rigorous definition of a manifold here.[5] But as the notion of 'local coordinates' is indispensable, we explain it here in detail. What we need is 'a stage, where we can do calculus', so let us recall the definition of a derivative of a function $f(z)$:

$$f'(z) = \lim_{h\to 0} \frac{f(z+h) - f(z)}{h}.$$

We want to consider such a thing on $\mathcal{R}$. In order for that we need 'a coordinate z', to which we can 'add a small number h', and moreover 'the limit as $h \to 0$'.

For such operations at a point P in $\mathcal{R}$ it is sufficient that a neighbourhood of P has the same structure as a part of the complex plane $\mathbb{C}$. (Corresponding to the fact that we need only the limit $h \to 0$, such a structure is necessary only for a 'neighbourhood' of P.) Mathematically, it is enough to have an open set U including P and a map from U to $\mathbb{C}$, $\phi : U \to \mathbb{C}$, so that U and the subset $\phi(U)$ of $\mathbb{C}$ can be identified via ϕ (Fig. 6.6).

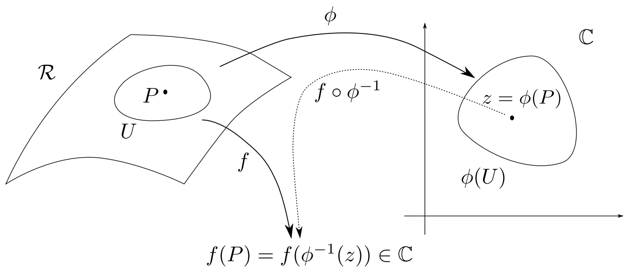

Fig. 6.6 Local coordinate system.

As the limits in U and in $\phi(U)$ should correspond via ϕ, we impose the following conditions:

- $\phi : U \to \phi(U)$ is a bijection (a one-to-one correspondence).
- $\phi : U \to \phi(U)$ as well as its inverse $\phi^{-1} : \phi(U) \to U$ is continuous.

Such a map is called a *homeomorphism* in topology. We call the pair (U, ϕ) a *local coordinate system* or simply a *local coordinate* and the open set U its *coordinate neighbourhood*. An important point is that it is sufficient to define a coordinate system not on the whole set $\mathcal{R}$ but only *locally on a neighbourhood U of each point P*.

[4] Those who know this notion well might say, "You are defining a *differentiable* manifold, not just a (topological) manifold". Yes, but as we use only complex manifolds later, we omit the word 'differentiable'.

[5] For the definition of a manifold we refer to [Na], [ST], [Tu].

Using a local coordinate (U, ϕ), we can define 'holomorphy (or differentiability) of a function f at a point P' as 'the function $f(\phi^{-1}(z))$ on $\phi(U) \subset \mathbb{C}$ is holomorphic (or differentiable) at the point $\phi(P)$' (cf. Fig. 6.6).

Now we require an important property of local coordinate systems. Since coordinate neighbourhoods are open sets, they may intersect. For example, assume that two local coordinates (U_0, ϕ_0) and (U_1, ϕ_1) intersect: $U_0 \cap U_1 \neq \varnothing$ (Fig. 6.7). If a function were holomorphic in one local coordinate system (U_0, ϕ_0) but not holomorphic in another local coordinate system (U_1, ϕ_1), we could not do complex analysis on the manifold. If $f(\phi_0^{-1}(z_0))$ is holomorphic in z_0, $f(\phi_1^{-1}(z_1))$ should be holomorphic in z_1, too. Note that z_0 is a function of z_1 through the composition $z_1 \mapsto \phi_1^{-1}(z_1) \mapsto z_0(z_1) = \phi_0(\phi_1^{-1}(z_1))$ via a point in $U_0 \cap U_1$. (Conversely, if we trace back this composition, z_1 is regarded as a function of z_0.) Therefore,

$$f(\phi_1^{-1}(z_1)) = f(\phi_0^{-1}(z_0(z_1))),$$

which means that $f(\phi_1^{-1}(z_1))$ is a composition of two functions, $f(\phi_0^{-1}(z_0))$ and $z_0(z_1)$. As we argued, when $f(\phi_0^{-1}(z_0))$ is holomorphic in z_0, we want $f(\phi_1^{-1}(z_1))$ to be holomorphic in z_1 automatically. So, we *require* that $z_0(z_1) = \phi_0(\phi_1^{-1}(z_1))$ should be a holomorphic function of z_1.

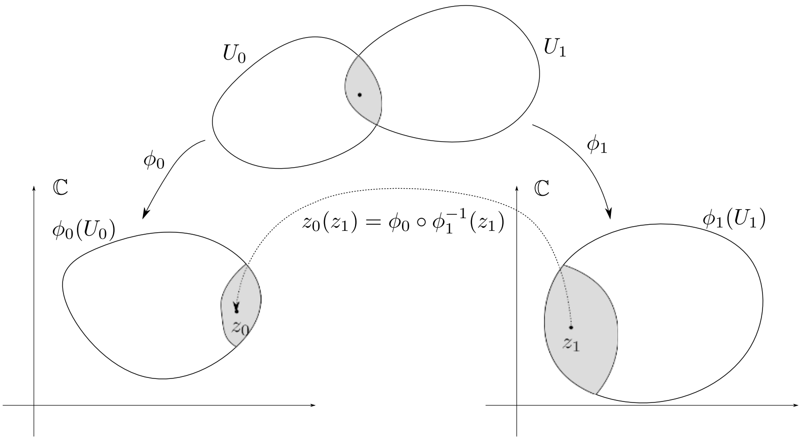

Fig. 6.7 Coordinate change on a manifold.

Then, as a composition of two holomorphic functions gives a holomorphic function, $f(\phi_1^{-1}(z_1)) = f(\phi_0^{-1}(z_0(z_1)))$ is holomorphic in z_1.

This requirement, 'When two local coordinate systems intersect, the coordinate change between them is holomorphic', is the main point in the definition of a complex manifold.[6]

[6] There are other requirements like 'being a Hausdorff space' or 'existence of sufficiently many local coordinate systems' but they are not used explicitly in this book. Replacing $\mathbb{C}$ by $\mathbb{C}^N$ and z

Example 6.1 When we learned complex analysis, we certainly encountered the Riemann sphere $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$. It is the union of the complex plane $\mathbb{C}$ and the point at infinity $\{\infty\}$, around which we use a coordinate $w = \dfrac{1}{z}$, where z is the coordinate on $\mathbb{C}$. The point at infinity $z = \infty$ naturally corresponds to $w = 0$. This is exactly the coordinate change which we explained above. The Riemann sphere $\mathbb{P}^1(\mathbb{C})$ is a basic example of a one-dimensional complex manifold.

Let us denote the original $\mathbb{C}$ by U_0 and the set obtained by subtracting 0 from $\mathbb{C}$ and attaching ∞ instead by U_∞: $U_\infty := (\mathbb{C} \setminus \{0\}) \cup \{\infty\}$.

We use z as a local coordinate in U_0 and w in U_∞ (Fig. 6.8). On the intersection $U_0 \cap U_\infty = \mathbb{C} \setminus \{0\}$ there is a relation $z = \dfrac{1}{w}$, $w = \dfrac{1}{z}$ between z and w, which are holomorphic functions. (Note that $U_0 \cap U_\infty$ does not include $z = 0$ and $w = 0$.)

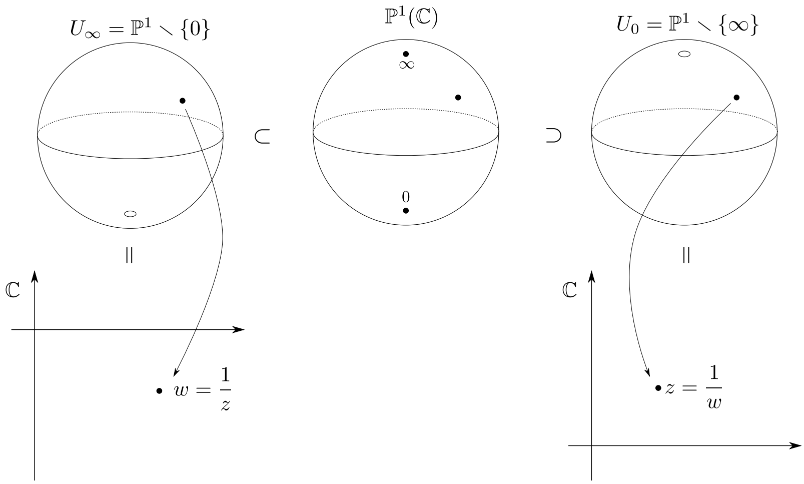

Fig. 6.8 The Riemann sphere $\mathbb{P}^1(\mathbb{C})$.

Thus we have prepared the terminology. Let us get back to the Riemann surface of $\sqrt{z}$. The set defined by (6.2) is a one-dimensional complex manifold, as is guaranteed by the following theorem.

Theorem 6.2 *Assume that a polynomial $F(z,w)$ of two variables z and w with complex coefficients satisfies* $\left(F, \dfrac{\partial F}{\partial z}, \dfrac{\partial F}{\partial w}\right) \neq (0,0,0)$ *on a domain U in $\mathbb{C}^2$.*

Then $\{(z,w) \mid F(z,w) = 0\} \cap U$ is a one-dimensional complex manifold.

Proof (Idea) The most important part of the proof of this theorem is to show that 'a holomorphic coordinate change is possible', which is essentially due to the following lemma.

by $(z_1, \ldots, z_N)$ we obtain the definition of an N-dimensional complex manifold. We study only one-dimensional complex manifolds in this book.

Lemma 6.3 (Implicit function theorem)

Let $F(z,w)$ be a function of two variables z and w on a domain $U \subset \mathbb{C}^2$, which is holomorphic[7] *in z as well as in w. Assume that $F(z_0,w_0) = 0$ and $\dfrac{\partial F}{\partial w}(z_0,w_0) \neq 0$ at a point $(z_0,w_0) \in U$. Then,*

- *for sufficiently small $r, \rho > 0$ the projection*

$$\left\{(z,w) \,\middle|\, \begin{array}{c} |z-z_0| < r,\ |w-w_0| < \rho, \\ F(z,w) = 0 \end{array} \right\} \ni (z,w) \mapsto z \in \{z \mid |z-z_0| < r\} \tag{6.3}$$

  is bijective.
- *The component $\varphi(z)$ of its inverse $z \mapsto (z, \varphi(z))$ is a holomorphic function of z.* □

We postpone the proof of Lemma 6.3 to the end of this subsection. Using this lemma, we can prove Theorem 6.2 as follows. For example, assume that $F(z_0,w_0) = 0$ and $\dfrac{\partial F}{\partial w}(z_0,w_0) \neq 0$. Then, Lemma 6.3 gives a one-to-one correspondence between a neighbourhood of z_0 in $\mathbb{C}$ and a neighbourhood of (z_0,w_0) in $\mathcal{R} = \{F(z,w) = 0\}$. We can take z as a local coordinate of $\mathcal{R}$ in the neighbourhood of (z_0,w_0) by this correspondence (Fig. 6.9).

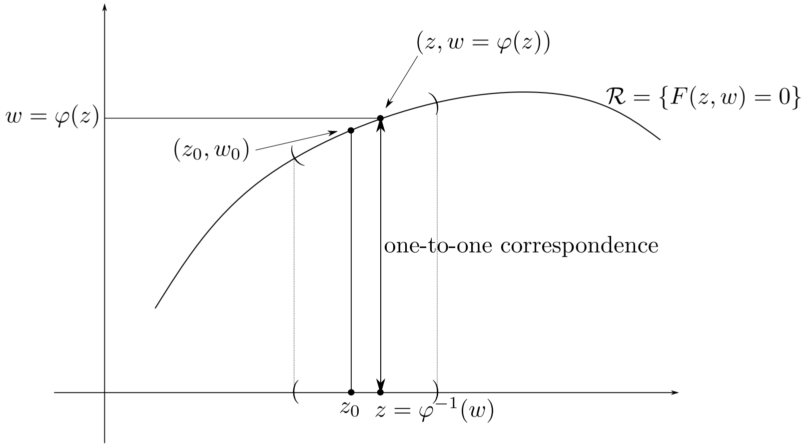

Fig. 6.9 Coordinate z of $\mathcal{R} = \{F(z,w) = 0\}$. (This is a schematic picture. The true picture would be four-dimensional over $\mathbb{R}$, as z and w are complex numbers.)

In a neighbourhood of a point (z_0,w_0), at which $\dfrac{\partial F}{\partial z}(z_0,w_0) \neq 0$, let us use w as a local coordinate of $\mathcal{R}$. If both $\dfrac{\partial F}{\partial w}(z_0,w_0) \neq 0$ and $\dfrac{\partial F}{\partial z}(z_0,w_0) \neq 0$ hold, we

[7] Here in Theorem 6.2 $F(z,w)$ is a polynomial but we shall use this lemma for more general functions later.

can use either z or w as a local coordinate. In this case, Lemma 6.3 says that both $z \mapsto w(z) := (w \text{ satisfying } F(z,w) = 0)$ and $w \mapsto z(w) := (z \text{ satisfying } F(z,w) = 0)$ are holomorphic functions. This means that two local coordinates z and w are connected via *holomorphic* coordinate changes $z \mapsto w(z)$ and $w \mapsto z(w)$.

We omitted details[8] but essentially we have defined the structure of a one-dimensional complex manifold on the set $\mathcal{R} = \{(z,w) \mid F(z,w) = 0\}$.

This is the end of the proof of Theorem 6.2. □

In the terminology of algebraic geometry $\mathcal{R}$ is also called a '*non-singular algebraic curve*', because

- 'non-singular': There is no singular point on $\mathcal{R} = \{(z,w) \mid F(z,w) = 0\}$. This means that at any point of $\mathcal{R}$ Lemma 6.3 is applicable. (A singular point is characterised by the condition $\dfrac{\partial F}{\partial w} = \dfrac{\partial F}{\partial z} = 0$.)
- 'algebraic': F is a polynomial.
- 'curve': $\mathcal{R}$ is one-dimensional (a coordinate is one variable) over the complex field $\mathbb{C}$.

Example 6.4 Let us return to the Riemann surface of $\sqrt{z}$. In this case $F(z,w) = w^2 - z$ and $\mathcal{R} = \{(z,w) \mid w^2 = z\}$. Since

$$\frac{\partial F}{\partial w} = 2w, \qquad \frac{\partial F}{\partial z} = -1,$$

the conditions of Theorem 6.2 are satisfied. In addition, it follows from the discussion in the proof of the theorem that

- z can be taken as a local coordinate except at $(z,w) = (0,0)$.
- w serves as a local coordinate everywhere.

Having defined the Riemann surface in this way, the function $\sqrt{z}$ on $\mathcal{R}$ can be defined as a projection $(z,w) \mapsto w$. A merit of this definition is that it is obvious that $\sqrt{z}$ is defined *naturally everywhere on* $\mathcal{R}$ and becomes *a holomorphic function* of coordinates. Recall that the real function $\sqrt{x}$ ($x \in [0,+\infty)$) is not differentiable at $x = 0$. In contrast, our $\sqrt{z} : \mathcal{R} \to \mathbb{C}$ is holomorphic even at the origin $(0,0)$. In fact we use w as a coordinate at the origin and $\sqrt{z} = w$ is indeed a holomorphic function of w.

Before closing this subsection, let us prove the implicit function theorem, Lemma 6.3.

Proof (of Lemma* 6.3*) Perhaps the reader who has already encountered a similar theorem in a calculus course on real functions might ask "Isn't this implicit function theorem a trivial consequence of that theorem?" Definitely no. In order to prove

[8] In particular, we did not prove the conditions which we omitted in the definition of the complex manifold.

Lemma 6.3 from the implicit function theorem for real functions, we at least need to pay attention to the following points.

- To apply the real implicit function theorem, we have to regard a complex function as two real functions, the real part and the imaginary part. Accordingly, the condition 'the derivative does not vanish' becomes 'the Jacobian determinant does not vanish'.
- As the real implicit function theorem guarantees only 'differentiability of the resulting function as a real function', we have to check the holomorphicity of $\varphi(z)$ separately.

These gaps can be filled by using the Cauchy–Riemann equations.

Here, not using the real implicit function theorem, we prove Lemma 6.3 with the help of the *argument principle* and its generalisation in complex analysis.

Let $f(w)$ be the value of $F(z,w)$ at (z_0,w): $f(w) := F(z_0,w)$. By the assumption, $f(w_0) = 0$ and $f'(w_0) \neq 0$, which means that w_0 is a zero of $f(w)$ of the first order. Therefore, if we take sufficiently small ρ, the only zero of f in $\{w \mid |w-w_0| \leq \rho\}$ is w_0. Applying the argument principle (Theorem A.10) to f, we have

$$\text{(6.4)} \qquad \frac{1}{2\pi i}\oint_{|w-w_0|=\rho} \frac{f'(w)}{f(w)}\,dw = (\text{number of zeros in } |w-w_0| < \rho) = 1.$$

Now, let us count the number of zeros, varying z. Since $F(z_0,w) = f(w)$ does not vanish on the circle $\{w \mid |w-w_0| = \rho\}$ and $F(z,w)$ is continuous, $F(z,w) \neq 0$ on the same circle, $\{w \mid |w-w_0| = \rho\}$, if z is close to z_0 (if $|z-z_0| < r$ for certain small $r > 0$). Applying the argument principle (Theorem A.10) to $F(z,w)$ as a function of w, we have

$$\frac{1}{2\pi i}\oint_{|w-w_0|=\rho} \frac{\dfrac{\partial F}{\partial w}(z,w)}{F(z,w)}\,dw = \begin{pmatrix} \text{number of } w \text{ in } |w-w_0| < \rho \\ \text{such that } F(z,w) = 0 \end{pmatrix}.$$

Let us denote this number by $N(z)$, which is a non-negative integer, because it is a number of points. On the other hand, it is a continuous function of z, because it is expressed as an integral of a continuous function of z as in the left-hand side. (See also Lemma A.6.) It follows that $N(z)$ is an 'integer-valued continuous function', which should be constant in a neighbourhood of z_0. From (6.4) it follows that $N(z_0) = 1$, and, consequently, we obtain $N(z) \equiv 1$. In other words, 'there exists one and only one w for each z such that $F(z,w) = 0$'. Thus we have proved that the map (6.3) is a one-to-one correspondence.

Let us denote its inverse map by

$$z \mapsto (z, \varphi(z)).$$

This is equivalent to $F(z,\varphi(z)) = 0$. We can apply the generalised argument principle (Theorem A.11) to $F(z,w)$ as $g(w)$, w as $\varphi(w)$ and the circle $\{w \mid |w-w_0| = \rho\}$ as the contour C. There is only one zero and no pole of $g(w) = F(z,w)$ inside this circle. Therefore the right-hand side of (A.3) becomes the position of the zero of $g(w) = F(z,w)$, which is nothing but $\varphi(z)$:

$$\varphi(z) = \frac{1}{2\pi i} \oint_{|w-w_0|=\rho} w \, \frac{\dfrac{\partial F}{\partial w}(z,w)}{F(z,w)} \, dw.$$

The integrand being a holomorphic function of z, the integral depends on z holomorphically. This means that $\varphi(z)$ is a holomorphic function of z. □

6.1.4 The Riemann surface of $\sqrt{1-z^2}$

Now let us take a small step from $\sqrt{z}$ toward an elliptic integral. When we study elliptic integrals, we need $\sqrt{\varphi(z)}$, where $\varphi(z)$ is a polynomial of degree 3 or 4, but here, as an intermediate step, we consider a function $f(z) = \sqrt{1-z^2}$ and its Riemann surface.

In the case of $\sqrt{z}$ a problem of sign occurs when we go around the origin = the point where the inside of the radical symbol vanishes. In the present case of $\sqrt{1-z^2}$, too, we first examine the situation around the points $z = \pm 1$ where the inside of the radical symbol vanishes. For example, let us take a point $z = 1 - r\, e^{i\varphi}$ (r is a positive constant) on a circle with centre 1 and small radius r. This point goes around 1 when the argument φ changes from 0 to 2π. The value of the function f at this point is

$$f(z) = \sqrt{r\, e^{i\varphi}(2 - r\, e^{i\varphi})} = \sqrt{r}\, e^{i\varphi/2} \sqrt{2 - r\, e^{i\varphi}}.$$

The last factor, $\sqrt{2 - r\, e^{i\varphi}}$, is a single-valued continuous function if $r \geq 0$ is small, since $2 - r\, e^{i\varphi}$ does not differ much from 2 and the argument of this complex number can be chosen uniquely (the situation of Fig. 6.4, left). But the factor $e^{i\varphi/2}$ causes a problem of multi-valuedness, as was the case with the function $\sqrt{z}$. Hence every time $z = 1 - r\, e^{i\varphi}$ goes around 1, $f(z)$ changes its sign and cannot be a single-valued function in a neighbourhood of 1. Similarly $f(z)$ changes its sign when z goes around -1 (Fig. 6.10).

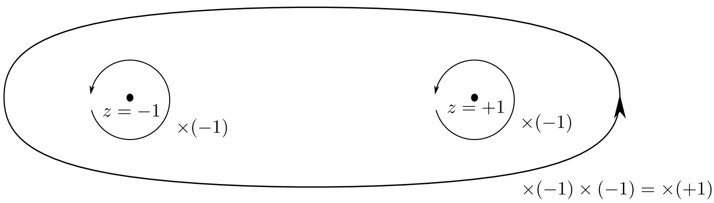

Fig. 6.10 Change of sign of $\sqrt{1-z^2}$.

So, in this case we have two *branch points*, around which we cannot take a neighbourhood to make the function single-valued, and $f(z)$ changes as follows:

- When z goes around near $+1$ or -1, $f(z)$ changes its sign.
- When z goes around *both* $+1$ and -1, $f(z)$ does not change sign.

Taking this observation into account, let us construct the Riemann surface of $f(z) = \sqrt{1-z^2}$. It turns out that the above problem is remedied by gluing two $\mathbb{C}$'s with slits along the interval $[-1,+1]$ (Fig. 6.11).

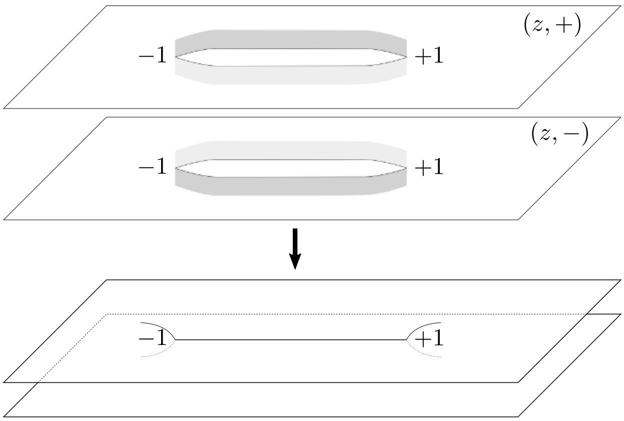

Fig. 6.11 The Riemann surface of $\sqrt{1-z^2}$.

This means that using two complex planes without ± 1, $\mathbb{C} \setminus \{\pm 1\}$, we construct the Riemann surface $\mathcal{R}$ as

$$\mathcal{R} = (\mathbb{C} \setminus \{\pm 1\})_+ \cup \{-1,+1\} \cup (\mathbb{C} \setminus \{\pm 1\})_-.$$

Here we followed the first method of the construction of the Riemann surface of $\sqrt{z}$ in the previous subsection.

Another construction of the Riemann surface of $\sqrt{z}$ used the equation $w^2 - z = 0$, which $w = \sqrt{z}$ satisfies. In the case of $f(z) = \sqrt{1-z^2}$ the equation is $F(z,w) := z^2 + w^2 - 1 = 0$, as $f(z)$ satisfies $z^2 + f(z)^2 = z^2 + (1-z^2) = 1$. Thus we obtain the second description of the Riemann surface of $\sqrt{1-z^2}$ as follows.

$$\mathcal{R} = \{(z,w) \mid F(z,w) := z^2 + w^2 - 1 = 0\}. \tag{6.5}$$

Since

$$\frac{\partial F}{\partial w} = 2w, \qquad \frac{\partial F}{\partial z} = 2z,$$

the implicit function theorem (Lemma 6.3) tells us the following:

- We can use z as a local coordinate around a point (z_0, w_0), where $w_0 \neq 0$ (i.e., $z_0 \neq \pm 1$).

- Around the excluded two points $(\pm 1, 0)$ we can take w as a local coordinate.

The Riemann surface $\mathcal{R}$ defined by (6.5) becomes a one-dimensional complex manifold with these coordinates. The function $f(z) = \sqrt{1-z^2}$ can be redefined on $\mathcal{R}$ by the projection

$$f : \mathcal{R} \ni (z, w) \mapsto w \in \mathbb{C},$$

which is a *single-valued holomorphic* function.

Well, then, what is the shape of the Riemann surface $\mathcal{R}$ constructed in this way? You might think, "What are you asking now? We have already drawn a picture of $\mathcal{R}$, Fig. 6.11." However, if we take this picture as it is, we will be tricked. In Fig. 6.11 the two sheets intersect along the interval $[-1, 1]$. But, in reality, *there is no intersection*. In fact, at each $z \in [-1, 1]$ there are two points $(z, w) = (z, \pm\sqrt{1-z^2})$ with different signs on $\mathcal{R}$ and we have to distinguish two sheets even along the apparent intersection, as was the case with $\sqrt{z}$.

To express the Riemann surface visually a little more accurate, it is enough to change the orientation of the sheets when we glue them. In Fig. 6.11 we glued the sheets with the same orientation, so that it was evident that we were gluing 'two same complex planes'. If we glue one sheet as it was and another flipped upside down, we have the same but different-looking Riemann surface as in Fig. 6.12.

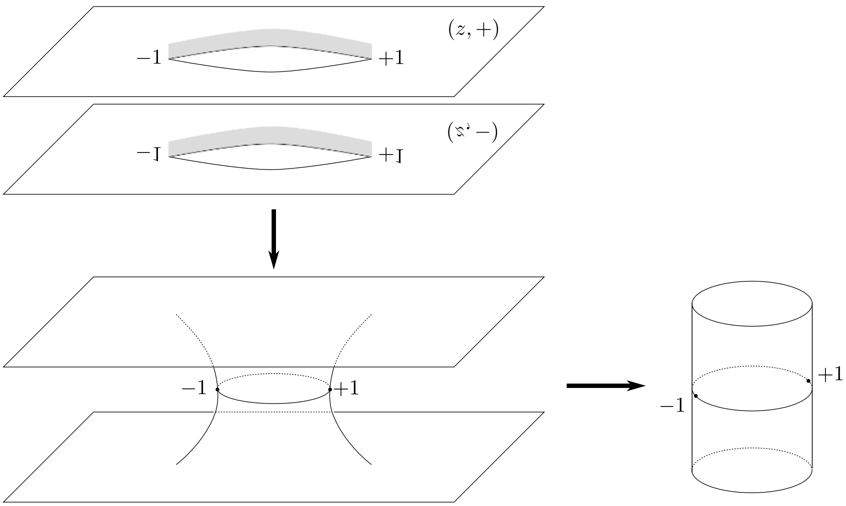

Fig. 6.12 Another construction of the Riemann surface of $\sqrt{1-z^2}$.

As you see, the shape of $\mathcal{R}$ is a cylinder. Using the terminology of topology, we can say that $\mathcal{R}$ is *homeomorphic* to a cylinder.

6.2 Analysis on Riemann Surfaces

6.2.1 Integrals on Riemann surfaces

Let us recall our original goal: we want to consider elliptic integrals, for example

$$\int \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}},$$

with complex variables.

What is an integral with a complex variable? The integral of a complex-valued function $f(z)$ along a smooth curve γ in $\mathbb{C}$ is defined by

$$\int_\gamma f(z)\,dz := \int_a^b f\bigl(z(t)\bigr)\,\frac{dz}{dt}(t)\,dt,$$

where $[a,b] \ni t \mapsto z(t) \in \mathbb{C}$ is a parametrisation of γ. We can use any smooth parametrisation and obtain the same result so far as the orientation is preserved, which is a direct consequence of the formula of change of variables for integration over a segment. The formula of change of integration variables for complex integration,

$$\int_{g(\gamma)} f(z)\,dz = \int_\gamma f\bigl(g(\tilde{z})\bigr)\,g'(\tilde{z})\,d\tilde{z},$$

follows from this definition. Here $g(\tilde{z})$ is a holomorphic function defined in a neighbourhood of γ. We can interpret the last parts in both sides as '$dz = \dfrac{dz}{d\tilde{z}}d\tilde{z}$', which is parallel to the real case. Any course on complex analysis contains such definitions and formulae.

In view of this formula of change of variables, it is more natural to consider that the 'object which is integrated' is not a function $f(z)$ itself but a function with something extra, a combination '$f(z)\,dz$', which changes as

$$f(z)\,dz \mapsto f\bigl(g(\tilde{z})\bigr)\,g'(\tilde{z})\,d\tilde{z}$$

when the integration variable is changed. Such an object $f(z)\,dz$ is called a *one-form* (exactly speaking, a *differential one-form*).

We want to consider elliptic integrals as complex integrals, but before that let us consider $\displaystyle\int \frac{dz}{\sqrt{1-z^2}}$ with complex variables as a prototype.

As a matter of course, we cannot determine values of the integrand, unless the function $\sqrt{1-z^2}$ is single-valued. So we think of the one-form $\omega = \dfrac{dz}{\sqrt{1-z^2}}$ as an object on the Riemann surface of the function $\sqrt{1-z^2}$. The Riemann surface has complex coordinates, which gives a local identification of a subset of the Riemann surface with a subset of $\mathbb{C}$. Therefore we can integrate ω by fixing a coordinate and using the above reviewed complex integration.

"But you said that a local coordinate cannot be fixed, didn't you?", you might ask. Yes, you are right. On a Riemann surface, i.e., a one-dimensional complex manifold, we might need to use different coordinates at different places. This is why we introduced 'one-forms' above. The definition of one-forms on a Riemann surface is the same as that of one-forms on $\mathbb{C}$: locally a one-form has the form $f(z)\,dz$ in a coordinate z. If the coordinate is changed to $\tilde{z}$, it changes to $f(z)\dfrac{dz}{d\tilde{z}}\,d\tilde{z}$. This matches the formula of change of integration variable, and hence the value of the integral of a one-form on a Riemann surface is uniquely determined.

Now let us return from the general theory to the integral $\displaystyle\int \frac{dz}{\sqrt{1-z^2}}$. First we examine the integrand, or, exactly speaking, the one-form $\omega = \dfrac{dz}{\sqrt{1-z^2}}$. The coefficient function $\dfrac{1}{\sqrt{1-z^2}}$ of dz is a holomorphic function of z when $z \neq \pm 1$. Therefore it is a holomorphic one-form on $\mathbb{C} \smallsetminus \{\pm 1\}$.

How about at $z = \pm 1$? Since the denominator vanishes, ω is not defined there? Well, yes, if we are considering ω as a one-form on $\mathbb{R}$. Things are different on the Riemann surface $\mathcal{R}$. As we discussed in the previous section, we cannot take z as a coordinate in neighbourhoods of $(z, w) = (\pm 1, 0) \in \mathcal{R}$. Instead we can use w as a coordinate. How can we express ω by this coordinate? Here the formula of change of variables of the one-form plays an essential role. Applying it to ω, we have

$$\omega = \frac{1}{\sqrt{1-z^2}}\,dz = \frac{1}{w}\frac{dz}{dw}\,dw.$$

Since $z = \sqrt{1-w^2}$, we can substitute $\dfrac{dz}{dw} = \dfrac{-w}{\sqrt{1-w^2}}$ into the above formula. Or, differentiating the relation $z^2 + w^2 - 1 = 0$, we have

$$2z\frac{dz}{dw} + 2w = 0, \text{ which means } \frac{1}{w}\frac{dz}{dw} = -\frac{1}{z}.$$

Anyway, the result is

$$\omega = -\frac{dw}{z} = -\frac{dw}{\sqrt{1-w^2}}.$$

In a neighbourhood of $(z, w) = (\pm 1, 0)$ the denominator of the right-hand side does not vanish and, as a consequence, the coefficient function of dw, $-\dfrac{1}{z} = -\dfrac{1}{\sqrt{1-w^2}}$ is a holomorphic function there.

Thus $\omega = \dfrac{dz}{w} = -\dfrac{dw}{z}$ is defined and holomorphic everywhere on $\mathcal{R}$.

6.2.2 Homology groups (a very short crash course)

We return to the complex plane $\mathbb{C}$ from the Riemann surface for a moment. Let $f(z)\,dz$ be an everywhere holomorphic one-form on $\mathbb{C}$ and define a function $F(z)$ by the following integral:

$$F(z) := \int_{z_0}^{z} f(z')\,dz'. \tag{6.6}$$

We fix the lower bound z_0 and regard this integral as a function of the upper bound z. As we recalled in the previous subsection, such a complex line integral is defined only when an integration contour from z_0 to z is specified. However the Cauchy integral theorem guarantees that the above integral does not depend on the intermediate contour. If the starting and end points of the paths C and C' in Fig. 6.13 coincide, the values of the integrals are the same:

$$\int_C f(z)\,dz = \int_{C'} f(z')\,dz'.$$

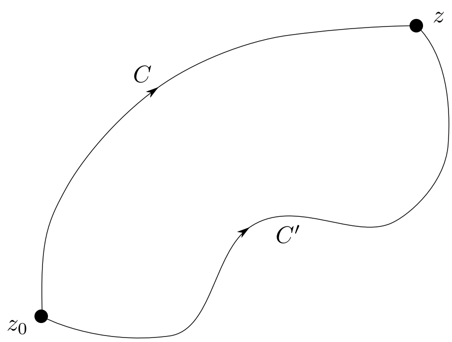

Fig. 6.13 Integrals on the plane.

Thus the expression (6.6) can define a single-valued function without specifying a contour explicitly.

Then, how about the integral of the one-form $\omega = \dfrac{dz}{\sqrt{1-z^2}}$, which is holomorphic everywhere on $\mathcal{R}$? In this case, values of the integral $\displaystyle\int_C \omega$ *may change depending on the contour* C because the *shape* of $\mathcal{R}$ is different from that of the complex plane.

For example, let us take two contours C_0 (Fig. 6.14) and C_1 (Fig. 6.15) connecting two points P and Q on $\mathcal{R}$. The left picture in each figure is the z-plane (the sheet to be pasted in the construction of $\mathcal{R}$) and the right is the cylinder version in Fig. 6.12.

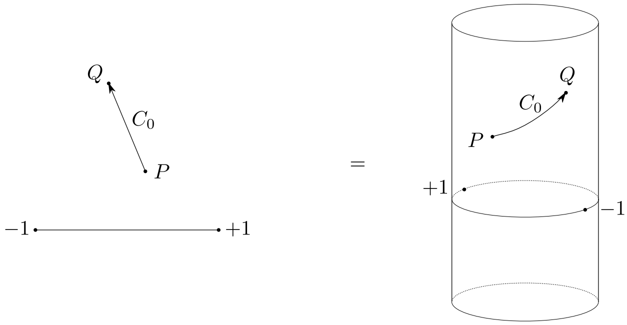

Fig. 6.14 Contour C_0 on $\mathcal{R}$.

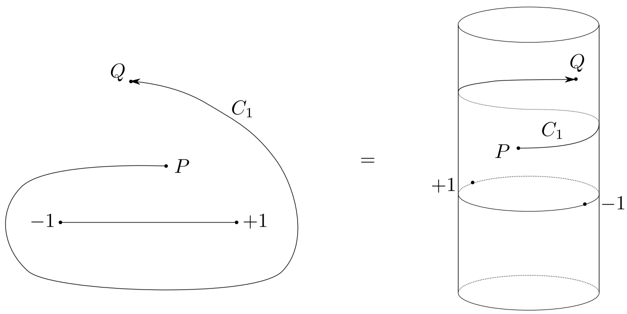

Fig. 6.15 Contour C_1 on $\mathcal{R}$.

When we deform the contour C_1 to the contour C_1' in Fig. 6.16 continuously, we can apply the Cauchy integral theorem and the value of the integral does not change:

$$\int_{C_1} \omega = \int_{C_1'} \omega.$$

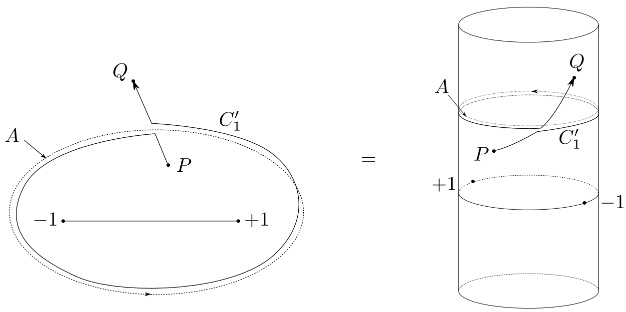

Fig. 6.16 Contour C_1 is deformed to C_1' on $\mathcal{R}$.

The curve consisting of the first and the last parts of C_1' is the same as C_0. Thus the difference of the integrals along C_1' and along C_0, which is equal to the difference of the integrals along C_1 and along C_0, is the integral along the intermediate loop A:

$$\int_{C_1} \omega - \int_{C_0} \omega = \int_{C_1'} \omega - \int_{C_0} \omega = \int_A \omega.$$

Let us take another, slightly more complicated example C_2 in Fig. 6.17. The dotted curve in the left picture here is the part in the lower sheet of the two sheets glued to form the Riemann surface. Comparing the left and the right pictures clarifies what this means.

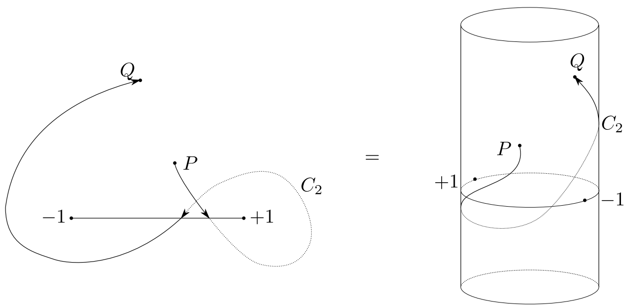

Fig. 6.17 Contour C_2 on $\mathcal{R}$.

By deforming the curve C_2 we obtain the curve C_2' in Fig. 6.18. In this case the difference of the integrals is the integral along A but in the opposite direction. (We call it 'the integral on $-A$'.) The value of the integral is multiplied by -1, as the direction is opposite:

$$\int_{C_2} \omega - \int_{C_0} \omega = \int_{C_2'} \omega - \int_{C_0} \omega = \int_{-A} \omega = -\int_A \omega.$$

Now let us study the general case. For that purpose it is convenient to use terminology from topology, in particular, *homology*. We explain very briefly the part of homology theory which is necessary for us. What will be important later is not the definitions but the examples. So, you do not have to pay attention to every detail of the following definitions. For details, we refer, for example, to [Ha], [ST], or §4.3 of [Ah].

Let $C_1, \ldots, C_N$ be closed curves in $\mathcal{R}$ and $n_1, \ldots, n_N$ be integers. An 'object' of the form

$$C = n_1 C_1 + \cdots + n_N C_N \tag{6.7}$$

is called a *1-dimensional cycle* or a *1-cycle*, or just a *cycle*, when there is no risk of confusion. Please do not worry about the meaning of 'integer multiple of curves' or

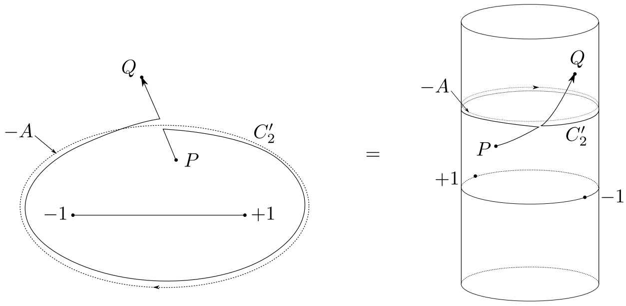

Fig. 6.18 Contour C_2 is deformed to C_2' on $\mathcal{R}$.

'sums of curves'. Here we take the above expression just as a combined symbol. As you can guess from this notation, we introduce a natural summation of cycles just by 'summing up coefficients'. We regard a trivial curve ('a curve not moving from a point') as a zero-element of the addition. This means that we can remove or add such curves freely to the above linear combination of curves. We denote this abelian group by Z_1.

An equivalence relation $\sim$ on Z_1 is defined by the following:

- If two closed curves C and C' can be deformed to each other continuously (in other words, *C and C' are homotopically equivalent*)[9], they are equivalent: $C \sim C'$,
- If the start point = the end point of C and the start point = the end point of C' coincide, we can connect them to get a new curve $C \cdot C'$, which, by definition, is equivalent to $C + C'$: $C \cdot C' \sim C + C'$,
- Given a closed curve C, we can reverse the orientation and obtain a curve $-C$, which is, by definition, equivalent to $(-1)C$: $-C \sim (-1)C$.

(See Fig. 6.19.)

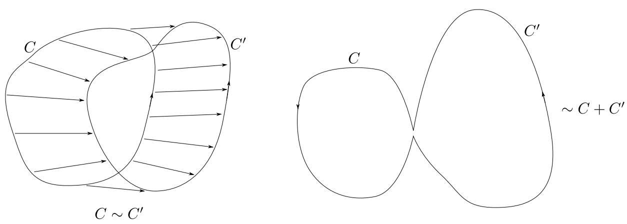

Fig. 6.19 The equivalence in Z_1.

The *first homology group* $H_1(\mathcal{R}, \mathbb{Z})$ of the Riemann surface $\mathcal{R}$ is defined as the factor additive group of Z_1 by $\sim$:

[9] The main point here is that we can apply the Cauchy integral theorem to the integrals along C and along C', which gives the equality of those integrals.

$$H_1(\mathcal{R},\mathbb{Z}) := Z_1/\sim .$$

(As we use only the first homology group in this book, we call it the *homology group*.) The equivalence class containing $C \in Z_1$ is an element of $H_1(\mathcal{R},\mathbb{Z})$ and denoted by $[C]$.

Example 6.5 Any closed curve in the plane $\mathbb{C}$ continuously shrinks to a point, which means that any cycle is equivalent to the zero element: $H_1(\mathbb{C},\mathbb{Z}) = 0$.

Example 6.6 The homology group of the Riemann surface $\mathcal{R}$ of $\sqrt{1-z^2}$ constructed in the previous section is generated by the equivalence class of the closed curve A in Fig. 6.16:

$$H_1(\mathcal{R},\mathbb{Z}) = \mathbb{Z}[A] = \{n[A] \mid n \in \mathbb{Z}\}.$$

Any closed curve on the cylinder can be deformed to the curve which iterates the closed curve A (or $-A$) several times. Therefore closed curves are classified by the integer which describes how many times they go around the cylinder.

The integral of a holomorphic 1-form α along a 1-cycle (6.7) is defined by

$$\int_C \alpha = n_1 \int_{C_1} \alpha + \cdots + n_N \int_{C_N} \alpha.$$

The integrals of α along equivalent cycles C and C', $C \sim C'$, coincide because of the Cauchy integral theorem and the definition of the equivalence relation $\sim$, which guarantees the well-definedness of the integral $\int_{[C]} \alpha$ for an element $[C]$ of $H_1(\mathcal{R},\mathbb{Z})$. In this way the notion of homology is compatible with the integrals.

6.2.3 Periods of one-forms

We return to the integral of $\omega = \dfrac{dz}{\sqrt{1-z^2}}$. Using the terminology of homology, we can restate the examples in Section 6.2.1 as follows. The closed curve $C_1 - C_0$ obtained by connecting C_0 with reverse orientation to C_1 is equivalent to A in the homology group $H_1(\mathcal{R},\mathbb{Z})$: $[C_1 - C_0] = [A]$. Similarly, $[C_2 - C_0] = -[A]$ and correspondingly we have equations for integrals:

$$\int_{C_1} \omega - \int_{C_0} \omega = \int_A \omega, \quad \int_{C_2} \omega - \int_{C_0} \omega = -\int_A \omega.$$

In the same way, we associate to an arbitrary curve C from P to Q a closed curve $C - C_0$ obtained by connecting C_0 with reverse orientation. Since $H_1(\mathcal{R},\mathbb{Z}) = \mathbb{Z}[A]$ (Example 6.6), there exists an integer n such that $[C - C_0] = n[A]$. Thus,

$$\int_C \omega - \int_{C_0} \omega = n \int_A \omega.$$

This means that 'the integral $\int_P^Q \omega$' is not determined by the start point P and the end point Q only and has an ambiguity of an integral multiple of $\int_A \omega$. We call $\int_A \omega$ the *period* of the 1-form ω on A.

Now, let us compute this period explicitly. First, we shrink A to the boundary of the slit $[-1, 1]$ (Fig. 6.20).

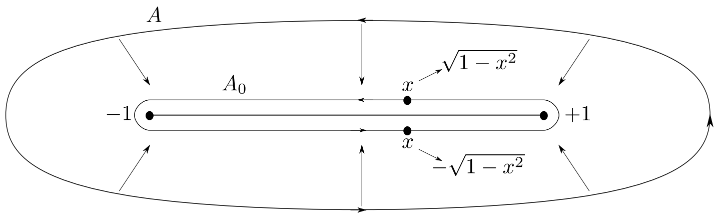

Fig. 6.20 Shrink A to the slit.

Of course this does not change the value of the integral: $\int_A \omega = \int_{A_0} \frac{dx}{\sqrt{1-x^2}}$. Here it is important that the square root $\sqrt{1-x^2}$ changes its sign when x goes around ± 1. So, if we take the positive square root on the slit $[-1, 1]$, when x moves from -1 to 1 in Fig. 6.20,

$$\begin{aligned}\int_{A_0} \omega &= \int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}} + \int_{1}^{-1} \frac{-dx}{\sqrt{1-x^2}} \\ &= \arcsin x|_{x=-1}^{x=1} - \arcsin x|_{x=1}^{x=-1} \\ &= \left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) - \left(\left(-\frac{\pi}{2}\right) - \frac{\pi}{2}\right) = 2\pi.\end{aligned}$$

Thus we have obtained 2π as the period of ω on A.

"Oh? I know something which has 'period 2π'." You have a keen eye. We can rephrase what we have computed as follows. Fix the start point P of the integral to 0. The integral

$$u(z) = \int_0^z \omega$$

(the integration contour of which is intensionally unfixed) changes its value by an integral multiple of 2π when the point z moves over the plane from $x \in \mathcal{R}$ and returns to x: $u(x) \rightsquigarrow u(x) + 2\pi n$, $n \in \mathbb{Z}$.

Hence, roughly speaking, the "inverse function" $x(u)$ of $u(x)$ does not change its value when u is shifted by an integral multiple of 2π:

$$x(u + 2\pi n) = x(u), \quad n \in \mathbb{Z}.$$

In fact,

$$u(x) = \int_0^x \frac{dx}{\sqrt{1-x^2}} = \arcsin x, \quad x(u) = \sin u,$$

and, as you know well, $\sin u$ is a periodic function with period 2π. We can say that we have provided a geometric derivation of the periodicity of $\sin u$ from the topological properties of the cylinder.

Now we are ready to consider complex elliptic integrals. In the next chapter we introduce a Riemann suface called an ‘elliptic curve’, on which elliptic integrals live.

Chapter 7
Elliptic Curves

In the previous chapter we studied Riemann surfaces of functions like $\sqrt{z}$ and $\sqrt{1-z^2}$ and an integral of the form $\displaystyle\int \frac{dz}{\sqrt{1-z^2}}$ as a prototype of an integral on a Riemann surface. It is time to come back to our main topic, elliptic integrals. As we discussed in Chapter 1 and Chapter 2, elliptic integrals are of the form $\displaystyle\int R(z, \sqrt{\varphi(z)})\, dz$, where $R(z,s)$ is a rational function of two variables and $\varphi(z)$ is a polynomial of degree three or four without multiple roots. In order to consider such integrals with complex variables, we need a Riemann surface, on which $\sqrt{\varphi(z)}$ is a single-valued function. Moreover, in contrast to the integrals studied in the previous chapter, we can extend contours of elliptic integrals to infinity, which motivates us to add infinities to the Riemann surfaces and to introduce one-dimensional complex manifolds called 'elliptic curves'.

7.1 The Riemann Surface of $\sqrt{\varphi(z)}$

Let $\varphi(z)$ be a polynomial of degree three or four without multiple roots. The construction of the Riemann surface $\mathcal{R}$ of $\sqrt{\varphi(z)}$ is, in principle, the same as those of $\sqrt{z}$ and $\sqrt{1-z^2}$, which we considered in Chapter 6.

• <u>the case $\deg \varphi(z) = 3$</u>: Let us factorise $\varphi(z)$ as

$$\varphi(z) = a(z-\alpha_1)(z-\alpha_2)(z-\alpha_3). \tag{7.1}$$

From the assumption on $\varphi(z)$ it follows that $a \neq 0$ and $\alpha_1, \alpha_2, \alpha_3$ are distinct from each other.

The function $\sqrt{z}$ changes its sign when z goes around 0, and this is the reason why the Riemann surface of $\sqrt{z}$ has a branch point at $z = 0$. For the same reason the points $z = \pm 1$ are the branch points of the Riemann surface of the function $\sqrt{1-z^2}$.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_8

For the function $\sqrt{\varphi(z)}$ the same is true. The function $\varphi(z)$ vanishes at three points $z = \alpha_i$ $(i = 1,2,3)$ which become the branch points: If z goes around one of these points, $\sqrt{\varphi(z)}$ gets multiplied by -1. If z goes all the way around a simple (Jordan) curve which encircles two of those three points, the sign does not change and $\sqrt{\varphi(z)}$ gets back to its original value. If the inside of the curve contains all three points, the round trip along the curve changes the sign.

To realise this behaviour of the function $\sqrt{\varphi(z)}$ around α_i, we glue two sheets $\mathbb{C} \smallsetminus \{\alpha_1, \alpha_2, \alpha_3\}$ together, and then add three points $\{\alpha_1, \alpha_2, \alpha_3\}$ to obtain the Riemann surface $\mathcal{R}$ of $\sqrt{\varphi(z)}$.

$$\mathcal{R} = (\mathbb{C} \smallsetminus \{\alpha_1, \alpha_2, \alpha_3\})_+ \cup \{\alpha_1, \alpha_2, \alpha_3\} \cup (\mathbb{C} \smallsetminus \{\alpha_1, \alpha_2, \alpha_3\})_-. \tag{7.2}$$

How should we glue the sheets? Let us recall the construction of the Riemann surfaces of $\sqrt{z}$ and $\sqrt{1-z^2}$. The Riemann surface of $\sqrt{z}$ is obtained by gluing two sheets cut along one infinitely long half line. In the case of $\sqrt{1-z^2}$, we cut sheets along a segment $[-1,1]$.

Similarly to these cases the Riemann surface of $\sqrt{\varphi(z)}$ is obtained by gluing sheets $(\mathbb{C} \smallsetminus \{\alpha_1, \alpha_2, \alpha_3\})_\pm$ cut along an infinitely long half line between α_1 and ∞ and along a segment $[\alpha_2, \alpha_3]$. You can confirm that this construction is compatible with the changes of the sign stated above and gives a single-valued function $\sqrt{\varphi(z)}$ on $\mathcal{R}$ by carefully chasing the value of the function on Fig. 7.1.

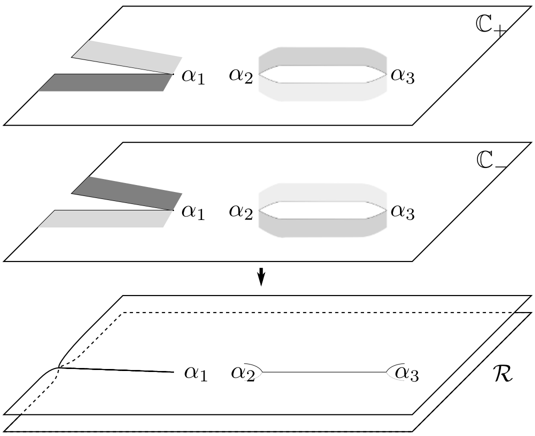

Fig. 7.1 Riemann surface of $\sqrt{\varphi(z)}$ ($\deg \varphi = 3$).

This construction does not seem to treat the three points α_1, α_2, α_3 equally. However, as was the case with $\sqrt{z}$ or $\sqrt{1-z^2}$, the "cuts" or "gluing" or "self-intersection" are nothing more than appearances on the figure. The points α_i *are* on an equal footing. This can be understood naturally by constructing $\mathcal{R}$ as

$$\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\} \tag{7.3}$$

by means of the implicit function theorem.

• the case $\deg\varphi(z) = 4$: Again we factorise $\varphi(z)$ as

$$\varphi(z) = a(z-\alpha_0)(z-\alpha_1)(z-\alpha_2)(z-\alpha_3). \tag{7.4}$$

($a \neq 0$ and the four complex numbers $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ are distinct.) The Riemann surface $\mathcal{R}$ of $\sqrt{\varphi(z)}$ is obtained by gluing two sheets (= planes without four branch points) $\mathbb{C} \setminus \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$ and adding the set of branch points $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$:

$$\mathcal{R} = (\mathbb{C} \setminus \{\alpha_0, \dots, \alpha_3\})_+ \cup \{\alpha_0, \dots, \alpha_3\} \cup (\mathbb{C} \setminus \{\alpha_0, \dots, \alpha_3\})_-. \tag{7.5}$$

When gluing together, the four roots of $\varphi(z)$ are divided into two sets of two points (for example, $\{\alpha_0, \alpha_1\}$ and $\{\alpha_2, \alpha_3\}$) and the cuts are made between two points of each set. In this case both cuts are of finite length. (See Fig. 7.2.)

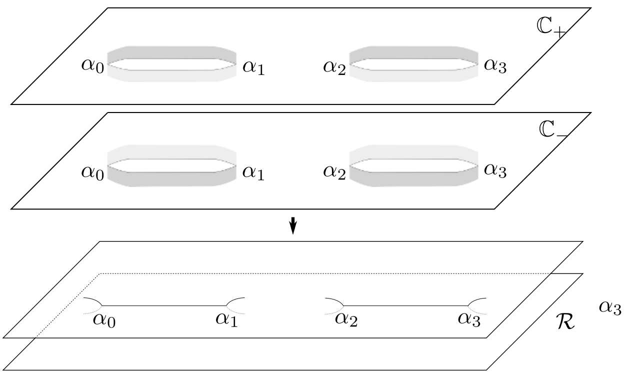

Fig. 7.2 Riemann surface of $\sqrt{\varphi(z)}$ ($\deg\varphi = 4$).

Similarly to the previously considered cases ($\sqrt{z}$, $\sqrt{1-z^2}$, $\sqrt{\varphi(z)}$ ($\deg\varphi = 3$)), we can express $\mathcal{R}$ as

$$\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}. \tag{7.6}$$

The following properties of the Riemann surface of $\sqrt{\varphi(z)}$ are proved in the same way as in Chapter 6.

Proposition 7.1 *In both cases* $\deg\varphi(z) = 3$ *and* $\deg\varphi(z) = 4$ *we have*

(i) $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$ *is a non-singular algebraic curve. (This means that* $\left(F, \dfrac{\partial F}{\partial z}, \dfrac{\partial F}{\partial w}\right) \neq (0,0,0)$, *where* $F(z,w) = w^2 - \varphi(z)$.*)*
(ii) $\sqrt{\varphi(z)} = w$ *is a holomorphic function on* $\mathcal{R}$.
(iii) The 1*-form* $\omega = \dfrac{dz}{\sqrt{\varphi(z)}} = \dfrac{dz}{w}$ *is holomorphic on* $\mathcal{R}$.

Exercise 7.2 Prove the above proposition. (Hint: (i) follows from the fact that $\varphi(z)$ does not have multiple roots. For (ii) and (iii) use local coordinates z (around a point which is not a branch point) and w (around branch points), after checking that they really are local coordinates.)

7.2 Compactification and Elliptic Curves

Up to this point we have seen no essential difference from the cases of $\sqrt{z}$ and $\sqrt{1-z^2}$. The only difference is the numbers of cuts in a sheet. But, when we consider elliptic integrals, we need to extend the Riemann surfaces constructed in the previous section a little bit.

The reason for this is that, if the degree of $\varphi(z)$ is more than two, improper integrals like $I = \displaystyle\int_{z_0}^{\infty} \frac{dz}{\sqrt{\varphi(z)}}$ converge. In fact, the absolute value of the integrand $\left|\dfrac{1}{\sqrt{\varphi(z)}}\right|$ decays as $\dfrac{1}{|z|^{\deg\varphi/2}}$ when $|z| \to \infty$. As is well known, an improper integral $\displaystyle\int_1^{\infty} \frac{dx}{x^{\sigma}}$ converges when $\sigma > 1$. Therefore the integral I converges when $\deg\varphi > 2$.

Hence it is natural to add '∞' to $\mathcal{R}$ and consider elliptic integrals there. This is similar to the construction of the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ (which is also called the *projective line*; we abbreviate it to $\mathbb{P}^1$) by adding a point at infinity to $\mathbb{C}$.

Our Riemann surfaces are subsets of $\mathbb{C}^2$ defined by (7.3) or (7.6). We want to add a point $(z, w) = (\infty, \infty)$ at infinity to them. Our strategy is to add infinities to the 'container' $\mathbb{C}^2$ (the *ambient space*) of $\mathcal{R}$ in such a way that they become infinities of $\mathcal{R}$. There are two methods. The first one uses the extension of $\mathbb{C}^2$ called the *projective plane*. This is a standard method in algebraic geometry. However a problem occurs when $\deg\varphi = 4$. So, as the second method, we show another way, which can be applied to both cases, $\deg\varphi = 3$ and $\deg\varphi = 4$.

7.2.1 Embedding of $\mathcal{R}$ into the projective plane

Before considering the projective *plane* $\mathbb{P}^2$, let us recall the construction of the projective *line*, $\mathbb{P}^1$. Of course, set-theoretically it is constructed just as a union, $\mathbb{C} \cup \{\infty\}$. But here the important thing is, as we have already discussed in Example 6.1, it has a manifold structure. In other words, there is a natural coordinate around the point at infinity. Denoting the coordinate on the complex plane $\mathbb{C}$ by z, we can use the coordinate $w = \dfrac{1}{z}$ in a neighbourhood of $z = \infty$.

For example, a fraction $z = \dfrac{x_1}{x_0}$ becomes $w = \dfrac{x_0}{x_1}$ in the coordinate around the point at infinity. If the denominator x_0 of z is not 0, we can divide both the numerator and the denominator by x_0 and we obtain just z, but if x_0 is 0, $z = \dfrac{x_1}{x_0}$ does not make sense as a complex number and, on the other hand, $w = \dfrac{x_0}{x_1}$ is 0. Moreover we can multiply both the numerator and the denominator by the same number λ ($\neq 0$), which does not change the fractions: $z = \dfrac{x_1}{x_0} = \dfrac{\lambda x_1}{\lambda x_0}$, $w = \dfrac{x_0}{x_1} = \dfrac{\lambda x_0}{\lambda x_1}$. (Sorry, this is too trivial.)

Based on this point of view, the projective line $\mathbb{P}^1$ is defined as a quotient set of the set $\mathbb{C}^2 \setminus \{(0,0)\}$ of pairs of complex numbers ($\neq (0,0)$) by the following equivalence relation:

$$(x_0,x_1) \sim (y_0,y_1) \overset{\text{def.}}{\iff} \begin{array}{l}\text{There exists a } \lambda\ (\neq 0) \text{ such} \\ \text{that } (\lambda x_0, \lambda x_1) = (y_0,y_1).\end{array}$$

Hereafter we denote the equivalence class including (x_0,x_1) by $[x_0 : x_1]$. This element is designated by two complex numbers, but thanks to the equivalence relation it is essentially reduced to one complex number $z = \dfrac{x_1}{x_0}$ or $w = \dfrac{x_0}{x_1}$. Each part of the decomposition $\mathbb{P}^1 = \mathbb{C} \cup \{\infty\}$ is interpreted as follows:

$$\mathbb{C} \ni z \leftrightarrow [1 : z] \in \mathbb{P}^1, \quad \infty \leftrightarrow [0 : 1] \in \mathbb{P}^1. \tag{7.7}$$

So far this is a review of the construction of the Riemann sphere = the projective line $\mathbb{P}^1$.

Let us consider a two-dimensional version of this construction. The quotient set of the set $\mathbb{C}^3 \setminus \{(0,0,0)\}$ of triples of complex numbers distinct from $(0,0,0)$ by the following equivalence relation is called the *projective plane* $\mathbb{P}^2(\mathbb{C})$ (hereafter $\mathbb{P}^2$):

$$(x_0,x_1,x_2) \sim (y_0,y_1,y_2) \overset{\text{def.}}{\iff} \begin{array}{l}\text{There exists a } \lambda\ (\neq 0) \text{ such that} \\ (\lambda x_0, \lambda x_1, \lambda x_2) = (y_0,y_1,y_2).\end{array}$$

The equivalence class containing (x_0,x_1,x_2) is denoted by $[x_0 : x_1 : x_2]$ as in the case of the projective line. As the complex plane $\mathbb{C}$ is embedded into $\mathbb{P}^1$ by (7.7), the subset

$$U_0 := \{[x_0 : x_1 : x_2] \mid x_0 \neq 0\} \subset \mathbb{P}^2 \tag{7.8}$$

of $\mathbb{P}^2$ is identified with $\mathbb{C}^2$ by the following correspondence:

$$\begin{aligned} \mathbb{C}^2 \ni (z,w) &\mapsto [1 : z : w] \in U_0, \\ U_0 \ni [x_0 : x_1 : x_2] &\mapsto \left(\frac{x_1}{x_0}, \frac{x_2}{x_0}\right) \in \mathbb{C}^2. \end{aligned} \tag{7.9}$$

The points of the form $[0 : x_1 : x_2]$, which do not belong to U_0, are not 'visible' in $\mathbb{C}^2$, as z and w in (7.9) diverge. This part is called the *line at infinity* added to $\mathbb{C}^2$. (The pair $[x_1 : x_2]$ can be regarded as a point of the projective line $\mathbb{P}^1$, hence *line*.)

Since the Riemann surface $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$ is a subset of $\mathbb{C}^2$, it can be regarded as a subset of $U_0 \subset \mathbb{P}^2$ through the above identification. Exactly speaking, we substitute $z = \dfrac{x_1}{x_0}$ and $w = \dfrac{x_2}{x_0}$ into the equation $w^2 = \varphi(z)$ defining $\mathcal{R}$ and regard the set of all $[x_0 : x_1 : x_2]$ which satisfy $x_0 \neq 0$ and the equation

$$\left(\frac{x_2}{x_0}\right)^2 = \varphi\left(\frac{x_1}{x_0}\right) \tag{7.10}$$

as $\mathcal{R}$. Let us add the points at infinity to $\mathcal{R}$ by extending $\mathcal{R}$ as a subset of $\mathbb{P}^2$ up to the line at infinity $\{x_0 = 0\}$. The situation varies, depending on the degree of the polynomial φ.

• The case $\deg \varphi = 3$.

When $\deg \varphi(z) = 3$, we can write down the equation in the projective plane (7.10) in explicit form, using (7.1) as

$$\left(\frac{x_2}{x_0}\right)^2 = a\left(\frac{x_1}{x_0} - \alpha_1\right)\left(\frac{x_1}{x_0} - \alpha_2\right)\left(\frac{x_1}{x_0} - \alpha_3\right),$$

or, by multiplying by x_0^3,

$$x_0 x_2^2 = a(x_1 - \alpha_1 x_0)(x_1 - \alpha_2 x_0)(x_1 - \alpha_3 x_0). \tag{7.11}$$

To extend $\mathcal{R}$ to the whole $\mathbb{P}^2$ it is natural to extend this equation to $x_0 = 0$:

$$\bar{\mathcal{R}} := \{[x_0 : x_1 : x_2] \mid \text{Equation (7.11)}\} \subset \mathbb{P}^2. \tag{7.12}$$

The original Riemann surface $\mathcal{R}$ is included in $\bar{\mathcal{R}}$ as

$$\mathcal{R} \ni (z, w) \leftrightarrow [1 : z : w] \in \bar{\mathcal{R}}.$$

Then what are the points of $\bar{\mathcal{R}}$ not included in $\mathcal{R}$? As a point at infinity of $\mathbb{P}^2$ has the form $[0 : x_1 : x_2]$, let us find a point of this form satisfying (7.11). By substituting $x_0 = 0$ into (7.11), we obtain $0 = a x_1^3$. Since $a \neq 0$, this gives $x_1 = 0$, which means that a point of the form $[0 : 0 : x_2]$ is a point in $\bar{\mathcal{R}}$ not in $\mathcal{R}$. Recall that, by the definition of $\mathbb{P}^2$, the triple $[0 : 0 : 0]$ is excluded and hence $x_2 \neq 0$. Taking the equivalence relation defining $\mathbb{P}^2$ into account, any triple $(0, 0, x_2)$ ($x_2 \neq 0$) is equivalent to $(0, 0, 1)$. Summarising, the set $\bar{\mathcal{R}} \setminus \mathcal{R}$ of the points $\bar{\mathcal{R}}$ not included in $\mathcal{R}$ consists of only one point $[0 : 0 : 1]$. We call this point the *point at infinity* of $\mathcal{R}$ and denote it by $\infty = [0 : 0 : 1]$: $\bar{\mathcal{R}} = \mathcal{R} \cup \{\infty\}$.

As $x_0 = 0$ at the point ∞, $\left(z = \dfrac{x_1}{x_0}, w = \dfrac{x_2}{x_0}\right)$ cannot be used as local coordinates of $\mathbb{P}^2$ around it. Instead we can divide the coordinates by x_2 and use $(\xi, \eta) := \left(\dfrac{x_0}{x_2}, \dfrac{x_1}{x_2}\right)$ as local coordinates. Dividing (7.11) by x_2^3, we obtain

$$\xi = a(\eta - \alpha_1 \xi)(\eta - \alpha_2 \xi)(\eta - \alpha_3 \xi). \tag{7.13}$$

This is the equation of $\bar{\mathcal{R}}$ in a neighbourhood of ∞.

Exercise 7.3 Show that equation (7.13) defines a non-singular algebraic curve in a neighbourhood of $(\xi,\eta) = (0,0)$. (Hint: It is enough to check the conditions for the implicit function theorem. Rewrite equation (7.13) in the form $G(\xi,\eta) = 0$ and show $\left(G, \frac{\partial G}{\partial \xi}, \frac{\partial G}{\partial \eta}\right) \neq (0,0,0)$ at $(\xi,\eta) = (0,0)$.)

Thus we have extended the Riemann surface $\mathcal{R}$ to a one-dimensional complex manifold $\bar{\mathcal{R}}$ in $\mathbb{P}^2$.

• The case $\deg\varphi = 4$.

Assume that the polynomial $\varphi(z)$ is of degree four and has a factorisation (7.4). Let us add points at infinity to the Riemann surface $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$ in $\mathbb{P}^2$. The equation for $\mathcal{R}$ in $\mathbb{P}^2$ is obtained by multiplying (7.10) by x_0^4:

$$x_0^2 x_2^2 = a(x_1 - \alpha_0 x_0)\cdots(x_1 - \alpha_3 x_0). \tag{7.14}$$

Taking this as a defining equation, we can extend $\mathcal{R}$ in $\mathbb{P}^2$. The same calculation as for the case $\deg\varphi = 3$ shows that the set of points satisfying (7.14) is $\mathcal{R} \cup \{\infty = [0:0:1]\}$.

"Oh, the calculation is the same? Then there is no problem." Unfortunately, there *is*. In this construction the point $\infty = [0:0:1]$ is singular.

Exercise 7.4 Check this. (Hint: The calculation is the same as in Exercise 7.3. The equation corresponding to (7.13) is $\xi^2 = a(\eta - \alpha_0\xi)\cdots(\eta - \alpha_3\xi)$.)

Since there is no "good coordinate" around a singular point, we have a problem with calculus there. What can we do?

7.2.2 Another way. (Embedding into $\mathcal{O}(2)$)

In fact, we can extend $\mathcal{R}$ with $\deg\varphi = 4$ to a non-singular complex manifold in some other way. In order to do that, we add points at infinity to $\mathbb{C}^2$ by embedding it into a space different from $\mathbb{P}^2$.

Let W and W' be two copies of $\mathbb{C}^2$. We identify W with the 'original' $\mathbb{C}^2 = \{(z,w) \mid z,w \in \mathbb{C}\}$, in which the Riemann surface $\mathcal{R}$ lives. Assume that W' has a coordinate system (ξ,η): $W' = \{(\xi,\eta) \mid \xi,\eta \in \mathbb{C}\}$. This W' 'carries points at infinity' as the set $\{[x_0 : x_1 : x_2] \mid x_2 \neq 0\}$ in $\mathbb{P}^2$.

In $\mathbb{P}^2$ coordinate systems (z,w) and (ξ,η) are connected by relations $\xi = \frac{1}{w}\left(= \frac{x_0}{x_2}\right)$ and $\eta = \frac{z}{w}\left(= \frac{x_1}{x_2}\right)$. Instead of these relations, we glue W and W' by the following relations: a point (z,w) in W and a point (ξ,η) in W' are identified when

$$\xi = \frac{1}{z},\ \eta = \frac{w}{z^2}, \quad \text{if } z \neq 0,$$
$$z = \frac{1}{\xi},\ w = \frac{\eta}{\xi^2}, \quad \text{if } \xi \neq 0. \tag{7.15}$$

'Identification' means, as always, taking the quotient space of $W \cup W'$ by the equivalence relation $\sim$ defined by (7.15):

$$X := W \cup W'/\sim .$$

(See Fig. 7.3.)

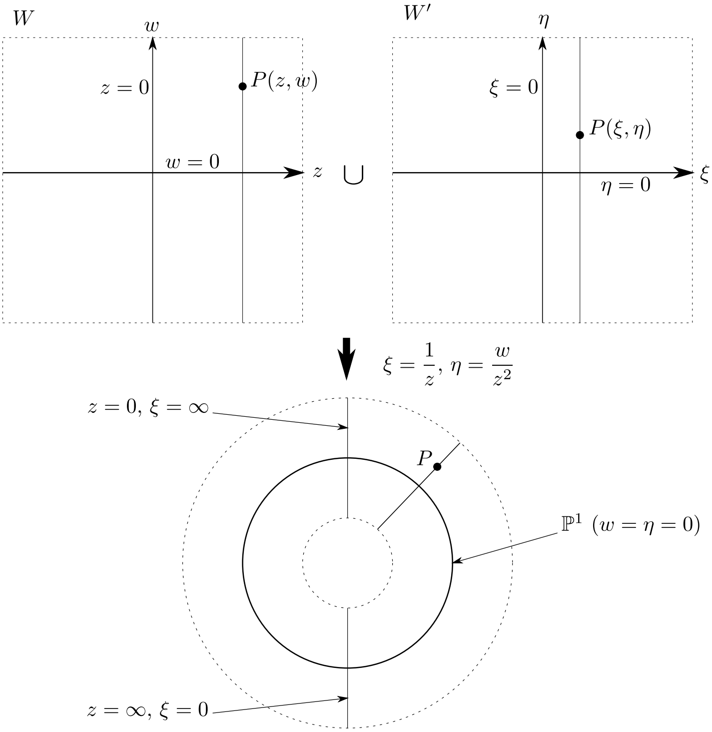

Fig. 7.3 Gluing of W and W'. (Schematic diagram; $\mathbb{C}$ is expressed by a segment.)

When w and η are fixed to 0, $(z,0)$ and $(\xi,0)$ are identified by $z = \frac{1}{\xi}$ or $\xi = \frac{1}{z}$, which means that this subset of X is nothing but the union of $\mathbb{C} = \{z \mid z \in \mathbb{C}\}$ and $\{z = \infty\} = \{\xi = 0\}$ (or the union of $\mathbb{C} = \{\xi \mid \xi \in \mathbb{C}\}$ and $\{\xi = \infty\} = \{z = 0\}$), i.e., the Riemann sphere $\mathbb{P}^1$. We can interpret X as a space based on $\mathbb{P}^1$ ('circle' in Fig. 7.3) with additional structure $\mathbb{C}$ ('radial segments' in Fig. 7.3) whose coordinate is w or η. (This space X is called the *line bundle* $O(2)$ on $\mathbb{P}^1$ in algebraic geometry.)

• The case $\deg\varphi = 4$.

Let us consider the case $\deg\varphi = 4$ first. We regard $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$ as a subset of $W = \{(z,w) \mid z,w \in \mathbb{C}\}$. When $z \neq 0$, the point $(z,w) \in \mathcal{R}$ belongs also to W' and thus is expressed by the coordinate system (ξ,η) of W'. Rewriting the equation $w^2 = \varphi(z)$ using the factorisation (7.4) and the relations (7.15), we obtain the equation for (ξ,η):

$$\left(\frac{\eta}{\xi^2}\right)^2 = a\left(\frac{1}{\xi} - \alpha_0\right)\left(\frac{1}{\xi} - \alpha_1\right)\left(\frac{1}{\xi} - \alpha_2\right)\left(\frac{1}{\xi} - \alpha_3\right).$$

Multiplying by ξ^4, we have a polynomial equation,

$$\eta^2 = a(1-\alpha_0\xi)(1-\alpha_1\xi)(1-\alpha_2\xi)(1-\alpha_3\xi). \tag{7.16}$$

This equation's left-hand side is a square of one variable and its right-hand side is a quartic polynomial of another variable (or a cubic polynomial when one of α_i's is zero). In this sense it has the same form as $w^2 = \varphi(z)$. Therefore we can use the results in Section 7.1, which shows that the subset of W' defined by

$$\mathcal{R}' := \{(\xi,\eta) \in W' \mid \eta^2 = a(1-\alpha_0\xi)(1-\alpha_1\xi)(1-\alpha_2\xi)(1-\alpha_3\xi)\}$$

is a non-singular algebraic curve. So we define

$$\bar{\mathcal{R}} := \mathcal{R} \cup \mathcal{R}' \subset W \cup W' \tag{7.17}$$

by combining $\mathcal{R}$ and $\mathcal{R}'$.

What is added to $\mathcal{R}$? Let us study the newly added part $\bar{\mathcal{R}} \setminus \mathcal{R} = \mathcal{R}' \setminus \mathcal{R}$, the set of points in $\mathcal{R}'$ not included in $\mathcal{R}$. The subset of W' which is excluded from W is $W' \setminus W = \{(\xi = 0,\eta) \mid \eta \in \mathbb{C}\}$. Therefore as for the z-coordinate, we have added 'points at infinity $z = \infty$ ($\xi = 0$)'. We can compute the corresponding η-coordinate by substituting $\xi = 0$ into (7.16), which gives $\eta^2 = a$. Thus we have

$$\mathcal{R}' \setminus \mathcal{R} = \{(\xi,\eta) = (0, \pm\sqrt{a})\} \subset W' \setminus W,$$

which means that two points $(0,\pm\sqrt{a})$ are added to $\mathcal{R}$. We denote these 'points at infinity' by $\infty_\pm$:

$$\bar{\mathcal{R}} = \mathcal{R} \cup \{\infty_+, \infty_-\}.$$

Since $\mathcal{R}$ and $\mathcal{R}'$ do not have singular points and are one-dimensional complex manifolds, it is easy to show that $\bar{\mathcal{R}}$ naturally becomes a one-dimensional complex manifold.

• The case $\deg\varphi(z) = 3$.

This case is similar. Factorising $\varphi(z)$ as in (7.1) and rewriting the equation $w^2 = \varphi(z)$ for $\mathcal{R}$ in terms of the coordinate system (ξ, η) of W' under the assumption $z \neq 0$, we obtain

$$\eta^2 = a\xi(1-\alpha_1\xi)(1-\alpha_2\xi)(1-\alpha_3\xi). \tag{7.18}$$

This equation has the structure 'η^2 is equal to a quartic polynomial in ξ (a cubic polynomial if one of the α_i's is 0)'. Therefore, the subset of W'

$$\mathcal{R}' := \{(\xi,\eta) \in W' \mid \eta^2 = a\xi(1-\alpha_1\xi)(1-\alpha_2\xi)(1-\alpha_3\xi)\}$$

is a non-singular algebraic curve and

$$\bar{\mathcal{R}} := \mathcal{R} \cup \mathcal{R}' \subset W \cup W' \tag{7.19}$$

is a one-dimensional complex manifold.

The difference from the case $\deg \varphi = 4$ is that the set $\mathcal{R}' \setminus \mathcal{R}$ of points of $\mathcal{R}'$ not included in $\mathcal{R}$ consists of only one point. In fact, substituting $\xi = 0$ into the equation (7.18) ($z = \infty$), we obtain $\eta^2 = 0$ and

$$\mathcal{R}' \setminus \mathcal{R} = \{(\xi,\eta) = (0,0)\} \subset W' \setminus W.$$

So, the point $(0,0)$ is the unique 'point at infinity' ∞ of $\mathcal{R}$:

$$\bar{\mathcal{R}} = \mathcal{R} \cup \{\infty\}.$$

We can show that $\bar{\mathcal{R}}$ constructed here is the same as that constructed in Section 7.2.1. The proof is not difficult but requires lengthy computation, so we omit it here.[1]

What we have constructed now has a special name.

Definition 7.5 The one-dimensional complex manifold $\bar{\mathcal{R}}$ ((7.12) or (7.17)) obtained by adding points at infinity to the Riemann surface $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$ ($\deg \varphi = 3$ or 4) is called an *elliptic curve*.

Remark 7.6 Although $\bar{\mathcal{R}}$ is a 'surface', it is called a 'curve' because a local coordinate is only one complex number. In algebraic geometry an elliptic curve over a field K is defined by replacing the field $\mathbb{C}$ of complex numbers by another field K (for example, a finite field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$).

Remark 7.7 If the degree of the polynomial $\varphi(z)$ is more than 4, we can also construct a one-dimensional complex manifold adding one point at infinity (when $\deg \varphi$ is odd) or two points at infinity (when $\deg \varphi$ is even) to the Riemann surface of $\sqrt{\varphi(z)}$. Such a manifold is called a *hyperelliptic curve*.

[1] See, for example, [U], Appendix J.

7.3 The Shape of $\bar{\mathcal{R}}$

In this section we study the 'shape' of an elliptic curve $\bar{\mathcal{R}}$. First we consider the case $\deg\varphi = 4$ and then we explain why an elliptic curve of the case $\deg\varphi = 3$ has the same shape.

In Section 7.1 we constructed the Riemann surface $\mathcal{R}$ in two ways: gluing of two sheets and the description $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z))\}$. Let us use gluing to construct $\mathcal{R}'$ for a quartic $\varphi(z)$ in the previous section. For the moment we assume that all α_i $(i = 0,\dots,3)$ are different from 0 for simplicity. Defining $\beta_i := \alpha_i^{-1}$, we have

$$
\begin{aligned}
\mathcal{R}' &= (\text{Riemann surface of } \eta = \sqrt{a(1-\alpha_0\xi)\cdots(1-\alpha_3\xi)}) \\
&= (\mathbb{C}\smallsetminus\{\beta_0,\dots,\beta_3\})_+ \cup \{\beta_0,\dots,\beta_3\} \cup (\mathbb{C}\smallsetminus\{\beta_0,\dots,\beta_3\})_-.
\end{aligned}
\tag{7.20}
$$

It is easy to see that we have the following connection between (7.5) and (7.20), which comes essentially from (7.15):

$$
\begin{array}{rcl}
\mathring{\mathbb{C}}_{\alpha,\pm} \ni z = \xi^{-1} & \longleftrightarrow & \xi = z^{-1} \in \mathring{\mathbb{C}}_{\beta,\pm}, \\
z = \infty & \longleftrightarrow & \xi = 0, \\
z = 0 & \longleftrightarrow & \xi = \infty, \\
\text{branch points} : \alpha_i = \beta_i^{-1} & \longleftrightarrow & \beta_i = \alpha_i^{-1},
\end{array}
$$

where $\mathring{\mathbb{C}}_{\alpha,\pm} := (\mathbb{C}\smallsetminus\{0,\alpha_0,\dots,\alpha_3\})_\pm$, $\mathring{\mathbb{C}}_{\beta,\pm} := (\mathbb{C}\smallsetminus\{0,\beta_0,\dots,\beta_3\})_\pm$.

When we constructed $\bar{\mathcal{R}}$, we first glued $(\mathbb{C}\smallsetminus\{\alpha_0,\dots,\alpha_3\})_\pm$ and $\{\alpha_0,\dots,\alpha_3\}$ to $\mathcal{R}$ and then $(\mathbb{C}\smallsetminus\{\beta_0,\dots,\beta_3\})_\pm$ and $\{\beta_0,\dots,\beta_3\}$ to $\mathcal{R}'$. Finally we glued $\mathcal{R}$ and $\mathcal{R}'$ to get $\bar{\mathcal{R}}$.

Now, let us change the order of this gluing operation: first, glue only the upper sheets of $\mathcal{R}$ and $\mathcal{R}'$, $(\mathbb{C}\smallsetminus\{\alpha_0,\dots,\alpha_3\})_+$ and $(\mathbb{C}\smallsetminus\{\beta_0,\dots,\beta_3\})_+$. By the above mentioned connection we obtain the Riemann sphere without branch points, $(\mathbb{C}\cup\{\infty\})\smallsetminus\{\alpha_0,\dots,\alpha_3\} = \mathbb{P}^1\smallsetminus\{\alpha_0,\dots,\alpha_3\}$. We denote this by $(\mathbb{P}^1\smallsetminus\{\alpha_0,\dots,\alpha_3\})_+$ for obvious reasons. Similarly, by gluing the lower sheets, $(\mathbb{C}\smallsetminus\{\alpha_0,\dots,\alpha_3\})_-$ and $(\mathbb{C}\smallsetminus\{\beta_0,\dots,\beta_3\})_-$, we obtain $(\mathbb{P}^1\smallsetminus\{\alpha_0,\dots,\alpha_3\})_-$. Finally by gluing these punctured spheres $(\mathbb{P}^1\smallsetminus\{\alpha_0,\dots,\alpha_3\})_\pm$ and adding the missing branch points $\{\alpha_0,\dots,\alpha_3\}$, we get the elliptic curve $\bar{\mathcal{R}}$.

Summarising, $\bar{\mathcal{R}}$ is constructed by gluing two Riemann spheres together, each of which has two slits between branch points as in Fig. 7.4.

$$
\bar{\mathcal{R}} = (\mathbb{P}^1\smallsetminus\{\alpha_0,\dots,\alpha_3\})_+ \cup \{\alpha_0,\dots,\alpha_3\} \cup (\mathbb{P}^1\smallsetminus\{\alpha_0,\dots,\alpha_3\})_-.
$$

Thus the constructed manifold has the shape of a torus.

The argument for the case $\deg\varphi = 3$ is almost the same. Assuming $\alpha_i \neq 0$ $(i = 1,2,3)$ again, we define $\beta_i := \alpha_i^{-1}$. Then

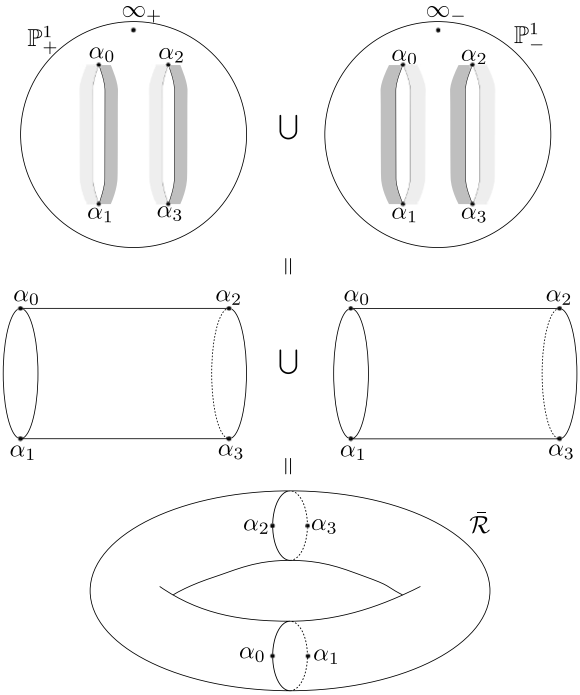

Fig. 7.4 Gluing two Riemann spheres to construct $\bar{\mathcal{R}}$.

$$(7.21)\qquad \begin{aligned} \mathcal{R}' &= (\text{Riemann surface of } \eta = \sqrt{a\xi(1-\alpha_1\xi)\cdots(1-\alpha_3\xi)}) \\ &= (\mathbb{C}\smallsetminus\{0,\beta_1,\beta_2,\beta_3\})_+ \cup \{0,\beta_1,\beta_2,\beta_3\} \cup (\mathbb{C}\smallsetminus\{0,\beta_1,\beta_2,\beta_3\})_-. \end{aligned}$$

There are four branch points of $\mathcal{R}'$; 0, β_1, β_2 and β_3. Following the recipe in Section 7.1, the Riemann surface $\mathcal{R}'$ is constructed by gluing two sheets, each of which has cuts between 0 and β_1 and between β_2 and β_3. On the other hand, we cut sheets between ∞ and α_1 and between α_2 and α_3 when constructing $\mathcal{R}$. These cuts coincide with the cuts for $\mathcal{R}'$. Shortly speaking, if we regard ∞ as α_0, the case $\deg\varphi = 3$ does not differ from the case $\deg\varphi = 4$. (In Fig. 7.4 two α_0's are replaced with $\infty_\pm$ when $\deg\varphi = 3$.) Therefore the elliptic curve for $\deg\varphi = 3$ case is also a torus.

Up to this point we have assumed that $\alpha_i \neq 0$. Even when one of the α_i's vanishes (for example, $\alpha_3 = 0$), the construction does not change much. If α_3 in (7.20) or (7.21) is 0, then erase the β_3's in the following expressions. That's all. The Riemann surface $\mathcal{R}'$ is defined by a cubic polynomial as $\eta^2 = (\text{cubic polynomial of } \xi)$. As before, we have only to identify cuts of $\mathcal{R}$ and $\mathcal{R}'$ pairwise.

Thus we have shown that in both cases, $\deg\varphi = 3$ and $\deg\varphi = 4$, elliptic curves have the same shape as a torus, or, in terms of topology, are *homeomorphic* to a torus.

The operation of 'adding points at infinity', which we described in detail, is often called *compactification*. In fact, the resulting elliptic curves are homeomorphic to a torus, which is compact, as it is realised as a bounded closed set in $\mathbb{R}^3$.

When we classified elliptic integrals in Section 2.2, we showed that $\sqrt{\varphi(z)}$ in elliptic integrals can be reduced to $\sqrt{(1-z^2)(1-k^2z^2)}$ or $\sqrt{z(1-z)(1-\lambda z)}$. We can say the same thing for elliptic curves. According to Lemma 2.3, for any four distinct points $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ on $\mathbb{P}^1$ there is a fractional linear transformation $T(z) = \dfrac{Az+B}{Cz+D}$ such that

$$T(1) = \alpha_0, \qquad T(-1) = \alpha_1,$$
$$T(k^{-1}) = \alpha_2, \quad T(-k^{-1}) = \alpha_3.$$

A fractional linear transformation is a holomorphic bijection between two Riemann spheres (in other words, an isomorphism between them as complex manifolds). Hence, moving $\alpha_0, \ldots, \alpha_3$ in Fig. 7.4 to ± 1, $\pm k^{-1}$ by T, we can transform the corresponding elliptic curve to the elliptic curve obtained by compactification of the Riemann surface

$$w^2 = (1-z^2)(1-k^2z^2)$$

(Legendre–Jacobi standard form). By using other fractional linear transformations, we can transform the same elliptic curve to the compactification of the Riemann surface

$$w^2 = z(1-z)(1-\lambda z)$$

(Riemann's standard form).

In this way, elliptic curves are classified by one parameter (k or λ). There is a deep theory on this classification, but we do not touch upon it in this book.

Exercise 7.8 Fill in the omitted details in the above explanation and show that any elliptic curve can be reduced to a standard form.

At last we are ready to discuss complex elliptic integrals. They will be considered as integrals on elliptic curves in the next chapter.

Chapter 8
Complex Elliptic Integrals

In the previous chapter we constructed an elliptic curve, a stage on which the elliptic integrals $\int R(x, \sqrt{\varphi(x)})\, dx$ live. Points at infinity were added to the Riemann surface of $\sqrt{\varphi(z)}$ to get an elliptic curve, which is nothing but a torus as a surface (over $\mathbb{R}$). In this chapter we define elliptic integrals on this elliptic curve. As before, $\varphi(z)$ is a polynomial of degree three or four without multiple roots.

8.1 Complex Elliptic Integrals of the First Kind

Let us begin with an elliptic integral of the first kind, $\int \frac{dz}{\sqrt{\varphi(z)}}$. The original elliptic integral of the first kind defined in Section 2.2 is for a specific polynomial $\varphi(z) = (1-z^2)(1-k^2z^2)$. However, for a reason which we shall discuss later (see Definition 8.9), we shall call any integral of this form with an arbitrary $\varphi(z)$ the 'elliptic integral of the first kind'. We denote the one-form in the integral by

$$\omega_1 := \frac{dz}{\sqrt{\varphi(z)}} = \frac{dz}{w}.$$

As was explained in Section 7.1, the one-form ω_1 is holomorphic on the Riemann surface $\mathcal{R} = \bar{\mathcal{R}} \smallsetminus \{\infty_\pm\}$ (when $\deg\varphi = 4$) or $\mathcal{R} = \bar{\mathcal{R}} \smallsetminus \{\infty\}$ (when $\deg\varphi = 3$). (See Proposition 7.1 (iii).) Then, what does it look like around the points at infinity $\infty_\pm$ or ∞ which were added to $\mathcal{R}$ in the compactification procedure?

First we examine the case $\deg\varphi = 4$. As usual, we factorise $\varphi(z)$ as $\varphi(z) = a(z-\alpha_0)(z-\alpha_1)(z-\alpha_2)(z-\alpha_3)$. We saw in Section 7.2 that in this case $\bar{\mathcal{R}}$ has two points at infinity, $\infty_\pm = (\xi = 0, \eta = \pm\sqrt{a})$, and in a neighbourhood of each of them

- ξ is a local coordinate.
- $\bar{\mathcal{R}}$ is defined by the equation

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_9

$$\eta^2 = a(1-\alpha_0\xi)(1-\alpha_1\xi)(1-\alpha_2\xi)(1-\alpha_3\xi).$$

At points $\infty_\pm$ we need to express ω_1 in terms of (ξ,η) instead of (z,w). The formulae for the coordinate change are $\xi = \frac{1}{z}$, $\eta = \frac{w}{z^2}$, from which follows $d\xi = -\frac{dz}{z^2}$. Thus we have

$$\omega_1 = \frac{dz}{w} = -\frac{d\xi}{\eta}. \tag{8.1}$$

The denominator

$$\eta(\xi) = \sqrt{a(1-\alpha_0\xi)(1-\alpha_1\xi)(1-\alpha_2\xi)(1-\alpha_3\xi)}$$

is holomorphic in ξ in a neighbourhood of $\infty_\pm$, at which $\xi = 0$, and $\eta(\xi) \neq 0$. Thus we have shown that the coefficient of $d\xi$ in the expression (8.1) for ω_1 at $\infty_\pm$ is a holomorphic function of ξ and hence ω_1 is a holomorphic differential form at $\infty_\pm$, too.

As a consequence ω_1 is holomorphic everywhere on $\bar{\mathcal{R}}$. Moreover, as the coefficients in the expressions $\omega_1 = \frac{dz}{w} = -\frac{d\xi}{\eta}$ ($\frac{1}{w}$ except at $\infty_\pm$, $-\frac{1}{\eta}$ at $\infty_\pm$) do not vanish, we have also shown that $\omega_1 \neq 0$ everywhere on $\bar{\mathcal{R}}$. (We shall use this fact later.)

Up to now we have assumed that $\deg\varphi = 4$, but the case $\deg\varphi = 3$ is similar, so we leave the details to the reader as an exercise.

Exercise 8.1 Show that ω_1 is holomorphic everywhere on $\bar{\mathcal{R}}$ and never vanishes also in the case $\deg\varphi = 3$. (Hint: In this case ∞ is a branch point, which means that we have to use η, not ξ, as a local coordinate. Rewrite $\frac{dz}{w} = -\frac{d\xi}{\eta}$ in the form (holomorphic function) $\times d\eta$, using (7.18).)

Now complex elliptic integrals come into play. Let us fix any point $P_0 \in \bar{\mathcal{R}}$ on the elliptic curve as the starting point of integration contours. Taking a curve C connecting a point P on $\bar{\mathcal{R}}$ and P_0, we can define a function 'of P' by integrating ω_1:

$$F(P) := \int_{P_0\to P} \omega_1 = \int_{C:\text{curve from } P_0 \text{ to } P} \omega_1. \tag{8.2}$$

We have used quotation marks because, by this 'definition', $F(P)$ depends not only on P but also on C.

But, if P wanders around only in a small neighbourhood of a fixed point, say P_1, we can fix a part C_{01} of the path C from P_0 to P_1 and assume that a part C_1 from P_1 to P lies in a small neighbourhood of P_1 (Fig. 8.1). For example, let us take C_1 in an ε-neighbourhood of P_1, $\{P \mid |z(P) - z(P_1)| < \varepsilon\}$, where z is any local coordinate at P_1. (One can use w or ξ or whatever instead of z.)

Since ω_1 is a holomorphic differential form, the integral (8.2) does not depend on the choice of C_1 thanks to the Cauchy integral theorem. In this sense the function

$F(P)$ is 'locally' (which means that P does not move far) defined as a single-valued function. In addition, it is a holomorphic function, as the integrand is a holomorphic differential.

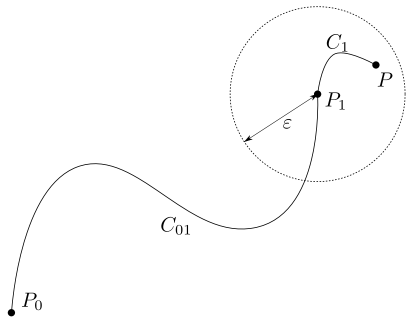

Fig. 8.1 Connect P_0 and P via P_1. The value of the integral is determined only by the position of the point P as long as the part C_1 of the curve connecting P_1 and P is within the small neighbourhood.

Thus we have shown that the integral (8.2) defines a holomorphic function locally. Then, how about 'globally'? What if we move the point P far away? To answer this, we need to know how many globally different contours exist on $\bar{\mathcal{R}}$. This means that we have to study the topology of an elliptic curve = a torus.

In topology the following fact is well known.

$$H_1(\bar{\mathcal{R}}, \mathbb{Z}) = \mathbb{Z}[A] \oplus \mathbb{Z}[B]. \tag{8.3}$$

This means that any element $[C]$ of the first homology group $H_1(\bar{\mathcal{R}}, \mathbb{Z})$ of a torus $\bar{\mathcal{R}}$ is expressed as

$$[C] = m[A] + n[B] \tag{8.4}$$

with uniquely determined integers m and n. Here A and B are closed curves on the elliptic curve as in Fig. 8.2 and often called the *A-cycle* and the *B-cycle*. (We do not need to take the curves exactly as in this picture, but not all choices are valid.)

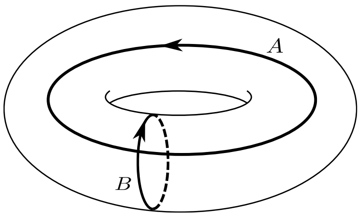

Fig. 8.2 *A*-cycle and *B*-cycle on an elliptic curve.

The expression (8.4) means that 'in the homology group any closed curve is equivalent to a curve which goes around A several times and then goes around B several times'. This can be explained in the following way.

Assume that there is a closed curve C on a torus as in Fig. 8.3. When we cut this torus along the A-cycle and the B-cycle, we get a rectangle, on which segments of the chopped curve C lie. By arranging these rectangles vertically and horizontally on a plane, the originally connected curve C is reproduced on it.

On the right picture in Fig. 8.3 the whole curve C appears when we arrange two rectangles horizontally and three vertically. On this 'plane map' we can deform the curve C continuously to a polygonal line which moves by two edges of the rectangle horizontally and by three edges vertically. This deformation on the plane corresponds to a deformation on a torus.

You might think that this is not so obvious. In that case, first consider a small (local) deformation which deforms only a small portion of the curve. It is not so difficult to see that such a local deformation on the plane corresponds to a local deformation on the torus. We can deform the whole curve by a sequence of such local moves. So the global deformation on the plane is equivalent to a global deformation on the torus.

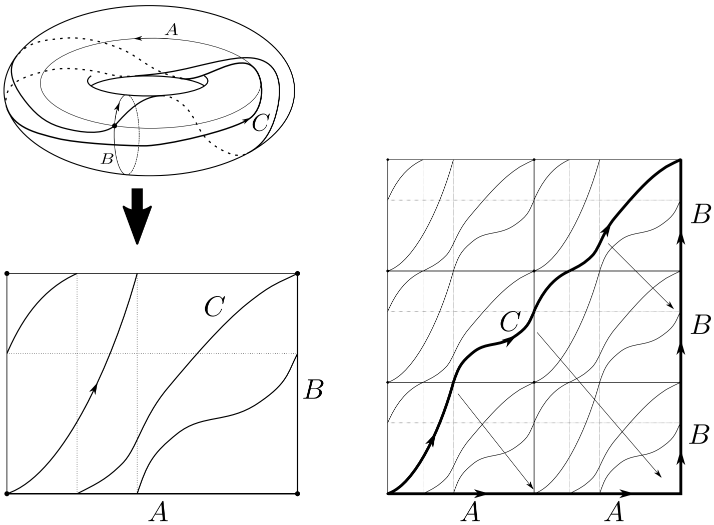

Fig. 8.3 Cut open an elliptic curve (left) and deform a closed curve C to a sum of A-cycles and B-cycles (right).

Moving along the horizontal edge of the rectangle corresponds to moving along the A-cycle and moving along the vertical edge corresponds to moving along the B-cycle. Thus the closed curve C in Fig. 8.3 is deformed to a curve which goes around the A-cycle twice and around the B-cycle three times. Therefore in the homology group $H_1(\bar{\mathcal{R}},\mathbb{Z})$ we have

$$[C] = 2[A] + 3[B].$$

This is the idea of the proof of (8.3).

If two curves C and C' are connecting P_0 and P on $\bar{\mathcal{R}}$, we obtain a closed curve by going along C' and coming back along C. The corresponding element $[C'-C]$ in the homology group is expressed as $[C'-C] = m[A]+n[B]$ for some $m,n \in \mathbb{Z}$ by (8.3). Hence the integral of ω_1 along C' and along C differ by

$$\int_{C'} \omega_1 - \int_C \omega_1 = m\int_A \omega_1 + n\int_B \omega_1. \tag{8.5}$$

We call the integrals $\int_A \omega_1$ and $\int_B \omega_1$ in the right-hand side the *A-period* and the *B-period* of ω_1 respectively.

Thus we have found:

- The function $F(P)$ defined by the elliptic integral (8.2) is locally a single-valued holomorphic function.
- However, unfortunately (or, rather *fortunately*), it is multi-valued globally.
- Its multi-valuedness is not random but controlled as in (8.5) by two numbers, the A- and B-periods of ω_1.

Summarising, we have obtained the following theorem.

Theorem 8.2 *A complex elliptic integral of the first kind* (8.2) *gives a multi-valued holomorphic function $F(P)$ on the elliptic curve. The multi-valuedness is given by a linear combination of the A- and the B-periods of ω_1 with integral coefficients.*

Let us compute the A-period and the B-period explicitly in the case $\varphi(z) = (1-z^2)(1-k^2z^2)$:

$$\omega_1 = \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}.$$

For simplicity we assume $k \in \mathbb{R}$, $0<k<1$. (This is the situation from Chapter 1 till Chapter 3, where we considered real elliptic integrals.)

For explicit computation we need a more concrete description of the integration contours (the A- and B-cycles), for which we have to reexamine the construction of an elliptic curve $\bar{\mathcal{R}}$. As we found in Section 7.3, an elliptic curve is obtained by gluing two projective lines $\mathbb{P}^1$ with two cuts on each of them. The cuts connect roots of the polynomial $\varphi(z)$, ± 1 and $\pm\frac{1}{k}$. (See Fig. 8.4.) We take the A-cycle and the B-cycle as in this picture. (Here the A-cycle is composed of two parts, A_+ and A_-.)

Proposition 8.3 *In this case the periods of ω_1 are*

$$\int_A \omega_1 = 4K(k), \qquad \int_B \omega_1 = 2iK'(k). \tag{8.6}$$

Here

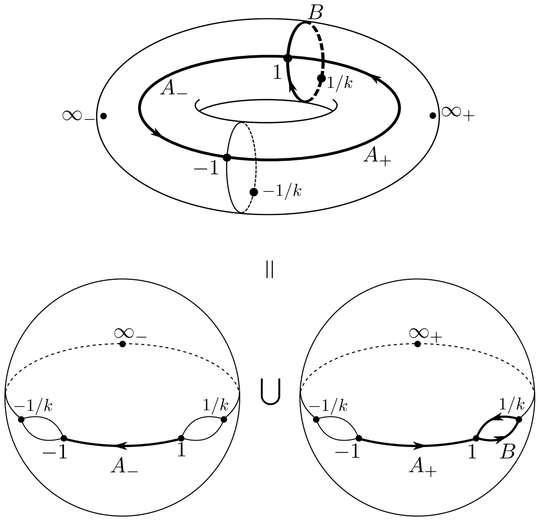

Fig. 8.4 A-cycle and B-cycle for $\varphi(z) = (1 - z^2)(1 - k^2 z^2)$.

$$K(k) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} \tag{8.7}$$

is the complete elliptic integral of the first kind and $K'(k)$ *is defined by*

$$K'(k) := K(k'), \qquad k' := \sqrt{1-k^2}. \tag{8.8}$$

(k' is the complementary modulus of the modulus k.)

Proof A-period: If we take the A-cycle as in Fig. 8.4, the integral on the A-cycle is decomposed into two parts, from -1 to 1 (A_+ in Fig. 8.4) and from 1 to -1 (A_- in Fig. 8.4),

$$\int_A \omega_1 = \int_{-1}^{1} \frac{dx}{+\sqrt{(1-x^2)(1-k^2x^2)}} + \int_1^{-1} \frac{dx}{-\sqrt{(1-x^2)(1-k^2x^2)}}.$$

Since the integration contours are on the real line, the integrals in the right-hand side are usual Riemann integrals. Notice that the signs of the integrands are opposite. This is because the integration contour turns around the branch points ± 1 when we move from the first term to the second term or back.

Using usual tricks like interchanging the upper and the lower ends of the second integral or rewriting integrals of even functions on $[-1, 1]$ as integrals on $[0, 1]$, eventually we obtain

$$\int_A \omega_1 = 4 \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} = 4K(k).$$

This is the first equation in (8.6).

<u>B-period:</u> Rewriting the integral as in the computation of the A-period, we obtain

$$\begin{aligned}\int_B \omega_1 &= \int_1^{\frac{1}{k}} \frac{dx}{+\sqrt{(1-x^2)(1-k^2x^2)}} + \int_{\frac{1}{k}}^1 \frac{dx}{-\sqrt{(1-x^2)(1-k^2x^2)}} \\ &= 2\int_1^{\frac{1}{k}} \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.\end{aligned} \tag{8.9}$$

In the last integral, the inside of the radical symbol, $(1-x^2)(1-k^2x^2)$, becomes negative when the variable x is in $\left(1, \dfrac{1}{k}\right)$, as $0 < k < 1$. Hence the integrand takes a purely imaginary value and

$$\int_1^{\frac{1}{k}} \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} = i\int_1^{\frac{1}{k}} \frac{dx}{\sqrt{(x^2-1)(1-k^2x^2)}}.$$

Then we change the variable from x to $t = \dfrac{1}{k'}\sqrt{1-\dfrac{1}{x^2}}$, which means $x^2 = \dfrac{1}{1-k'^2t^2}$. This turns the inside of the square root of the integrand into

$$(x^2-1)(1-k^2x^2) = \frac{k'^4t^2(1-t^2)}{(1-k'^2t^2)^2}.$$

As $x = \dfrac{1}{\sqrt{1-k'^2t^2}}$, we have

$$dx = \frac{k'^2t}{(1-k'^2t^2)^{3/2}}dt.$$

When x moves from 1 to $\dfrac{1}{k}$, t increases from 0 to 1 monotonically. Using these facts, we can rewrite (8.9) as

$$\begin{aligned}\int_B \omega_1 &= 2i\int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k'^2t^2)}} \\ &= 2iK(k') = 2iK'(k).\end{aligned}$$

Thus we have obtained the second equation of (8.6). □

Remark 8.4 In order to simplify the proof and make it easier to read, some details were omitted.

1. All the integrals in the proof are improper in the sense that the integrands diverge at the end of the interval. We discussed details of their convergence in Section 4.1.

2. We pulled out $\sqrt{-1}$ from the integrand in (8.9). Why is this $\sqrt{-1}$ and not $-\sqrt{-1}$? In the computation of the A-period we took a branch of the square root in the integrand such that the square root is a positive real number on A_+ (from -1 to $+1$). Taking this into account and carefully chasing the change of the argument of $\sqrt{1-x}$ in

$$\sqrt{(1-x^2)(1-k^2x^2)} = \sqrt{(1-x)(1+x)(1-k^2x^2)},$$

as we did in Chapter 6, we can conclude that we should pull out $\sqrt{-1}$ when we consider $\sqrt{(1-x^2)(1-k^2x^2)}$ on $\left[1, \frac{1}{k}\right]$. □

For a general complex modulus $k \in \mathbb{C}$ the result is the same. We omit the details here. (We have only to continue the result analytically as a function of k.)

At the end of Section 6.2 we found that the period of the one-form $\dfrac{dz}{\sqrt{1-z^2}}$ gives the period of $\sin u = 2\pi$ by computing the integral $\displaystyle\int \frac{dz}{\sqrt{1-z^2}}$. The same is true for the above computation of the elliptic integral. One of the results above is

$$A\text{-period of } \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = 4K(k),$$

and the right-hand side is equal to the period of $\operatorname{sn}(u)$ introduced as the inverse function of the incomplete elliptic integral of the first kind in Chapter 4.

Then what role does the B-period, $2iK'(k)$, play for $\operatorname{sn}(u)$? We shall show that it is another period of the function $\operatorname{sn}(u)$. As a complex function the function $\operatorname{sn}(u)$ has *two* periods! Well, the story has gone a little ahead. Let us come back to this theme later. Stay tuned!

8.2 Complex Elliptic Integrals of the Second Kind

The next complex elliptic integral which we consider is the elliptic integral of the second kind,

$$\int \sqrt{\frac{1-k^2z^2}{1-z^2}}\, dz = \int \frac{1-k^2z^2}{\sqrt{\varphi(z)}}\, dz.$$

The polynomial $\varphi(z)$ is $(1-z^2)(1-k^2z^2)$ as before. The stage on which this integral lives is the Riemann surface $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$, or its compactification, the elliptic curve $\bar{\mathcal{R}}$, as before. The only difference from the previous case is the one-form being integrated. Here we denote it by ω_2:

$$\omega_2 := \sqrt{\frac{1-k^2z^2}{1-z^2}}\, dz = \frac{1-k^2z^2}{\sqrt{\varphi(z)}}\, dz = \frac{1-k^2z^2}{w}\, dz.$$

It is not hard to show that this is holomorphic on $\mathcal{R}$.

Exercise 8.5 Check that ω_2 is holomorphic on $\mathcal{R}$. (Hint: Except at branch points ± 1, $\pm\dfrac{1}{k}$ we can take z as a local coordinate and at branch points we should take w instead. Express ω_2 in the form (function of z)$\times dz$ or (function of w)$\times dw$ in each case and check that the coefficients are holomorphic.)

Then how about at $\{\infty_\pm\}$, which are not contained in $\mathcal{R}$?

As $\xi = \dfrac{1}{z}$ is a local coordinate at $\infty_\pm$, let us express ω_2 in terms of ξ, using the coordinate transformation (7.15). Since $dz = d\left(\dfrac{1}{\xi}\right) = -\dfrac{d\xi}{\xi^2}$, we have

$$
\begin{aligned}
\omega_2 &= \frac{1-k^2/\xi^2}{\eta/\xi^2}\, d(\xi^{-1}) = \frac{\xi^2-k^2}{\eta}\cdot\left(-\frac{d\xi}{\xi^2}\right) \\
&= \frac{1}{\xi^2}\frac{k^2-\xi^2}{\eta}\, d\xi = \frac{1}{\xi^2}\frac{k^2-\xi^2}{\pm k(1+O(\xi^2))}\, d\xi \\
&= \left(\frac{\pm k}{\xi^2} + (\text{holomorphic function at } \xi = 0)\right) d\xi.
\end{aligned} \tag{8.10}
$$

In the second line we used $\eta = \pm k(1+O(\xi^2))$, which follows from the square root of the equation (7.16). In fact, equation (7.16) becomes

$$
\begin{aligned}
\eta^2 &= k^2(1-\xi)(1+\xi)\left(1-\frac{\xi}{k}\right)\left(1+\frac{\xi}{k}\right) \\
&= k^2\left(1-\left(1+\frac{1}{k^2}\right)\xi^2+\frac{\xi^4}{k^2}\right)
\end{aligned}
$$

in the present case. The sign of $\pm k$, which appears when we open the square root $\sqrt{k^2}$, is $+$ at ∞_+ and $-$ at ∞_- by the definition of $\infty_\pm$ in Section 7.2.2. The appearance of $1+O(\xi^2)$ is due to, for example, the generalised binomial theorem,

$$
(1+t)^{1/2} = 1 + \sum_{n=1}^{\infty} \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-n+1\right)}{n!} t^n,
$$

applied to $1-\left(1+\dfrac{1}{k^2}\right)\xi^2+\dfrac{\xi^4}{k^2}$. (Substitute $-\left(1+\dfrac{1}{k^2}\right)\xi^2+\dfrac{\xi^4}{k^2}$ for t.)

The expression (8.10) says that the one-form ω_2 has second order poles at $\infty_\pm$ and that the residues there are zero, $\mathrm{Res}_{\infty_\pm}\, \omega_2 = 0$, because there is no term of the form $\dfrac{(\text{coefficient})}{\xi}\, d\xi$.

Now let us consider a 'function'

$$
G(P) := \int_{P_0 \to P} \omega_2 = \int_{C:\text{curve from } P_0 \text{ to } P} \omega_2 \tag{8.11}
$$

by integrating ω_2, as we did in (8.2). (P_0 is an arbitrary point of the elliptic curve $\bar{\mathcal{R}}$.) As was the case with $F(P)$ defined by (8.2), a priori this integral depends not only on P but also the curve C.

We want to show that this $G(P)$, too, does not depend on the choice of C 'locally' like $F(P)$. In fact, in a neighbourhood of a point P_1 which is not at infinity, $P_1 \neq \infty_\pm$, we can prove single-valuedness of $G(P)$ by decomposing the contour into two parts, from P_0 to P_1 and from P_1 to P, and moving only the second part.

The situation at singularities $\infty_\pm$ of ω_2 is a little different. In general, if a one-form has a singularity (in the present case, a second-order pole), the Cauchy integral theorem cannot be applied and the integral might change its value when the contour moves across the singularity. Hence, in a neighbourhood of a singularity it might be possible for an integral like (8.11) to be a multi-valued function of P. (In fact, we shall soon meet such an example.)

However, fortunately, the one-form ω_2 has residue 0 at its singularities. The difference of the integral of ω_2 along the contour passing to the right of the singularity (C_{right} in Fig. 8.5) and the integral along the contour passing to the left (C_{left} in Fig. 8.5) is equal to the integral along the contour around the singularity (C_0 in Fig. 8.5) $= 2\pi i \times$ (residue of ω_2 at $\infty_\pm$) $= 0$.

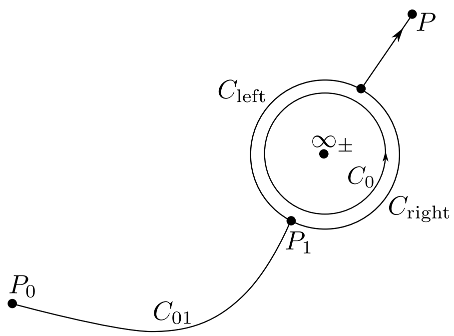

Fig. 8.5 To go right or to go left of $\infty_\pm$.

Hence the value of the integral (8.11) does not depend on the contour C in a neighbourhood of points at infinity, which means that $G(P)$ is a locally single-valued function of P.

As ω_2 is holomorphic except at $\infty_\pm$, $G(P)$ is holomorphic except at points at infinity. At $\infty_\pm$, ω_2 has an expansion (8.10), from which it follows that $G(P)$ behaves like

$$G(P) = \mp \frac{k}{\xi} + (\text{holomorphic function at } \xi = 0)$$

in a neighbourhood of $\infty_\pm$. So $G(P)$ has simple poles at $\infty_\pm$.

Global multi-valuedness is a problem of homology of $\bar{\mathcal{R}}$ again. If both C and C' are curves from P_0 to P, there exist integers $m, n \in \mathbb{Z}$ such that the element $[C' - C]$ of $H_1(\bar{\mathcal{R}}, \mathbb{Z})$ represented by the closed curve consisting of C' and reversed C is expressed as $[C' - C] = m[A] + n[B]$. Therefore the difference of integrals of ω_2 along C' and C is expressed as

$$\int_{C'} \omega_2 - \int_C \omega_2 = m \int_A \omega_2 + n \int_B \omega_2.$$

Similarly to the periods of ω_1, we call $\displaystyle\int_A \omega_2$ the *A-period* of ω_2 and $\displaystyle\int_B \omega_2$ the *B-period*.

We summarise the results on the complex elliptic integrals of the second kind as follows.

Theorem 8.6 *A complex elliptic integral of the second kind* (8.11) *gives a multi-valued meromorphic function $G(P)$ on the elliptic curve. The poles of $G(P)$ are at $\infty_\pm$ and of the first order with residues $\mp k$. Multi-valuedness is expressed by a linear combination of the A-period and the B-period of ω_2 with integral coefficients.*

Exercise 8.7 Assume that $0 < k < 1$ and that the A-cycle and the B-cycle are as in Fig. 8.4.

(i) Show that the A-period of ω_2 is expressed as $4E(k)$, where $E(k)$ is the complete elliptic integral of the second kind, $E(k) = \displaystyle\int_0^1 \sqrt{\frac{1-k^2z^2}{1-z^2}}\, dz$.

(ii) Express the B-period of ω_2, using the complete elliptic integrals of the first and the second kinds. (This is a little difficult.)

(iii) Show $d\left(\dfrac{zw}{1-k^2z^2}\right) = \dfrac{k^2z^4 - 2z^2 + 1}{1-k^2z^2}\omega_1$. (Hint: Compute dw, using the relation $w^2 = \varphi(z)$.)

(iv) Show that the complete elliptic integral of the first kind $K(k)$ and the complete elliptic integral of the second kind $E(k)$ satisfy the following system of linear differential equations:

$$\frac{dE}{dk} = \frac{E}{k} - \frac{K}{k}, \qquad \frac{dK}{dk} = \frac{E}{kk'^2} - \frac{K}{k}.$$

(Hint: To show the first equation, differentiate ω_2 with respect to k and then integrate it over $[0,1]$. The second equation is obtained by comparing $\dfrac{\partial}{\partial k}\omega_1 - \dfrac{1}{kk'^2}\omega_2 + \dfrac{1}{k}\omega_1$ with (iii) and computing the A-period. Note that the integral of an exact form over a cycle is zero.)

8.3 Complex Elliptic Integrals of the Third Kind

Lastly we consider the elliptic integral of the third kind, but we touch on it only briefly. As we discussed in Section 2.2, it is an integral of the form

$$\int \frac{dz}{(z^2-a^2)\sqrt{\varphi(z)}},$$

where $\varphi(z) = (1-z^2)(1-k^2z^2)$ and a is a parameter.

The one-form

$$\omega_3 := \frac{dz}{(z^2-a^2)\sqrt{\varphi(z)}} = \frac{dz}{(z^2-a^2)w}$$

in the integral can be studied in a similar way as in the previous sections. As a result we can show that ω_3 is holomorphic on $\bar{\mathcal{R}}$ (including points at infinity, $\infty_\pm$) except at four points

$$(z,w) = (\pm a, \pm\sqrt{(1-a^2)(1-k^2a^2)}),$$

where it has *simple poles*. (Any combination of the signs $\pm$ of z and w are possible.)

Exercise 8.8 (i) Check these facts (holomorphicity, positions and order of poles).

(ii) Find the residue at each pole.

For example, when we integrate a one-form $\dfrac{dz}{z-a}$ on $\mathbb{C}$ with a simple pole at $z=a$, we obtain $\log(z-a)$ (+ constant), which is infinitely multi-valued in a neighbourhood of $z=a$. Similarly to this case the function

$$H(P) := \int_{P_0\to P} \omega_3$$

defined by the elliptic integral of the third kind is locally holomorphic except at the poles of ω_3, $(z,w) = (\pm a, \pm\sqrt{\varphi(a)})$, and infinitely multi-valued around the poles. The function $H(P)$ also has global multi-valuedness by the A- and the B-periods like the elliptic integrals of the first and the second kinds. We do not go further into this complicated object.

Definition 8.9 A meromorphic differential one-form ω on a Riemann surface is called an *Abelian differential*. It is classified as follows:

- the *first kind*, if ω is holomorphic everywhere;
- the *second kind*, if ω has poles but all residues are zero;
- the *third kind*, otherwise (ω has a pole with non-zero residue).

According to this classification ω_1 is an Abelian differential of the first kind, ω_2 is of the second kind and ω_3 is of the third kind. The naming of the Abelian differential is consistent with the naming of the elliptic integrals.[1]

Remark 8.10 There are several variations of this naming. Sometimes

- 'the second kind' means that the one-form has only one pole without residue;
- 'the third kind' means that the one-form has only simple poles.

[1] I guess that historically the classes of Abelian differentials were named after the elliptic integrals, but I have not yet confirmed this hypothesis by reference.

So you have to pay attention to the definition when you read or write books and articles.

In the next chapter, as a natural example of the elliptic integral of the first kind, we study holomorphic functions mapping a rectangle to the upper half plane.

Chapter 9
Mapping the Upper Half Plane to a Rectangle

In the previous chapter we considered elliptic integrals

$$\int R(x, \sqrt{\varphi(x)})\,dx$$

on elliptic curves and saw that a incomplete elliptic integral of the first kind defines a multi-valued holomorphic function on the elliptic curve, that of the second kind defines a multi-valued meromorphic function and that of the third kind defines a complicated multi-valued function. In this chapter we study a problem in which an elliptic integral of the first kind appears naturally. The elliptic integral is on the stage from some moment but reveals its identity only at the end of the story. It might be interesting to ask "Who is the elliptic integral?" while reading the following, like a detective story.

9.1 The Riemann Mapping Theorem

How many kinds of domains are there in a plane? Here by the word 'domain' we mean a connected open set (or, equivalently, an arcwise-connected open set).

The criterion of classification is, first of all, homeomorphic or not, which means, whether there exists a continuous bijection $f : D_1 \to D_2$ between two domains whose inverse $f^{-1} : D_2 \to D_1$ is also continuous. This guarantees that all topological properties of D_1 and D_2 are the same.

For example, the two domains in Fig. 9.1 are not homeomorphic, because they have different topological properties: in the left domain any closed curve can be deformed continuously to a point, but in the right domain there is a closed curve which gets caught by a 'hole' in the domain and cannot be reduced to a point.

A domain like the left one, in which any closed curve can be reduced to a point, is called a *simply-connected domain* and others are called multiply-connected domains. Here we consider the classification of simply-connected domains only. Even when

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_10

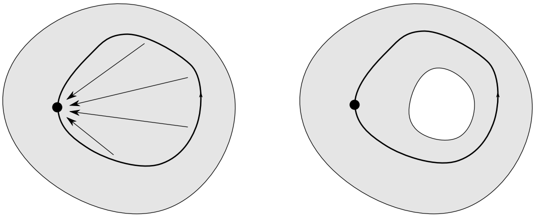

Fig. 9.1 Simply-connected and multiply-connected domains.

we restrict ourselves to this case, there are various simply-connected domains: the whole plane, half planes, polygons, disks, fractal figures and so on. Can we really classify them?

Surprisingly the answer is *yes*. We have the following simple classification.

Theorem 9.1 (Riemann mapping theorem) [1]

Let D_1 and D_2 be two simply-connected domains in $\mathbb{C}$ and assume that they are not $\mathbb{C}$ itself. Then there exists a holomorphic function on D_1 which is a bijection from D_1 to D_2.

In the usual formulation of the Riemann mapping theorem D_2 is set to the unit disk $\Delta = \{w \mid |w| < 1\} \subset \mathbb{C}$. Both formulations are equivalent.

What is remarkable about this theorem is that the proper domains of $\mathbb{C}$ are only required to be simply-connected, and yet the domains are connected by a very good continuous map, a holomorphic function.[2] It is easy to see that the whole of $\mathbb{C}$ and, for example, the unit disk Δ are homeomorphic.[3] Therefore, according to the Riemann mapping theorem there is only one type of simply-connected domain classified by homeomorphism, and two types, one class containing just $\mathbb{C}$ and the other class containing all other simply-connected domains $\subsetneq \mathbb{C}$, classified by holomorphic homeomorphism.

We shall use the following theorem on the correspondence between boundaries of the domains later.

Theorem 9.2 (Carathéodory's theorem)

In addition to the assumptions of the Riemann mapping theorem, we assume that the boundaries of D_1 and D_2 are simple closed curves (Jordan curves; closed curves without self-intersections). Then the function f in Theorem 9.1 *can be extended to a homeomorphism between the closure of D_1 and the closure of D_2. (Hence it gives a homeomorphism between the boundaries.)*

[1] Riemann stated this theorem in his thesis at Göttingen University (article I (1851) in [Ri]). His version of this theorem requires some conditions on the boundary of the domain and his proof had a logical flaw.

[2] Recall that the inverse map of a holomorphic function is holomorphic. See Corollary A.12.

[3] However the homeomorphism $\psi : \mathbb{C} \to \Delta$ cannot be holomorphic, because, if it were holomorphic, it should be a constant function by Liouville's theorem for a bounded entire function.

The following are simple examples.

Example 9.3 A holomorphic bijection from the upper half plane H to the unit disk Δ $\varphi_1 : z \mapsto \dfrac{z-i}{z+i}$. The inverse mapping is $\varphi_1^{-1}(w) = i\dfrac{1+w}{1-w}$.

Exercise 9.4 Check that the above φ_1 and φ_1^{-1} are really bijections between H and Δ. (Hint: As for φ_1, compare $|z-i|$ and $|z+i|$, considering their geometric meaning. As for φ_1^{-1}, express the imaginary part of φ_1^{-1}, using $|w|$.)

Example 9.5 A holomorphic bijection from the upper half plane H to an angular domain $\{w \mid 0 < \arg w < \alpha\}$: $\varphi_2(z) = z^{\frac{\alpha}{\pi}}$. This is a direct consequence of the fact that a complex number z in the upper half plane is expressed as $z = re^{i\theta}$ $(r > 0, 0 < \theta < \pi)$ in polar coordinates and a complex number in the angular domain is expressed as $w = \rho e^{i\phi}$ $(\rho > 0, 0 < \phi < \alpha)$. This is a very simple example, but we shall use it later.

Although the existence of a holomorphic bijection between two simply-connected domains is guaranteed by the Riemann mapping theorem, *explicit* construction is another story. The proof of the Riemann mapping theorem is almost useless for constructing the map concretely. We have to make use of the characteristics of the domain in consideration to find a mapping function.

In Section 9.3, we construct a holomorphic bijection from the upper half plane H to a rectangle explicitly. For that purpose we prepare a useful instrument in the next section.

9.2 The Reflection Principle

What we want to construct is a holomorphic function from the upper half plane to a rectangle. In this case both the domain of definition and the range of the function have 'straight' boundaries, a line or segments of lines. In such cases there is a standard way to extend the holomorphic function beyond that boundary.

Let D be a domain symmetric with respect to the real axis: $z \in D \iff \bar{z} \in D$. (See the left picture of Fig. 9.2.) We denote the intersection of D with the upper half plane $\{z \mid \operatorname{Im} z > 0\}$ by D_+, the intersection with the lower half plane $\{z \mid \operatorname{Im} z < 0\}$ by D_- and the intersection with the real axis by δ: $D = D_+ \cup \delta \cup D_-$. We assume that a holomorphic function on D_+ is extended to δ continuously and $f(\delta) \subset \mathbb{R}$.

You might think that assumptions are too much, but in fact, Carathéodory's theorem guarantees that our goal, the holomorphic bijection from the upper half plane to a rectangle, satisfies similar conditions. (The side of the rectangle might not be in the real axis, but this is not an essential point. The rectangle can be moved by rotation (multiplication by $e^{i\theta}$, $\theta \in \mathbb{R}$) and parallel translation (addition by α, $\alpha \in \mathbb{C}$) so that a side lies in the real axis.)

In such a situation, we can naturally extend the domain of definition of f from $D_+ \cup \delta$ to D_- in the lower half plane. Precisely speaking, if we define

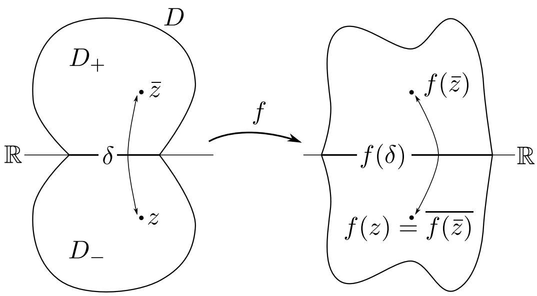

Fig. 9.2 Reflection principle.

$$f(z) := \begin{cases} \text{original } f(z), & (\text{when } z \in D_+ \cup \delta), \\ \overline{f(\bar{z})} & (\text{when } z \in D_-), \end{cases} \tag{9.1}$$

the newly defined function f is holomorphic on D. This is called the *(Schwarz) reflection principle*.

It is easy to check that the function defined by (9.1) is holomorphic in D_-. For example, let us take a point z_0 in D_-. Since $\overline{z_0}$ is in D_+, f has a Taylor series at $\overline{z_0}$:

$$f(z) = a_0 + a_1(z - \overline{z_0}) + \cdots + a_n(z - \overline{z_0})^n + \cdots$$

Applying the definition (9.1), f has the following Taylor series at z_0:

$$f(z) = \overline{f(\bar{z})} = \overline{a_0} + \overline{a_1}(z - z_0) + \cdots + \overline{a_n}(z - z_0)^n + \cdots,$$

which shows that f is holomorphic at $z_0 \in D_-$. (This can also be proved directly from the definition of the derivative, or by using the Cauchy–Riemann equations.)

It remains to show holomorphicity on the joint boundary δ of the two regions $D_\pm$. The function f defined by (9.1) is holomorphic on $D_\pm$, hence continuous there. Continuity on δ follows from continuity of the original f on $D_+ \cup \delta$, the assumption $f(\delta) \subset \mathbb{R}$ and the definition (9.1). But, is it holomorphic (or differentiable) on δ? Let us recall an example in real analysis: $|x|$ is differentiable on $x < 0$ and on $x > 0$ and continuous everywhere on $\mathbb{R}$. However it is not differentiable at $x = 0$. Can we say that such things do not occur for the function f defined by (9.1)?

We can. It is one of the miraculously good properties of holomorphic functions.[4]

Let D_1 and D_2 be interiors of rectifiable Jordan closed curves C_1 and C_2. Assume that $D_1 \cap D_2 = \emptyset$ and $C_1 \cap C_2 = \Gamma$ is a connected arc. It is easy to see that $D = D_1 \cup \mathring{\Gamma} \cup D_2$ is a domain with the boundary $C_1 \cup C_2 \setminus \mathring{\Gamma}$, where $\mathring{\Gamma} = \Gamma \setminus$ endpoints.

Theorem 9.6 (Painlevé's theorem)

[4] I would nominate 'differentiability leads to Taylor expansion' and 'identity theorem (if two holomorphic functions agree on a part, they agree everywhere)' as other *miraculously good* properties of holomorphic functions, which cannot be true for real differentiable functions.

If each $f_i(z)$ $(i = 1, 2)$ is holomorphic on D_i, continuous on $D_i \cup C_i$ and $f_1(z) = f_2(z)$ on Γ, then

$$f(z) := \begin{cases} f_1(z), & \text{if } z \in D_1 \cup C_1, \\ f_2(z), & \text{if } z \in D_2 \cup C_2, \end{cases}$$

is holomorphic on D and continuous on its closure.

This theorem says that a function obtained by continuously gluing two holomorphic functions is holomorphic.

The proof of this theorem is rather complicated, but the following simple version is sufficient for our purposes.

Theorem 9.7 *Let D be a domain in $\mathbb{C}$ with non-empty intersection with $\mathbb{R}$: $\delta := D \cap \mathbb{R} \neq \emptyset$. Assume that a continuous function f on D is holomorphic on $D \setminus \delta$. Then it is holomorphic everywhere on D, even on δ.*

The holomorphicity of f defined by (9.1) on $D = D_+ \cup \delta \cup D_-$ follows directly from this theorem.

Let us prove Theorem 9.7. In order to show holomorphicity at a point x in δ, we have only to examine f in a neighbourhood of x. In Fig. 9.3 we magnify the neighbourhood of x. The contour C is a rectangle around x, which is decomposed into the upper part C_+ and the lower part C_-, which are completed to rectangles by adding the part of the real axis. (The lower side of C_+ and the upper side of C_- are in the real axis, and the other sides of them are in C.)

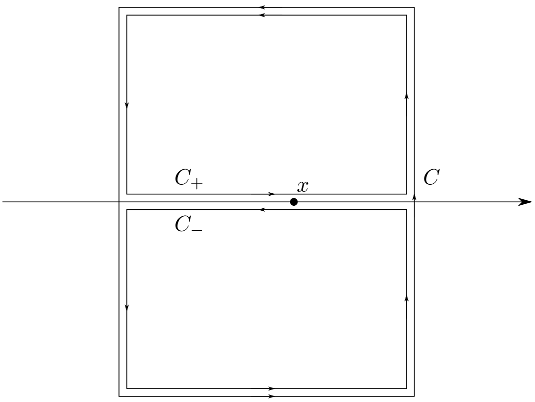

Fig. 9.3 The contour C around x and its decomposition into two parts, C_+ and C_-.

As f is continuous, it is continuous on C and the integral

$$\tilde{f}(z) := \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{\zeta - z}\, d\zeta \tag{9.2}$$

is well-defined, where z is a point inside C. Since the integrand is holomorphic in z, the function $\tilde{f}(z)$ is a holomorphic function inside C thanks to Lemma A.6.

We are now going to show the equality $\tilde{f}(z) = f(z)$ inside C. As $\tilde{f}$ is holomorphic inside C, this means that f is holomorphic inside C and, hence, in a neighbourhood of x. The point x being arbitrary in δ, holomorphicity of f in D has been proved.

First we show $\tilde{f}(z) = f(z)$, assuming that z is in the upper half plane. Since the integrals on $C_+ \cap \mathbb{R}$ and $C_- \cap \mathbb{R}$ cancel due to the opposite directions, the integral (9.2) is rewritten as

$$\tilde{f}(z) = \frac{1}{2\pi i}\int_{C_+} \frac{f(\zeta)}{\zeta - z}\,d\zeta + \frac{1}{2\pi i}\int_{C_-} \frac{f(\zeta)}{\zeta - z}\,d\zeta. \tag{9.3}$$

If we can show

$$\frac{1}{2\pi i}\int_{C_+} \frac{f(\zeta)}{\zeta - z}\,d\zeta = f(z), \qquad \frac{1}{2\pi i}\int_{C_-} \frac{f(\zeta)}{\zeta - z}\,d\zeta = 0, \tag{9.4}$$

by substituting them into (9.3) we have $\tilde{f}(z) = f(z)$ for $z \in D_+$.

Let us prove equations (9.4). "Aren't they obvious from the Cauchy integral formula?" Well, there is a subtle point. Can we really apply the Cauchy formula? Probably not everybody pays attention to the following condition for the usual Cauchy integral formula (and the usual Cauchy integral theorem): the function $f(z)$ is supposed to be holomorphic in a domain which *contains the integration contour*. In other words, f should be holomorphic on the contour, too. As we are now proving holomorphicity of f on δ, we cannot apply the Cauchy integral formula in this form.

To tell the truth, we *can* apply the Cauchy integral formula/theorem to a function which is holomorphic inside a rectifiable Jordan closed curve and continuous inside and on that curve. This is called the *stronger form* of the Cauchy integral theorem/formula. Their proofs for general rectifiable Jordan curves are not simple. But in our case the contours are rectangles and the formulae (9.4) can be proved as follows.

Let us define two rectangles $C_{\pm\varepsilon}$ as in Fig. 9.4: $C_{+\varepsilon}$ (respectively, $C_{-\varepsilon}$) is almost the same as C_+ (respectively, C_-), but the lower side (respectively, the upper side) is separated from the real axis by distance $\varepsilon > 0$.

Since $f(z)$ is holomorphic in D_+, we can apply the Cauchy integral formula to $C_{+\varepsilon}$ and obtain

$$f(z) = \frac{1}{2\pi i}\int_{C_{+\varepsilon}} \frac{f(\zeta)}{\zeta - z}\,d\zeta, \tag{9.5}$$

if z is inside $C_{+\varepsilon}$. On the other hand, since $f(z)$ is continuous on D_+ and on δ, $\dfrac{f(\zeta)}{\zeta - z}$ is continuous as a function of ζ on $D_+ \cup \delta$ except at z. Hence, changing the order of the integration and the uniformly convergent limit ($\varepsilon \to 0$), we have

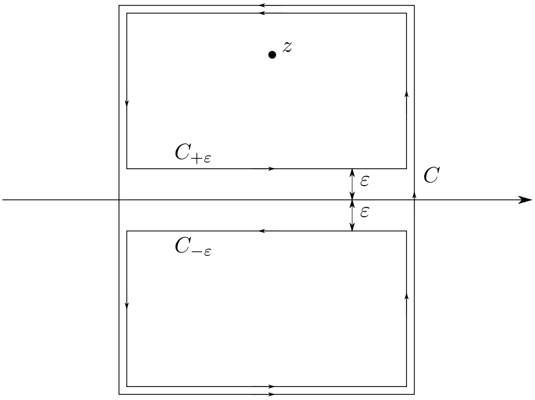

Fig. 9.4 Contours $C_{+\varepsilon}$ and $C_{-\varepsilon}$ are apart from the real axis.

$$\lim_{\varepsilon \to 0} \frac{1}{2\pi i} \int_{C_{+\varepsilon}} \frac{f(\zeta)}{\zeta - z}\, d\zeta = \frac{1}{2\pi i} \int_{C_+} \frac{f(\zeta)}{\zeta - z}\, d\zeta. \tag{9.6}$$

Thus, the first equation of (9.4) follows from (9.5) and (9.6).

The point z is outside of C_- and $C_{-\varepsilon}$. Therefore by a similar argument we can prove the second equation of (9.4) as

$$\frac{1}{2\pi i} \int_{C_-} \frac{f(\zeta)}{\zeta - z}\, d\zeta = \lim_{\varepsilon \to 0} \frac{1}{2\pi i} \int_{C_{-\varepsilon}} \frac{f(\zeta)}{\zeta - z}\, d\zeta = 0. \tag{9.7}$$

Thus we have proved $\tilde{f}(z) = f(z)$ for $z \in D_+$. The argument for a point $z \in D_-$ is similar.

It remains to show $\tilde{f}(x) = f(x)$ for $x \in \delta$. This follows immediately from continuity of these two functions: for example, take a sequence $\left\{ z_n := x + \frac{i}{n} \right\}_{n=1,2,\ldots}$, which approaches x from the upper half plane. On the upper half plane, we have already shown the equality $\tilde{f}(z_n) = f(z_n)$. Since $\tilde{f}$ and f are continuous on $D_+ \cup \delta$, the limit $n \to \infty$ of this equality gives $\tilde{f}(x) = f(x)$.

This is the end of the proof of the reflection principle. □

Aside from the proof, the statement of the reflection principle is quite clear. As the name of the theorem tells us, 'in order to extend the domain of definition of a holomorphic function $f(z)$ to a region reflected by the real axis, we take the value reflected by the real axis'. Moreover, by combining rotation (= multiplication by $e^{i\theta}$ ($\theta \in \mathbb{R}$)) and parallel translation (= addition by $\alpha \in \mathbb{C}$) we can extend the domain of definition by reflection with respect to an arbitrary straight line as in Fig. 9.5. The exact formulation of the statement and the proof are left to the reader. To understand the discussion below, you have only to keep in mind the key word 'reflection'.

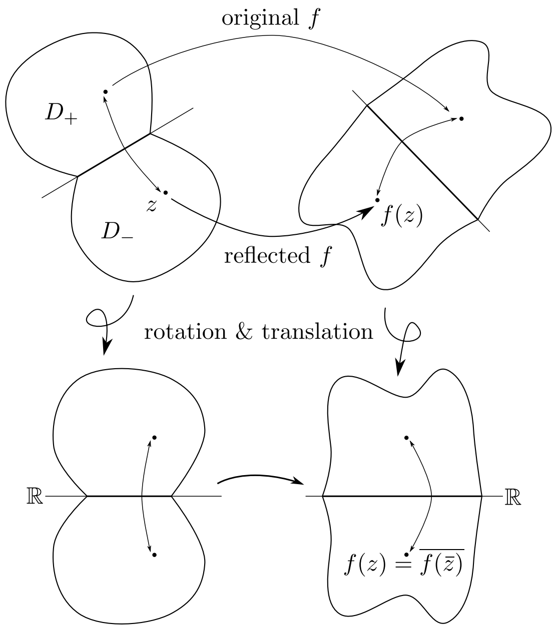

Fig. 9.5 Reflection with respect to arbitrary lines.

9.3 Holomorphic Mapping from the Upper Half Plane to a Rectangle

Now it is time to construct a holomorphic bijection $F : \mathsf{H} \to D$ from the upper half plane $\mathsf{H} = \{z \mid \operatorname{Im} z > 0\}$ to a rectangle D (without its boundary) explicitly. Its existence is guaranteed by the Riemann mapping theorem. Our strategy to construct F, whose existence is known, is as follows:

- Extend the domain of definition of F by the reflection principle.
- From properties of F derive an ordinary differential equation, which is satisfied by F.
- Find the explicit form of F by solving the differential equation.

(I) Extension of F by the reflection principle.

It follows from Carathéodory's theorem (Theorem 9.2) that the holomorphic bijection F from the upper half plane to a rectangle D has following properties.

- F can be extended to a continuous function from the closure $\bar{\mathsf{H}} = \mathsf{H} \cup \mathbb{R} \cup \{\infty\}$ of H in the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ to the closure of the rectangle $\bar{\mathsf{D}} = \mathsf{D} \cup \partial \mathsf{D}$ ($\partial \mathsf{D}$ is the boundary of D) as a homeomorphism.[5]

[5] Carathéodory's theorem cannot be applied directly to a domain whose boundary contains $\mathbb{R}$. But the composition of the map φ_1 from the unit disk to the upper half plane considered in Example 9.3

We denote this extended map also by F.

- The restriction of F to the boundary $\partial\mathsf{H} = \mathbb{R} \cup \{\infty\}$ of H (which is regarded as a circle in $\mathbb{P}^1(\mathbb{C})$ = the equator of the Riemann sphere) is a homeomorphism between $\partial\mathsf{H}$ and $\partial\mathsf{D}$.

For explicit computation we assume that the real numbers α_0, α_1, α_2 and α_3 correspond to the vertices A_0, A_1, A_2 and A_3 ($A_j \in \mathbb{C}$) of the rectangle (Fig. 9.6): $F(\alpha_j) = A_j$. The image of the point at infinity ∞ is a point on the side A_3A_0.

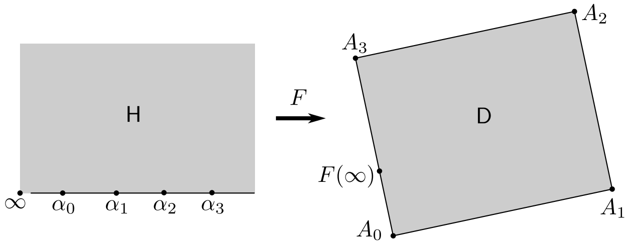

Fig. 9.6 A holomorphic bijection from the upper half plane to a rectangle.

Let us apply the reflection principle to this situation. For example, the open interval (α_0, α_1), which is a part of the boundary of H, is mapped to a side A_0A_1 of the rectangle D. Therefore, using the reflection principle, we can extend F by reflection with respect to these segments (Fig. 9.7). We denote the extended map by $F^{(01)}$:

$$F^{(01)} : \mathsf{H} \cup (\alpha_0, \alpha_1) \cup \mathsf{H}_- \to \mathsf{D} \cup (A_0A_1) \cup \mathsf{D}^{(01)}. \tag{9.8}$$

($\mathsf{H}_- := \{z \mid \operatorname{Im} z < 0\}$ is the lower half plane.) Note that the image of the part $\partial\mathsf{H} \setminus [\alpha_0, \alpha_1]$ of the boundary $\partial\mathsf{H} = \mathbb{R} \cup \{\infty\}$ of the upper half plane by F is not on the same line as A_0A_1. Hence $F^{(01)}$ is not defined on $\partial\mathsf{H} \setminus (\alpha_0, \alpha_1)$.

However, we can extend the restriction of $F^{(01)}$ on the lower half plane H_- to its own boundary $\partial\mathsf{H}_- = \mathbb{R} \cup \{\infty\}$. (Sorry for this complication.) This means that we forget about H and D in Fig. 9.7 and focus only on the newly constructed correspondence $F^{(01)}|_{\mathsf{H}_-} : \mathsf{H}_- \to \mathsf{D}^{(01)}$, which is nothing but the reflection of the original map $F : \mathsf{H} \to \mathsf{D}$. We can apply Carathéodory's theorem to $F^{(01)}|_{\mathsf{H}_-}$ to extend it to the boundary of H_-.

Now, we apply the reflection principle to this map $F^{(01)}|_{\mathsf{H}_-}$ on the lower half plane. Of course, if we use the same segments (α_0, α_1) and A_0A_1, which we have used to define $F^{(01)}$, the resulting map just returns to the original F and nothing interesting happens. So, let us use other segments, for example, (α_1, α_2) of $\partial\mathsf{H}_-$ and $A_1A_2^{(01)}$ of the new rectangle $\mathsf{D}^{(01)}$ corresponding to each other by $F^{(01)}$ (cf.

with F maps the unit disk to the rectangle, which can be extended to a homeomorphism between closures by Carathéodory's theorem. We can obtain the desired map by composing the inverse map of φ_1 to the extended homeomorphism.

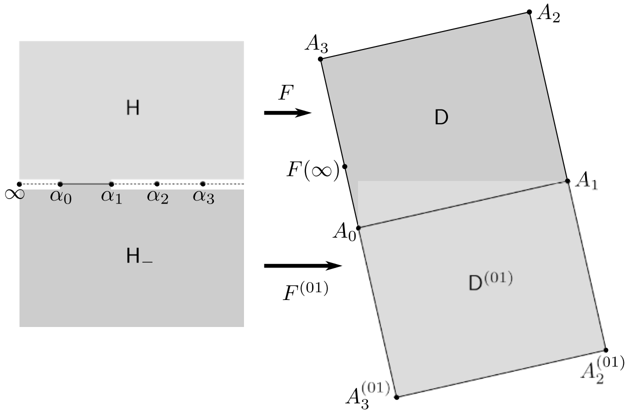

Fig. 9.7 Apply the reflection principle with respect to (α_0, α_1) and A_0A_1.

Fig. 9.8). When we apply the reflection principle to $F^{(01)}$ with respect to them, we obtain a new bijection in Fig. 9.8,

(9.9) $$F^{(01)(12)} : \mathsf{H}_- \cup (\alpha_1, \alpha_2) \cup \mathsf{H} \to \mathsf{D}^{(01)} \cup (A_1 A_2^{(01)}) \cup \mathsf{D}^{(01)(02)}.$$

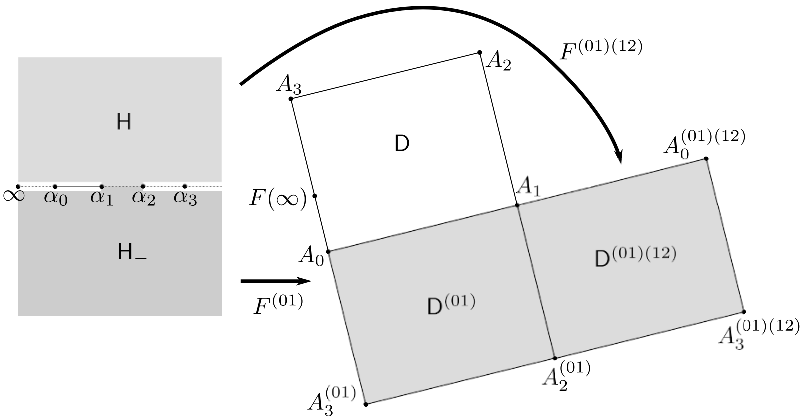

Fig. 9.8 Apply the reflection principle with respect to (α_1, α_2) and $A_1 A_1^{(01)}$.

Since we have applied reflections twice, $F \to F^{(01)} \to F^{(01)(12)}$, the domain of definition of $F^{(01)(02)}$ has returned to the upper half plane H. But, as shown in Fig. 9.8, the value of the function has not returned to the original value, as we have applied reflections to different directions. In other words, the values of F and $F^{(01)(12)}$ at the same point $z \in \mathsf{H}$ are different. We have learned in Chapter 6 that 'if the value does not come back to the original one after a round trip, we should consider the function on a Riemann surface', but that can wait. For now let us work out a single-valued function from these F and $F^{(01)(12)}$.

Comparing the range of the function $F^{(01)(12)}$ and that of the function F in Fig. 9.8, we have

$$F^{(01)(12)}(z) = -F(z) + (\text{constant}), \tag{9.10}$$

because we take reflections twice, which rotates the rectangle by 180°. Differentiating (9.10) gives

$$(F^{(01)(12)})'(z) = -F'(z),$$

whose logarithm is

$$\log(F^{(01)(12)})'(z) = \log F'(z) + \pi i + 2\pi i n,$$

where n is an integer. The additional term $2\pi i n$ comes from the multi-valuedness of the logarithm. The constant terms $\pi i + 2\pi i n$ vanish by differentiation:

$$(\log(F^{(01)(12)})')'(z) = (\log F'(z))'. \tag{9.11}$$

Thus we obtain one and the same value for both $F^{(01)(02)}$ and F. This argument can be applied to any reflection along the sides of the rectangle and gives a single-valued holomorphic function

$$\Phi(z) := (\log F'(z))' \tag{9.12}$$

for $z \in \mathbb{C} \setminus \{\alpha_0, \dots, \alpha_3\}$.

One thing which we have to care about is that the image of ∞ is on the side A_3A_0. When we apply the reflection principle along this side, we have to use the coordinate $\zeta = \dfrac{1}{z}$ around the point at infinity of $\mathbb{P}^1(\mathbb{C})$ to show holomorphicity. The remaining arguments are the same as reflection along other sides.

(II) Local properties of $F(z)$ (or $\Phi(z)$).

Here we examine the properties of $F(z)$ in a neighbourhood of $z = \alpha_j$ $(j = 0, \dots, 3)$.

In a neighbourhood of $F(\alpha_j) = A_j$ the image of F is an angular domain as described in Example 9.5 with the angle $\alpha = \dfrac{\pi}{2}$. The inverse to the map in that example, namely, raising to the $\dfrac{\pi}{\alpha}$-power = squaring maps this angular domain back to a half plane. Exactly speaking, the function

$$g_j : \mathsf{H} \ni z \mapsto (F(z) - A_j)^2$$

maps the intersection of a neighbourhood of α_j and the upper half plane to one side of a neighbourhood of 0 divided by a straight line (Fig. 9.9).

Since the boundary of the image of g_j is a part of a straight line, we can apply the reflection principle once more. The result is an extension of g_j to the lower half

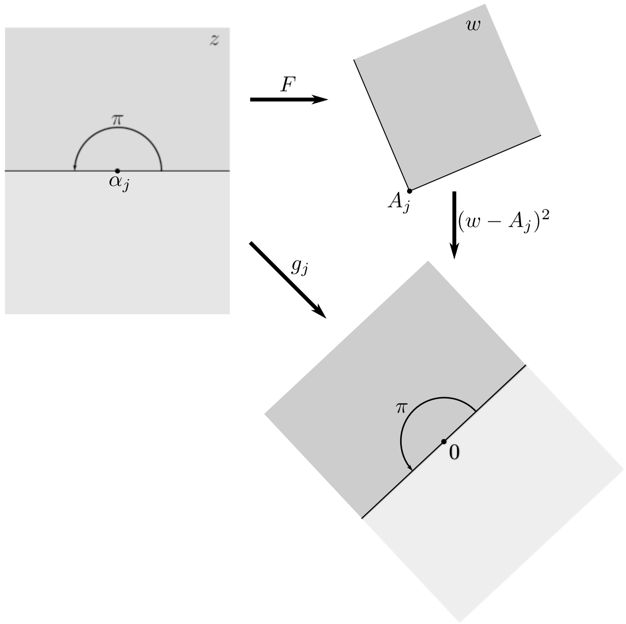

Fig. 9.9 Local behaviour of F.

plane (the light shaded part in Fig. 9.9), which is a holomorphic bijection between a neighbourhood of α_j and a neighbourhood of 0.

As $g_j(\alpha_j) = 0$, g_j has an expansion

$$g_j(z) = c_1^{(j)}(z-\alpha_j) + c_2^{(j)}(z-\alpha_j)^2 + \cdots .$$

Suppose the first N coefficients, $c_1^{(j)}, \dots, c_N^{(j)}$, vanish: $g_j(z) = c_{N+1}^{(j)}(z-\alpha_j)^{N+1} + \cdots$. If we factorise $g_j(z)$ as $g_j(z) = (z-\alpha_j)^{N+1}h(z)$, the function $h(z) = c_{N+1}^{(j)} + c_{N+2}^{(j)}z + \cdots$ is holomorphic in a neighbourhood of $z = \alpha_j$ and $h(\alpha_j) \neq 0$. Therefore there exists a holomorphic $(N+1)$-st root $k(z) := h(z)^{1/(N+1)}$. (We discussed the existence of the square root in Section 6.1.2. The logic is the same for the $(N+1)$-st root.) Then $\tilde{F}(z,w) := w - (z-\alpha_j)k(z)$ satisfies $\dfrac{\partial \tilde{F}}{\partial z}(\alpha_j, 0) \neq 0$, which makes it possible to apply the implicit function theorem Lemma 6.3. As a result we have a holomorphic bijection $z(w)$ in a neighbourhood of $w = 0$, which satisfies $w = (z(w) - \alpha_j)k(z(w))$. Thus we obtain $w^{N+1} = (z(w) - \alpha_j)^{N+1}k(z(w))^{N+1} = g_j(z(w))$, which means that w^{N+1} is a bijection in a neighbourhood of $w = 0$ as a composition of two local bijections $g_j(z)$ and $z(w)$. However w^{N+1} can be bijective around $w = 0$ only when $N = 0$. This shows[6] $c_1^{(j)} \neq 0$.

Thus the function $g_j(z)$ has a decomposition into a product

[6] We can prove $c_1^{(j)} \neq 0$ also by Rouché's theorem.

$$g_j(z) = (F(z) - A_j)^2 = (z - \alpha_j) h_j(z),$$

where $h_j(z) = c_1^{(j)} + c_2^{(j)}(z - \alpha_j) + \cdots$ and $h_j(\alpha_j) = c_1^{(j)} \neq 0$. Taking the square root, we obtain a local expression for $F(z)$ as

$$F(z) = A_j + \sqrt{z - \alpha_j}\, \tilde{h}_j(z), \quad \tilde{h}_j(z) := \sqrt{h_j(z)}. \tag{9.13}$$

The signs of $\sqrt{z - \alpha_j}$ and $\sqrt{h_j(z)}$ are determined so that the above expression recovers $F(z)$. Since $h_j(\alpha_j) \neq 0$, $\tilde{h}_j(z)$ is holomorphic in a neighbourhood of $z = \alpha_j$ (cf. Section 6.1). Substituting (9.13) into (9.12), we can express $\Phi(z)$ in a neighbourhood of $z = \alpha_j$ as follows:

$$\begin{aligned} \Phi(z) &= -\frac{1}{2(z - \alpha_j)} + \frac{\frac{3}{2}\tilde{h}_j'(z) + (z - \alpha_j)\tilde{h}_j''(z)}{\frac{1}{2}\tilde{h}_j(z) + (z - \alpha_j)\tilde{h}_j'(z)} \\ &= -\frac{1}{2(z - \alpha_j)} + (\text{a holomorphic function around } z = \alpha_j). \end{aligned} \tag{9.14}$$

(Note that the denominator of the second term in the first line does not vanish at $z = \alpha_j$.)

The coefficient $-\frac{1}{2}$ of $\frac{1}{z - \alpha_j}$ comes from the fact that each angle of the rectangle is equal to $\frac{\pi}{2}$ in the following sense: if we use the mapping function $\varphi_2(z)$ in Example 9.5 instead of $F(z)$ in the derivation of (9.12), we have

$$\left(\log \varphi_2'(z)\right)' = \frac{\alpha - \pi}{\pi} \frac{1}{z},$$

as $\varphi_2'(z) = \frac{\alpha}{\pi} z^{\frac{\alpha - \pi}{\pi}}$. When the angle α is the right angle $\frac{\pi}{2}$, the coefficient of $\frac{1}{z}$ in this equation is $-\frac{1}{2}$. This is the reason why $-\frac{1}{2}$ appears in the second line of (9.14). The expression (9.13) of $F(z)$ means 'the mapping function $F(z)$ is obtained from $(z - \alpha_j)^{\frac{1}{2}}$, which corresponds to $\varphi_2(z) = z^{\frac{\alpha}{\pi}}$ $(\alpha = \frac{\pi}{2})$, by multiplication of $\tilde{h}_j(z)$ ($\tilde{h}_j(\alpha_j) \neq 0$), which does not affect local properties, and parallel translation (the part '$A_j +$')'. Locally the factor $(z - \alpha_j)^{\frac{1}{2}}$ in (9.13) determines the form of $\Phi(z)$ as (9.14).

Having local expression (9.14) at each α_j ($j = 0, \ldots, 3$), the function $\Phi(z)$ has the form

$$\Phi(z) = -\frac{1}{2} \sum_{j=0}^{3} \frac{1}{z - \alpha_j} + \psi(z). \tag{9.15}$$

Since $\Phi(z)$ is holomorphic on $\mathbb{C} \setminus \{\alpha_0, \dots, \alpha_3\}$ by construction, $\psi(z)$ in this expression is a holomorphic function everywhere on $\mathbb{C}$.

In fact, $\psi(z)$ is identically 0. Let us show this by examining the local behaviour of $\Phi(z)$ at $z = \infty$. We analytically continued the original mapping function F to a function which maps a neighbourhood of $z = \infty$ to a neighbourhood of $F(\infty)$ bijectively. Applying the same argument as above, we have a Taylor expansion of $F(z)$ around $z = \infty$ (a Taylor expansion of $F(1/\zeta)$ around $\zeta = 0$),

$$F(z) = F(\infty) + c_1^{(\infty)} \zeta + c_2^{(\infty)} \zeta^2 + \cdots .$$

Here $c_1^{(\infty)} \neq 0$, because F is a one-to-one correspondence in a neighbourhood of $z = \infty$ ($\zeta = 0$). A computation similar to (9.14) leads to

$$\Phi(z) = -2\zeta + \zeta^2 \times (\text{a holomorphic function at } \zeta = 0),$$

which tends to 0 when $z \to \infty$ ($\zeta \to 0$). Therefore $\psi(z)$ also tends to 0 when $z \to \infty$ by (9.15). In particular, $\psi(z)$ is bounded on $\mathbb{C}$. Liouville's theorem says that a bounded holomorphic function on $\mathbb{C}$ is a constant. Hence $\psi(z)$ is a constant, which is equal to $\lim_{z \to \infty} \psi(z) = 0$. Thus we have proved $\psi(z) \equiv 0$ and, as a result,

$$\Phi(z) = \frac{d}{dz}\left(\log \frac{dF}{dz}\right) = -\frac{1}{2} \sum_{j=0}^{3} \frac{1}{z - \alpha_j}, \tag{9.16}$$

which is the differential equation for F.

(III) Solving the differential equation for $F(z)$.

Lastly we solve the differential equation (9.16) and derive an explicit form of F. This only requires two simple integrations.

Integrating (9.16) by z once, we obtain

$$\log \frac{dF}{dz} = -\frac{1}{2} \sum_{j=0}^{3} \log(z - \alpha_j) + c_1,$$

where c_1 is an integration constant. Taking the exponential, we can remove logarithms:

$$\frac{dF}{dz} = \frac{C}{\sqrt{\varphi(z)}},$$

where $C = e^{c_1}$ and $\varphi(z) = (z - \alpha_0) \cdots (z - \alpha_3)$.

By integrating this expression from an arbitrarily fixed point z_0 to z, we finally obtain the formula of a mapping function of the upper half plane to a rectangle,

$$F(z) = C \int_{z_0}^{z} \frac{dz}{\sqrt{\varphi(z)}} + F(z_0). \tag{9.17}$$

This is, as you see, the *elliptic integral of the first kind*.

The undetermined constants C and $F(z_0)$ have the following meaning: note that we can parallelly translate, rotate or dilate the rectangle. Parallel translation corresponds to the additional term '$+F(z_0)$', while rotation ('$\times e^{i\theta}$', $\theta \in \mathbb{R}$) and dilation ('$\times A$', $A > 0$) correspond to multiplication by $C \in \mathbb{C}$.

Remark 9.8 There is an expression in integral form (not an elliptic integral) for a mapping function from the upper half plane to a general polygon, which is called the *Schwarz–Christoffel formula*. The formula (9.17) is a special case of this general formula, which is proved in the same way as we proved (9.17). See §2.2, Chapter 6 of [Ah].

9.4 Elliptic Integrals on an Elliptic Curve

In the previous section, when we constructed a multi-valued holomorphic function $F(z)$, we extended a holomorphic function on the upper half plane to the lower half plane, and then to the upper half plane again and so on, repeatedly. Let us review this construction from the viewpoint of Riemann surfaces.

As we know that the true character of F is the elliptic integral of the first kind, it is natural to take an elliptic curve as the Riemann surface. For example, in step (I) of the construction in Section 9.3, we extended $F(z)$ on the upper half plane to the lower half plane by reflection with respect to (α_0, α_1), which gives $F^{(01)}$. Then we constructed another function $F^{(01)(12)}$ on the upper half plane by reflection with respect to (α_1, α_2). In the context of the construction of a Riemann surface of $\sqrt{\varphi(z)}$ (cf. Chapter 7), this corresponds to moving a point z from the upper half plane of the upper sheet $\mathbb{C}_+$ to the lower half plane of the lower sheet $\mathbb{C}_-$ through the slit (α_0, α_1), and then coming back to the upper half plane but not of the original sheet $\mathbb{C}_+$ but of the sheet $\mathbb{C}_-$, passing through the interval (α_1, α_2) between two slits (α_0, α_1) and (α_2, α_3). The function $F(z)$ lives on the upper half plane of the sheet $\mathbb{C}_+$ and $F^{(01)(12)}(z)$ lives on the upper half plane of the sheet $\mathbb{C}_-$ (Fig. 9.10).

The extension procedure by the reflection principle can also be restated as a procedure on the Riemann surface or on the elliptic curve. Let us name the upper and lower half planes on each sheet as

$$\begin{aligned} \mathsf{H}_{\pm+} &: \text{the upper half plane of } \mathbb{C}_\pm, \\ \mathsf{H}_{\pm-} &: \text{the lower half plane of } \mathbb{C}_\pm, \end{aligned}$$

and the intervals (α_i, α_{i+1}) $(i = 0, 1, 2, 3)$ $A_\pm$, $B_\pm$, $C_\pm$, $D_\pm$ as in Fig. 9.11. The elliptic curve is divided into four pieces corresponding to $\mathsf{H}_{\pm\pm}$.

On the other hand, the images of $\mathsf{H}_{\pm\pm}$ by F are four rectangles of the same size as in Fig. 9.12, which are reflections of D. The large rectangle obtained by combining four copies of D corresponds to the whole elliptic curve by the holomorphic mapping. The sides of the large rectangle correspond to closed curves (C_+, C_-) and (D_+, D_-) on the elliptic curve. (We can also take other 'large rectangles' whose sides are

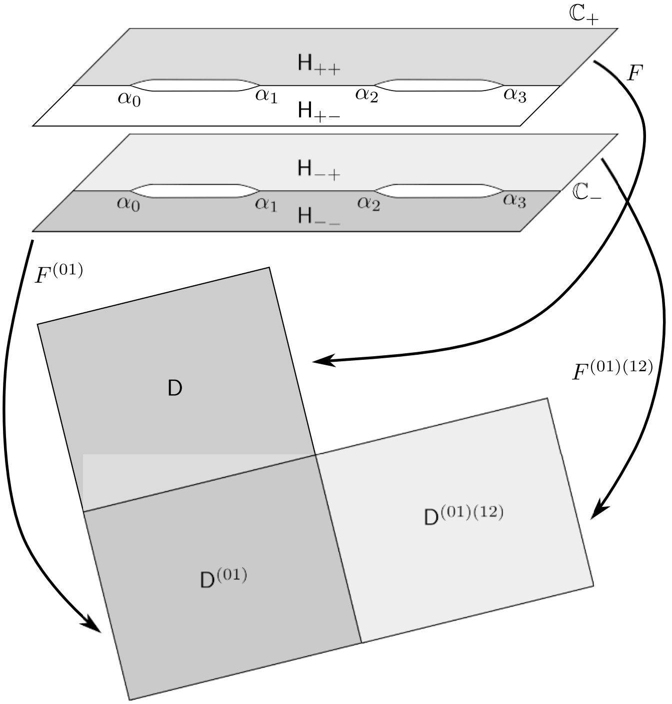

Fig. 9.10 F, $F^{(01)}$ and $F^{(01)(12)}$ on the Riemann surface.

images of (A_+, A_-) and (B_+, B_-) or images of (A_+, A_-) and (D_+, D_-) or images of (C_+, C_-) and (B_+, B_-).)

As the function F is an elliptic integral of the first kind, each side of the large rectangle is a period of the Abelian differential of the first kind,

$$\int_{A_+ + A_-} \frac{dz}{\sqrt{\varphi(z)}} = \int_{C_+ + C_-} \frac{dz}{\sqrt{\varphi(z)}}, \quad \int_{B_+ + B_-} \frac{dz}{\sqrt{\varphi(z)}} = \int_{D_+ + D_-} \frac{dz}{\sqrt{\varphi(z)}}.$$

Here integrals along $A_+ + A_-$ and along $C_+ + C_-$ are equal, because the closed curves $A_+ + A_-$ and $C_+ + C_-$ are equivalent in $H_1(\bar{\mathcal{R}}, \mathbb{Z})$. Since the Abelian differential $\dfrac{dz}{\sqrt{\varphi(z)}}$ is holomorphic everywhere on the elliptic curve $\bar{\mathcal{R}}$, the Cauchy integral theorem implies $\displaystyle\int_{A_+ + A_-} \frac{dz}{\sqrt{\varphi(z)}} = \int_{C_+ + C_-} \frac{dz}{\sqrt{\varphi(z)}}$. The equation $\displaystyle\int_{B_+ + B_-} \frac{dz}{\sqrt{\varphi(z)}} = \int_{D_+ + D_-} \frac{dz}{\sqrt{\varphi(z)}}$ holds for the same reason.

When we apply the reflection principle to F, the images of D and its copies are reflected with respect to their sides. Consequently the large rectangle is translated

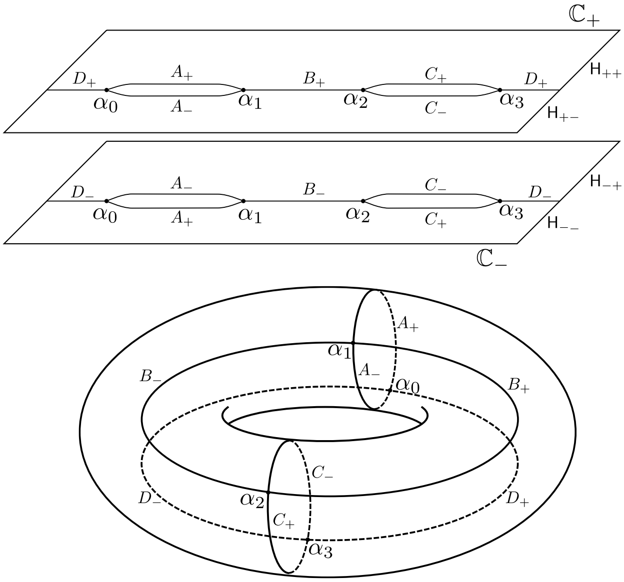

Fig. 9.11 Divide the elliptic curve into four parts.

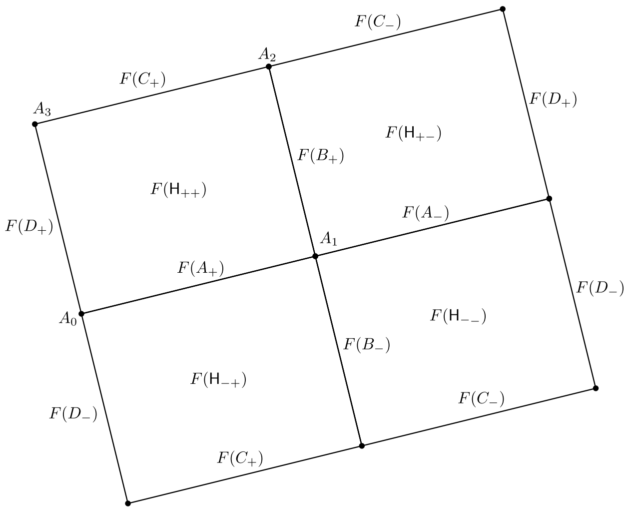

Fig. 9.12 Combine four images of F into a large rectangle.

parallely. The extended F gives a correspondence between the elliptic curve and the translated rectangles.

In fact, a general elliptic integral of the first kind gives a correspondence between an elliptic curve and a parallelogram, the sides of which are periods of the Abelian differential of the first kind. This is the Abel–Jacobi theorem, which is the theme of the following chapter.

Chapter 10
The Abel–Jacobi Theorem

In the previous chapter we showed that a holomorphic bijection from the upper half plane to a rectangle is expressed by means of an elliptic integral of the first kind. Moreover, extending this bijection, we constructed a bijection between an elliptic curve and a larger rectangle with its edges pairwise identified. In fact, via an elliptic integral of the first kind any elliptic curve can be identified with a parallelogram whose edges are pairwise identified. This is the theme of this chapter.

10.1 Statement of the Abel–Jacobi Theorem

Before discussing general elliptic curves, let us review the results of Chapter 9 from a different viewpoint.

The holomorphic correspondence between an elliptic curve and a rectangle constructed in Section 9.3 is given by a function defined by an elliptic integral of the first kind

$$F(P) = \int_{P_0}^{P} \omega_1, \qquad \omega_1 = \frac{dz}{\sqrt{\varphi(z)}}, \tag{10.1}$$

on an elliptic curve as shown in Section 9.4. Here the elliptic curve is the compactification of the Riemann surface $\{(z,w) \mid w^2 = \varphi(z)\}$, where the polynomial $\varphi(z) = \prod_{i=0}^{3}(z-\alpha_i)$ has *real* roots, $\alpha_0, \dots, \alpha_3$ and the edges of the rectangle are the A- and the B-periods of the differential form $\omega_1 = \dfrac{dz}{\sqrt{\varphi(z)}}$,

$$\Omega_A := \int_A \omega_1, \qquad \Omega_B := \int_B \omega_1. \tag{10.2}$$

In the notation of the previous chapter, the A-cycle is $A_+ + A_-$ or $C_+ + C_-$ and the B-cycle is $B_+ + B_-$ or $D_+ + D_-$.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_11

The function F is a priori defined on the upper half plane H by (10.1). Then it is extended by repeated use of the reflection principle to a correspondence between the elliptic curve and the rectangle with edges Ω_A and Ω_B. Further it defines a correspondence between the elliptic curve and parallelly translated rectangles.

We can regard F as a correspondence between the elliptic curve and a quotient set of $\mathbb{C}$ factored by the following equivalence relation. Let us define the set of periods by

$$\Gamma := \mathbb{Z}\Omega_A + \mathbb{Z}\Omega_B = \{m\Omega_A + n\Omega_B \mid m, n \in \mathbb{Z}\}, \tag{10.3}$$

which is called the *period lattice*. An equivalence relation[1] $\sim$ is defined by

$$a \sim b \overset{\text{def.}}{\iff} a - b \in \Gamma.$$

This equivalence relation pairwise identifies edges of the large rectangle mentioned above.

We denote the set of all complex numbers which are equivalent to a (the equivalence class of a) by $[a]$:

$$[a] := \{z \in \mathbb{C} \mid z \sim a\}.$$

Recall that the quotient set $\mathbb{C}/\!\sim$ of $\mathbb{C}$ by $\sim$ is the set of all such classes: $\mathbb{C}/\!\sim := \{[a] \mid a \in \mathbb{C}\}$. Since the equivalence relation is defined by the additive group Γ, it is usually denoted by $\mathbb{C}/\Gamma$. Any equivalence class is represented by an element of one fixed rectangle: for example, in Fig. 10.1 we take a rectangle with vertices 0, Ω_A, Ω_B and $\Omega_A + \Omega_B$. Thus as a concrete set $\mathbb{C}/\Gamma$ is 'a torus obtained by identifying opposite edges of a rectangle'. As a consequence F gives a holomorphic correspondence between the elliptic curve and the torus $\mathbb{C}/\Gamma$.

Now, let us consider a similar map for a general elliptic curve $\bar{\mathcal{R}}$. I mean, take a polynomial $\varphi(z)$ of degree three or four without multiple roots and compactify the Riemann surface $\mathcal{R} = \{(z, w) \mid w^2 = \varphi(z)\}$ (adding one point or two points at infinity) to get an elliptic curve $\bar{\mathcal{R}}$, and then define F on it by (10.1). We have shown in Section 8.1 that this function is a multi-valued function, the ambiguity (multi-valuedness) of which belongs to Γ.

In general the set Γ is a parallelogramic lattice as in Fig. 10.2 and in a special case (for example, $\Omega_A \in \mathbb{R}$, $\Omega_B \in i\mathbb{R}$) becomes a rectangular lattice as in the previous chapter.

The quotient set $\mathbb{C}/\Gamma$ is a torus obtained by the pairwise identification of opposite edges of the parallelogram (Fig. 10.3).

It is easy to see that the multi-valued function F defines a map from the elliptic curve to the torus $\mathbb{C}/\Gamma$, since we identify all possible values of F at a point P ($[F(P)] = F(P) + \Gamma$).

The non-trivial fact is that F gives not only a map but a holomorphic one-to-one correspondence, as was the case with an elliptic curve and a rectangle in the previous

[1] It really defines an equivalence relation because Γ is a group with respect to addition.

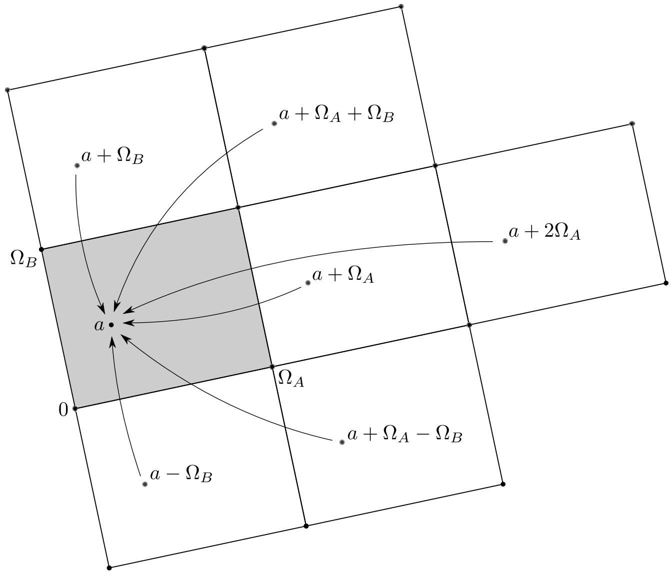

Fig. 10.1 The equivalence class $[a]$ comprising the elements $a+m\Omega_A+n\Omega_B$ $(m,n\in\mathbb{Z}\ a)$ is represented by an element a of a fixed rectangle.

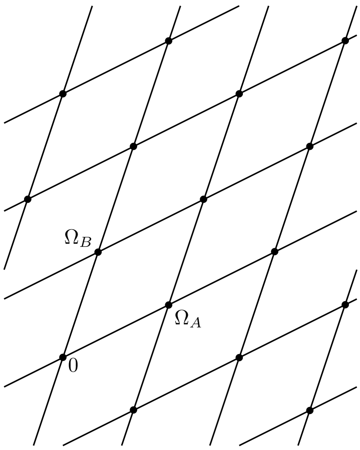

Fig. 10.2 Period lattice Γ (the set of dots).

chapter. In other words, we can 'identify' the elliptic curve $\bar{\mathcal{R}}$ and the torus $\mathbb{C}/\Gamma$. This is the Abel–Jacobi theorem:

Theorem 10.1 (Abel–Jacobi theorem)

The Abel–Jacobi map *defined by the elliptic integral of the first kind $F(P)$,*

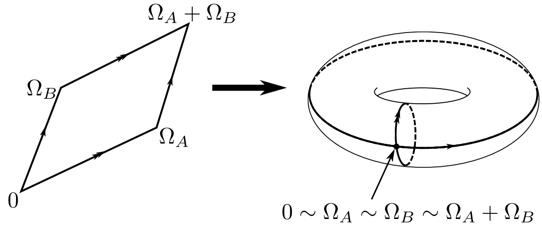

Fig. 10.3 Regard $\mathbb{C}/\Gamma$ as a torus.

$$AJ : \bar{\mathcal{R}} \ni P \mapsto \left[\int_{P_0}^{P} \omega_1 \right] \in \mathbb{C}/\Gamma, \tag{10.4}$$

is

(i) bijective.
(ii) holomorphic and its inverse is also holomorphic.

Thus AJ gives an isomorphism of complex manifolds between $\bar{\mathcal{R}}$ and $\mathbb{C}/\Gamma$.

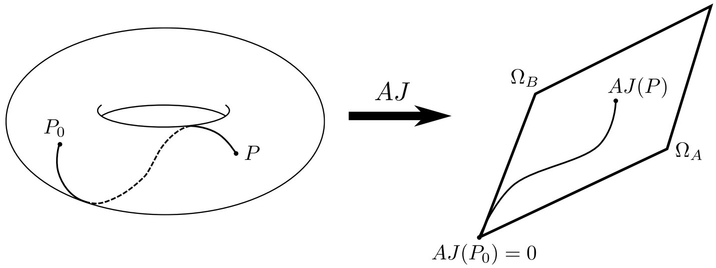

Fig. 10.4 The Abel–Jacobi map $AJ : \bar{\mathcal{R}} \to \mathbb{C}/\Gamma$.

The statement (ii) follows immediately from (i) by using general complex analysis: holomorphicity of AJ is a direct consequence of holomorphicity of $F(P)$, or that of the Abelian differential $\omega_1 = \dfrac{dz}{\sqrt{\varphi(z)}}$. The inverse function is holomorphic by virtue of Corollary A.12, which guarantees that the inverse function of a holomorphic bijection is holomorphic.

The important and non-trivial part of this theorem is the bijectivity, which will be proved in Section 10.3 and Section 10.4.

10.2 Abelian Differentials and Meromorphic Functions on an Elliptic Curve

In this section we prepare several facts on abelian differentials and meromorphic functions on elliptic curves. Corresponding facts for general compact Riemann surfaces can be found in textbooks like [FK], [Ko] and [Wy]. For elliptic curves proofs often become simple and explicit.

10.2.1 Abelian differentials of the first kind

First we study Abelian differentials of the first kind on an elliptic curve.

Let us recall that we used the fact that $\mathbb{C}/\Gamma$ is a torus before we stated Theorem 10.1. Actually, this requires a proof. For example, if $\Omega_A = \Omega_B = 0$, then $\mathbb{C}/\Gamma = \mathbb{C}$. If $\Omega_B = n\Omega_A$ ($n \in \mathbb{Z}$), then $\Gamma = \mathbb{Z}\Omega_A$ and $\mathbb{C}/\Gamma$ would be a cylinder. The fact that the lattice Γ consists of points in a parallelogram array is guaranteed by the following lemma.

Lemma 10.2 *The periods Ω_A, Ω_B do not vanish and* $\operatorname{Im}\dfrac{\Omega_B}{\Omega_A} \neq 0$.

Since the argument of $\dfrac{\Omega_B}{\Omega_A}$ is equal to $\arg\Omega_B - \arg\Omega_A$, the second half of this lemma says that 0, Ω_A and Ω_B are not on the same line. Consequently four points, 0, Ω_A, Ω_B and $\Omega_A+\Omega_B$ form a non-degenerate parallelogram as in Fig. 10.3, left.

This property of the Abelian differential ω_1 is proved by a technique which will be used in the proof of another theorem in Section 10.3. So we postpone the proof of Lemma 10.2 until then.

Next, note that the 'ratio' of two (not identically vanishing) Abelian differentials $\omega^{(1)}$ and $\omega^{(2)}$ is a meromorphic function and that, conversely, if we multiply an Abelian differential ω by a meromorphic function f, the result $f\omega$ is also an Abelian differential. In fact, in a neighbourhood of a point of a Riemann surface with a local coordinate z an Abelian differential is expressed as $g(z)\,dz$, where $g(z)$ is a meromorphic function. Therefore, for $\omega^{(1)} = g^{(1)}(z)\,dz$ and $\omega^{(2)} = g^{(2)}(z)\,dz$ their ratio, $\dfrac{\omega^{(1)}}{\omega^{(2)}} = \dfrac{g^{(1)}(z)}{g^{(2)}(z)}$, is a meromorphic function. On the other hand, multiplication of ω by a meromorphic function f, $\omega = g(z)\,dz \mapsto f\omega = (f(z)\,g(z))\,dz$, is naturally defined and gives an Abelian differential.

The only thing which we have to check is that the 'ratio' and the 'multiplication by a meromorphic function' do not depend on the choice of local coordinates. Let z and $\tilde{z}$ be two local coordinates on an open set. The rule of coordinate change of a one-form is the following (cf. Section 6.2): $g(z)\,dz = \tilde{g}(\tilde{z})\,d\tilde{z}$, if and only if $g(z) = \tilde{g}(\tilde{z}(z))\dfrac{d\tilde{z}}{dz}$. (In the right-hand side we regard $\tilde{z}$ as a function of z.) Thanks to this rule, the ratio of $\omega^{(1)}$ and $\omega^{(2)}$ does not depend on the choice of local coordinates because $\dfrac{g^{(1)}(z)}{g^{(2)}(z)} = \dfrac{\tilde{g}^{(1)}(\tilde{z})}{\tilde{g}^{(2)}(\tilde{z})}$. The multiplication of ω by f is also independent of the choice of local coordinates, as $f\omega = (f(z)\,g(z))\,dz = (f(z(\tilde{z}))\,\tilde{g}(\tilde{z}))\,d\tilde{z}$.

We proved in Section 8.1 that the one-form $\omega_1 = \dfrac{dz}{w}$ is an Abelian differential of the first kind, in other words, holomorphic everywhere on $\bar{\mathcal{R}}$, and never vanishes (if we express it as $\omega_1 = f(\tilde{z})\,d\tilde{z}$ in a neighbourhood of a point by a local coordinate $\tilde{z}$, $f(\tilde{z})$ does not vanish). Therefore there is a natural correspondence between meromorphic functions and Abelian differentials as follows:

$$\begin{array}{ccc} \text{meromorphic function} & \longleftrightarrow & \text{Abelian differential} \\ f & \longrightarrow & \omega := f\omega_1 \\ f := \dfrac{\omega}{\omega_1} & \longleftarrow & \omega \end{array} \tag{10.5}$$

Since ω_1 never vanishes, zeros and poles are preserved by this correspondence: zeros of f and zeros of ω coincide and poles of f and poles of ω coincide.

The following lemma is fundamental.

Lemma 10.3 *Any Abelian differential of the first kind is a constant multiple of* ω_1.

Proof By the above correspondence, the statement means 'an everywhere holomorphic function f on $\bar{\mathcal{R}}$ is a constant function', which follows from compactness of $\bar{\mathcal{R}}$. (Recall that the elliptic curve $\bar{\mathcal{R}}$ is homeomorphic to a torus, which is a bounded closed set in $\mathbb{R}^3$ and therefore is compact.)

In fact, the absolute value $|f|$ of the holomorphic function f is a real-valued continuous function, which attains a maximum somewhere, by the theorem in analysis. (A real-valued function on a compact set attains a maximum.) If $|f(P_0)|$ ($P_0 \in \bar{\mathcal{R}}$) is a maximum, it is a maximum of $|f|$ in a coordinate neighbourhood U of P_0. The maximum principle[2] in complex analysis claims that a holomorphic function $f(z)$ on a connected open set is a constant function, if $|f(z)|$ attains a maximum. Therefore f is constant on U. By the identity theorem a holomorphic function f is constant on $\bar{\mathcal{R}}$, as it is constant on an open set. □

That is all we need for Abelian differentials of the first kind.

10.2.2 Abelian differentials of the second/third kinds and meromorphic functions

Next we turn to Abelian differentials of the second and the third kinds on an elliptic curve. As defined in Definition 8.9 we call a meromorphic one-form which has

[2] For example, §3.4 of Chapter 4 in [Ah].

poles without residues an *Abelian differential of the second kind* and a meromorphic one-form which has poles with non-zero residues an *Abelian differential of the third kind*.

The following is a fundamental fact about residues of Abelian differentials defined on $\bar{\mathcal{R}}$.

Proposition 10.4 *If ω is an Abelian differential on $\bar{\mathcal{R}}$, the sum of its residues at all poles is equal to zero.*

Proof Let us cut open the elliptic curve $\bar{\mathcal{R}}$ along an A-cycle and a B-cycle, both of which do not pass through poles of ω, to obtain a quadrilateral S_0 as in Fig. 10.5.[3]

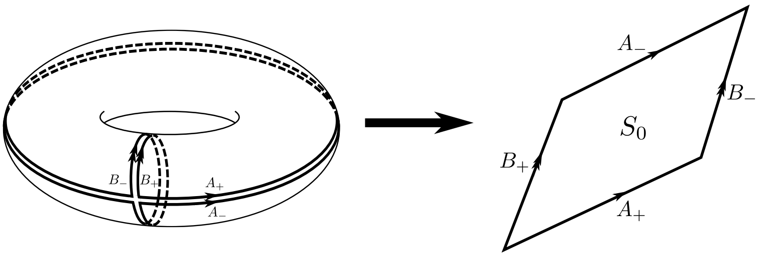

Fig. 10.5 Cut open $\bar{\mathcal{R}}$.

The residue of the Abelian differential ω at a point P is equal to the integral $\dfrac{1}{2\pi i}\displaystyle\int_{C_P} \omega$ of ω along a simple closed curve C_P around P. Expressing all the residues of ω in this way and deforming the integration contours by Cauchy's integral theorem, we can express the sum of the residues by the integral of ω along the boundary ∂S_0 of the quadrilateral S_0 as follows:

$$\sum_{P:\,\text{pole of }\omega} \operatorname{Res}_P \omega = \frac{1}{2\pi i}\int_{\partial S_0} \omega.$$

Since A_+ and A_- are one and the same closed curve on $\bar{\mathcal{R}}$, ω on A_+ and ω on A_- coincide. Therefore the integral of ω on A_+ and that on A_- are equal. For the same reason the integrals on B_+ and on B_- coincide. Thus, taking the directions of the integration contour into account, we have

$$\int_{\partial S_0} \omega = \left(\int_{A_+} - \int_{A_-} + \int_{B_+} - \int_{B_-}\right)\omega = 0.$$

This shows that the sum of residues is equal to 0. □

Using this fact, we can show the following.

[3] Here the symbols $A_\pm$ and $B_\pm$ are not those in the previous chapter.

Lemma 10.5 *There is no meromorphic function defined on $\bar{\mathcal{R}}$ with only one pole which is simple.*

Proof Suppose that such a meromorphic function f existed. The Abelian differential $\omega = f\,\omega_1$ corresponding to f by (10.5) would have "one simple pole with residue 0" because of Proposition 10.4. Saying 'P is a simple pole of ω' means that ω has an expansion like

$$\left(\frac{c_{-1}}{z-z_P} + \text{(holomorphic function)}\right) dz, \qquad c_{-1} \neq 0,$$

in a neighbourhood of P, where z is a local coordinate such that $z = z_P$ at P. However 'the residue at P is 0' means that $c_{-1} = 0$, which is a contradiction. Therefore such f cannot exist. □

Lemma 10.6 *Let f be a meromorphic function defined on $\bar{\mathcal{R}}$. Then the number of zeros of f with multiplicity (= the sum of multiplicity of zeros) is equal to the number of poles of f with multiplicity (= the sum of multiplicity of poles).*

Proof This is a direct consequence of the argument principle, Theorem A.10, which states that (the number of zeros with multiplicity) − (the number of poles with multiplicity) is the sum of all residues of the meromorphic one-form $\dfrac{df}{f} = \dfrac{f'}{f}dz$. Such a sum is equal to zero by Proposition 10.4.

Here we used the fact that $df := \dfrac{df}{dz}dz$ does not depend on the local coordinate z and therefore is globally defined on $\bar{\mathcal{R}}$. In fact, if $\tilde{z}$ is another local coordinate, we have

$$\frac{d}{d\tilde{z}} f(z(\tilde{z}))\, d\tilde{z} = \frac{d}{dz} f(z)\, \frac{dz}{d\tilde{z}}\, d\tilde{z} = \frac{d}{dz} f(z)\, dz$$

by the chain rule and the rule of coordinate change of a one-form (cf. Section 6.2). □

10.2.3 Construction of special meromorphic functions and Abelian differentials

According to the correspondence (10.5) between Abelian differentials and meromorphic functions, Abelian differentials of the second kind and of the third kind correspond to 'meromorphic functions whose poles have residue 0' and 'meromorphic functions which have a pole with non-zero residue' respectively.

Keeping this in mind, let us construct the following meromorphic functions and Abelian differentials, which we use in the proof of the Abel–Jacobi theorem: We fix two points P and Q on the elliptic curve $\bar{\mathcal{R}}$ and an integer N ($N \geq 2$).

(I) A meromorphic function $f_{P,N}$ and an Abelian differential of the second kind $\omega(P,N) = f_{P,N}\,\omega_1$, each of which has a pole of order N at P and is holomorphic everywhere else.

(II) An Abelian differential $\omega_3(P,Q)$ of the third kind, which has simple poles at P and Q, is holomorphic everywhere else and satisfies the normalisation conditions:

- $\operatorname{Res}_P \omega_3(P,Q) = 1$, $\operatorname{Res}_Q \omega_3(P,Q) = -1$.
- $\displaystyle\int_{A^0} \omega_3(P,Q) = 0$.

Here A^0 is a fixed curve representing the A-cycle, which does not pass through P or Q.

Why do we not consider the case $N = 1$ for $\omega(P,N)$? It is because of Lemma 10.5. In fact, the construction discussed later does not work for $N = 1$. Note also that the residue of $f_{P,N}$ at P is automatically 0 for the same reason as in Lemma 10.5, namely by Proposition 10.4, since it has only one pole.

(I) Construction of $f_{P,N}$, or, equivalently, an Abelian differential $\omega(P,N)$ of the second kind.

Remark 10.7 An Abelian differential of the second kind has already appeared in Section 8.2 as an integrand of an elliptic integral of the second kind. This does not fit our purpose, as it has its second-order pole at infinity and not at an arbitrarily fixed point P. Moreover, if $\varphi(z)$ is a polynomial of degree four, there are two points at infinity, so it has two second-order poles.

The elliptic curve $\bar{\mathcal{R}}$ is the union of the Riemann surface $\mathcal{R} = \{(z,w) \mid w^2 = \varphi(z)\}$ and points at infinity. At first we construct $f_{P,N}$ for $P = (z_P, w_P)$, which is not at infinity (i.e., z_P is finite) and not a branch point (i.e., $w_P \neq 0$, $\varphi(z_P) \neq 0$)). In a neighbourhood of such a point we can use z in the equation $w^2 = \varphi(z)$ as a local coordinate. The desired function $f_{P,N}$ should have the form

$$\frac{1}{(z-z_P)^N} \times \begin{pmatrix} \text{a holomorphic function not} \\ \text{vanishing at } z = z_P \end{pmatrix}, \tag{10.6}$$

in this local coordinate. “Then, can we use the factor $\dfrac{1}{(z-z_P)^N}$ itself as $f_{P,N}$?” Sorry, no. Indeed this function *has* a pole of order N at $P = (z_P, w_P)$, but it has another pole of order N at $P' = (z_P, -w_P)$, which is on a different sheet from the sheet on which P is lying. (Note that, if $(z_P, w_P) \in \mathcal{R}$, then $(z_P, -w_P) \in \mathcal{R}$ and, since $w_P \neq 0$, w_P and $-w_P$ are distinct.)

Can we erase the pole at P' of $\dfrac{1}{(z-z_P)^N}$, keeping the pole at P? For that purpose, it is enough to multiply by a function which does not vanish at (z_P, w_P) but vanishes at $(z_P, -w_P)$ in the N-th order ($O((z-z_P)^N)$). When we use only z, we construct a function which behaves in the same way both at P and at P'. So, we need to use w as well. A difference of P and P' is that the square root $\sqrt{\varphi(z)}$ takes different signs at P and P': $\sqrt{\varphi(z_P)} = w_P$ at P and $\sqrt{\varphi(z_P)} = -w_P$ at P'. Let us make use of this difference of signs.

The Taylor expansion of $w(z) = \sqrt{\varphi(z)}$ in a neighbourhood of P is

$$w(z) = w_P + w_{P,1}(z - z_P) + \cdots + w_{P,N-1}(z - z_P)^{N-1} + O((z - z_P)^N). \tag{10.7}$$

We denote the sum of the first N terms by $w_P^{(N)}(z)$:

$$w_P^{(N)}(z) := w_P + w_{P,1}(z - z_P) + \cdots + w_{P,N-1}(z - z_P)^{N-1}. \tag{10.8}$$

At P' the sign of $w(z) = \sqrt{\varphi(z)}$ changes and the expansion becomes

$$w(z) = -w_P^{(N)}(z) + O((z - z_P)^N).$$

Hence, $w(z) + w_P^{(N)}(z)$ is divisible by $(z - z_P)^N$ at P'. On the other hand we have $w(z_P) + w_P^{(N)}(z_P) = 2w_P \neq 0$ at P because of (10.7). As we discussed in Section 7.1, z and w are holomorphic functions on $\mathcal{R}$ (including branch points). Thus

$$f_{P,N}(z) := \frac{w(z) + w_P^{(N)}(z)}{(z - z_P)^N} \tag{10.9}$$

has a Laurent expansion of the form (10.6) in a neighbourhood of P and is holomorphic at any other point of $\mathcal{R}$. "Done." Uh, sorry, it is too early to think so. What we want is a meromorphic function with a pole only at P over the *elliptic curve* $\bar{\mathcal{R}}$. So, we have to check that $f_{P,N}$ above is holomorphic not only on $\mathcal{R}$ but also at the points at infinity.

To examine the behaviour of $f_{P,N}$ at points at infinity, let us rewrite (10.9) with a local coordinate in a neighbourhood of these points. First we assume that $\underline{\varphi(z) \text{ is of degree four}}$. We know that we can use $\zeta = \dfrac{1}{z}$ as a local coordinate at the points at infinity in this case. (See Section 7.2.2; at infinities $\zeta = 0$.) In this coordinate $f_{P,N}$ is expressed as

$$f_{P,N}(\zeta^{-1}) = \frac{w(\zeta^{-1}) + w_P^{(N)}(\zeta^{-1})}{(\zeta^{-1} - z_P)^N} = \frac{\zeta^N \sqrt{\varphi(\zeta^{-1})} + \zeta^N w_P^{(N)}(\zeta^{-1})}{(1 - z_P \zeta)^N}.$$

The denominator becomes 1 at $\zeta = 0$. The second term in the numerator, $\zeta^N w_P^{(N)}(\zeta^{-1})$, converges to 0 when $\zeta \to 0$, as $w_P^{(N)}(z)$ is a polynomial in z of degree $N - 1$ (cf. (10.8)). What remains is the term $\zeta^N \sqrt{\varphi(\zeta^{-1})}$ in the numerator. Since $\varphi(z)$ is of degree four in z, $\sqrt{\varphi(\zeta^{-1})}$ diverges with order ζ^{-2} when $\zeta \to 0$. Hence, under the assumption $N \geq 2$, $\zeta^N \sqrt{\varphi(\zeta^{-1})}$ has a finite limit when $\zeta \to 0$. (If $N = 1$, ζ^N cannot suppress the divergence of $\sqrt{\varphi(\zeta^{-1})}$ and, as we have mentioned already, this construction does not work.)

Thus we have proved that $f_{P,N}$ in (10.9) satisfies the desired properties, when $\varphi(z)$ is of degree four. The idea of the proof for $\underline{\varphi(z) \text{ of degree three}}$ is the same, so we leave the proof to the reader.

Exercise 10.8 Show that $f_{P,N}$ in (10.9) is holomorphic at infinity when $\varphi(z)$ is of degree three. (Hint: In this case the point at infinity is a branch point. Take $\eta = \dfrac{w}{z^2}$ as a local coordinate there.)

Let us proceed to the construction of $f_{P,N}$, when P is a branch point. (This means $(z_P, w_P) = (\alpha_i, 0)$, where $\varphi(\alpha_i) = 0$; $i = 0,1,2,3$ for $\varphi(z)$ of degree four, $i = 1,2,3$ for $\varphi(z)$ of degree three.)

Exercise 10.9 We cannot use the construction above when P is a branch point, $(z_P, w_P = 0)$ $(\varphi(z_P) = 0)$. Explain why. (Hint: Compute the coefficient of $w_P^{(N)}(z)$ in (10.8).)

Around a branch point P we use w as a local coordinate, in which $f_{P,N}$ should have a form like $\dfrac{1}{w^N} + \cdots$. So, one might bet that $\dfrac{1}{w^N}$ would be a candidate for $f_{P,N}$, but it has a pole at each point, where $w = 0$. This means that $\dfrac{1}{w^N}$ has poles at all branch points. This is not good. We need other candidates.

Those who have solved Exercise 10.9 might have noticed that $\dfrac{dw}{dz}$ diverges at P. (This is one of the reasons why w cannot be a holomorphic function of z.) By inverting this formula, we have $\dfrac{dz}{dw} = 0$, which means that z has the form (constant) + (terms of degree ≥ 2 in w) as a function of w. To be more precise, let us compute the Taylor expansion of $z = z(w)$ at P up to the second order. Differentiating the relation $w^2 = \varphi(z)$ by w and substituting $(z, w) = (\alpha_i, 0)$, we have

$$0 = \varphi'(\alpha_i) \frac{dz}{dw}(0).$$

(Here φ' is the derivative of φ by z.) Since the polynomial $\varphi(z)$ does not have multiple roots, $\varphi'(\alpha_i) \neq 0$. Therefore $\dfrac{dz}{dw}(0) = 0$. Moreover, Differentiating $w^2 = \varphi(z)$ twice by w, substituting $(z, w) = (\alpha_i, 0)$ and using $\dfrac{dz}{dw}(0) = 0$, we obtain

$$\left.\frac{d^2 z}{dw^2}\right|_{(z,w)=(\alpha_i,0)} = \frac{2}{\varphi'(\alpha_i)} (\neq 0),$$

from which follows the Taylor expansion of $z(w)$:

$$z(w) = \alpha_i + \frac{1}{\varphi'(\alpha_i)} w^2 + O(w^3). \tag{10.10}$$

Since the first-order term in w is zero and the second-order term does not vanish in this expansion, the function

$$\frac{1}{z(w) - \alpha_i}$$

has a second-order pole at the point $w = 0$, in other words, at P. Furthermore, if $z \neq \alpha_i$, the denominator does not vanish, which means that this function is holomorphic at any point of $\mathcal{R}$ except at P. This *is* the candidate for $f_{P,2}$. It remains to check holomorphicity at infinity, which can be proved in the same way as the non-branch point case. We leave the detailed proof to the reader together with the verification of the answer for general N: When N is even,

$$f_{P,N} = \frac{1}{(z-\alpha_i)^{N/2}}, \tag{10.11}$$

and when N is odd and ≥ 3,

$$f_{P,N} = \frac{w}{(z-\alpha_i)^{(N+1)/2}}. \tag{10.12}$$

Exercise 10.10 Check that (10.11) and (10.12) have a pole of order N at $P = (\alpha_i, 0)$ and are holomorphic at any other point of the elliptic curve $\bar{\mathcal{R}}$.

The case when P is at infinity still remains. In this case, we have only to rewrite the equation $w^2 = \varphi(z)$ by using the coordinate change rule $(\zeta, \eta) = \left(\frac{1}{z}, \frac{w}{z^2}\right)$ at infinity and construct a meromorphic function with a pole of order N at $\zeta = 0$. As explained in Section 7.2.2, the equation for (ζ, η) is of the form $\eta^2 =$ (a polynomial in ζ of degree three or four). So we can just replace (z, w) by (ζ, η) in the discussion above to construct $f_{P,N}$.

(II) Construction of an Abelian differential $\omega_3(P,Q)$ of the third kind.

Now take two distinct points P and Q. We are going to construct an Abelian differential $\omega_3(P,Q)$ of the third kind, which satisfies the following:

- $\omega_3(P,Q)$ is holomorphic except at P and Q, and has simple poles at P and Q.
- $\mathrm{Res}_P\, \omega_3(P,Q) = 1$, $\mathrm{Res}_Q\, \omega_3(P,Q) = -1$.
- $\displaystyle\int_{A^0} \omega_3(P,Q) = 0$.

Here the integration contour A^0 is one of closed curves which represent the homology class of the A-cycle and does not pass through P or Q. (Since $\omega_3(P,Q)$ has residues, contour integrals depend not only on the homology class but also on curves representing the class.)

Among the three conditions the last one is not very important in the sense that, if the existence of $\tilde{\omega}_3(P,Q)$ satisfying all of the above conditions but the last is shown, the existence of $\omega_3(P,Q)$ satisfying all the conditions follows from it. In fact, notice that

- For any $\lambda \in \mathbb{C}$ $\tilde{\omega}_3(P,Q) + \lambda\omega_1$ also satisfies the conditions which $\tilde{\omega}_3(P,Q)$ satisfies.
- $\displaystyle\int_A \omega_1 = \Omega_A \neq 0$ (Lemma 10.2).

Hence it is easy to show that $\omega_3(P,Q) := \tilde{\omega}_3(P,Q) + \lambda\omega_1$ satisfies all the conditions, including the normalisation condition $\displaystyle\int_{A^0} \omega_3(P,Q) = 0$, if we set $\lambda = -\dfrac{1}{\Omega_A}\displaystyle\int_{A^0} \tilde{\omega}_3(P,Q)$.

The basic idea of the construction of $\tilde{\omega}_3(P,Q)$ is the same as that of the construction of $f_{P,N}$; find a candidate meromorphic expression in z and w, check the properties by Laurent expansions at P, Q and at points at infinity. Here we give the answer to the most fundamental case. Checking the properties and construction of the other cases is left to the reader as an exercise.

Let $\varphi(z)$ be a polynomial of degree four. In this case the elliptic curve $\bar{\mathcal{R}}$ has two points $\infty_\pm$ at infinity.

(a) Case $P, Q \neq \infty_\pm$.

Assume that P and Q have coordinates $P = (z_P, w_P = \sqrt{\varphi(z_P)})$ and $Q = (z_Q, w_Q = \sqrt{\varphi(z_Q)})$ respectively. (The signs of $\sqrt{\ }$ are irrelevant here.) In this case

$$\tilde{\omega}_3(P,Q) := \frac{1}{2}\left(\frac{w+w_P}{z-z_P} - \frac{w+w_Q}{z-z_Q}\right)\frac{dz}{w}. \tag{10.13}$$

Exercise 10.11 Check that $\tilde{\omega}_3(P,Q)$ in (10.13) satisfies all the conditions (holomorphic on $\bar{\mathcal{R}} \smallsetminus \{P,Q\}$; simple poles at P and Q; $\mathrm{Res}_P = 1$, $\mathrm{Res}_Q = -1$).

The other cases,

(b) Case $P = \infty_+$, $Q \neq \infty_\pm$,

(c) Case $P = \infty_+$, $Q = \infty_-$,

are limits of the case considered above.[4]

Exercise 10.12 Construct $\tilde{\omega}_3(P,Q)$ for cases (b) and (c). (Hint: When z_P diverges to $\infty_\pm$, $w_P = \sqrt{\varphi(z_P)}$ is almost proportional to z_P^2. Therefore $\tilde{\omega}_3(P,Q)$ in (a) itself diverges, when $P \longrightarrow \infty_\pm$. Instead, find an appropriate parameter $\lambda = \lambda(z_P)$ depending on z_P and make $\lim\limits_{z_1\to\infty} (\tilde{\omega}_3(P,Q) - \lambda\omega_1)$ have a limit. After finding $\tilde{\omega}_3(P,Q)$ in this way, check that it really satisfies the required conditions separately.)

Exercise 10.13 Find $\tilde{\omega}_3(P,Q)$ for the case $\deg\varphi = 3$. (Hint: When P and Q are not at infinity, the answer is the same as (10.13). When P or Q is at infinity, one should take a local coordinate at infinity carefully. Note that there is only one point at infinity and, consequently, it is impossible that both P and Q are at infinity. This problem is a little difficult.)

Remark 10.14 Meromorphic functions satisfying similar conditions as $f_{P,N}$ or Abelian differentials satisfying similar conditions as $\omega_3(P,Q)$ exist on any compact Riemann surface. But their proofs require advanced analysis. However, for the

[4] Later we use only the ‘existence’ and not the explicit form of them. So, if you look at the formula (10.13) and can accept the existence of $\tilde{\omega}_3(P,Q)$ for other cases by analogy, it is sufficient to read the remaining part of this book.

hyperelliptic curves defined from the Riemann surfaces of $w^2 = \varphi(z)$ ($\varphi(z)$ is a multiplicity-free polynomial of degree more than four) we can construct $f_{P,N}$ and $\omega_3(P,Q)$ explicitly by the same method discussed here.

Our introduction of the 'cast' in the proof of the Abel–Jacobi theorem is over. In the next two sections they play important roles to prove the theorem.

10.3 Surjectivity of AJ (Jacobi's Theorem)

In this section we prove the surjectivity of the Abel–Jacobi map AJ (10.4) (Jacobi's theorem).[5]

There are two ways to show surjectivity of the Abel–Jacobi map. The first one is 'finding an inverse image' of each point of $\mathbb{C}/\Gamma$, honestly following the definition of surjectivity. The other is a method which uses topological properties of holomorphic functions. The latter is shorter, but for those who are not familiar with topological spaces it might be hard to read.

10.3.1 Finding an inverse image

Here we show the existence of an inverse image of an arbitrary point $[u] \in \mathbb{C}/\Gamma$. The strategy is as follows:

(I) For a certain neighbourhood U_0 of $[0] \in \mathbb{C}/\Gamma$ we show that for each $[u] \in U_0$ there exists a $P \in \bar{\mathcal{R}}$ such that $[u] = AJ(P)$.
(II) For any P' and $N \in \mathbb{N}$ there exists a $P \in \bar{\mathcal{R}}$ such that $AJ(P) = N\,AJ(P')$ [6] in $\mathbb{C}/\Gamma$.

Surjectivity of AJ is a direct consequence of these two facts: for any $u \in \mathbb{C}$, $\dfrac{1}{N}u$ is close enough to 0 for sufficiently large $N \in \mathbb{N}$ and thus $\left[\dfrac{1}{N}u\right]$ belongs to the neighbourhood U_0 of $[0]$. As shown in (I) there exists a $P' \in \bar{\mathcal{R}}$ such that $\left[\dfrac{1}{N}u\right] = AJ(P')$. Thanks to (II) there exists a $P \in \bar{\mathcal{R}}$ such that $AJ(P) = N\,AJ(P') = N\left[\dfrac{1}{N}u\right] = [u]$. This is nothing but surjectivity. □

---

[5] Jacobi considered this theorem for hyperelliptic integrals in his article [Ja1].

[6] For a point $[w]$ ($w \in \mathbb{C}$) in $\mathbb{C}/\Gamma$ a point $N[w]$ is defined by $N[w] := [Nw]$. It is an easy exercise to show that this definition does not depend on the choice of the representative of $[w]$; if $[w] = [w']$, then $[Nw] = [Nw']$.

(I) A neighbourhood of $[0]$ is included in the image of AJ.

We show that the image of the map $F(P)$ defined by the elliptic integral of the first kind (10.1) covers a neighbourhood of $0 \in \mathbb{C}$. This means that the image of AJ covers a neighbourhood of $[0]$ in the quotient set $\in \mathbb{C}/\Gamma$.

Let us take a local coordinate z in a neighbourhood of P_0 in $\bar{\mathcal{R}}$ such that $z = 0$ at P_0, the starting point of the integration contour of the Abel–Jacobi map, (10.4). In this coordinate we denote $F(P)$ simply by $F(z)$. We want to show that, if u is close enough to 0, equation

$$u - F(z) = 0$$

always has a solution z. Of course, if $u = 0$, it has a solution, as $F(0) = 0$. Recall that ω_1 does not vanish on $\bar{\mathcal{R}}$, from which follows $\dfrac{\partial}{\partial z}(u - F(z)) = -\dfrac{d}{dz}\displaystyle\int_0^z \omega_1 \neq 0$. Therefore the implicit function theorem (Lemma 6.3) is applicable to a holomorphic function $u - F(z)$ of u and z. Thus we can claim that there exists a neighbourhood U_0 of 0 such that equation $u - F(z) = 0$ has a solution z in a neighbourhood of P_0 for any $u \in U_0$.

This is what we wanted to show.

(II) For any P' and $N \in \mathbb{Z}$, $N > 0$, there exists a $P \in \bar{\mathcal{R}}$ such that

$$AJ(P) = N\, AJ(P').$$

Without taking the quotient by Γ, this means that for any $P' \in \bar{\mathcal{R}}$ and $N \in \mathbb{Z}$ ($N > 0$) there exists $P \in \bar{\mathcal{R}}$ and $m, n \in \mathbb{Z}$ such that

$$N \int_{P_0}^{P'} \omega_1 = \int_{P_0}^{P} \omega_1 + m\Omega_A + n\Omega_B. \tag{10.14}$$

Here the integration contours are not important, so we do not write them explicitly. If $P' = P_0$, the left-hand side is zero, so we can take $P = P_0$, $m = n = 0$. If $N = 1$, then we can take $P = P'$, $m = n = 0$. Hereafter we assume that $P' \neq P_0$ and $N \geq 2$.

As a preparation we show the existence of a meromorphic function f on $\bar{\mathcal{R}}$ which satisfies

(i) f is everywhere holomorphic or has a pole only at P', the order of which is not greater than N.
(ii) f has a zero at P_0 of order not less than $N - 1$.

In the previous section we constructed functions $f_{P',k}$ ($k \geq 2$) such that $f_{P',k}$ has a pole of order k at P' and is holomorphic except at P'. Any linear combination of these functions and 1,

$$f = c_0 + c_2 f_{P',2} + \cdots + c_N f_{P',N}, \tag{10.15}$$

satisfies condition (i). Such functions form an N-dimensional linear space with basis $\{1, f_{P',2}, \dots, f_{P',N}\}$. In fact, 1, $f_{P',2}, \dots, f_{P',N}$ are linearly independent because $f_{P',k}$ has a pole of order k.

It remains to pick up a function satisfying condition (ii) from this linear space. The condition (ii) is equivalent to

$$\frac{d^k f}{dz^k}(P_0) = 0, \qquad (k = 0, \dots, N-2). \tag{10.16}$$

Substituting expression (10.15) into it, we have a homogeneous system of linear equations for $c_0, c_2, \dots, c_N$, the coefficients of which are derivatives of $f_{P',k}$ at P_0. We are imposing $N-1$ linear conditions on elements of an N-dimensional linear space, hence we have at least a one-dimensional solution space, which means that there exists a *non-zero* meromorphic function satisfying both conditions (10.15) and (10.16). This is what we were searching for.

Let f be one of such functions. According to Lemma 10.6 it has the same number of zeros and poles. Thus there are three possible cases:

(a) P' is a pole of order N, P_0 is a zero of order $N-1$ and there exists a zero P of the first order, which is distinct from P_0.There are no other zeros..
(b) P' is a pole of order $N-1$, P_0 is a zero of order $N-1$. There are no other zeros.
(c) P' is a pole of order N, P_0 is a zero of order N and there are no other zeros.

First we consider the case (a). The one-form $\omega_f := \dfrac{df}{f}$ has three simple poles at P', P_0 and P, where f has a pole and zeros, and the residues are $\mathrm{Res}_{P'} \omega_f = -N$, $\mathrm{Res}_{P_0} \omega_f = N-1$ and $\mathrm{Res}_P \omega_f = 1$ respectively. (See the proof of the argument principle in Section A.2.4.)

In the previous section we constructed an Abelian differential $\omega_3(P,Q)$ of the third kind, which has simple poles at P and Q, is holomorphic except at these points and satisfies normalisation conditions,

$$\begin{gathered} \mathrm{Res}_P\, \omega_3(P,Q) = 1, \ \mathrm{Res}_Q\, \omega_3(P,Q) = -1, \\ \int_{A^0} \omega_3(P,Q) = 0. \end{gathered} \tag{10.17}$$

Since all residues cancel in the linear combination $\omega_f + N\omega_3(P',P_0) - \omega_3(P,P_0)$ by these conditions, this is a holomorphic one-form on $\bar{\mathcal{R}}$, i.e., an Abelian differential of the first kind. Hence according to Lemma 10.3 there exists a constant $c \in \mathbb{C}$ such that

$$\begin{gathered} \omega_f + N\omega_3(P',P_0) - \omega_3(P,P_0) = c\,\omega_1, \\ \text{or, equivalently,} \\ \omega_f = -N\omega_3(P',P_0) + \omega_3(P,P_0) + c\,\omega_1. \end{gathered} \tag{10.18}$$

Let us integrate this over the closed curve A^0 representing the A-cycle, which we used in the normalisation condition (10.17). The result is

$$\int_{A^0} \omega_f = c \int_{A^0} \omega_1 = c\,\Omega_A. \tag{10.19}$$

On the other hand, by the definition of ω_f it is expressed as $\omega_f = d(\log f)$. Its integral along the curve A^0 is

$$\int_{A^0} d(\log f) = \log f(\text{terminal point of } A^0) - \log f(\text{initial point of } A^0).$$

The curve A^0 is closed, which means that its initial point and its terminal point are the same. "Therefore the right-hand side is zero." Don't jump to this conclusion. The logarithm $\log z$ is defined by $\log z = \log|z| + i \arg z$ as a complex function. The real part $\log|z|$ is determined uniquely without any problem, but $\arg z$ is a multi-valued function. The argument $\arg z$ is shifted by an integral multiple of 2π when z goes around the origin $z = 0$. So, the imaginary part of $\log f(\text{terminal point of } A^0)$ and $\log f(\text{initial point of } A^0)$ may differ by an integral multiple of 2π, and we can only claim that 'there exists an integer m such that $\int_{A^0} d(\log f) = 2\pi i m$'. From this fact and (10.19) it follows that

$$2\pi i m = c\,\Omega_A. \tag{10.20}$$

To integrate (10.18) along a curve representing the B-cycle we need the following important formula on the integral of $\omega_3(P,Q)$. We fix a closed curve B^0 representing the B-cycle which does not pass through P or Q.

Lemma 10.15 *There exists a curve C from Q to P such that*

$$\int_{B^0} \omega_3(P,Q) = \frac{2\pi i}{\Omega_A} \int_C \omega_1.$$

We prove this lemma later. For the present let us just use it to compute the integral of (10.18) as follows.

$$\begin{aligned} \int_{B^0} \omega_f &= \int_{B^0} (-N\,\omega_3(P',P_0) + \omega_3(P,P_0) + c\,\omega_1) \\ &= -\frac{2\pi i N}{\Omega_A} \int_{P_0}^{P'} \omega_1 + \frac{2\pi i}{\Omega_A} \int_{P_0}^{P} \omega_1 + c\,\Omega_B. \end{aligned} \tag{10.21}$$

As in the case of the integral along A, the left-hand side is expressed as $\int_{B^0} \omega_f = 2\pi i n$ for an integer n:

$$2\pi i n = -\frac{2\pi i N}{\Omega_A} \int_{P_0}^{P'} \omega_1 + \frac{2\pi i}{\Omega_A} \int_{P_0}^{P} \omega_1 + c\,\Omega_B. \tag{10.22}$$

Subtracting (10.20) multiplied by $\dfrac{\Omega_B}{2\pi i}$ from (10.22) multiplied by $\dfrac{\Omega_A}{2\pi i}$, we obtain

$$N \int_{P_0}^{P'} \omega_1 = \int_{P_0}^{P} \omega_1 - n\,\Omega_A + m\,\Omega_B, \tag{10.23}$$

which shows (10.14).

By similar computations, we can show (10.23), or (10.14) for cases (b) and (c). In fact, for the case (b), when f has a pole of order $N-1$ at P' and a zero of order $N-1$ at P_0, we can take $P = P'$. For the case (c), when f has a pole of order N at P' and a zero of order N at P_0, we can take $P = P_0$.

Exercise 10.16 Check these facts.

Thus we have proved surjectivity of the Abel–Jacobi map AJ. □

Now let us prove Lemma 10.15, which we have used. As in the previous section we cut open the elliptic curve $\bar{\mathcal{R}}$ along the curves A^0 and B^0 to obtain a quadrilateral S_0. In Fig. 10.6 $A_\pm$ are the curve A^0 and $B_\pm$ are the curve B^0.

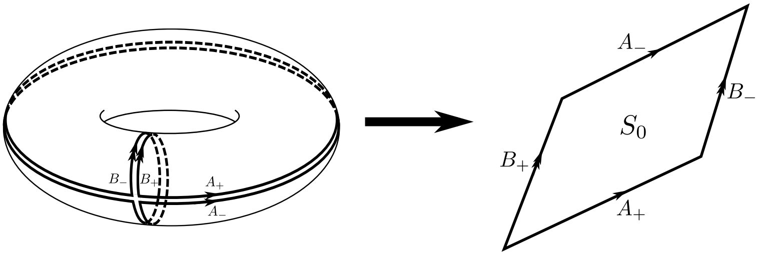

Fig. 10.6 Cut open $\bar{\mathcal{R}}$.

By the assumption, the curves $A_\pm$ and $B_\pm$ do not pass through the points P and Q.

The value of the integral of the one-form $F(z)\,\omega_3(P,Q)$ along the boundary ∂S_0 of the quadrilateral is given by the residue theorem:

$$
\begin{aligned}
&\frac{1}{2\pi i}\int_{\partial S_0} F(z)\,\omega_3(P,Q) \\
&= \mathrm{Res}_P\, F(z)\,\omega_3(P,Q) + \mathrm{Res}_Q\, F(z)\,\omega_3(P,Q) \qquad (10.24)\\
&= F(P) - F(Q) = \int_{P_0}^{P} \omega_1 - \int_{P_0}^{Q} \omega_1 = \int_{Q}^{P} \omega_1
\end{aligned}
$$

as in the proof of Proposition 10.4. Here the last line follows from the definition of $\omega_3(P,Q)$ (P is a simple pole with residue 1 and Q is a simple pole with residue -1) and from the definition (10.1) of $F(z)$.

On the other hand this integral is expressed as the sum of integrals along sides $A_\pm$, $B_\pm$ of the quadrilateral S_0 (Fig. 10.6):

$$
\int_{\partial S_0} F(z)\,\omega_3(P,Q) = \left(\int_{A_+} + \int_{B_-} - \int_{A_-} - \int_{B_+}\right) F(z)\,\omega_3(P,Q).
$$

Note that to move from A_+ to A_- we have to go around the B-cycle on the elliptic curve. When another B-cycle is added to the integration contour defining the function $F(z)$, the value of $F(z)$ increases by a constant $\Omega_B = \int_B \omega_1$. Therefore

$$\begin{aligned}
&\int_{A_+} F(z)\,\omega_3(P,Q) - \int_{A_-} F(z)\,\omega_3(P,Q) \\
&= \int_{A_+} (F(z) - F(z \circlearrowleft_B))\,\omega_3(P,Q) \\
&= \int_{A_+} (-\Omega_B)\,\omega_3(P,Q) = -\Omega_B \left(\int_{A^0} \omega_3(P,Q) \right).
\end{aligned}$$

($F(z \circlearrowleft_B)$ means that a B-cycle is added to the integration contour defining $F(z)$.) The above expression is zero because of the normalisation condition (10.17) for $\omega_3(P,Q)$.

Similarly the difference of integrals along $B_\pm$ is

$$\int_{B_-} F(z)\,\omega_3(P,Q) - \int_{B_+} F(z)\,\omega_3(P,Q) = \Omega_A \int_{B^0} \omega_3(P,Q).$$

Substituting these results into (10.24), we have

$$2\pi i \int_Q^P \omega_1 = \Omega_A \int_{B^0} \omega_3(P,Q),$$

which proves Lemma 10.15. □

Using a similar technique as in this proof, we can show Lemma 10.2, which we have left unproved. It asserts: Ω_A and Ω_B do not vanish and

$$\operatorname{Im} \frac{\Omega_B}{\Omega_A} \neq 0.$$

Proof This time we integrate a one-form $F(P)\,\bar{\omega}_1$ along the boundary ∂S_0 of S_0 in Fig. 10.6. Here $\bar{\omega}_1$ is the complex conjugate of ω_1. "What is a *complex conjugate* of a one-form?" Don't be scared. For a one-form $\omega = f(z)\,dz$ we define its complex conjugate just by taking complex conjugates of the coefficient function $f(z)$ and dz:

$$\bar{\omega} = \overline{f(z)}\,\overline{dz} = \overline{f(z)}\,(dx - i\,dy).$$

We can express ω as a real one-form, $\omega = (u_1(x,y)\,dx + u_2(x,y)\,dy) + i(v_1(x,y)\,dx + v_2(x,y)\,dy)$ by real coordinates (x,y) and real functions $u_1,\dots,v_2$. Using this expression we can define $\bar{\omega}$ also as

$$\bar{\omega} = (u_1(x,y)\,dx + u_2(x,y)\,dy) - i(v_1(x,y)\,dx + v_2(x,y)\,dy).$$

These definitions are obviously equivalent. We can easily show that the definition does not depend on the choice of local coordinates, which we leave to the reader as an exercise.

The integral of $F(P)\,\bar{\omega}_1$ is decomposed as

$$\int_{\partial S_0} F(z)\,\bar{\omega}_1 = \left(\int_{A_+} - \int_{A_-} + \int_{B_-} - \int_{B_+}\right) F(z)\,\bar{\omega}_1. \tag{10.25}$$

The first two terms become

$$\int_{A_+} F(z)\,\bar{\omega}_1 - \int_{A_-} F(z)\,\bar{\omega}_1 = \int_A (-\Omega_B)\,\bar{\omega}_1 = -\Omega_B \left(\int_A \bar{\omega}_1\right)$$

by a similar argument in the previous proof. Since $\int_\gamma \bar{\omega} = \overline{\int_\gamma \omega}$ according to the definition of the complex conjugate of a one-form, the factor in the last parentheses is equal to $\overline{\int_A \omega_1} = \bar{\Omega}_A$. Hence,

$$\int_{A_+} F(z)\,\bar{\omega}_1 - \int_{A_-} F(z)\,\bar{\omega}_1 = -\Omega_B \bar{\Omega}_A.$$

Similarly the difference of integrals on $B_\pm$ is

$$\int_{B_-} F(z)\,\bar{\omega}_1 - \int_{B_+} F(z)\,\bar{\omega}_1 = \Omega_A\,\bar{\Omega}_B.$$

Substituting the results into (10.25), we obtain

$$\int_{\partial S_0} F(z)\,\bar{\omega}_1 = \Omega_A\,\bar{\Omega}_B - \bar{\Omega}_A\,\Omega_B. \tag{10.26}$$

On the other hand, expressing ω_1 as $\omega_1 = \psi(z)(dx + i\,dy)$ and applying Green's formula,

$$\int_{\partial D} (u(x,y)\,dx + v(x,y)\,dy) = \int_D \left(-\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) dx\,dy, \tag{10.27}$$

(or, equivalently, using the generalised Stokes theorem for a one-form), we can rewrite the left-hand side of (10.25) as

$$\int_{\partial S_0} F(z)\,\bar{\omega}_1 = -2i \int_{S_0} |\psi(z)|^2\,dx\,dy. \tag{10.28}$$

In fact, applying (10.27) to the left-hand side, we have

$$
\begin{aligned}
\int_{\partial S_0} F(z)\,\bar{\omega}_1 &= \int_{\partial S_0} F(z)\,\overline{\psi(z)}\,(dx - i\,dy) \\
&= \int_{S_0} \left(-\frac{\partial}{\partial y}\bigl(F(z)\,\overline{\psi(z)}\bigr) - i\frac{\partial}{\partial x}\bigl(F(z)\,\overline{\psi(z)}\bigr) \right) dx\,dy.
\end{aligned}
\tag{10.29}
$$

On the other hand, differentiating the definition $F(z) = \int_{P_0}^{z} \psi(z)\,(dx + i\,dy)$, we have $\dfrac{\partial F}{\partial x} = \psi(z)$ and $\dfrac{\partial F}{\partial y} = i\psi(z)$. Hence,

$$
\begin{aligned}
\frac{\partial}{\partial y}\bigl(F(z)\,\overline{\psi(z)}\bigr) &= i\psi(z)\,\overline{\psi(z)} + F(z)\,\overline{\left(\frac{\partial \psi}{\partial y}\right)} \\
&= i|\psi(z)|^2 + F(z)\,\overline{\left(\frac{\partial^2 F}{\partial y\,\partial x}\right)}, \\
\frac{\partial}{\partial x}\bigl(F(z)\,\overline{\psi(z)}\bigr) &= \psi(z)\,\overline{\psi(z)} + F(z)\,\overline{\left(\frac{\partial \psi}{\partial x}\right)} \\
&= |\psi(z)|^2 + F(z)\,\overline{\left(\frac{1}{i}\,\frac{\partial^2 F}{\partial x\,\partial y}\right)}.
\end{aligned}
$$

Substituting them into (10.29), we obtain (10.28).

Combining the two results, (10.26) and (10.28), we have

$$
\frac{1}{2i}(\bar{\Omega}_A\,\Omega_B - \Omega_A\,\bar{\Omega}_B) = \int_{S_0} |\psi(z)|^2\,dx\,dy.
\tag{10.30}
$$

Since $\psi(z)$ is a non-zero holomorphic function, the integral in the right hand side is not zero. Hence, in particular, neither Ω_A nor Ω_B is zero. Moreover, as $\Omega_A\,\bar{\Omega}_B = \overline{\bar{\Omega}_A\,\Omega_B}$, the left-hand side of (10.30) is equal to $\operatorname{Im}\bar{\Omega}_A\,\Omega_B = \operatorname{Im}|\Omega_A|^2\dfrac{\Omega_B}{\Omega_A} = |\Omega_A|^2 \operatorname{Im}\dfrac{\Omega_B}{\Omega_A}$. Thus we have $\operatorname{Im}\dfrac{\Omega_B}{\Omega_A} \neq 0$ and the proof of Lemma 10.2 is over. □

10.3.2 Topological proof of surjectivity of AJ

Here we use the general theory of topological spaces in order to prove surjectivity of the Abel–Jacobi map. What we use is usually encountered in the first two years of university (mathematics major). If you are not familiar with topological spaces, you can skip this part and jump to the proof of injectivity in Section 10.4.

Recall the fundamental properties of compact sets:[7]

- The image of a compact set by a continuous map is compact.

[7] For example, Chapter 3, §1.4 and §1.5 in [Ah].

- A compact set in a Hausdorff space is closed.

The Abel–Jacobi map AJ is holomorphic, in particular, continuous. The elliptic curve $\bar{\mathcal{R}}$ is a torus, so it is compact. Since $\mathbb{C}/\Gamma$ is a Hausdorff space, we can apply the above theorems, which guarantees that the image $AJ(\bar{\mathcal{R}})$ of AJ is a *closed* subset of $\mathbb{C}/\Gamma$.

Now recall a theorem from complex analysis (Theorem A.13):

- A non-constant holomorphic function is open, which means that the image of an open set is open.

The elliptic curve $\bar{\mathcal{R}}$ itself is trivially open in $\bar{\mathcal{R}}$, so its image $AJ(\bar{\mathcal{R}})$ is *open*.

Thus $AJ(\bar{\mathcal{R}})$ is simultaneously *closed and open* in $\mathbb{C}/\Gamma$.

Lastly we use theorems on connectedness of topological spaces:[8]

- The image of a connected set by a continuous map is connected.
- A closed and open subset of a connected topological space is the whole space.

As there is a natural continuous map $\mathbb{C} \to \mathbb{C}/\Gamma$ and $\mathbb{C}$ is connected, $\mathbb{C}/\Gamma$ is connected. We have shown that $AJ(\bar{\mathcal{R}})$ is closed and open, hence it coincides with the whole space $\mathbb{C}/\Gamma$:

$$AJ(\bar{\mathcal{R}}) = \mathbb{C}/\Gamma.$$

This is the end of another proof of surjectivity of AJ. □

We can prove Lemma 10.2 as a corollary of this lemma.

Proof (Another proof of Lemma* 10.2.*) This is a direct consequence of compactness of $AJ(\bar{\mathcal{R}}) = \mathbb{C}/\Gamma$, which we used in the above proof.

Suppose that Ω_A and Ω_B are linearly *dependent* over $\mathbb{R}$. Then, the lattice $\Gamma = \mathbb{Z}\Omega_A + \mathbb{Z}\Omega_B$ is contained in $\ell = \mathbb{R}\Omega_A$ or $\ell = \mathbb{R}\Omega_B$. (See Fig. 10.7.) The set ℓ is a line (if $\Omega_A \neq 0$ or $\Omega_B \neq 0$) or the origin $\{0\}$ (if $\Omega_A = \Omega_B = 0$).

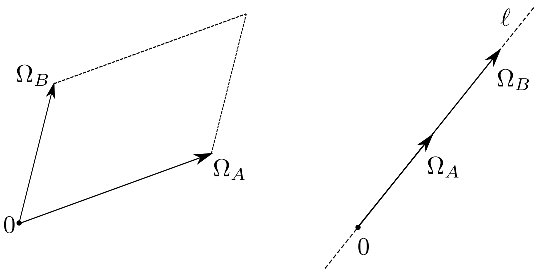

Fig. 10.7 Ω_A and Ω_B are linearly independent (left) or dependent (right) over $\mathbb{R}$.

The difference of two complex numbers α and β in a line ℓ' $(\neq \ell)$ does not belong to Γ, which means that α and β are not identified by Γ. Therefore $\mathbb{C}/\Gamma$ contains ℓ' as it is (not factored by Γ), and hence is not compact. This contradicts what we have shown above. Thus linear independence of Ω_A and Ω_B over $\mathbb{R}$ has been proved. □

[8] For example, Chapter 3, §1.3 and §1.5 of [Ah].

10.4 Injectivity of AJ (Abel's Theorem)

We prove injectivity by contradiction: Suppose that there are two distinct points P_1 and P_2, $P_1 \neq P_2$, whose images are the same: $AJ(P_1) = AJ(P_2)$. Under this assumption we construct a meromorphic function with only one pole on $\bar{\mathcal{R}}$, which is simple and located at P_2.

However such a meromorphic function cannot exist as we have shown in Lemma 10.5. This means that $AJ(P_1)$ and $AJ(P_2)$ should be distinct, if $P_1 \neq P_2$. Thus injectivity is proved.

This "impossible" meromorphic function $f(z)$ is defined by

$$f(z) := \exp \int_{Q_0}^{z} \left(\omega_3(P_1,P_0) - \omega_3(P_2,P_0) - \frac{2\pi i N}{\Omega_A}\omega_1 \right). \tag{10.31}$$

The symbols in this expression are defined as follows:

- Q_0 is a point in $\bar{\mathcal{R}}$ distinct from P_0, P_1, P_2.
- $\omega_3(P,Q)$ is the Abelian differential of the third kind, which was defined in Section 10.2.3.
- N is an integer to be determined later.

It is sufficient to show that f defined by (10.31) has the following properties.

(I) The function f is holomorphic except at P_2, where f has a simple pole.
(II) f is a single-valued meromorphic function on $\bar{\mathcal{R}}$.

(I) <u>f is holomorphic except at P_2, where f has a simple pole.</u>

The function in the exponential in the right-hand side of (10.31) may have singularities at P_1, P_0 and P_2, where the integrated Abelian differentials of the third kind have poles. Let us examine the behaviour of f in a neighbourhood of each point.

At P_2, $\omega_3(P_2,P_0)$ has a simple pole with residue 1, which has the form[9]

$$\omega_3(P_2,P_0) = \left(\frac{1}{z-P_2} + (\text{a holomorphic function}) \right) dz$$

in a local coordinate z. Integrating this we have

$$\int_{Q_0}^{z} \omega_3(P_2,P_0) = \log(z-P_2) + (\text{a holomorphic function}). \tag{10.32}$$

In this expression the logarithm of the first term in the right-hand side determines the behaviour of the integral when z approaches P_2.

Any other terms in the parentheses in the right-hand side of (10.31) are holomorphic in a neighbourhood of P_2, as they are Abelian differentials holomorphic around P_2. Hence in a neighbourhood of P_2, $f(z)$ behaves as

[9] We denote the value of the local coordinate z at P by the same P for simplicity.

$$\begin{aligned} f(z) &= \exp\bigl(-\log(z-P_2) + (\text{a holomorphic function})\bigr) \\ &= \frac{1}{z-P_2} \times (\text{a non-zero holomorphic function}), \end{aligned}$$

which shows that f has a simple pole at P_2.

By exactly the same argument we can show that f has a simple zero at P_1. In particular, it is holomorphic at P_1.

At P_0 the residues of $\omega_3(P_1,P_0)$ and $\omega_3(P_2,P_0)$ cancel. Therefore the Abelian differential in the parentheses in the right-hand side of (10.31) does not have a singularity at P_0, which implies that f is holomorphic also at this point.

(II) Single-valuedness of $f(z)$.

The possibility of multi-valuedness arises because the integration contour in (10.31) is not fixed. Since the integrated one-form is meromorphic, the value of the integral does not change by continuous deformation of the contour unless its end points change or the contour crosses over the poles of the one-form. (The Cauchy integral theorem!) The following three kinds of deformation of the integration contour can result in multi-valuedness.

(i) Adding a loop around a pole of the one-form.
(ii) Adding an A-cycle.
(iii) Adding a B-cycle.

(i) Adding a loop around poles of ω_3.

Assume that we add a loop around P_1 to the integration contour in (10.31). (The cases when we add a loop around P_0 or add a loop around P_2 are similar.) We denote this addition of the loop by $z \circlearrowleft_{P_1}$. The integral on the added loop picks up the residue of $\omega_3(P_1,P_0)$ at P_1. So we have

$$\int_{Q_0}^{z \circlearrowleft_{P_1}} \omega_3(P_1,P_0) = \int_{Q_0}^{z} \omega_3(P_1,P_0) + 2\pi i.$$

In the definition (10.31) of $f(z)$ this integral is exponentiated. Therefore the change of f caused by this deformation of the integration contour is $f \mapsto f(z) \times e^{2\pi i} = f(z)$, which is trivial.

(ii) Adding an A-cycle

Next let us consider the case when the integration contour goes around an extra A-cycle.

If the added part of the contour is the curve A^0 in the normalisation condition (10.17), then the integral of ω_3 does not change:

$$\int_{Q_0}^{z \circlearrowleft_{A^0}} \omega_3(P_i,P_0) = \int_{Q_0}^{z} \omega_3(P_i,P_0),$$

because $\displaystyle\int_{A^0} \omega_3(P_i,Q) = 0$. (Here $i = 1$ or 2; $z \circlearrowleft_{A^0}$ means the addition of the curve A^0.)

If the added curve A^1 representing the A-cycle is not A^0, we have to deform it continuously to A^0, which is possible because A^1 and A^0 are of the same homology class. During this deformation the integral may pick up residues of $\omega_3(P_i, P_0)$. Since $\operatorname{Res}_{P_i} \omega_3(P_i, P_0) = 1$ and $\operatorname{Res}_{P_0} \omega_3(P_i, P_0) = -1$, the difference of integrals of $\omega_3(P_i, P_0)$ along A^0 and along A^1 is equal to $2\pi i n_i$ for some integer n_i. Thus

$$\int_{Q_0}^{z \circlearrowleft_{A^1}} \omega_3(P_i, P_0) = \int_{Q_0}^{z \circlearrowleft_{A^0}} \omega_3(P_i, P_0) + 2\pi i n_i = \int_{Q_0}^{z} \omega_3(P_i, P_0) + 2\pi i n_i.$$

On the other hand, $\displaystyle\int_{Q_0}^{z \circlearrowleft_A} \omega_1 = \int_{Q_0}^{z} \omega_1 + \Omega_A$, where $z \circlearrowleft A$ means that a curve representing the A-cycle is added to the integration contour. (We do not have to specify a concrete curve, since ω_1 is a holomorphic Abelian differential.) Hence, if z goes around an A-cycle,

$$f(z) \mapsto f(z) \exp\left(2\pi i n_1 - 2\pi i n_2 - \frac{2\pi i N}{\Omega_A}\Omega_A\right) = f(z),$$

and the value f returns to the original value also in this case.

(iii) Adding a B-cycle.

In this case, at last the condition $AJ(P_1) = AJ(P_2)$ plays its role.

First we add the curve B^0 to the integration contour, where B^0 is the curve fixed in Lemma 10.15. Lemma 10.15 implies

$$\begin{aligned}
&\int_{Q_0}^{z \circlearrowleft_{B^0}} (\omega_3(P_1, P_0) - \omega_3(P_2, P_0)) - \int_{Q_0}^{z} (\omega_3(P_1, P_0) - \omega_3(P_2, P_0)) \\
&= \int_{B^0} (\omega_3(P_1, P_0) - \omega_3(P_2, P_0)) \\
&= \frac{2\pi i}{\Omega_A}\left(\int_{P_0}^{P_1} \omega_1 - \int_{P_0}^{P_2} \omega_1\right).
\end{aligned}$$

(Here $z \circlearrowleft_{B^0}$ has the same meaning as above.) In the last expression, the integration contours from P_0 to P_1 and P_2 are those whose existence is guaranteed by Lemma 10.15.

From assumption $AJ(P_1) = AJ(P_2)$ it follows that there exist integers M and N such that

$$\int_{P_0}^{P_1} \omega_1 = \int_{P_0}^{P_2} \omega_1 + M\Omega_A + N\Omega_B.$$

We take this 'N' as the integer N not fixed in the definition (10.31) of $f(z)$.

Having made this preparation, let us compute the value of $f(z)$, when z moves additionally around a curve B^1 representing the B-cycle. As in the case (ii), the deformation of B^1 to B^0 picks up residues of $\omega_3(P_i, P_0)$ as

$$\int_{Q_0}^{z \circlearrowleft_{B^1}} \omega_3(P_i, P_0) = \int_{Q_0}^{z \circlearrowleft_{B^0}} \omega_3(P_i, P_0) + 2\pi i m_i,$$

where the m_i's are integers. Therefore,

$$\begin{aligned}
f(z) \mapsto f(z) &\exp\left(2\pi i m_1 - 2\pi i m_2 + \frac{2\pi i}{\Omega_A}\left(\int_{P_0}^{P_1} \omega_1 - \int_{P_0}^{P_2} \omega_1\right) - \frac{2\pi i N}{\Omega_A}\int_B \omega_1\right) \\
&= f(z)\exp\left(\frac{2\pi i}{\Omega_A}(M\Omega_A + N\Omega_B) - \frac{2\pi i N}{\Omega_A}\int_B \omega_1\right) \\
&= f(z)\exp\left(2\pi i M + \frac{2\pi i N \Omega_B}{\Omega_A} - \frac{2\pi i N}{\Omega_A}\Omega_B\right) \\
&= f(z).
\end{aligned}$$

Again the value of $f(z)$ returns to its initial value.

Thus we have shown that the function $f(z)$ is well-defined as a single-valued function.

This is the end of the proof of the Abel–Jacobi theorem. □

Remark 10.17 The Abel–Jacobi map is defined not only for elliptic curves but also for any compact Riemann surface. In general, both its domain of definition and image are higher- dimensional complex manifolds. The Abel–Jacobi theorem also holds. In fact, the proof in this chapter is a rewrite of the proof for a general compact Riemann surface in the case of an elliptic curve.

Usually, 'the Abel–Jacobi theorem' refers to the theorem for general compact Riemann surfaces.

In the next chapter, keeping this Abel–Jacobi theorem in mind, we define elliptic functions as complex functions. The main actors enter the stage at last. Well, they have already been on the stage for a long time, as will be explained.

Chapter 11
The General Theory of Elliptic Functions

In the previous chapter we showed that an elliptic curve and $\mathbb{C}/\Gamma$ ($\Gamma = \mathbb{Z}\Omega_A + \mathbb{Z}\Omega_B$) are identified by the Abel–Jacobi map AJ, namely, by the elliptic integral of the first kind. Through this identification the definition of an elliptic function as the inverse function of an elliptic integral introduced in Chapter 4 and the definition as a doubly periodic meromorphic function in Chapter 0 are connected.

In this chapter we explain the definition and properties of elliptic functions given in standard textbooks of complex analysis, including the above mentioned equivalence of two definitions, from our standpoint.

11.1 Definition of Elliptic Functions

We are approaching the 'usual' definition of elliptic functions at last, after a long journey starting from elliptic integrals. However, the following definition would be most natural at this moment.

Definition 11.1 A meromorphic function on an elliptic curve is called an *elliptic function*.

This means that all those meromorphic functions (with poles or zeros at specified points) on elliptic curves which appeared in Chapter 10 *are* elliptic functions. They were actively playing on the stage even before they were defined.

Now, let us put this definition together with the Abel–Jacobi theorem,

$$\text{elliptic curve} \cong \mathbb{C}/\Gamma. \tag{11.1}$$

Here Γ is the lattice $\mathbb{Z}\Omega_A + \mathbb{Z}\Omega_B$, whose bases Ω_A and Ω_B are the A- and B-periods of the Abelian differential $\omega_1 = \dfrac{dz}{\sqrt{\varphi(z)}}$ of the first kind.

Identifying both sides of (11.1), we can regard elliptic functions as meromorphic functions on $\mathbb{C}/\Gamma$, which is obtained by factoring the set $\mathbb{C}$ of complex numbers by

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_12

the equivalence relation $u \sim u' \iff u - u' \in \Gamma = \mathbb{Z}\Omega_A + \mathbb{Z}\Omega_B$. Therefore a 'function f on $\mathbb{C}/\Gamma$' assigns a complex number $f([u])$ to each equivalence class $[u]$. In other words, it defines a value for each set of complex numbers of the form $u + m\Omega_A + n\Omega_B$ ($m, n \in \mathbb{Z}$). This means that we are considering a function f on $\mathbb{C}$ as follows:

> f assigns a value $f(u)$ to each complex number u and satisfies the condition that, if $u \sim u'$ (or, equivalently, if there exist integers m and n such that $u' = u + m\Omega_A + n\Omega_B$), $f(u) = f(u')$.

So, apart from the elliptic curve $\bar{\mathcal{R}}$, let us assume that two complex numbers Ω_A and Ω_B, which are linearly independent over the real field $\mathbb{R}$ (i.e., they span a non-degenerate parallelogram), are given and define elliptic functions as follows:

Definition 11.2 A meromorphic function $f(u)$ on $\mathbb{C}$ satisfying

$$f(u + \Omega_A) = f(u), \qquad f(u + \Omega_B) = f(u)$$

is called an *elliptic function* with periods Ω_A and Ω_B.

This is today's *standard* definition of elliptic functions. Regarding the elliptic curve as $\mathbb{C}/\Gamma$ by identification (11.1) of the Abel–Jacobi theorem, and then regarding a function on $\mathbb{C}/\Gamma$ as a function on $\mathbb{C}$, the two definitions, Definition 11.1 and Definition 11.2, are connected. Thus an elliptic function in the sense of Definition 11.1 is an elliptic function in the sense of Definition 11.2.

Remark 11.3 Then, a natural question is, "Conversely, is an elliptic function in the sense of Definition 11.2 an elliptic function defined by Definition 11.1?" Actually this is also true, but we need to prove that for arbitrary complex numbers Ω_A, Ω_B, which are linearly independent over $\mathbb{R}$, there exists an elliptic curve $\bar{\mathcal{R}} =$ (compactification of $\{(z, w) \mid w^2 = \varphi(z)\}$), whose A- and B-periods are Ω_A and Ω_B. ($\varphi(z)$ is a polynomial of degree three or four without multiple roots.)

In the next chapter we construct an example of an elliptic function in the sense of Definition 11.2 explicitly and show that it satisfies a differential equation, from which the existence of such an elliptic curve follows.

We give several examples of elliptic functions.

Example 11.4 First, let us make the statement 'an elliptic function in the sense of Definition 11.1 is an elliptic function in the sense of Definition 11.2' more explicit. As before, we take a polynomial $\varphi(z)$ of degree three or four without multiple roots. We define a projection pr from the elliptic curve to the projective line $\mathbb{P}^1$ by

$$\begin{array}{ccc} \bar{\mathcal{R}} = \overline{\{(z, w) \mid w^2 = \varphi(z)\}} & \xrightarrow{\text{pr}} & \mathbb{P}^1 \\ (z, w) & \mapsto & z \\ \infty & \mapsto & \infty. \end{array}$$

This map just selects the z-coordinate from the pair (z,w).

It might sound a little bit verbose, but let us check that this defines a meromorphic function on $\bar{\mathcal{R}}$: We can use z as a local coordinate of $\bar{\mathcal{R}}$ except at branch points $(z,w) = (\alpha_i, 0)$ (α_i is a root of $\varphi(z)$) and at infinity. So, z is trivially holomorphic except at those points. At a branch point we need to use w as a local coordinate, with respect to which z is a holomorphic function as a consequence of the implicit function theorem, Lemma 6.3. In a neighbourhood of points at infinity we have to use another local coordinate ($\xi = z^{-1}$ for the case $\deg\varphi = 4$ and $\eta = wz^{-2}$ for the case $\deg\varphi = 3$; See Section 7.2). We can show that z has a pole with respect to this coordinate.

Exercise 11.5 Show that the above defined pr has a pole of the second order at ∞ when $\deg\varphi = 3$ and simple poles at $\infty_\pm$ when $\deg\varphi = 4$. (Hint: In both cases express z as ξ^{-1} and then rewrite it in terms of the local coordinate. It is trivial when $\deg\varphi = 4$. When $\deg\varphi = 3$, show the relation $\xi^{-1} = \eta^{-2}\times$ (non-vanishing holomorphic function), using the relation of η and ξ.)

Therefore, each map in the sequence

$$\mathbb{C} \xrightarrow{\ \pi\ } \mathbb{C}/\Gamma \xrightarrow{AJ^{-1}} \bar{\mathcal{R}} \xrightarrow{\ \mathrm{pr}\ } \mathbb{P}^1$$

is a meromorphic function, where $\pi : \mathbb{C} \to \mathbb{C}/\Gamma$ is a natural projection ($u \in \mathbb{C}$ is mapped to its equivalence class $[u]$ by Γ). Hence the composition $f(u) = \mathrm{pr} \circ AJ^{-1} \circ \pi(u)$ is an elliptic function on $\mathbb{C}$ in the sense of Definition 11.2:

Since AJ is defined by an elliptic integral, the above statement is a reformulation of what was found two hundred years ago,

'The inverse function of an elliptic integral gives an elliptic function.'

from our viewpoint.

Example 11.6 Let us consider Example 11.4 more explicitly for $\varphi(z) = 4z^3 - g_2 z - g_3$. ($g_2, g_3 \in \mathbb{C}$; $\varphi(z)$ does not have multiple roots.) In the definition of the Abel–Jacobi map AJ, i.e., the definition of the elliptic integral of the first kind, there is an ambiguity of the base point P_0 of the integral. Let us fix it to infinity:

$$AJ(z) := \int_\infty^z \frac{dz}{w} = \int_\infty^z \frac{dz}{\sqrt{4z^3 - g_2 z - g_3}}.$$

Then the elliptic function constructed in Example 11.4 is called *Weierstrass's $\wp$-function*, $\wp(u) := \mathrm{pr} \circ AJ^{-1} \circ \pi(u)$.

Since $AJ(\infty) = 0$, $\wp(0) = \infty$ and $u = 0$ is a pole of $\wp(u)$. It follows from Exercise 11.5 that this pole is of the second order.

In the next chapter we construct this $\wp$-function in another way.

Example 11.7 When $\varphi(z) = (1 - z^2)(1 - k^2 z^2)$ ($k \in \mathbb{C}$, $k \neq 0, \pm 1$), it is convenient to take 0 as the base point of the Abel–Jacobi map:

$$AJ(z) := \int_0^z \frac{dz}{w} = \int_0^z \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}.$$

In this case the elliptic function constructed in Example 11.4 is *Jacobi's* sn-*function*, $\mathrm{sn}(u) := \mathrm{pr} \circ AJ^{-1} \circ \pi(u)$.

In Proposition 8.3 we computed periods of an Abelian differential of the first kind $\omega_1 = \dfrac{dz}{\sqrt{\varphi(z)}}$ for $\varphi(z) = (1-z^2)(1-k^2z^2)$ and obtained

$$\Omega_A = 4K(k), \qquad \Omega_B = 2iK'(k).$$

Therefore the periods of $\mathrm{sn}(u)$ are $4K(k)$ and $2iK'(k)$.

In Chapter 4 we defined sn on $\mathbb{R}$ as the inverse function of an incomplete elliptic integral integral of the first kind. The above construction is its extension to $\mathbb{C}$. When we considered sn on $\mathbb{R}$, we first defined it on the interval $[-K(k), K(k)]$ and then extended it to $\mathbb{R}$ so that it is a smooth periodic function with period $4K(k)$, which is eventually justified by the addition theorem. As a complex function we justify the extension by means of the Abel–Jacobi theorem.

In the final chapter of this book sn will be defined in yet another way.

11.2 General Properties of Elliptic Functions

This section is devoted to fundamental properties of an elliptic function on $\mathbb{C}$ with periods Ω_A and Ω_B. Of course, as an elliptic function is a 'meromorphic function on an elliptic curve', we have already discussed some properties in Chapter 10. However, it is important to give several proofs of one theorem, as each proof provides a new perspective. So we give new proofs of theorems already proved in Chapter 10, based on the definition of an elliptic function as a 'doubly periodic meromorphic function'. Moreover, as the notion of a 'doubly periodic meromorphic function' is clear and easy to handle in complex analysis, it makes the proofs simpler. For example, when we start from the definition 'a meromorphic function on an elliptic curve $\bar{\mathcal{R}}$', we often need to consider points at infinity and branch points separately, while we can avoid such a case-by-case check by adopting the definition 'a doubly periodic meromorphic function'.

First we start from the following simple fact.

Lemma 11.8 *Let f and g be elliptic functions with periods Ω_A and Ω_B.*

(i) *$f \pm g$, fg and f/g (if $g \not\equiv 0$) are also elliptic functions. Thus the set of all elliptic functions with the same periods forms a field (in the algebraic sense).*
(ii) *The derivative $f'(u)$ of $f(u)$ is also an elliptic function with the same periods.*

Thus the set of all elliptic functions with the same periods is a differential field.

Proof The first statement (i) follows directly from the periodicity of f and g. For example $(f \pm g)(u+\Omega_A) = f(u+\Omega_A) \pm g(u+\Omega_A) = f(u) \pm g(u) = (f \pm g)(u)$. Notice also that the set of meromorphic functions is closed under arithmetic operations.

The second statement (ii) follows from the facts that the derivative of a meromorphic function is meromorphic and that the derivative of the periodicity condition $f(u+\Omega_A) = f(u+\Omega_B) = f(u)$ is $f'(u+\Omega_A) = f'(u+\Omega_B) = f'(u)$, which shows the periodicity of f'.

This lemma can be easily proved also by Definition 11.1, because arithmetic operations and differentiation preserve the set of meromorphic functions on an elliptic curve. □

We call a parallelogram spanned by Ω_A and Ω_B a *period parallelogram* of the elliptic function $f(u)$ (Fig. 11.1).

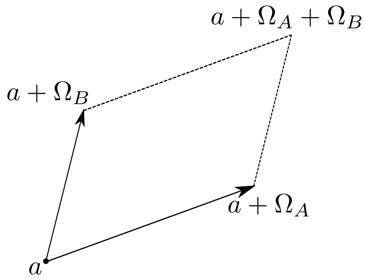

Fig. 11.1 Period parallelogram.

The following four theorems are fundamental in the theory of elliptic functions and sometimes called *Liouville's theorems*.[1]

The first theorem below is the base of Liouville's theory. It is equivalent to Lemma 10.3: any Abelian differential of the first kind is a constant multiple of $\omega_1 = \dfrac{dz}{\sqrt{\varphi(z)}}$.

Theorem 11.9 *If an elliptic function $f(u)$ is entire, which means that it is holomorphic on $\mathbb{C}$ and has no singularity, then it is constant.*

Proof Let $f(u)$ be an elliptic function without a singularity. We denote the period parallelogram with vertices at 0, Ω_A, $\Omega_A + \Omega_B$ and Ω_B ($a = 0$ in Fig. 11.1) by Π_0. (Here we take the closure of the parallelogram as the period parallelogram.)

[1] These theorems appeared in Liouville's lecture notes [Li] (Liouville, Joseph, 1809–1882). As it took more than thirty years for Liouville's lectures to be published, they might have appeared somewhere else before [Li]. Liouville's proofs are completely different from today's proofs, which we show later. It seems that in [Li] there is no explicit statement corresponding to Theorem 11.10 except for the case with two simple poles.

Any complex number $u \in \mathbb{C}$ can be moved into Π_0 by an appropriate element of the period lattice Γ: there exist integers m, $n \in \mathbb{Z}$ such that $u + m\Omega_A + n\Omega_B \in \Pi_0$. Since $f(u)$ is doubly periodic, $f(u) = f(u + m\Omega_A + n\Omega_B)$. Therefore the set of values of f on $\mathbb{C}$ is equal to the set of values of f on Π_0: $f(\mathbb{C}) = f(\Pi_0)$.

On the other hand, the absolute value $|f|$ of f is a continuous function on the complex plane, as a holomorphic function is continuous. Moreover Π_0 is a bounded closed set, i.e., a compact set. Hence a continuous function $|f|$ takes a maximum and a minimum on Π_0, and consequently is bounded on the whole complex plane.

Liouville's theorem in complex analysis says that a bounded entire function is constant.[2] Thus f is a constant function. □

We used the maximum principle in the proof of Lemma 10.3, which is the key in the proof of the above theorem. In this context, too, Lemma 10.3 and Theorem 11.9 are parallel.

The next theorem corresponds to Proposition 10.4.

Theorem 11.10 *Let f be an elliptic function and Π be its period parallelogram. We assume that there are no poles of f on the boundary of Π.*

Then the sum of all residues of f at its poles[3] in Π is zero.

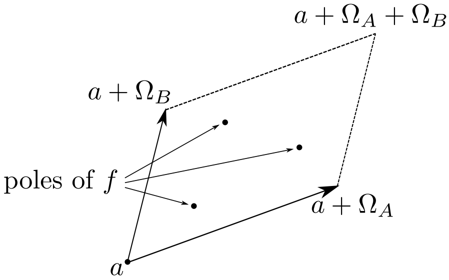

Fig. 11.2 Poles in a period parallelogram.

Proof Let a, $a + \Omega_A$, $a + \Omega_A + \Omega_B$, $a + \Omega_B$ be vertices of Π (Fig. 11.2). By the residue theorem in complex analysis, we have

$$\int_{\partial\Pi} f(u)\,du = 2\pi i(\text{sum of residues in } \Pi).$$

[2] According to §5.63 of [WW] this theorem was proved by Cauchy (*Comptes Rendus*, **19** (1844), pp. 1377–1378) and not by Liouville. Liouville used a different method to prove Theorem 11.9 in [Li].

[3] Since the residue at a point P is defined by an integral $\dfrac{1}{2\pi i}\displaystyle\int_{C_P} f(u)\,du$ (C_P is a small circle around P), it is assigned rather to a one-form $f(u)\,du$ than to a function $f(u)$. However in this chapter we are considering only functions on $\mathbb{C}$, not functions on Riemann surfaces or at infinity and the coordinate is fixed to u. Therefore we can consider the residue to be assigned to a function.

On the other hand, the integral around the boundary of the parallelogram is decomposed as

$$\int_{\partial\Pi} f(u)\,du = \left(\int_a^{a+\Omega_A} + \int_{a+\Omega_A}^{a+\Omega_A+\Omega_B} + \int_{a+\Omega_A+\Omega_B}^{a+\Omega_B} + \int_{a+\Omega_B}^{a}\right) f(u)\,du.$$

By virtue of periodicity integrals on segments $[a+\Omega_A, a+\Omega_A+\Omega_B]$ and $[a+\Omega_A+\Omega_B, a+\Omega_B]$ are reduced to integrals on the opposite sides:

$$\begin{aligned}\int_{a+\Omega_A}^{a+\Omega_A+\Omega_B} f(u)\,du &= \int_a^{a+\Omega_B} f(u+\Omega_A)\,du \\ &= \int_a^{a+\Omega_B} f(u)\,du = -\int_{a+\Omega_B}^{a} f(u)\,du, \\ \int_{a+\Omega_A+\Omega_B}^{a+\Omega_B} f(u)\,du &= \int_{a+\Omega_A}^{a} f(u+\Omega_B)\,du \\ &= \int_{a+\Omega_A}^{a} f(u)\,du = -\int_a^{a+\Omega_A} f(u)\,du.\end{aligned}$$

Summing them up, we obtain

$$2\pi i(\text{sum of residues in }\Pi) = \int_{\partial\Pi} f(u)\,du = 0,$$

which proves the theorem. □

Lemma 10.5, which can be restated as follows, is a direct consequence of this theorem.

Corollary 11.11 *There is no elliptic function with exactly one pole of order one in its period parallelogram.*

Proof If such an elliptic function existed, the sum of its residues in a period parallelogram would be the residue of its unique pole, which cannot be zero (if it were zero, the pole would not be a singularity). This contradicts Theorem 11.10. □

Definition 11.12 The number of poles of an elliptic function $f(u)$ in its period parallelogram is called the *order* of f and denoted by $\operatorname{ord} f$. The number of poles is counted with multiplicity: the pole of order r is counted as 'r poles'. Or, equivalently,

$$\operatorname{ord} f := \sum_{P:\text{ pole in }\Pi} (\text{order of pole } P),$$

where Π is a period parallelogram. Thanks to the periodicity $\operatorname{ord} f$ does not depend on the choice of Π.

Using this terminology, Corollary 11.11 is stated simply as 'there is no elliptic function of order one'.

Exercise 11.13 Let $f(u)$ be an elliptic function of order n and $P(X)$ be a polynomial of degree N. Show that $P(f(u))$ is an elliptic function of order Nn. In particular $P(f(u))$ is not identically 0. (Hint: Poles of $P(f(u))$ are located at the poles of $f(u)$. The order of a pole of $P(f(u))$ can be found by substituting the Laurent expansion of $f(u)$ into $P(X)$.)

Theorem 11.14 *For any $\alpha \in \mathbb{C}$ and a period parallelogram Π the number of roots of $f(u) = \alpha$ in Π is equal to* $\operatorname{ord} f$. *Here the number of roots is counted with multiplicity (double root = two roots, triple root = three roots, and so on) and we assume that $f(u) \neq \alpha$ on the boundary $\partial\Pi$ of Π.*

Proof Since the order of an elliptic function is defined by the number of poles, $f(u) - \alpha$ has the same order as $f(u)$. So, we can assume $\alpha = 0$ without loss of generality by considering $f(u) - \alpha$ instead of $f(u)$. In other words we have only to show that the number of zeros of f in Π is equal to $\operatorname{ord} f$. (This is equivalent to Lemma 10.6.)

According to the argument principle (Theorem A.10)

$$\text{(number of zeros of } f(u) \text{ in } \Pi) - \text{(number of poles of } f(u) \text{ in } \Pi) = \frac{1}{2\pi i} \int_{\partial\Pi} \frac{f'(u)}{f(u)}\, du,$$

where both numbers of zeros and of poles are counted with multiplicities. The right-hand side is equal to the sum of residues of $\dfrac{f'(u)}{f(u)}$ in Π because of the residue theorem. As $\dfrac{f'(u)}{f(u)}$ is also an elliptic function by Lemma 11.8, the sum of its residues in Π is zero according to Theorem 11.10. Therefore the number of zeros of $f(u)$ in Π is equal to the number of poles of $f(u)$ in Π, which is the order of f by definition. □

If we start from the definition of elliptic functions as 'meromorphic functions on an elliptic curve', the proof of the next theorem is rather complicated, because we have to 'add' positions of zeros and poles. Actually we can define 'addition of two points of an elliptic curve', which we shall comment on at the end of the next chapter.

Theorem 11.15 *Let N be the order of an elliptic function f and α be an arbitrary complex number. According to Theorem* 11.14 *there are N roots of the equation $f(u) = \alpha$ in a period parallelogram Π. We denote them by $a_1, \dots, a_N$. The poles of f in Π are denoted by $b_1, \dots, b_N$. We assume that there are neither roots nor poles on the boundary $\partial\Pi$ of Π.*

Then, the difference of the sum of positions of zeros and the sum of positions of poles is a linear combination of Ω_A and Ω_B with integer coefficients:

$$a_1 + \cdots + a_N \equiv b_1 + \cdots + b_N \mod \Gamma.$$

Here $u \equiv v \mod \Gamma$ *means*[4] $u - v \in \Gamma$.

Proof By the same argument as in the proof of Theorem 11.14 we may assume that $\alpha = 0$. (Note that the poles of $f(u)$ and the poles of $f(u) - \alpha$ are located at the same points.)

Applying the generalised argument principle (Theorem A.11) to $g = f$, $D = \Pi$ and $\varphi(u) = u$, we have

$$\frac{1}{2\pi i}\int_{\partial\Pi} u\frac{f'(u)}{f(u)}\,du = \sum_{j=1}^{N} a_j - \sum_{j=1}^{N} b_j. \tag{11.2}$$

We want to show that this quantity belongs to Γ. Let us compute the integral in the left-hand side. (Caution: Although this integral is similar to the one which appeared in the proof of Theorem 11.14, the integrand of (11.2) is *not* an elliptic function. The factor $\dfrac{f'(u)}{f(u)}$ is an elliptic function but u is not.)

To compute the integral, we decompose it to integrals on each side of the parallelogram:

$$\int_{\partial\Pi} u\frac{f'(u)}{f(u)}\,du = \left(\int_a^{a+\Omega_A} + \int_{a+\Omega_A}^{a+\Omega_A+\Omega_B} + \int_{a+\Omega_A+\Omega_B}^{a+\Omega_B} + \int_{a+\Omega_B}^{a}\right) u\frac{f'(u)}{f(u)}\,du. \tag{11.3}$$

Shifting the integration variable by Ω_A, we can rewrite the second term in the right-hand side as

$$\begin{aligned}\int_{a+\Omega_A}^{a+\Omega_A+\Omega_B} u\frac{f'(u)}{f(u)}\,du &= \int_a^{a+\Omega_B} (u+\Omega_A)\frac{f'(u+\Omega_A)}{f(u+\Omega_A)}\,du \\ &= -\int_{a+\Omega_B}^{a} u\frac{f'(u)}{f(u)}\,du - \Omega_A\int_{a+\Omega_B}^{a}\frac{f'(u)}{f(u)}\,du,\end{aligned} \tag{11.4}$$

because $\dfrac{f'(u)}{f(u)}$ has period Ω_A.

The integrand in the last term is a logarithmic derivative $\dfrac{f'(u)}{f(u)} = \dfrac{d}{du}\log f(u)$, which makes it possible to compute the integral explicitly as follows:

$$\begin{aligned}\int_{a+\Omega_B}^{a}\frac{f'(u)}{f(u)}\,du &= \int_{a+\Omega_B}^{a} d\log f(u) \\ &= \log f(a) - \log f(a+\Omega_B).\end{aligned} \tag{11.5}$$

We have already discussed in Section 10.3.1, when we computed the integral of $\omega_f = d(\log f)$ in Section 10.3, p. 182, that we cannot draw the following type of conclusion from (11.5): "As Ω_B is one of the periods of f, $f(a+\Omega_B) = f(a)$. So this

[4] We have been writing this equivalence relation as '$\sim$' up to now, but '$\sim$' might be used with a different meaning. To avoid confusion we use '$\equiv \mod \Gamma$' hereafter.

integral vanishes." Indeed, taking the multi-valuedness of logarithm into account, what we can claim is that there exists an integer n satisfying

$$\log f(a+\Omega_B) - \log f(a) = 2\pi i n.$$

Substituting this into (11.5), we have

$$\int_{a+\Omega_B}^{a} \frac{f'(u)}{f(u)}\,du = -2\pi i n.$$

Hence from (11.4) it follows that

$$\int_{a+\Omega_A}^{a+\Omega_A+\Omega_B} u\frac{f'(u)}{f(u)}\,du = -\int_{a+\Omega_B}^{a} u\frac{f'(u)}{f(u)}\,du + 2\pi i n\Omega_A. \tag{11.6}$$

Similarly, there exists an integer m such that

$$\int_{a+\Omega_A+\Omega_B}^{a+\Omega_B} u\frac{f'(u)}{f(u)}\,du = -\int_{a}^{a+\Omega_A} u\frac{f'(u)}{f(u)}\,du + 2\pi i m\Omega_B. \tag{11.7}$$

We can compute the left-hand side of (11.2) from (11.6), (11.7) and (11.3):

$$\sum_{j=1}^{N} a_j - \sum_{j=1}^{N} b_j = \frac{1}{2\pi i}\int_{\partial\Pi} u\frac{f'(u)}{f(u)}\,du = n\Omega_A + m\Omega_B,$$

which proves the theorem. □

In the next chapter, keeping these properties in mind, we construct Weierstrass's $\wp$-function explicitly. As we mentioned in Remark 11.3, we can prove the equivalence of the two definitions of elliptic functions using some properties of this function.

Chapter 12
The Weierstrass $\wp$-Function

In the previous chapter we defined elliptic functions as meromorphic functions on an elliptic curve = doubly periodic meromorphic functions on $\mathbb{C}$ and studied their properties. In particular, we gave several examples of elliptic functions which are obtained immediately from the definitions. In this chapter we construct one of them, Weierstrass's $\wp$-function,[1] in a more explicit way than the method we used in the previous chapter.

12.1 Construction of the $\wp$-Function

In the previous chapter we have shown that

1. a holomorphic elliptic function on $\mathbb{C}$ is constant;
2. there is no elliptic function of order one (an elliptic function which does not have a singularity except for one simple pole in a period parallelogram).

From these facts it follows that the 'simplest' non-constant elliptic function has either

- one double pole; or
- two simple poles,

in a period parallelogram. The $\wp$-function is the former, while Jacobi's sn function as a complex function, which will be discussed in the last chapter, is the latter.

In the previous chapter we defined $\wp(u)$ as the inverse function of the elliptic integral $u(z) = \displaystyle\int_\infty^z \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}}$. This function has a double pole at the origin $u = 0$ and, therefore, has a double pole at each point of the form $m\Omega_A + n\Omega_B$ $(m, n \in \mathbb{Z})$

[1] Weierstrass (Weierstraß, Karl Theodor Wilhelm, 1815–1897) introduced this function in the fourteenth article (1882) in the collected works [Wi]. His definition is different from the one we have given here.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_13

by periodicity. Here Ω_A and Ω_B are the A- and B-periods of the Abelian differential of the first kind $\dfrac{dz}{\sqrt{4z^3 - g_2 z - g_3}}$ on an elliptic curve.

In this chapter we define $\wp(u)$, using its double periodicity. This means that, forgetting the story "an elliptic curve is defined by compactifying the Riemann surface of $\sqrt{\varphi(z)}$ and ...", we construct an elliptic function with periods Ω_1 and Ω_2, which are arbitrary complex numbers linearly independent over $\mathbb{R}$. As before we denote the period lattice by $\Gamma := \mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2$. Our goal now is

Construct an elliptic function having second order poles only on Γ.

In general an elliptic function $f(u)$ with poles of order n at each point $m_1\Omega_1 + m_2\Omega_2$ of Γ ($m_1, m_2 \in \mathbb{Z}$) has the Laurent expansion,

$$f(u) = \frac{c}{(u - m_1\Omega_1 - m_2\Omega_2)^n} + \cdots$$

at that point. Here $\cdots$ denotes the terms whose degree in $(u - m_1\Omega_1 - m_2\Omega_2)$ is larger than $-n$. By the periodicity $f(u+\Omega_1) = f(u+\Omega_2) = f(u)$ the coefficient c does not depend on m_1 and m_2. Based on this observation, we take

$$f_n(u) := \sum_{m_1, m_2 \in \mathbb{Z}} \frac{1}{(u - m_1\Omega_1 - m_2\Omega_2)^n} = \sum_{\Omega \in \Gamma} \frac{1}{(u - \Omega)^n} \tag{12.1}$$

as a candidate for the simplest elliptic function of order n. We have to check that this series converges and really does define a doubly periodic function.

Theorem 12.1 *Assume $n \geq 3$.*

(1) *The series $f_n(u)$ converges uniformly and absolutely on any compact set K not including Γ: $K \subset \mathbb{C} \setminus \Gamma$.*
(2) *$f_n(u)$ is an elliptic function with poles of order n on Γ.*
(3) *$f_n(u)$ is an even function if n is even and an odd function if n is odd.*

Proof Let $K \subset \mathbb{C} \setminus \Gamma$ be a compact set. Since K is bounded, there is an $R > 0$ such that the closed disk $D_R := \{z \in \mathbb{C} \mid |z| \leq R\}$ completely contains K: $K \subset D_R$ (Fig. 12.1).

We want to prove that the functional series (12.1) converges absolutely and uniformly on K. Since the convergence property is not affected by omitting finitely many terms of the series, we omit terms for Ω such that $|\Omega| < 2R$ and prove convergence of the series

$$f_{n,R}(u) := \sum_{\Omega \in \Gamma,\ |\Omega| \geq 2R} \frac{1}{(u - \Omega)^n}. \tag{12.2}$$

In fact, each term of this series is defined on the whole disk D_R, even when $u \in \Gamma$. So, let us prove uniform and absolute convergence of (12.2) on D_R, from which follows uniform and absolute convergence of the series defined by (12.1) on K.

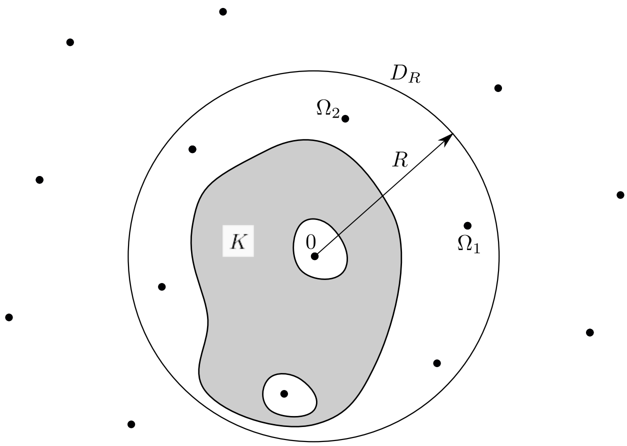

Fig. 12.1 A compact set $K \subset \mathbb{C} \setminus \Gamma$. Dots designate elements of Γ.

Each term of (12.2) is rewritten as

$$\frac{1}{(u-\Omega)^n} = \frac{1}{\Omega^n} \frac{1}{\left(\dfrac{u}{\Omega} - 1\right)^n}.$$

As $|u| \le R$, $|\Omega| \ge 2R$, the absolute value of $\dfrac{u}{\Omega} - 1$ is estimated from below as

$$\left|\frac{u}{\Omega} - 1\right| \ge \left|\frac{|u|}{|\Omega|} - 1\right| \ge \frac{1}{2}.$$

Thus we obtain an estimate,

$$\frac{1}{\left|\left(\dfrac{u}{\Omega} - 1\right)^n\right|} \le \left(\frac{1}{2}\right)^{-n} = 2^n,$$

and each term of the right-hand side of (12.2) does not exceed $\dfrac{2^n}{|\Omega|^n}$:

$$\sum_{\Omega\in\Gamma,\ |\Omega|\ge 2R} \left|\frac{1}{(u-\Omega)^n}\right| \le \sum_{\Omega\in\Gamma,\ |\Omega|\ge 2R} \frac{2^n}{|\Omega|^n}. \tag{12.3}$$

Hence, in order to prove uniform and absolute convergence of (12.2) on D_R, we have only to show convergence of the right-hand side of (12.3), which is bounded by a constant ($= 2^n$) multiple of

$$\sum_{\Omega\in\Gamma,\ \Omega\ne 0} \frac{1}{|\Omega|^n}. \tag{12.4}$$

(This is nothing but Weierstrass's M-test. Notice that the right-hand side of (12.3) does not depend on u.)

We replace the condition '$|\Omega| \geq 2R$' in the sum (12.3) by a weaker condition '$\Omega \neq 0$' in (12.4) to make the proof simpler. This adds finitely many terms to the series (12.3), which does not affect convergence.

Let us decompose $\Omega \setminus \{0\}$ into onion-like layers as shown in Fig. 12.2.

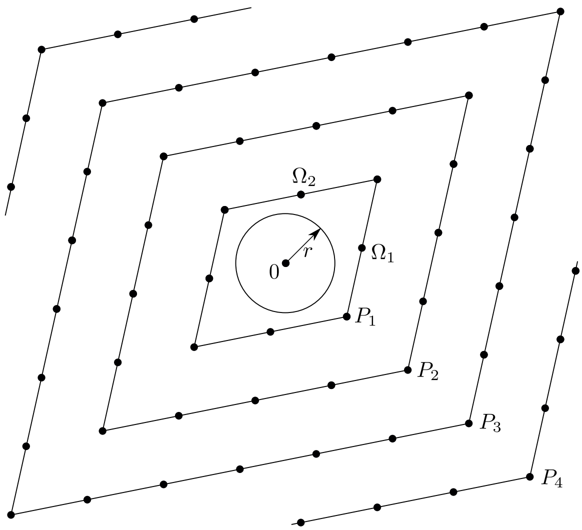

Fig. 12.2 Decomposition of $\Gamma \setminus \{0\}$ into parallelograms.

For a positive integer k let P_k be the set of elements in Γ on one of the layers, the sides of parallelogram with vertices $\pm k\Omega_1 \pm k\Omega_2$. The set $\Gamma \setminus \{0\}$ is decomposed as

$$\Gamma \setminus \{0\} = \bigcup_{k=1}^{\infty} P_k.$$

Let r be the radius of a circle which has its centre at 0 and lies strictly inside the smallest parallelogram, i.e., the parallelogram which contains P_1. (See Fig. 12.2.) The distance between a point of P_1 and the origin is larger than r: if $\Omega \in P_1$, then $|\Omega| > r$. The parallelogram containing P_k is k times larger than P_1. Therefore if $\Omega \in P_k$, then $|\Omega| > kr$.

Since each P_k has $8k$ lattice points, we can estimate (12.4) as

$$\begin{aligned}\sum_{\Omega \in \Gamma,\ \Omega \neq 0} \frac{1}{|\Omega|^n} &= \sum_{k=1}^{\infty} \sum_{\Omega \in P_k} \frac{1}{|\Omega|^n} \\ &< \sum_{k=1}^{\infty} 8k \frac{1}{k^n r^n} = \frac{8}{r^n} \sum_{k=1}^{\infty} \frac{1}{k^{n-1}}.\end{aligned}$$

As we learned in calculus, the series $\sum_{k=1}^{\infty} \frac{1}{k^{n-1}}$ converges if $n \geq 3$. Thus (12.4) converges when $n \geq 3$, and consequently (12.2) converges uniformly and absolutely on D_R.

As we mentioned after (12.2), this shows that the series (12.1) converges uniformly and absolutely on any compact subset of $\mathbb{C} \setminus \Gamma$ as a series of functions of u, which is the first statement of Theorem 12.1.

As each term of (12.1) is a holomorphic function on $\mathbb{C} \setminus \Gamma$, the series gives a holomorphic function on $\mathbb{C} \setminus \Gamma$ by Weierstrass's double series theorem (Theorem A.9).

For any $\Omega_0 \in \Gamma$ by a similar argument the series $\displaystyle\sum_{\Omega \in \Gamma, \Omega \neq \Omega_0} \frac{1}{(u-\Omega)^n}$ is holomorphic even in a neighbourhood of Ω_0. Hence

$$f_n(u) = \frac{1}{(u-\Omega_0)^n} + \sum_{\Omega \in \Gamma, \Omega \neq \Omega_0} \frac{1}{(u-\Omega)^n}$$

has a pole of order n at Ω_0. Thus we have shown that $f_n(u)$ is a meromorphic function on $\mathbb{C}$ with a pole of order n at each $\Omega \in \Gamma$.

Moreover it follows immediately from the definition (12.1) that for any $\Omega_0 \in \Gamma$,

$$f(u+\Omega_0) = \sum_{\Omega \in \Gamma} \frac{1}{(u+\Omega_0-\Omega)^n} = \sum_{\Omega' \in \Gamma} \frac{1}{(u-\Omega')^n} = f(u),$$

where $\Omega' = \Omega - \Omega_0$. This means that $f(u)$ is an elliptic function with periods Ω_1 and Ω_2.

Lastly we show the third statement of Theorem 12.1. This is also a direct consequence of the definition:

$$\begin{aligned} f_n(-u) &= \sum_{\Omega \in \Gamma} \frac{1}{(-u-\Omega)^n} \\ &= \sum_{\Omega' \in \Gamma} \frac{(-1)^n}{(u-\Omega')^n} = (-1)^n f_n(u), \end{aligned}$$

where $\Omega' = -\Omega$. □

Thus we have constructed elliptic functions of order n ($n \geq 3$). "Then why is the case $n = 2$ excluded? Can we obtain what we want just by putting $n = 2$?" Sorry, no. The series $\displaystyle\sum_{\Omega \in \Gamma,\ \Omega \neq 0} \frac{1}{|\Omega|^n}$ in the above proof diverges for $n = 2$ and the claims of the theorem do not hold.

So, let us 'modify' the series (12.1) to construct an elliptic function of order 2.

Theorem 12.2 *The series*

$$\wp(u) := \frac{1}{u^2} + \sum_{\Omega\in\Gamma,\ \Omega\neq 0} \left(\frac{1}{(u-\Omega)^2} - \frac{1}{\Omega^2} \right) \tag{12.5}$$

converges uniformly and absolutely on any compact subset of $\mathbb{C} \setminus \Gamma$ *and gives an even elliptic function of order* 2 *with poles at* Γ.

The elliptic function $\wp(u)$ defined by (12.5) is called the *Weierstrass* $\wp$*-function*. We shall show later that it coincides with the elliptic function introduced in Example 11.6 as the inverse function of the elliptic integral.

Proof The idea is to integrate the elliptic function

$$f_3(u) = \sum_{\Omega\in\Gamma} \frac{1}{(u-\Omega)^3}$$

of order 3 to obtain $\wp$.

To begin with, we integrate the part of the series f_3 without the term $\dfrac{1}{u^3}$ ($\Omega = 0$ in (12.1)) from 0 to u.

$$\begin{aligned} \int_0^u \left(f_3(v) - \frac{1}{v^3} \right) dv &= \int_0^u \left(\sum_{\Omega\in\Gamma\setminus\{0\}} \frac{1}{(v-\Omega)^3} \right) dv \\ &= \sum_{\Omega\in\Gamma\setminus\{0\}} \int_0^u \frac{1}{(v-\Omega)^3}\, dv = -\frac{1}{2} \sum_{\Omega\in\Gamma\setminus\{0\}} \left(\frac{1}{(u-\Omega)^2} - \frac{1}{\Omega^2} \right). \end{aligned}$$

Here we need to be careful at two points. The first one is the change of order of the integral and the summation (from the first line to the second), which is possible because of the uniform convergence of the series on a compact set in $\mathbb{C} \setminus (\Gamma \setminus \{0\})$. It also follows that the resulting series converges uniformly on the same compact set.

The other is that we do not specify the integration contour "from 0 to u". Note that singularities of the integrand $f_3(v) - \dfrac{1}{v^3}$ are poles of order 3 at $\Omega \in \Gamma \setminus \{0\}$. Since the principal parts of those poles are $\dfrac{1}{(v-\Omega)^3}$ and do not have residues, we can deform the integration contour freely by virtue of Cauchy's integral theorem, and thus we do not have to specify the integration contour. (Of course the contour should not pass through the points of Γ.)

The series obtained above is 'almost' the $\wp$-function defined by (12.5). It is meromorphic because of the uniform convergence and its poles are of order 2. Hence,

$$\wp(u) = \frac{1}{u^2} - 2 \int_0^u \left(f_3(v) - \frac{1}{v^3} \right) dv$$

is also a meromorphic function and has poles of order 2 at each point of Γ. Moreover it is easy to see that $\wp'(u) = -2 f_3(u)$ by differentiation.

Before proving that $\wp(u)$ is an elliptic function, let us prove the evenness first. The proof is a direct computation as in the proof of parity in Theorem 12.1.

$$\begin{aligned}
\wp(-u) &= \frac{1}{(-u)^2} + \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{(-u-\Omega)^2} - \frac{1}{\Omega^2} \right) \\
&= \frac{1}{u^2} + \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{(u+\Omega)^2} - \frac{1}{\Omega^2} \right) \\
&\overset{\Omega'=-\Omega}{=} \frac{1}{u^2} + \sum_{\Omega'\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{(u-\Omega')^2} - \frac{1}{\Omega'^2} \right) = \wp(u).
\end{aligned}$$

Now it is time to prove that $\wp(u)$ is an elliptic function with periods Ω_1 and Ω_2. Recall that $\wp'(u) = -2 f_3(u)$ is an elliptic function: $\wp'(u+\Omega_1) = \wp'(u+\Omega_2) = \wp'(u)$. Therefore the derivative of $\wp(u+\Omega_i) - \wp(u)$ $(i = 1,2)$ is 0 and thus $\wp(u+\Omega_i) - \wp(u)$ is constant independent of u. Let us denote these constants by C_i: $\wp(u+\Omega_i) - \wp(u) = C_i$. It remains to show that $C_1 = C_2 = 0$, which proves the periodicity of $\wp(u)$.

Let us set $u = -\dfrac{\Omega_i}{2}$ $(i = 1,2)$. Then we have

$$C_i = \wp\left(\frac{\Omega_i}{2}\right) - \wp\left(-\frac{\Omega_i}{2}\right),$$

which is 0 because of the evenness of $\wp(u)$.

This is the end of the proof of the theorem. □

Exercise 12.3 Show $\wp'\left(\dfrac{\Omega_i}{2}\right) = 0$ $(i = 1,2,3)$, where $\Omega_3 = \Omega_1 + \Omega_2$. (Hint: $\wp'$ is an odd periodic function with periods Ω_i $(i = 1,2,3)$.)

12.2 Properties of $\wp(u)$

Here we study properties of the $\wp$-function introduced in the previous section. By comparing properties we can show that the $\wp$-function constructed above is the same as the $\wp$-function defined as the inverse function of an elliptic integral.

One of the ‘musts’ with a meromorphic function is the Laurent expansion at its poles. The $\wp$-function has a pole of order 2 at each $\Omega \in \Gamma$. The Laurent expansion at Ω can be obtained from the expansion at $u = 0$ by parallel translation thanks to the periodicity. We expand the definition (12.5) at $u = 0$ as

$$\begin{aligned}
\wp(u) &= \frac{1}{u^2} + \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{(u-\Omega)^2} - \frac{1}{\Omega^2} \right) \\
&= \frac{1}{u^2} + c_0 + c_2 u^2 + \cdots + c_{2n} u^{2n} + \cdots .
\end{aligned}$$

Since $\wp(u)$ is an even function, the series contains only even powers of u. The series $c_0 + c_2 u^2 + \cdots$ is the Taylor expansion of the function defined by the sum in the first line. Hence the coefficients can be written explicitly by Taylor's formula:

$$c_{2n} = \frac{1}{(2n)!} \left. \frac{d^{2n}}{du^{2n}} \right|_{u=0} \sum_{\Omega \in \Gamma'} \left(\frac{1}{(u-\Omega)^2} - \frac{1}{\Omega^2} \right)$$
$$= \begin{cases} 0 & (n=0), \\ (2n+1) \displaystyle\sum_{\Omega \in \Gamma'} \frac{1}{\Omega^{2n+2}} & (n \neq 0), \end{cases}$$

where $\Gamma' := \Gamma \smallsetminus \{0\}$. (We can differentiate the series termwise, as the series converges uniformly on any compact set and the terms are holomorphic functions. See Theorem A.9.) For later convenience, we denote constant multiples of the first two non-zero coefficients c_2 and c_4 as follows:

$$\begin{aligned} g_2 &:= 20 c_2 = 60 \sum_{\Omega \in \Gamma'} \frac{1}{\Omega^4}, \\ g_3 &:= 28 c_4 = 140 \sum_{\Omega \in \Gamma'} \frac{1}{\Omega^6}. \end{aligned} \tag{12.6}$$

By this notation the Laurent expansions of $\wp(u)$ and its derivative $\wp'(u)$ up to the terms of order u^4 are

$$\wp(u) = \frac{1}{u^2} + \frac{g_2}{20} u^2 + \frac{g_3}{28} u^4 + O(u^6),$$
$$\wp'(u) = -\frac{2}{u^3} + \frac{g_2}{10} u + \frac{g_3}{7} u^3 + O(u^5).$$

Hence,

$$\begin{aligned} \wp'(u)^2 &= \frac{4}{u^6} - \frac{2g_2}{5} \frac{1}{u^2} - \frac{4g_3}{7} + O(u^2), \\ -4\wp(u)^3 &= -\frac{4}{u^6} - \frac{3g_2}{5} \frac{1}{u^2} - \frac{3g_3}{7} + O(u^2), \\ g_2 \wp(u) &= g_2 \frac{1}{u^2} + O(u^2). \end{aligned}$$

Summing up these three expansions, we have

$$\wp'(u)^2 - 4\wp(u)^3 + g_2 \wp(u) = -g_3 + O(u^2).$$

(The coefficients in (12.6) are chosen so that this equation looks simple.) The left-hand side of this equation is an elliptic function with period lattice Γ and has poles in Γ, as both $\wp(u)$ and $\wp'(u)$ are such elliptic functions.

On the other hand, the right-hand side, $-g_3 + O(u)$, does not have a singularity at the origin. Therefore the expression $\wp'(u)^2 - 4\wp(u)^3 + g_2 \wp(u)$ is an elliptic function without poles, which is nothing but a constant by Theorem 11.9. The value of the constant is determined by setting u to 0. Thus we obtain a differential equation $\wp'(u)^2 - 4\wp(u)^3 + g_2\,\wp(u) = -g_3$, or, equivalently,

$$\wp'(u)^2 = 4\wp(u)^3 - g_2\,\wp(u) - g_3, \tag{12.7}$$

which the $\wp$-function satisfies.

Exercise 12.4 (i) Show that the following relations hold for $e_i := \wp\left(\frac{\Omega_i}{2}\right)$ ($i = 1,2,3$; $\Omega_3 = \Omega_1 + \Omega_2$):

$$e_1 + e_2 + e_3 = 0,\ e_1e_2 + e_2e_3 + e_3e_1 = -\frac{g_2}{4},$$
$$e_1e_2e_3 = \frac{g_3}{4}.$$

(Hint: Use Exercise 12.3 and (12.7).)

(ii) Show that the e_i's are distinct from each other. (Hint: Since the $\wp$-function is of order two, there are two roots of the equation $\wp(u) = e_i$ with multiplicity in a period parallelogram. The multiplicity of $u = \dfrac{\Omega_i}{2}$ is known from Exercise 12.3.)

The differential equation (12.7) makes it possible to show the equivalence of the definitions of $\wp$ as the inverse function of the elliptic integral and as the series (12.5). Let us rewrite (12.7) as follows:

$$\frac{d\wp}{du} = \sqrt{4\wp^3 - g_2\,\wp - g_3},\ \text{i.e.,}\ \ 1 = \frac{1}{\sqrt{4\wp^3 - g_2\,\wp - g_3}}\frac{d\wp}{du}.$$

Integrating this from $u = 0$ ($\leftrightarrow \wp(0) = \infty$) to u ($\leftrightarrow \wp(u)$), we have

$$u = \int_\infty^{\wp(u)} \frac{dz}{\sqrt{4z^3 - g_2\,z - g_3}}, \tag{12.8}$$

by a change of the integration variable $z = \wp(u)$. As the right-hand side is the elliptic integral of the first kind, the above equation shows that $\wp(u)$ is the inverse function of the elliptic integral of the first kind.

In fact, the inverse of the Abel–Jacobi map AJ is given by

$$W : \mathbb{C}/\Gamma \ni u \mapsto (\wp(u), \wp'(u)) \in \bar{\mathcal{R}}. \tag{12.9}$$

Here $\bar{\mathcal{R}}$ is the elliptic curve defined as the compactification of

$$\mathcal{R} = \{(z,w) \mid w^2 = 4z^3 - g_2\,z - g_3\}.$$

Note that the polynomial $4z^3 - g_2\,z - g_3$ has roots e_1, e_2, e_3 because of Exercise 12.3 and (12.7). They are not multiple roots due to Exercise 12.4 (ii). Hence the equation

$w^2 = 4z^3 - g_2 z - g_3$ really defines an elliptic curve. Point $(z,w) = (\wp(u), \wp'(u))$ is on $\bar{\mathcal{R}}$ by the differential equation (12.7). The equation (12.8) means that the composition of W and the elliptic integral of the first kind is an identity map.

We leave the verification of the following two facts to the reader as an exercise.

Exercise 12.5 Show the following.

(i) W is *holomorphic* as a map from W to $\bar{\mathcal{R}}$, i.e., the coordinates of $W(u)$ in $\bar{\mathcal{R}}$ are holomorphic functions of u. (Hint: Do not forget to use local coordinates w and $\eta = wz^{-2}$ at branch points and at infinity, respectively.)

(ii) W is a bijection. (Hint: As the order of $\wp(u)$ is two, $\wp$ takes any value in $\mathbb{P}^1$ twice in a period parallelogram (Theorem 11.14). One has to show that, if the $\wp$-function takes the same value at distinct points u_1 and u_2, ($u_1 \neq u_2$, $\wp(u_1) = \wp(u_2)$), these points are distinguished by the values of $\wp'(u)$. Recall that $\wp(-u) = \wp(u)$, $\wp'(-u) = -\wp'(u)$ and $\wp'\left(\dfrac{\Omega_i}{2}\right) = 0$ (Exercise 12.3).)

Now we can show that 'if two complex numbers Ω_1 and Ω_2 linearly independent over $\mathbb{R}$ are given, there exists an elliptic curve, the A- and B-periods of which are Ω_1 and Ω_2' as we mentioned in Remark 11.3. We have only to define g_2 and g_3 by (12.6), using the given Ω_1 and Ω_2, and define the elliptic curve as above.

An elliptic function is a meromorphic function over an elliptic curve according to the definition at the beginning of Section 11.1, from which it follows that an elliptic function is a meromorphic function of coordinates z or w of $\bar{\mathcal{R}}$. In fact, it is expressed as a *rational function* (a ratio of polynomials) of z *and* w. We can prove this by showing that an elliptic function (a doubly periodic meromorphic function on $\mathbb{C}$) is expressed as a rational expression in $\wp(u)$ and $\wp'(u)$, and then interpreting it in terms of z and w through the map W defined by (12.9).

Exercise 12.6 An elliptic function $f(u)$ with period lattice Γ can be expressed in terms of $\wp(u)$ and $\wp'(u)$ as

$$f(u) = R_1(\wp(u)) + R_2(\wp(u))\, \wp'(u), \tag{12.10}$$

where $R_1(z)$ and $R_2(z)$ are rational functions determined by f. Show this fact as follows:

(i) Let $f(u)$ be an even elliptic function. Show that $f\left(u - \dfrac{\Omega}{2}\right)$ is also an even function for any $\Omega \in \Gamma$. Using this fact, show that if $f\left(\dfrac{\Omega}{2}\right) = 0$, then $\dfrac{\Omega}{2}$ are zeros of even order. Similarly, show that if $\dfrac{\Omega}{2}$ are poles, they are poles of even order.

(ii) Let f be an even elliptic function and $\{a_1, \dots, a_N\}$ be all its distinct zeros in a period parallelogram $\Pi := \{x\Omega_1 + y\Omega_2 \mid 0 \le x < 1,\ 0 \le y < 1\}$, which are not equal to 0 ($a_i \neq 0$). Since f is even, $-a_i$ ($i = 1, \dots, N$) are also zeros of f. Therefore for each i ($i = 1, \dots, N$) there is an i' ($i' = 1, \dots, N$), such that $a_{i'} \equiv -a_i \mod \Gamma$. Points a_i and $a_{i'}$ coincide if and only if $2a_i \in \Gamma$. So, we can renumber the a_i's so that

$$\begin{aligned}
i &= N'+1,\dots,N-N' &&\Rightarrow && 2a_i \in \Gamma, \\
i &= N-N'+1,\dots,N &&\Rightarrow && a_i \equiv -a_{N-i+1} \mod \Gamma.
\end{aligned}$$

In this way we decompose the set $\{a_1,\dots,a_N\}$ into the following two sets:

- N' pairs, $(a_1,a_N),\dots,(a_{N'},a_{N-N'+1})$ satisfying $a_i + a_{N-i+1} \equiv 0 \mod \Gamma$,
- $a_{N'+1},\dots,a_{N-N'}$, satisfying $2a_i \in \Gamma$.

Let $\{b_1,\dots,b_M\}$ be all distinct poles of f in the same parallelogram Π, which are not equal to zero ($b_j \neq 0$). Similarly to $\{a_1,\dots,a_N\}$, we may assume that the b_j's satisfy

$$\begin{aligned}
j &= M'+1,\dots,M-M' &&\Rightarrow && 2b_j \in \Gamma, \\
j &= M-M'+1,\dots,M &&\Rightarrow && b_j \equiv -b_{M-j+1} \mod \Gamma.
\end{aligned}$$

We denote the order of zero a_i by n_i and the order of pole b_j by k_j. Define integers m_i and l_j as follows.

$$m_i := \begin{cases} n_i & (2a_i \notin \Gamma), \\ \dfrac{n_i}{2} & (2a_i \in \Gamma), \end{cases} \qquad l_j := \begin{cases} k_j & (2b_j \notin \Gamma), \\ \dfrac{k_j}{2} & (2b_j \in \Gamma). \end{cases}$$

Show that $f(u)$ is expressed as

$$f(u) = c\, \frac{\displaystyle\prod_{i=1}^{N-N'} (\wp(u) - \wp(a_i))^{m_i}}{\displaystyle\prod_{j=1}^{M-M'} (\wp(u) - \wp(b_j))^{l_j}},$$

with an appropriate complex number c. (Hint: Show that the fraction in the right-hand side has the same zeros and poles as f does, and that the ratio of $f(u)$ and the fraction is constant (Theorem 11.9). Do not forget that $u = 0$ might be a zero or a pole of f.)

(iii) Let $f(u)$ be an odd elliptic function. Show that it is a product of $\wp'(u)$ and a rational function of $\wp(u)$. (Hint: Apply (ii) to the ratio of $f(u)$ and $\wp'(u)$.)

(iv) Using the results of (ii) and (iii), show that any elliptic function is expressed as (12.10). (Hint: $f(u)$ is the sum of an even function $\dfrac{f(u)+f(-u)}{2}$ and an odd function $\dfrac{f(u)-f(-u)}{2}$.)

Remark 12.7 This is an elliptic version of the theorem 'a meromorphic function on the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ is a rational function'.[2] In fact, a similar fact holds for any compact Riemann surface.

12.3 Weierstrass Zeta and Sigma Functions

When we constructed the $\wp$-function in Section 12.1, roughly speaking, we integrated $\wp'(u)$, which has been constructed in Theorem 12.1. "Then why not integrate $\wp(u)$ once more?", you might ask. Well, there is a reason why we cannot expect an elliptic function: if we integrate a pole of $\wp(u)$, which is of order two, we obtain a simple pole and the resulting function would have one simple pole in a period parallelogram. We have shown in Corollary 11.11 that such an elliptic function cannot exist.

However, there is no reason not to try to obtain something new. Let us mimic the argument of construction of $\wp(u)$. Instead of integrating $f_3(u) - \dfrac{1}{u^3} = \displaystyle\sum_{\Omega \in \Gamma \setminus \{0\}} \frac{1}{(v-\Omega)^3}$, we integrate $\wp(u) - \dfrac{1}{u^2} = \displaystyle\sum_{\Omega' \in \Gamma \setminus \{0\}} \left(\frac{1}{(u-\Omega')^2} - \frac{1}{\Omega'^2} \right)$. As we obtained this series by integrating $f_3(u) - \dfrac{1}{u^3}$, which is a series uniformly converging on any compact set in $\mathbb{C} \setminus \Gamma$, the series $\wp(u) - \dfrac{1}{u^2}$ also uniformly converges on any compact set in $\mathbb{C} \setminus \Gamma$. Moreover each term of this series $\dfrac{1}{(u-\Omega)^2} - \dfrac{1}{\Omega^2}$ ($\Omega \in \Gamma$, $\Omega \neq 0$) does not have residue. Hence, we can proceed just as in Section 12.1:

$$
\begin{aligned}
\text{(12.11)} \quad \int_0^u \left(\wp(v) - \frac{1}{v^2} \right) dv &= \int_0^u \sum_{\Omega \in \Gamma \setminus \{0\}} \left(\frac{1}{(v-\Omega)^2} - \frac{1}{\Omega^2} \right) dv \\
&= \sum_{\Omega \in \Gamma \setminus \{0\}} \int_0^u \left(\frac{1}{(v-\Omega)^2} - \frac{1}{\Omega^2} \right) dv = - \sum_{\Omega \in \Gamma \setminus \{0\}} \left(\frac{1}{u-\Omega} + \frac{1}{\Omega} + \frac{u}{\Omega^2} \right).
\end{aligned}
$$

This corresponds to the second part in the definition (12.5) of the $\wp$-function. So, it is reasonable to add a term corresponding to the first term of (12.5) and define

$$
\text{(12.12)} \qquad \zeta(u) := \frac{1}{u} + \sum_{\Omega \in \Gamma \setminus \{0\}} \left(\frac{1}{u-\Omega} + \frac{1}{\Omega} + \frac{u}{\Omega^2} \right),
$$

which is called the *Weierstrass zeta function*.

By the above construction, it is obvious that

[2] This is a consequence of Liouville's theorem 'a bounded entire function on $\mathbb{C}$ is constant'. See, for example, Exercise 4 of §3.2 in Chapter 4, [Ah].

$$\zeta'(u) = -\wp(u). \tag{12.13}$$

Because we integrated an even function $\wp(u) - \dfrac{1}{u^2}$, we should have obtained an odd function. Indeed,

$$\begin{aligned}
\zeta(-u) &= \frac{1}{-u} + \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{-u-\Omega} + \frac{1}{\Omega} + \frac{-u}{\Omega^2} \right) \\
&= -\frac{1}{u} - \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{u+\Omega} + \frac{1}{-\Omega} + \frac{u}{\Omega^2} \right) \\
&\overset{\Omega'=-\Omega}{=} -\frac{1}{u} - \sum_{\Omega'\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{u-\Omega'} + \frac{1}{\Omega'} + \frac{u}{\Omega'^2} \right) = -\zeta(u).
\end{aligned}$$

As was the case with $\wp(u+\Omega_i) - \wp(u)$, the difference $\zeta(u+\Omega_i) - \zeta(u)$ $(i = 1,2)$ is constant, since $\dfrac{d}{du}(\zeta(u+\Omega_i) - \zeta(u)) = -\wp(u+\Omega_i) + \wp(u) = 0$, because of the periodicity of $\wp(u)$.

The analogy with $\wp(u)$ stops here. As we have already mentioned at the beginning of this section, we cannot expect periodicity of $\zeta(u)$. But still, let us try to follow the computation with $\wp(u)$ in Section 12.1. We denote the constant $\zeta(u+\Omega_i) - \zeta(u)$ by η_i:

$$\eta_i := \zeta(u+\Omega_i) - \zeta(u), \qquad i = 1,2. \tag{12.14}$$

Since $\zeta(u)$ satisfies

$$\zeta(u+\Omega_i) = \zeta(u) + \eta_i, \qquad i = 1,2, \tag{12.15}$$

we say that $\zeta(u)$ is *additively quasi-periodic* with periods Ω_1 and Ω_2.

Hereafter we assume that $\operatorname{Im} \dfrac{\Omega_2}{\Omega_1} > 0$, so that Ω_1 and Ω_2 span a parallelogram with the orientation as in Fig. 12.3.

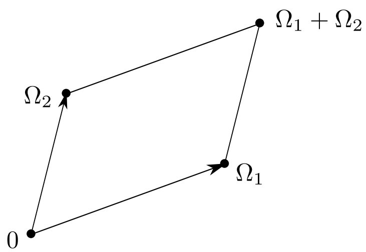

Fig. 12.3 Parallelogram spanned by Ω_1 and Ω_2.

The constants η_1 and η_2 satisfy the following important relation.

Proposition 12.8 (Legendre's relation)

$$\eta_1\Omega_2 - \eta_2\Omega_1 = 2\pi i. \tag{12.16}$$

This is a counterpart of Theorem 11.10, as the following proof shows.

Proof Let Π be a parallelogram with vertices a, $a+\Omega_1$, $a+\Omega_1+\Omega_2$ and $a+\Omega_2$, where $a \notin \Gamma$. The meromorphic function $\zeta(u)$ has one pole inside Π, which is simple and has residue 1. Hence by the residue theorem, we have

$$\int_{\partial\Pi} \zeta(u)\, du = 2\pi i.$$

On the other hand, the integral around the boundary of the parallelogram Π is decomposed as

$$\int_{\partial\Pi} \zeta(u)\, du = \left(\int_a^{a+\Omega_1} + \int_{a+\Omega_1}^{a+\Omega_1+\Omega_2} + \int_{a+\Omega_1+\Omega_2}^{a+\Omega_2} + \int_{a+\Omega_2}^{a}\right) \zeta(u)\, du.$$

By virtue of quasi-periodicity of $\zeta(u)$ integrals on segments $[a+\Omega_1, a+\Omega_1+\Omega_2]$ and $[a+\Omega_1+\Omega_2, a+\Omega_2]$ are reduced to integrals on the opposite sides:

$$\begin{aligned}
\int_{a+\Omega_1}^{a+\Omega_1+\Omega_2} \zeta(u)\, du &= \int_a^{a+\Omega_2} \zeta(u+\Omega_1)\, du \\
&= \int_a^{a+\Omega_2} (\zeta(u)+\eta_1)\, du = -\int_{a+\Omega_2}^{a} \zeta(u)\, du + \eta_1\Omega_2, \\
\int_{a+\Omega_1+\Omega_2}^{a+\Omega_2} \zeta(u)\, du &= \int_{a+\Omega_1}^{a} \zeta(u+\Omega_2)\, du \\
&= \int_{a+\Omega_1}^{a} (\zeta(u)+\eta_2)\, du = -\int_a^{a+\Omega_A} \zeta(u)\, du - \eta_2\Omega_1.
\end{aligned}$$

Summing them up, we obtain

$$2\pi i = \int_{\partial\Pi} \zeta(u)\, du = \eta_1\Omega_2 - \eta_2\Omega_1,$$

which proves the proposition. □

"We integrated $\wp(u)$ to obtain $\zeta(u)$. Then, why not once more?" OK, if you want. But, in this case we encounter another problem besides non-periodicity: multi-valuedness.

When we integrated $\dfrac{1}{(u-\omega)^n}$ with $n \geq 2$ in Section 12.1 and in (12.11), the results were nothing more than a pole $-\dfrac{1}{n-1}\dfrac{1}{(u-\Omega)^{n-1}}$. However, when we integrate

$\dfrac{1}{u-\Omega}$, it becomes a logarithmic function $\log(u-\Omega)$. Multi-valuedness comes from the ambiguity of the integration contour. If the contour goes around $\Omega \in \Gamma$, the integral picks up the residue of $\dfrac{1}{u-\Omega}$ and a multiple of $2\pi i$ is added to the integral.

Putting this problem aside, formally we can proceed as before:

$$
\begin{aligned}
\int_0^u \left(\zeta(v) - \frac{1}{v}\right) dv &= \int_0^u \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\frac{1}{v-\Omega} + \frac{1}{\Omega} + \frac{v}{\Omega^2}\right) dv \\
&= \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \int_0^u \left(\frac{1}{v-\Omega} + \frac{1}{\Omega} + \frac{v}{\Omega^2}\right) dv \\
&= \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\log(u-\Omega) - \log(-\Omega) + \frac{u}{\Omega} + \frac{u^2}{2\Omega^2}\right) \\
&= \sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\log\left(1-\frac{u}{\Omega}\right) + \frac{u}{\Omega} + \frac{u^2}{2\Omega^2}\right).
\end{aligned}
\tag{12.17}
$$

Recall that the ambiguity of the logarithm is an integer multiple of $2\pi i$. Therefore, we obtain a single-valued function by exponentiating it. This technique ('exponentiating a logarithmic multi-valued function to obtain a single-valued function') has already been used in the proof of the injectivity of the Abel–Jacobi map, Section 10.4, to construct a (nonexistent) single-valued meromorphic function f on an elliptic curve.

Thus it is reasonable to introduce the following function, which is called the *Weierstrass sigma function*:

$$
\begin{aligned}
\sigma(u) &:= u \exp\left(\sum_{\Omega\in\Gamma\smallsetminus\{0\}} \left(\log\left(1-\frac{u}{\Omega}\right) + \frac{u}{\Omega} + \frac{u^2}{2\Omega^2}\right)\right) \\
&= u \exp\left(\int_0^u \left(\zeta(v) - \frac{1}{v}\right) dv\right).
\end{aligned}
\tag{12.18}
$$

The first factor u corresponds to the exponential of the integral of $\dfrac{1}{v}$, which was subtracted from $\zeta(v)$ in the integrand of (12.17).

Since the infinite sum in the large parentheses converges absolutely and uniformly on any compact set in $\mathbb{C}\smallsetminus\Gamma$, as was the case with $\zeta(u)$, we can interchange exp and $\sum$ and rewrite the definition in the form of an infinite product:[3]

$$
\sigma(u) = u \prod_{\Omega\in\Gamma\smallsetminus\{0\}} \left(1-\frac{u}{\Omega}\right) \exp\left(\frac{u}{\Omega} + \frac{u^2}{2\Omega^2}\right).
\tag{12.19}
$$

[3] For convergence of infinite products, see Section 16.1.

It is obvious from this expression that $\sigma(u)$ is an everywhere holomorphic function[4] with simple zeros at each point of Γ. Therefore it cannot be an elliptic function, but it inherits good properties from $\zeta(u)$.

For example, replacing u in (12.18) by $-u$, we have

$$\begin{aligned}
(12.20)\quad \sigma(-u) &:= (-u)\exp\left(\sum_{\Omega\in\Gamma\setminus\{0\}}\left(\log\left(1-\frac{-u}{\Omega}\right)+\frac{-u}{\Omega}+\frac{(-u)^2}{2\Omega^2}\right)\right)\\
&\overset{\Omega'=-\Omega}{=} -u\exp\left(\sum_{\Omega'\in\Gamma\setminus\{0\}}\left(\log\left(1-\frac{u}{\Omega'}\right)+\frac{u}{\Omega'}+\frac{u^2}{2\Omega'^2}\right)\right)=-\sigma(u).
\end{aligned}$$

Hence $\sigma(u)$ is an odd function.

Let us rewrite the definition (12.18) of $\sigma(u)$ as follows.

$$(12.21)\qquad \log\sigma(u)=\log u+\int_0^u\left(\zeta(v)-\frac{1}{v}\right)dv,$$

where branches of logarithms should be appropriately taken. Differentiation of this equation gives

$$(12.22)\qquad \frac{d}{du}\log\sigma(u)=\frac{\sigma'(u)}{\sigma(u)}=\zeta(u).$$

The quasi-periodicity (12.15) of $\zeta(u)$ means

$$\frac{d}{du}\big(\log\sigma(u+\Omega_i)-\log\sigma(u)\big)=\eta_i,$$

because of (12.22). Integrating this, we obtain

$$\log\sigma(u+\Omega_i)-\log\sigma(u)=\eta_i u+\text{const.},$$

or, equivalently,

$$\sigma(u+\Omega_i)=C_i e^{\eta_i u}\sigma(u),$$

for some constant C_i. In order to determine the constant C_i it is enough to set $u=-\dfrac{\Omega_i}{2}$:

$$\sigma\left(\frac{\Omega_i}{2}\right)=C_i e^{-\eta_i\Omega_i/2}\sigma\left(-\frac{\Omega_i}{2}\right).$$

Since $\sigma(u)$ is an odd function, $\sigma\left(-\dfrac{\Omega_i}{2}\right)=-\sigma\left(\dfrac{\Omega_i}{2}\right)$. Hence $C_i=-e^{\eta_i\Omega_i/2}$. Thus we have obtained *multiplicative quasi-periodicity* of $\sigma(u)$,

$$(12.23)\qquad \sigma(u+\Omega_i)=-e^{\eta_i(u+\Omega_i/2)}\sigma(u).$$

---

[4] Exactly speaking, this is a result of uniform convergence of the infinite product. See Lemma 16.2.

We shall consider this kind of quasi-periodic function systematically in Chapter 15.

12.4 Addition Theorems for the $\wp$-Function.

We proved the addition theorems for Jacobi's elliptic functions in Section 4.2.3. We proved them only for real-valued functions on the real line there, but we shall show the complex version later in Section 17.3.1. Actually, as we shall prove in Chapter 13, a general elliptic function has an addition formula, which reflects the fact that $\mathbb{C}/\Gamma$ has a natural addition

$$[u_1 \bmod \Gamma] + [u_2 \bmod \Gamma] = [u_1 + u_2 \bmod \Gamma].$$

In this section we prove the addition theorem for the $\wp$-function. This will form the basis for the addition theorems of other elliptic functions in Chapter 13.

Theorem 12.9 (Addition theorem for the $\wp$-function)
If $u_1 + u_2 + u_3 = 0$ *(or,* $u_1 + u_2 + u_3 \equiv 0 \mod \Gamma$*),*

$$\begin{vmatrix} \wp'(u_1) & \wp(u_1) & 1 \\ \wp'(u_2) & \wp(u_2) & 1 \\ \wp'(u_3) & \wp(u_3) & 1 \end{vmatrix} = 0. \tag{12.24}$$

Since $\wp'(u)$ is an odd function, $\wp'(u_3) = -\wp'(u_1 + u_2)$ and, since $\wp(u)$ is an even function, $\wp(u_3) = \wp(u_1 + u_2)$. Therefore the relation (12.24) algebraically connects the values of the $\wp$-function and its derivative at $u_1 + u_2$ to those at u_1, u_2. In this sense the formula (12.24) is an 'addition theorem'.

Proof We assume $\wp(u_1) \neq \wp(u_2)$. By virtue of the identity theorem (Theorem A.7), if (12.24) holds in this case, it also holds for any u_1 and u_2.

Let (a, b) be a solution of the system of linear equations

$$a\wp(u_1) + b = \wp'(u_1), \quad a\wp(u_2) + b = \wp'(u_2). \tag{12.25}$$

In fact, this system is readily solved as follows, though we do not use these explicit expressions in this proof:

$$a = \frac{\wp'(u_1) - \wp'(u_2)}{\wp(u_1) - \wp(u_2)}, \qquad b = \frac{\wp(u_1)\wp'(u_2) - \wp'(u_1)\wp(u_2)}{\wp(u_1) - \wp(u_2)}. \tag{12.26}$$

These formulae are necessary for Exercise 12.11.

The function $f(u) := \wp'(u) - a\wp(u) - b$ has the following properties:

- f is an elliptic function, as it is a linear combination of elliptic functions.
- f has poles of order three at each point $u \in \Gamma$, since $\wp'(u)$ has poles of order three and $\wp(u)$ has poles of order two. There are no other poles of f.

This means that $f(u)$ is an elliptic function of order three, from which it follows that an equation $f(u) = 0$ has three roots in a period parallelogram by Theorem 11.14. Two of them are u_1 and u_2 because of (12.25).[5] Let us denote the third one by u_0.

By Theorem 11.15 we have

$$u_0 + u_1 + u_2 \equiv (\text{sum of positions of poles of } f) = 0 \mod \Gamma.$$

Hence $u_0 \equiv u_3 \mod \Gamma$, which means $f(u_3) = 0$. Thus we have found three (i.e., all modulo Γ) solutions of $f(u) = 0$: $f(u_1) = f(u_2) = f(u_3) = 0$.

By the definition $f(u) = \wp'(u) - a\wp(u) - b$ we can rewrite the result in a matrix form:

$$\begin{pmatrix} \wp'(u_1) & \wp(u_1) & 1 \\ \wp'(u_2) & \wp(u_2) & 1 \\ \wp'(u_3) & \wp(u_3) & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -a \\ -b \end{pmatrix} = 0. \tag{12.27}$$

The components of the left-hand side are $f(u_1)$, $f(u_2)$ and $f(u_3)$.

The vector equation (12.27) says that a matrix in the left-hand side annihilates the vector $\begin{pmatrix} 1 \\ -a \\ -b \end{pmatrix}$, which cannot be a zero vector, because the first component is 1. This means that the matrix in the left-hand side is degenerate and its determinant vanishes, which means (12.24). □

Corollary 12.10

$$\wp(u_1 + u_2) = -\wp(u_1) - \wp(u_2) + \frac{1}{4}\left(\frac{\wp'(u_1) - \wp'(u_2)}{\wp(u_1) - \wp(u_2)}\right)^2.$$

Perhaps this formula is more deserving of the name 'the addition formula' than (12.24), as it expresses $\wp(u_1 + u_2)$ without $\wp'(u_1 + u_2)$.

Exercise 12.11 Prove Corollary 12.10.

(Hint: Each of u_1, u_2, u_3 satisfies $\wp'(u) = a\wp(u) + b$ and (12.7). Therefore $\wp(u_1)$, $\wp(u_2)$ and $\wp(u_3) = \wp(u_1 + u_2)$ satisfy a cubic equation. Apply Vieta's formula to that equation.)

We can also describe the addition theorem geometrically. Essentially this is an interpretation of the additive group structure of $\mathbb{C}/\Gamma$ from the viewpoint of the elliptic curve $\bar{\mathcal{R}}$. This also explains the meaning of the linear equation (12.27).

Exercise 12.12 Confirm that the proof of Theorem 12.9 gives the following geometric addition of points of the elliptic curve

$$\bar{\mathcal{R}} := \overline{\{(z, w) \mid w^2 = 4z^3 - g_2 z - g_3\}} \subset \mathbb{P}^2.$$

[5] We may assume that u_1 and u_2 are in one period parallelogram by the periodicity of $f(u)$.

(Recall that the coordinates of $\bar{\mathcal{R}}$ are $(z, w) = (\wp(u), \wp'(u))$. (See Exercise 12.5.))

(i) The unit element $\mathbf{O}$ is the point at infinity: $\infty = [0 : 0 : 1] \in \mathbb{P}^2$.

(ii) Three points P_1, P_2 and P_3 on $\bar{\mathcal{R}}$ satisfy $P_1 + P_2 + P_3 = \mathbf{O}$ if and only if there exists a line in $\mathbb{C}^2$ (or in $\mathbb{P}^2$)[6] passing through P_1, P_2 and P_3. (When, for example, $P_1 = P_2$, the line should be considered as a tangent line at $P_1 = P_2$.)

Let us finish the story on one of the 'simplest elliptic functions', $\wp(u)$, here. To construct another class of 'simplest elliptic functions', Jacobi's complex elliptic functions, it is convenient to use theta functions, which are, so to speak, 'quasi-elliptic functions'. We shall introduce theta functions in Chapter 15 and redefine Jacobi's elliptic functions as complex functions in Chapter 17.

Before going there, in the next chapter we prove addition theorems for arbitrary elliptic functions, making use of properties of Weierstrass's $\wp$-function studied in this chapter (in particular, the expression of arbitrary elliptic functions in terms of the $\wp$-function (Exercise 12.6) and the addition theorem for the $\wp$-function (Theorem 12.9 and Corollary 12.10)).

[6] A line in $\mathbb{C}^2$ is a set of points (z, w) satisfying an equation of the form $az + bw + c = 0$. A line in $\mathbb{P}^2$ is a set of points $[x_1 : x_2 : x_3]$ satisfying an equation of the form $ax_1 + bx_2 + cx_3 = 0$.

Chapter 13
Addition Theorems

In the previous chapter we constructed Weierstrass's $\wp$-function, studied its properties and, in particular, derived its addition theorem. Moreover we proved (in an exercise) that the $\wp$-function is not a mere 'example of an elliptic function' but expresses all elliptic functions as its rational functions together with its derivative. In this chapter we derive addition formulae for arbitrary elliptic functions, using these facts.

As in the previous chapter we take two complex numbers Ω_1, Ω_2 linearly independent over $\mathbb{R}$ and denote the period lattice spanned by them by $\Gamma := \mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2$. All elliptic functions in this chapter are doubly periodic with respect to this period lattice.

13.1 Addition Theorems of the $\wp$-Function Revisited

To begin with, let us recall the addition theorem of Weierstrass's $\wp$-function.

One expression is

$$\begin{vmatrix} \wp'(u) & \wp(u) & 1 \\ \wp'(v) & \wp(v) & 1 \\ -\wp'(u+v) & \wp(u+v) & 1 \end{vmatrix} = 0, \tag{13.1}$$

which is Theorem 12.9 with variables u_1 and u_2 replaced by u and v. Since $u_3 \equiv -(u+v) \bmod \Gamma$, $\wp$ is even and $\wp'$ is odd, we have the above expression.

Another expression is Corollary 12.10,

$$\wp(u+v) = -\wp(u) - \wp(v) + \frac{1}{4}\left(\frac{\wp'(u) - \wp'(v)}{\wp(u) - \wp(v)}\right)^2. \tag{13.2}$$

Compared to the former (13.1), the latter (13.2) has a slightly 'better' form, as we mentioned after Corollary 12.10. Ideally, in an addition formula we want to express $\wp(u+v)$ in terms of $\wp(u)$ and $\wp(v)$ only. In (12.24) we use $\wp'(u)$, $\wp'(v)$ and $\wp'(u+v)$

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_14

besides them. The formula (13.2) looks more like an 'addition formula' as $\wp'(u+v)$ is absent.

But still $\wp'(u)$ and $\wp'(v)$ remain. This situation is similar to the addition theorem for the sine function. Indeed, in the addition formula for sine we usually use not only sine but also cosine as well:

$$\sin(u+v) = \sin u \cos v + \sin v \cos u. \tag{13.3}$$

But it *is* possible to write this formula using sines exclusively. In fact, using $\cos v = \sqrt{1-\sin^2 v}$, we have

$$\sin(u+v) = \sin u \sqrt{1-\sin^2 v} + \sin v \sqrt{1-\sin^2 u}.$$

"Well, it is true that only sines appear in this expression, but the square root always causes the problem of signs..." OK, then let us rewrite it in a 'square-root free' form. The square of the above expression is

$$\begin{aligned}\sin^2(u+v) = \sin^2 u(1-\sin^2 v) + \sin^2 v(1-\sin^2 u)& \\ + 2\sin u \sin v \sqrt{1-\sin^2 v}\sqrt{1-\sin^2 u}&.\end{aligned}$$

There still remain square roots. Moving the other terms to the left-hand side and squaring, we have

$$\begin{aligned}\left(\sin^2(u+v) - \sin^2 u - \sin^2 v + 2\sin^2 u \sin^2 v\right)^2& \\ = 4\sin^2 u \sin^2 v(1-\sin^2 u)(1-\sin^2 v)&.\end{aligned}$$

The expression is terrible, but at least we have 'a polynomial relation' among $\sin(u+v)$, $\sin u$ and $\sin v$.

In this way, if a function $f(u)$ satisfies a relation

$$P(f(u), f(v), f(u+v)) = 0 \tag{13.4}$$

for a non-zero three variable *polynomial* $P(X,Y,Z)$ for any u and v, we say that $f(u)$ has an *algebraic addition formula*. The above calculation shows that the sine function has an algebraic addition formula with

$$\begin{aligned}&P_{\sin}(X,Y,Z) \\ &= (Z^2 - X^2 - Y^2 + 2X^2Y^2)^2 - 4X^2Y^2(1-X^2)(1-Y^2) \\ &= X^4 + Y^4 + Z^4 - 2(X^2Y^2 + Y^2Z^2 + Z^2X^2) + 4X^2Y^2Z^2.\end{aligned}$$

"Why do we have to stick to polynomials?" If we allow *any* relation among $f(u)$, $f(v)$ and $f(u+v)$, we can take, for example, the trivial relation

$$f(u+v) = f\left(f^{-1}(f(u)) + f^{-1}(f(v))\right)$$

with the inverse function f^{-1} and insist that it is "an addition formula". The addition theorems of trigonometric functions are valuable because $f(u+v)$ is connected with $f(u)$ and $f(v)$ by 'simple' relations. If $\sin^{-1}$ appears in such a formula, it is worthless. That is why we consider *algebraic* addition formulae.

Of course, we can loosen the condition of polynomiality and take 'a rational function $R(X,Y,Z)$' instead of 'a polynomial $P(X,Y,Z)$', because we can reduce it to a polynomial relation by multiplying by the denominator of the rational function.

Now we return to elliptic functions. Does the $\wp$-function have an algebraic addition formula? If we can eliminate $\wp'(u)$ and $\wp'(v)$ from (13.2), it has. In the case of the addition theorem (13.3) of the sine function, we eliminated the cosine, using the relation $\cos^2 u = 1 - \sin^2 u$. In the case of the $\wp$-function, what corresponds to this relation? It should be a relation between $\wp'$ and $\wp$. Yes, it is the differential equation satisfied by the $\wp$-function (12.7),

$$\wp'(u)^2 = 4\wp(u)^3 - g_2\wp(u) - g_3. \tag{13.5}$$

Let us eliminate $\wp'(u)$ and $\wp'(v)$ from (13.2), using (13.5).[1] First we move $\wp(u)$ and $\wp(v)$ in (13.2) to the left-hand side and multiply by the denominator, $4(\wp(u) - \wp(v))^2$, to obtain

$$4(\wp(u+v) + \wp(u) + \wp(v))(\wp(u) - \wp(v))^2 = (\wp'(u) - \wp'(v))^2. \tag{13.6}$$

Subtracting this expression from $2\wp'(u)^2 + 2\wp'(v)^2$, we have

$$\begin{aligned} 2\wp'(u)^2 + 2\wp'(v)^2 - 4(\wp(u+v) + \wp(u) + \wp(v))(\wp(u) - \wp(v))^2 \\ = (\wp'(u) + \wp'(v))^2. \end{aligned} \tag{13.7}$$

On the other hand, subtracting the differential equation (13.5) with variable v instead of u from (13.5) itself, we obtain

$$\wp'(u)^2 - \wp'(v)^2 = (\wp(u) - \wp(v))(4\wp(u)^2 + 4\wp(u)\,\wp(v) + 4\wp(v)^2 - g_2). \tag{13.8}$$

The right-hand side of the product of (13.6) and (13.7) is equal to the square of the left-hand side of (13.8). Hence,

$$\begin{aligned} &4(\wp(u+v) + \wp(u) + \wp(v))(\wp(u) - \wp(v))^2 \\ &\quad \times (2\wp'(u)^2 + 2\wp'(v)^2 - 4(\wp(u+v) + \wp(u) + \wp(v))(\wp(u) - \wp(v))^2) \\ &\qquad = (\wp(u) - \wp(v))^2(4\wp(u)^2 + 4\wp(u)\,\wp(v) + 4\wp(v)^2 - g_2)^2. \end{aligned}$$

Cancelling the common factor $(\wp(u) - \wp(v))^2$, we have

[1] The following ingenious technique is due to Hiroyuki Ochiai.

$$\begin{aligned}4(\wp(u+v)+\wp(u)+\wp(v))&\\ \times(2\wp'(u)^2+2\wp'(v)^2-4(\wp(u+v)+\wp(u)+\wp(v))(\wp(u)-\wp(v))^2)&\\ =(4\wp(u)^2+4\wp(u)\,\wp(v)+4\wp(v)^2-g_2)^2.&\end{aligned}$$

Rewriting $\wp'(u)^2$ and $\wp'(v)^2$ in the left-hand side by the differential equation (13.5), we should obtain a relation of the form

$$P_0\big(\wp(u),\wp(v),\wp(u+v)\big)=0. \tag{13.9}$$

"Done! The $\wp$-function has an algebraic addition formula." Wait a moment. If the above P_0 is identically equal to 0, the equation (13.9) is nonsense. We should verify by explicit computation that the polynomial $P_0(a,b,c)$ does not vanish. In fact, the explicit form of $P_0(a,b,c)$ is

$$P_0(a,b,c)=(g_2+4(ab+ac+bc))^2+16(a+b+c)(g_3-4abc), \tag{13.10}$$

which is a non-zero polynomial. Thus we have finally obtained an algebraic addition formula (13.9) for the $\wp$-function.

13.2 Addition Formulae of General Elliptic Functions

We now know that Weierstrass's $\wp$-function has an algebraic addition formula. Then how about other elliptic functions? Recall that we proved addition theorems for Jacobi's elliptic functions sn, cn and dn in Section 4.2.3[2]. These formulae are mixtures of sn, cn and dn, but we can cook up an algebraic addition formula for each of sn, cn and dn, as was the case with sine.

The proofs of the addition theorems for Weierstrass's $\wp$-function and Jacobi's elliptic functions are completely different and make the most of their characteristic properties, which makes it difficult to generalise those methods to other elliptic functions. However, fortunately, we know that any elliptic function is expressed in terms of the $\wp$-function and its derivative (Exercise 12.6 in the previous chapter). Combining this fact with the algebraic addition formula of the $\wp$-function, we prove the following theorem.

Theorem 13.1 *For an elliptic function $f(u)$ there exists a three-variable polynomial $P(X,Y,Z)$ containing all variables (in particular, $\neq 0$) such that an algebraic addition formula*

$$P\big(f(u),f(v),f(u+v)\big)=0 \tag{13.11}$$

holds.

[2] That was a real-function version, but we shall prove the complex-function version in the last chapter (Chapter 17) of this book.

Proof According to Exercise 12.6 there are rational functions $R_1(a)$, $R_2(a)$ such that

$$f(u) = R_1(\wp(u)) + R_2(\wp(u))\,\wp'(u). \tag{13.12}$$

If $R_2 = 0$, we obtain a relation of the form (polynomial of $\wp(u)$) $\times f(u) =$ (polynomial of $\wp(u)$), multiplying (13.12) by the denominator of R_1. If $R_2 \neq 0$, we rewrite (13.12) in the form

$$\frac{f(u) - R_1(\wp(u))}{R_2(\wp(u))} = \wp'(u).$$

Then squaring it and rewriting $\wp'(u)^2$ in the right-hand side by (13.5), we obtain an expression containing $f(u)$ and $\wp(u)$ but not $\wp'(u)$. Since R_1 and R_2 are rational functions, we obtain a relation of the form

$$\bigl((\text{polynomial of } \wp(u)) \times f(u) - (\text{polynomial of } \wp(u))\bigr)^2 \\ = (\text{polynomial of } \wp(u)),$$

by cancelling denominators.

In any case we have found a polynomial of variables X and a

$$\tilde{Q}(X;a) = \tilde{\alpha}_0(X)a^N + \tilde{\alpha}_1(X)a^{N-1} + \cdots + \tilde{\alpha}_N(X),$$

($\tilde{\alpha}_i(X)$ are polynomials of X with complex coefficients), such that

$$\tilde{Q}(f(u), \wp(u)) = 0.$$

Dividing $\tilde{Q}(X;a)$ by the top coefficient $\tilde{\alpha}_0(X)$, we make it a monic polynomial of one variable a,

$$Q(X;a) = a^N + \alpha_1(X)a^{N-1} + \cdots + \alpha_N(X), \tag{13.13}$$

with coefficients $\alpha_i(X) \in \mathbb{C}(X) =$ (the set of rational functions of X). Thus we have a relation

$$Q(X;a) = 0 \tag{13.14}$$

for elliptic functions $X = f(u)$ and $a = \wp(u)$.

We have shown that there is at least one relation between $X = f(u)$ and $a = \wp(u)$. In fact, there are infinitely many such relations, because we may multiply $Q(x;a)$ by an arbitrary polynomial of a with coefficients in $\mathbb{C}(X)$. The resulting polynomial $\tilde{\tilde{Q}}(X;a)$ of a also satisfies $\tilde{\tilde{Q}}(f(u);\wp(u)) = 0$.

We choose[3] a monic (hence non-zero) polynomial *with the least degree in* a as $Q(X;a)$ from the set of polynomials of a with coefficients in $\mathbb{C}(X)$ which vanish by substitution $(X,a) = (f(u), \wp(u))$.

We can claim the following two facts on $Q(X;a)$.

Lemma 13.2 *(i)* $Q(X;a)$ *always contains* X*, i.e., there is a non-constant function of* X *among the coefficients* $\alpha_i(X)$ $(i = 1,\dots,N)$ *in the expression* (13.13).

(ii) $Q(X;a)$ *is* irreducible *as a polynomial in* a*. This means that* $Q(X;a)$ *cannot be factorised as* $Q(X;a) = Q_1(X;a)\,Q_2(X;a)$*, where* $Q_1(X;a)$ *and* $Q_2(X;a)$ *are polynomials in* a *of degree not less than* 1. □

Proof (i) If $Q(a) := Q(X;a)$ does not contain X, the condition $Q(\wp(u)) = 0$ means that a polynomial of $\wp(u)$ with complex coefficients is identically 0 as a function of u. This is not possible, as shown in Exercise 11.13. (In the present case we have only to check the order of the pole of $Q(\wp(u))$ at $u = 0$.)

(ii) Suppose that there is a non-trivial factorisation

$$Q(X;a) = Q_1(X;a)\,Q_2(X;a).$$

'Non-trivial' means that neither Q_1 nor Q_2 is constant: $0 < \deg_a Q_1 < \deg_a Q$, $0 < \deg_a Q_2 < \deg_a Q$. Dividing Q_1 and Q_2 by their coefficients of the top degree, if necessary, we may assume that they are monic polynomials of a.

Since $Q_1(f(u);\wp(u))$ and $Q_2(f(u);\wp(u))$ are both rational expressions of elliptic functions $f(u)$ and $\wp(u)$, they are elliptic functions in u by Lemma 11.8 and, therefore, holomorphic except at poles.

The equation $Q(f(u);\wp(u)) = 0$, (13.14), means that on the complement of the set of poles the product of two holomorphic functions $Q_1(f(u);\wp(u))$ and $Q_2(f(u);\wp(u))$ is identically equal to 0. Hence, by virtue of Theorem A.8, one of the factors must be identically 0. Let us assume $Q_1(f(u);\wp(u)) = 0$.

Recall that we have chosen $Q(X;a)$ so that it has the least degree in a among the monic polynomials satisfying $Q(f(u);\wp(u)) = 0$. However, Q_1 satisfies the same condition and $\deg_a Q_1 < \deg_a Q$, which contradicts the choice of $Q(X:a)$.

Thus we have proved the irreducibility of Q. □

The above proof might seem a little messy, but what we need first is the simple condition (13.14). The irreducibility will be used later.

As the relation $Q(f(u);\wp(u)) = 0$ holds for any u, we may change the variable to v or $u+v$: $Q(f(v);\wp(v)) = 0$, $Q(f(u+v);\wp(u+v)) = 0$. Let us denote $Y = f(v)$, $Z = f(u+v)$, $b = \wp(v)$ and $c = \wp(u+v)$ for brevity:

$$Q(Y;b) = 0, \qquad Q(Z;c) = 0. \tag{13.15}$$

[3] If you are an algebraist, you have probably noticed that we are talking about an ideal and its generator here. We do not use such notions explicitly, but knowledge of them will surely help to understand the rest of the proof.

Now we need some algebra. The essential tool is *elimination*, which 'eliminates a variable' from two equations. The concrete procedure is as follows.

(I) Eliminate the variable a from $Q(X;a)=0$ and the addition formula (13.9) of the $\wp$-function, $P_0(a,b,c)=0$, to obtain $P_1(X;b,c)=0$.
(II) Eliminate b from $Q(Y;b)=0$ and $P_1(X;b,c)=0$ to obtain $P_2(X,Y;c)=0$.
(III) Eliminate c from $Q(Z;c)=0$ and $P_2(X,Y;c)=0$ to obtain $P(X,Y,Z)=0$.

The result is an algebraic addition formula $P(f(u),f(v),f(u+v))=0$, (13.11).

We can roughly explain why such 'elimination of a variable' is possible, as follows. For example in the step (I),

- Solving the equation $Q(X;a)=0$, we can express a as a function of X: $a=a(X)$.
- Substituting it into $P_0(a,b,c)=0$, we obtain a relation among c, X and b, $P_0(a(X),b,c)=0$.

But this "explanation" has a problem. The function $a(X)$ obtained as a solution of $Q(X;a)=0$ is a complicated object. For example, even if $Q(X;a)=0$ is a quadratic equation in a, the function $a(X)$ is expressed by a complicated formula, as we know. In general, $Q(X;a)$ can be a higher degree polynomial and $a(X)$ is something very complicated, which cannot be explicitly written down. Therefore the above "explanation" does not answer the central question, 'Is there a polynomial relation?'

Here we illustrate elimination by a so-called resultant, which explicitly gives a polynomial relation.

Let $p(a)$ and $q(a)$ be polynomials of a which contain a, in other words, polynomials of degree not less than one in a:

$$p(a)=\alpha_0 a^l+\alpha_1 a^{l-1}+\cdots+\alpha_l, \qquad \alpha_0\neq 0,$$
$$q(a)=\beta_0 a^m+\beta_1 a^{m-1}+\cdots+\beta_m, \qquad \beta_0\neq 0,$$

where $l,m\geq 1$. The coefficients α_j $(j=0,\dots,l)$, β_k $(k=0,\dots,m)$ are not necessarily numbers and can be rational functions of other variables. (We shall use rational functions in X, Y, Z and polynomials in b, c.) For the moment we assume that the coefficients are rational functions of X for simplicity, but the discussion is the same even when other variables are present.

Definition 13.3 The following $(l+m)\times(l+m)$-matrix is called the *Sylvester matrix* of $p(a)$ and $q(a)$:

$$\mathrm{Syl}(p,q;a) := \begin{pmatrix}
\alpha_0 & & & & \beta_0 & & & \\
\alpha_1 & \alpha_0 & & & \beta_1 & \beta_0 & & \\
 & \alpha_1 & \ddots & & & \beta_1 & \ddots & \\
\vdots & & \ddots & \alpha_0 & \vdots & & & \beta_0 \\
 & \vdots & & \alpha_1 & & \vdots & & \beta_1 \\
\alpha_l & & & & \beta_m & & \ddots & \\
 & \alpha_l & & \vdots & & \beta_m & & \vdots \\
 & & \ddots & & & & \ddots & \\
 & & & \alpha_l & & & & \beta_m
\end{pmatrix}. \tag{13.16}$$

It has m columns consisting of α_j's and l columns consisting of β_k's. All elements which are not written above α_0 or β_0 and below α_l or β_m are zero.

The determinant of this matrix is called the *resultant* of p and q with respect to a and denoted by $\mathrm{res}(p,q;a)$:

$$\mathrm{res}(p,q;a) = \det \mathrm{Syl}(p,q;a). \tag{13.17}$$

Note that, although the notation contains the symbol a, it is a rational function in X free of a. □

Lemma 13.4 *(i) There exist polynomials $A(a)$ and $B(a)$ of a such that*

$$A(a)\,p(a) + B(a)\,q(a) = \mathrm{res}(p,q;a). \tag{13.18}$$

(The coefficients of $A(a)$ and $B(a)$ are rational functions of X.)

(ii) A necessary and sufficient condition for $\mathrm{res}(p,q;a) = 0$ *is that there exists a polynomial $r(a)$ in a (the coefficients are rational functions of X) which contains a and divides both $p(a)$ and $q(a)$.* □

We shall prove this lemma in the next section. Here we use these properties of the resultant to finish the proof of Theorem 13.1. Writing the variable X explicitly again like $p(a) = p(X;a)$, we can claim from (i) of Lemma 13.4 that, if a pair of variables (X,a) satisfies $p(X;a) = q(X;a) = 0$, X satisfies $\mathrm{res}(p,q;a)(X) = 0$. This means that, roughly speaking, 'the resultant $\mathrm{res}(p,q;a)(X)$ is obtained from $p(X;a)$ and $q(X;a)$ by eliminating the variable a'. Hence the proof of Theorem 13.1 is obtained by replacing the parts 'eliminate the variable a (or b, c)' in (I), (II) and (III) above by 'make a resultant'. More explicitly,

(I) Putting $P_1(X;b,c) := \mathrm{res}(Q(X;a), P_0(a,b,c);a)$, we can eliminate a from $Q(X;a) = 0$ and $P_0(a,b,c) = 0$ and obtain

$$P_1(X;b,c) = 0.$$

(II) Similarly, putting $P_2(X,Y;c) := \mathrm{res}(Q(Y;b), P_1(X;b,c);b)$, we obtain

$$P_2(X,Y;c) = 0$$

from $Q(Y;b) = 0$ and $P_1(X;b,c) = 0$.

(III) Finally, putting $P(X,Y,Z) := \mathrm{res}(Q(Z;c), P_2(X,Y;c); c)$, we obtain

$$P(X,Y,Z) = 0$$

from $Q(Z;c) = 0$ and $P_2(X,Y;c) = 0$, which gives an algebraic relation among $X = f(u)$, $Y = f(v)$ and $Z = f(u+v)$.

Recall that the resultant is a determinant of coefficients of polynomials and that a determinant is defined as a polynomial of elements of a matrix. Since $P_0(a,b,c)$ is a polynomial in a, b and c, its coefficients as a polynomial in a are polynomials of b and c. Therefore $P_1(X;b,c)$ defined in step (I) as a resultant is a polynomial in b and c whose coefficients are rational functions of X. In step (II) we consider it as a polynomial in b and construct a resultant $P_2(X,Y;c)$, which is a polynomial in c whose coefficients are rational functions of X and Y. We use it in step (III) to obtain $P(X,Y,Z)$ as a resultant. So, $P(X,Y,Z)$ is a rational function of X, Y and Z.

We are almost done, but there is one thing which we have to check: "Isn't this resultant equal to 0? If it is zero, what we have obtained is a trivial relation $0 = 0$." Here irreducibility of $Q(X;a)$ (Lemma 13.2 (ii)) plays its role.

For example, suppose that we had

$$P_1(X;b,c) = \mathrm{res}(Q(X;a), P_0(a,b,c); a) = 0$$

in step (I). We know from Lemma 13.4 (ii) that there is a polynomial $R(X,b,c;a)$ in a whose coefficients are rational functions of X, b and c such that

- $R(X,b,c;a)$ contains a: $\deg_a R(X,b,c;a) \geq 1$,
- $R(X,b,c;a)$ divides both $Q(X;a)$ and $P_0(a,b,c)$.

We may assume that $R(X,b,c;a)$ is monic in a (i.e., the coefficient of the highest degree term is 1). Since $Q(X;a)$ is irreducible,[4] $Q(X;a) = R(X,b,c;a)$, as both of them are monic. Thus $Q(X;a)$ should divide $P_0(a,b,c)$: $P_0(a,b,c) = Q(X;a)\,S(X,b,c;a)$, where $S(X,b,c;a)$ is a polynomial in a with coefficients in $\mathbb{C}(X,b,c)$.

But, as we saw in Lemma 13.2 (i), $Q(X;a)$ depends non-trivially on X, while $P_0(a,b,c)$ does not contain X. It follows from these facts that $Q(X;a)$ cannot divide $P_0(a,b,c)$, To see this,[5] let us fix $b = b_0 \in \mathbb{C}$ and $c = c_0 \in \mathbb{C}$ so that the highest degree term of $P_0(a,b,c)$ in a does not vanish: $N := \deg_a P_0(a,b,c) = \deg_a P_0(a,b_0,c_0)$. We denote $P_0(a,b_0,c_0)$ and $S(X,b_0,c_0;a)$ by $p(a)$ and $s(X;a)$ respectively for simplicity: $p(a) = Q(X;a)\,s(X;a)$.

[4] We proved the irreducibility of $Q(X;a)$ as a polynomial in a with coefficents in $\mathbb{C}(X)$ (= {rational functions of X}). Even if the coefficients are in $\mathbb{C}(X,b,c)$ (= {rational functions of X, b and c}), $Q(X;a)$ cannot be factorised; if $R_i(X,b,c;a)$ $(i = 1,2)$ are polynomials in a with coefficients in $\mathbb{C}(X,b,c)$ and $Q(X;a) = R_1(X,b,c;a)\,R_2(X,b,c;a)$, then fixing b and c to some complex numbers gives a factorisation of $Q(X;a)$ into polynomials in a with coefficients in $\mathbb{C}(X)$. Hence either R_1 or R_2 should be trivial.

[5] This immediately follows from the comparison of coefficients in $P_0(a,b,c) = Q(X;a)\,S(X,b,c;a)$, if $Q(X;a)$ and $S(X,b,c;a)$ are polynomials of X and a. The following proof is complicated, because $Q(X;a)$ is a rational function of X.

Let us factorise $p(a)$ as $p(a) = p_0 \prod_{j=1}^{N} (a - \alpha_j)$ ($p_0 \in \mathbb{C}$). The factor $a - \alpha_N$ is a factor of $Q(X;a)$ or $s(X;a)$. If it is a factor of $Q(X;a)$, $Q(X;a) = a - \alpha_N$ because of irreducibility of $Q(X;a)$. But this cannot happen, as $Q(X;a)$ non-trivially depends on X. Therefore $s(X;a)$ is divisible by $a - \alpha_N$: $s(X;a) = (a - \alpha_N) s_1(X;a)$, where $s_1(X;a)$ is a polymial of a with coefficients in $\mathbb{C}(X)$. Applying the same argument to $p_1(a) := p_0 \prod_{j=1}^{N-1} (a - \alpha_j) = Q(X;a)\, s_1(X;a)$, we can reduce the degree by one more: $p_2(a) := p_0 \prod_{j=1}^{N-2} (a - \alpha_j) = Q(X;a)\, s_2(X;a)$.

Repeating this procedure inductively, we come to a factorisation $p_{N-\deg_a Q}(a) := p_0 \prod_{j=1}^{\deg_a Q} (a - \alpha_j) = p_0 Q(X;a)$, which contradicts the non-trivial dependence of $Q(X;a)$ on X.

Thus we have proved $P_1(X;b,c) = \operatorname{res}(Q(X;a), P_0(a,b,c);a) \neq 0$. Similarly the resultants in the steps (II) and (III) do not vanish.

This is the end of the proof of Theorem 13.1. □

Are elliptic functions and trigonometric functions the only functions which have algebraic addition formulae? No, of course, the trivial linear function $l(u) = u$ satisfies a trivial addition formula $l(u+v) = l(u) + l(v)$. Starting from this relation, we can show that rational functions also have algebraic addition formulae: for example, a rational function $f(u) = \dfrac{p(u)}{q(u)}$ of u ($p(u)$ and $q(u)$ are polynomials of u) satisfies a relation

$$Q(f(u); l(u)) = 0,$$

where $Q(X;a) = q(a)\, X - p(a)$. This corresponds to (13.14) for an elliptic function. We have only to repeat the discussion for elliptic functions to prove the existence of an algebraic addition formula for $f(u)$.

An exponential function $\boldsymbol{e}_\alpha(u) = e^{\alpha u}$ ($\alpha \in \mathbb{C}$) also has an addition formula $\boldsymbol{e}_\alpha(u+v) = \boldsymbol{e}_\alpha(u)\, \boldsymbol{e}_\alpha(v)$. From this and the elimination algorithm it follows that rational functions of an exponential function also have algebraic addition formulae. The addition theorem for the sine function, which we considered in Section 13.1, is an example, as $\sin u = \dfrac{e^{iu} - e^{-iu}}{2i} = \dfrac{(e^{iu})^2 - 1}{2i\, e^{iu}}$.

Summarising, we have shown that

1. rational functions,
2. rational functions of $e^{\alpha u}$,
3. elliptic functions

have algebraic addition formulae.

How about other functions? — That is the main theme of the next chapter.

13.3 Resultants

In this section, as a digression, we give a proof of Lemma 13.4.

The base of the proof is the following simple correspondence. Given two polynomials

$$p(a) = \alpha_0 a^l + \alpha_1 a^{l-1} + \cdots + \alpha_l, \qquad \alpha_0 \neq 0, \tag{13.19}$$

$$r(a) = \gamma_0 a^k + \gamma_1 a^{k-1} + \cdots + \gamma_k, \qquad \gamma_0 \neq 0, \tag{13.20}$$

the coefficients of their product,

$$\begin{aligned} p(a)\,r(a) = \alpha_0\gamma_0\, a^{l+k} & \\ + (\alpha_1\gamma_0 + \alpha_0\gamma_1)\, a^{l+k-1} & \\ + (\alpha_2\gamma_0 + \alpha_1\gamma_1 + \alpha_0\gamma_2)\, a^{l+k-2} & \\ + \cdots & \\ + \alpha_l\gamma_k, & \end{aligned} \tag{13.21}$$

is given by the components of the vector,

$$\begin{pmatrix} \alpha_0 & & & \\ \alpha_1 & \alpha_0 & & \\ & \alpha_1 & \ddots & \\ \vdots & & \ddots & \alpha_0 \\ & \vdots & & \alpha_1 \\ \alpha_l & & & \\ & \alpha_l & & \vdots \\ & & \ddots & \\ & & & \alpha_l \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} = \begin{pmatrix} \alpha_0\gamma_0 \\ \alpha_1\gamma_0 + \alpha_1\gamma_0 \\ \alpha_2\gamma_0 + \alpha_1\gamma_1 + \alpha_0\gamma_2 \\ \vdots \\ \alpha_l\gamma_k \end{pmatrix}. \tag{13.22}$$

As you see, the matrix in the left-hand side of (13.22) is the left half of the Sylvester matrix $\mathrm{Syl}(p, q; a)$ defined by (13.16).

In view of this formula, the desired equation (13.18) is equivalent to

$$\mathrm{Syl}(p, q; a) \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{m-1} \\ \gamma_m \\ \vdots \\ \gamma_{l+m-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ \mathrm{res}(p, q; a) \end{pmatrix}, \tag{13.23}$$

where the γ_i's are coefficients of $A(a)$ and $B(a)$:

$$\begin{aligned} A(a) &= \gamma_0 a^{m-1} + \gamma_1 a^{m-2} + \cdots + \gamma_{m-1}, \\ B(a) &= \gamma_m a^{l-1} + \gamma_{m+1} a^{l-2} + \cdots + \gamma_{l+m-1}. \end{aligned} \tag{13.24}$$

Now we need some linear algebra. The *adjugate matrix* of an $N \times N$ matrix $T = (t_{ij})_{i,j=1,\dots,N}$ is defined by

$$\tilde{T} := (\tilde{t}_{ij}), \qquad \tilde{t}_{ij} = (-1)^{i+j} \det T_{\hat{j}\hat{i}}, \tag{13.25}$$

where $T_{\hat{j}\hat{i}}$ is the submatrix of T formed by deleting the j-th row and the i-th column from T. Its fundamental property is

$$T\tilde{T} = \tilde{T}T = (\det T) I_{N\times N} \tag{13.26}$$

($I_{N\times N}$ is the $N \times N$ identity matrix), which is a direct consequence of the cofactor expansion of the determinant.

In linear algebra this matrix appears as a part of the explicit formula of the inverse matrix, $T^{-1} = (\det T)^{-1}\tilde{T}$, when $\det T \neq 0$. What we need now is nothing more than the identity $T\tilde{T} = (\det T) I_{N\times N}$ in (13.26) or, more precisely, its rightmost column:

$$T(\text{the rightmost column of } \tilde{T}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \det T \end{pmatrix}.$$

By comparing this formula with (13.23) it is obvious that we can take the rightmost column vector of the adjugate matrix of the Sylvester matrix $\mathrm{Syl}(p, q; a)$ as the vector $\begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_{l+m-1} \end{pmatrix}$ in the left-hand side of (13.23).

This completes the proof of statement (i).

The proof of (ii) is also based on the correspondence of the product of polynomials (13.21) and the vector (13.22). Assume that polynomials $p(a)$ and $q(a)$ have a common divisor

$$r(a) = \gamma_0 a^k + \gamma_1 a^{k-1} + \cdots + \gamma_k, \qquad (\gamma_0 \neq 0,\ k \geq 1). \tag{13.27}$$

This means that there are polynomials

$$\begin{aligned} \tilde{p}(a) &= \tilde{\alpha}_0 a^{l-k} + \tilde{\alpha}_1 a^{l-k-1} + \cdots + \tilde{\alpha}_{l-k}, & \tilde{\alpha}_0 \neq 0, \\ \tilde{q}(a) &= \tilde{\beta}_0 a^{m-k} + \tilde{\beta}_1 a^{m-k-1} + \cdots + \tilde{\beta}_{m-k}, & \tilde{\beta}_0 \neq 0, \end{aligned}$$

such that

$$p(a) = r(a)\,\tilde{p}(a), \qquad q(a) = r(a)\,\tilde{q}(a). \tag{13.28}$$

Interpreting this by the formula (13.22), we have

$$
\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_l \end{pmatrix}
=
\begin{pmatrix}
\gamma_0 & & & \\
\gamma_1 & \gamma_0 & & \\
 & \gamma_1 & \ddots & \\
\vdots & & \ddots & \gamma_0 \\
 & \vdots & & \gamma_1 \\
\gamma_k & & & \\
 & \gamma_k & & \vdots \\
 & & \ddots & \\
 & & & \gamma_k
\end{pmatrix}
\begin{pmatrix} \tilde{\alpha}_0 \\ \tilde{\alpha}_1 \\ \vdots \\ \tilde{\alpha}_{l-k} \end{pmatrix},
\tag{13.29}
$$

(the matrix in the right-hand side is of size $(l+1)\times(l-k+1)$), and a similar equation for $\beta_0,\dots,\beta_m$, $\tilde{\beta}_0,\dots,\tilde{\beta}_{m-k}$. As the columns of the Sylvester matrix are obtained by shifting the vectors with components $\alpha_0,\dots,\alpha_l$ or $\beta_0,\dots,\beta_m$, we obtain the following expression for $\mathrm{Syl}(p,q;a)$ from (13.29):

$$
\begin{aligned}
&\mathrm{Syl}(p,q;a) \\
&=
\begin{pmatrix}
\gamma_0 & & & \\
\gamma_1 & \gamma_0 & & \\
 & \gamma_1 & \ddots & \\
\vdots & & \ddots & \gamma_0 \\
 & \vdots & & \gamma_1 \\
\gamma_k & & & \\
 & \gamma_k & & \vdots \\
 & & \ddots & \\
 & & & \gamma_k
\end{pmatrix}
\begin{pmatrix}
\tilde{\alpha}_0 & & & \tilde{\beta}_0 & & \\
\tilde{\alpha}_1 & \tilde{\alpha}_0 & & \tilde{\beta}_1 & \tilde{\beta}_0 & \\
 & \tilde{\alpha}_1 & \ddots & & \tilde{\beta}_1 & \ddots \\
\vdots & & \ddots \ \tilde{\alpha}_0 & \vdots & & \tilde{\beta}_0 \\
 & \vdots & \tilde{\alpha}_1 & & \vdots & \tilde{\beta}_1 \\
\tilde{\alpha}_{l-k} & & & \tilde{\beta}_{m-k} & & \ddots \\
 & \tilde{\alpha}_{l-k} & \vdots & & \tilde{\beta}_{m-k} & \vdots \\
 & & \ddots & & & \ddots \\
 & & \tilde{\alpha}_{l-k} & & & \tilde{\beta}_{m-k}
\end{pmatrix}.
\end{aligned}
$$

Visually the size of the matrix seems to be larger than the original one, (13.16), but it is merely the effect of the length of the indices. The sizes of the matrices in the right-hand side are $(l+m)\times(l+m-k)$ and $(l+m-k)\times(l+m)$. Hence the rank of the Sylvester matrix is not greater than $l+m-k$. (This is a result of the inequality[6] $\operatorname{rank} TT' \leq \min(\operatorname{rank} T, \operatorname{rank} T')$.)

If $k \geq 1$, the rank of $\mathrm{Syl}(p,q;a)$ is less than its size $l+m$, from which $\mathrm{res}(p,q;a) = \det \mathrm{Syl}(p,q;a) = 0$ follows.

Conversely, let us show that for polynomials $p(a)$ and $q(a)$ there exists a common divisor $r(a)$ ($\deg_a r(a) \geq 1$) as in (13.28) under the assumption $\det \mathrm{Syl}(p,q;a) = 0$.

The first thing we use is one of the most important theorems in linear algebra: 'If the determinant of a matrix T is 0, there exists a non-zero vector v such that $Tv = 0$'. Applying this to the Sylvester matrix $\mathrm{Syl}(p,q;a)$ with zero determinant, there exists a non-zero vector v such that $\mathrm{Syl}(p,q;a)v = 0$. This means that there is a set $\{\gamma_0,\dots,\gamma_{l+m-1}\}$ of rational functions of X, which contains at least one non-zero

[6] This estimate of the rank of the product of matrices follows from the definition of the rank (= number of linearly independent rows (or columns)) and the fact that rows (respectively, columns) of the product TT' are linear combinations of rows of T' (respectively, columns of T).

element and satisfies

$$\mathrm{Syl}(p,q;a)\begin{pmatrix}\gamma_0\\ \gamma_1\\ \vdots\\ \gamma_{m-1}\\ \gamma_m\\ \vdots\\ \gamma_{l+m-1}\end{pmatrix}=0. \tag{13.30}$$

Defining polynomials $A(a)$ and $B(a)$ in a by (13.24) with these γ_i's, we have

$$A(a)p(a)+B(a)q(a)=0, \tag{13.31}$$

as the correspondence (13.23) and (13.18).

Note that both $A(a)$ and $B(a)$ are not 0, since if one of them were zero, both of them would become 0 because of (13.31) and all γ_i's would vanish. Another important thing is that there are bounds of degrees,

$$\deg_a A(a) \le m-1, \qquad \deg_a B(a) \le l-1. \tag{13.32}$$

Now let us suppose that $p(a)$ and $q(a)$ do not have a common divisor. In this case there are polynomials $\tilde{A}(a)$ and $\tilde{B}(a)$ such that

$$\tilde{A}(a)\,p(a)+\tilde{B}(a)\,q(a)=1. \tag{13.33}$$

This is a consequence of *Euclid's algorithm*: suppose $l=\deg_a p(a)\ge m=\deg_a q(a)$ and put $p_0(a):=p(a)$, $p_1(a):=q(a)$. By dividing $p_0(a)$ by $p_1(a)$, we obtain a remainder of degree less than that of p_1:

$$p_0(a)=q_1(a)\,p_1(a)+p_2(a), \qquad \deg_a p_2(a)<\deg_a p_1(a). \tag{13.34}$$

Polynomials $p_1(a)$ and $p_2(a)$ do not have a common divisor, because such a divisor would divide both sides of (13.34) and become a common divisor of $p_0(a)$ and $p_1(a)$, which contradicts the assumption on $p_0(a)=p(a)$ and $p_1(a)=q(a)$.

Now by dividing $p_1(a)$ by $p_2(a)$, we have

$$p_1(a)=q_2(a)\,p_2(a)+p_3(a), \qquad \deg_a p_3(a)<\deg_a p_2(a). \tag{13.35}$$

By a similar argument we can show that p_2 and p_3 do not have a common divisor. Repeating this procedure as

$$p_k(a)=q_{k+1}(a)\,p_{k+1}(a)+p_{k+2}(a),\ \deg_a p_{k+2}(a)<\deg_a p_{k+1}(a), \tag{13.36}$$

we define a sequence $p_0(a), p_1(a), p_2(a), \dots$ of polynomials. This sequence has the following properties:

- $\deg_a p_k > \deg_a p_{k+1}$.
- $p_k(a)$ and $p_{k+1}(a)$ do not have a common divisor.

From the first property it follows that this sequence terminates after a finite number of steps and the final polynomial (say, $p_N(a)$) has degree 0: $\deg_a p_N(a) = 0$, i.e., $p_N(a) = c =$ (constant with respect to a) = (a rational function of X, which does not depend on a). This constant does not vanish, as $p_N(a) = c = 0$ means that p_{N-2} is divisible by p_{N-1}, which contradicts the second property of the sequence $\{p_k\}$. Thus we obtain

$$\begin{aligned} p_{N-2}(a) &= q_{N-1}(a)\, p_{N-1}(a) + c, \\ \text{or, } p_{N-2}(a) - q_{N-1}(a)\, p_{N-1}(a) &= c. \end{aligned} \tag{13.37}$$

Using (13.36) ($k = N-1$), $p_{N-1}(a) = p_{N-3}(a) - q_{N-2}(a)\, p_{N-2}(a)$, we can rewrite the left-hand side of the second equation of (13.37) as a linear combination of p_{N-2} and p_{N-3}. Repeating this procedure and going back to p_0 and p_1, we arrive at an equation of the form

$$\tilde{\tilde{A}}(a)\, p_0(a) + \tilde{\tilde{B}}(a)\, p_1(a) = c.$$

By dividing both sides by c, we obtain (13.33).

By multiplying (13.33) by $B(a)$ in (13.31), $B(a)$ can be expressed as

$$\begin{aligned} B(a) &= B(a)\big(\tilde{A}(a)\, p(a) + \tilde{B}(a)\, q(a)\big) \\ &= \big(\tilde{A}(a)\, B(a)\big)\, p(a) + \tilde{B}(a)\, \big(B(a)\, q(a)\big). \end{aligned}$$

Since $B(a)\, q(a) = -A(a)\, p(a)$ because of (13.31), this can also be written as

$$B(a) = \big(\tilde{A}(a)\, B(a) - \tilde{B}(a)\, A(a)\big)\, p(a).$$

Therefore $B(a)$ is divisible by $p(a)$. However the condition (13.32) implies $\deg_a B(a) \leq l - 1 = \deg_a p(a) - 1$ and $B(a)$ cannot be divisible by $p(a)$. This is a contradiction, which shows that the assumption '$p(a)$ and $q(a)$ do not have a common divisor' does not hold and that there exists a polynomial $\tilde{r}(a)$, $\deg_a \tilde{r} \geq 1$, such that

$$p(a) = \tilde{r}(a)\, \tilde{p}(a), \qquad q(a) = \tilde{r}(a)\, \tilde{q}(a).$$

□(End of the proof of Lemma 13.4.)

Although it is not directly related to elliptic functions, at this opportunity let us derive a discriminant formula, which is the simplest application of the resultant.

Exercise 13.5 (i) Show that a polynomial $P(z)$ with complex coefficients[7] has a multiple root α (i.e., divisible by $(z-\alpha)^2$) if and only if α is a common root of $P(z)$ and $P'(z) = \dfrac{d}{dz}P(z)$ (i.e., both $P(z)$ and $P'(z)$ are divisible by $(z-\alpha)$). (This is neither related to resultants nor to elliptic functions. It is nothing more than a preparation for the next problem.)

(ii) From the result of (i) it follows that a necessary and sufficient condition for $P(z)$ to have a multiple root is $\operatorname{res}(P, P'; z) = 0$. Write down this condition

[7] Coefficients of the polynomial can be taken from other fields. We take $\mathbb{C}$ for simplicity.

for $P(z) = az^2 + bz + c$ explicitly and confirm that we obtain the 'discriminant of a quadratic equation' which we know from our schooldays. How about the case $P(z) = az^3 + bz^2 + cz + d$?

In the general case, too, the resultant $\mathrm{res}(P, P'; z)$ is called the *discriminant* of a polynomial P.

Remark 13.6 As we have shown, in principle we can eliminate a variable from polynomials by computing resultants. But it requires computing determinants of large size and is not very efficient for practical use. Today the so-called Gröbner basis method is known to be more efficient for elimination and algorithms for the computation of Gröbner bases are implemented in many computer systems. For details on Gröbner bases see, for example, [CLO].

Chapter 14
Characterisation by Addition Formulae

In the previous chapter we proved that rational functions, rational functions of an exponential function and elliptic functions have addition theorems (algebraic addition formulae). Are there other functions which have algebraic addition formulae? The next Weierstrass–Phragmén theorem[1] answers this question.

Theorem 14.1 *If a meromorphic function $f(u)$ on $\mathbb{C}$ has an algebraic addition theorem $P(f(u), f(v), f(u+v)) = 0$ ($P(x, y, z)$ is a polynomial), it is either a rational function of u, a rational function of an exponential function $e^{\alpha u}$ ($\alpha \in \mathbb{C}$) or an elliptic function.*

The goal of this chapter is to prove this theorem. Our strategy is the following:

- To begin with, we study properties of a meromorphic function on $\mathbb{C}$ and show that the point at infinity is an essential singularity, unless it is a rational function.
- Then we study local properties in a neighbourhood of an essential singularity.
- Combining these properties at an essential singularity and the algebraic addition formula, we show that f has at least one period, unless f is a rational function.
- Finally, we show that such f is either an elliptic function with another period, or a rational function of an exponential function.

[1] Probably the first published proof is due to Phragmén (Phragmén, Lars Edvard, 1863–1937) [P]. According to the beginning of this article, 'Dans son enseignement à l'université de Berlin, M. Weierstrass a donné une théorie des fonctions elliptiques aussi remarquable par la beauté que par la simplicité des méthodes, où il prend pour point de départ le théorème suivant (In his teaching at the University of Berlin Mr Weierstrass has given a theory of elliptic functions as remarkable for its beauty as for the simplicity of its methods, where he took the following theorem as a starting point).' Exactly speaking, the theorem in [P] is for possibly multi-valued analytic functions. See Remark 14.9 at the end of this chapter.

The same theorem can be readily derived from an earlier work by Briot and Bouquet [BB], where a characterisation of elliptic functions by first-order ordinary differential equations is given.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_15

14.1 Behaviour of Meromorphic Functions at Infinity

First, let us review the definition of a meromorphic function: It is a complex function, all singularities of which are poles. We are considering a meromorphic function $f(u)$ on $\mathbb{C}$, so at any point $z \in \mathbb{C}$ it has a Laurent expansion with at most finite principal part. Exactly speaking: For any $z \in \mathbb{C}$ there exist an integer $N(z) \in \mathbb{Z}$, a positive real number $\rho(z) > 0$ and a sequence of complex numbers $\{a_n\}_{n \geq -N(z)}$ such that

$$f(u) = \sum_{n=-N(z)}^{\infty} a_n (u-z)^n, \qquad (a_{-N(z)} \neq 0), \tag{14.1}$$

holds on a punctured disk $\{u \mid 0 < |u-z| < \rho(z)\}$. When $N(z) \leq 0$, f is holomorphic at z, and when $N(z) > 0$, z is a pole of f of order $N(z)$.

The set of poles is discrete, i.e., does not have an accumulation point in $\mathbb{C}$, because a function cannot have a Laurent expansion (14.1) at an accumulation point of poles.

From the condition $\rho(z) > 0$ it follows that there are only a countable number of poles. In fact, let S be the set of all poles, $S := \{z \mid N(z) > 0\}$. For each $z \in S$ take a disk $D(z) := \{\zeta \mid |\zeta - z| < \rho(z)/2\}$. Since there is no pole in the disk $\{\zeta \mid |\zeta - z| < \rho(z)\}$ except for z, $D(z)$ and $D(z')$ do not intersect, $D(z) \cap D(z') = \emptyset$, if $z, z' \in S$ and $z \neq z'$. Therefore we can choose a distinct rational point $\kappa(z) \in \mathbb{Q} + i\mathbb{Q} \subset \mathbb{C}$ from each $D(z)$: $\kappa(z) \in (\mathbb{Q}+i\mathbb{Q}) \cap D(z)$, $\kappa(z) \neq \kappa(z')$ $(z \neq z')$. This means that there is an injection $z \mapsto \kappa(z)$ from S to a countable set $\mathbb{Q}+i\mathbb{Q}$, from which it follows that S is a countable set.

How does such $f(u)$ behave in a neighbourhood of $u = \infty$? Using the coordinate change $w = u^{-1}$, we regard f as a meromorphic function on $\mathbb{C} \setminus \{w = 0\}$: $\tilde{f}(w) := f(w^{-1})$. Let us examine its behaviour around $w = 0$ (i.e., $u = \infty$).

Lemma 14.2 *If $f(u)$ is not a rational function, $w = 0$ is an essential singularity of $\tilde{f}(w)$, which means that $u = \infty$ is an essential singularity of $f(u)$.*

Here we mean by 'an essential singularity' either an isolated singular point, at which a Laurent expansion (14.1) *is infinite in the negative direction ($N_0 := N(\infty) = \infty$), or a non-isolated singularity, which is an accumulation point of poles.*

Proof What we have to show is that, if a meromorphic function on $\mathbb{C}$ does not have an essential singularity at infinity, it is a rational function. Let us assume that $u = \infty$ is not an essential singularity of $f(u)$, or, equivalently, $\tilde{f}(w)$ has the form

$$\tilde{f}(w) = \frac{b_{-N_0}}{w^{N_0}} + \cdots + \frac{b_{-1}}{w} + (\text{holomorphic function at } w = 0)$$

in a neighbourhood of $w = 0$. Returning to $f(u)$, we can rewrite the condition as

$$f(u) = b_{-N_0} u^{N_0} + \cdots + b_{-1} u + (\text{holomorphic function at } u = \infty). \tag{14.2}$$

Suppose that $f(u)$ has infinitely many poles in $\mathbb{C}$. Although the number of poles is infinite, there are only a finite number of poles in a bounded subset of $\mathbb{C}$ (= subset of $\mathbb{C}$ contained in a disk $\{u \mid |u| \leq R\}$ for a fixed R). Indeed, if there are infinitely many poles in a bounded region, there is an accumulation point of poles in $\mathbb{C}$ by virtue of the Bolzano–Weierstrass theorem, which contradicts our assumption.

Thus $f(u)$ has infinitely many poles in total, but only a finite number of them in a bounded region, from which it follows that for any $N \in \mathbb{N}$ there exists a pole in the region $\{u \mid |u| > N\}$. For each N we denote one such pole by u_N: $|u_N| > N$. In the w-coordinate, this sequence of poles has the coordinates $\{w_N = u_N^{-1}\}$, which converges to 0 when $N \to \infty$. Therefore $w = 0$ $(u = \infty)$ is an accumulation point of poles and becomes an essential singularity, which contradicts our assumption.

So, the number of poles of $f(u)$ is finite. We denote these poles by $u_1, \dots, u_M$, at which $f(u)$ has the Laurent expansion

$$f(u) = \sum_{n=-N_i}^{\infty} a_{i,n}(u-u_i)^n, \qquad (a_{i,-N_i} \neq 0). \tag{14.3}$$

Subtracting the principal part

$$P_i(u) := \sum_{n=-N_i}^{-1} a_{i,n}(u-u_i)^n$$

of this Laurent expansion at $u = u_i$ from (14.3), we obtain a power series in $(u-u_i)$ without negative powers, i.e., a Taylor series, which is holomorphic at $u = u_i$. Since a term $(u-u_i)^{-n}$ does not have a singularity except at $u = u_i$, the principal parts of poles of $f(u) - P_i(u)$ at $u = u_j$ $(j \neq i)$ are the same as those of $f(u)$. Therefore, subtracting all the principal parts at poles, we obtain a holomorphic function on $\mathbb{C} \cup \{\infty\}$,

$$g(u) := f(u) - \sum_{i=1}^{M} P_i(u) - (b_{-N_0}u^{N_0} + \cdots + b_{-1}u).$$

The last part in the parentheses is the principal part at $u = \infty$, i.e., the positive power terms in the expansion (14.2), which diverges at $u = \infty$. According to Liouville's theorem (not those proved in Section 11.2 but that in complex analysis; 'a bounded entire function is constant'), $g(u)$ is a constant function $g(u) \equiv c$. This means that $f(u)$ has the form

$$f(u) = c + \sum_{i=1}^{M} P_i(u) + (b_{-N_0}u^{N_0} + \cdots + b_{-1}u),$$

which is a rational function of u. □

14.2 Properties of Essential Singularities

It is clear from Lemma 14.2 that properties of essential singularities are important in the proof of Theorem 14.1. There are deep theories and complicated theorems on essential singularities, but here we restrict ourselves to the minimum, about which we can show the full proof.

The following theorem is classical. We shall apply it to an essential singularity at infinity, but, in order to unburden the notation, we take an essential singularity at $z = 0$. Of course, we can restate it for an essential singularity at $u = \infty$ by the coordinate change $u = z^{-1}$.

Proposition 14.3 (Casorati–Sokhotskii–Weierstrass theorem)

If a meromorphic function $f(z)$ on a domain $\{z \mid 0 < |z| < R\}$ has an essential singularity at $z = 0$, for any $\varepsilon > 0$, $\delta \in (0, R)$ and $c \in \mathbb{C}$, then there exists a $z_{\varepsilon,\delta}$ satisfying

$$0 < |z_{\varepsilon,\delta}| < \delta, \qquad |f(z_{\varepsilon,\delta}) - c| < \varepsilon. \tag{14.4}$$

This theorem has several names:[2] ‘Casorati–Weierstrass theorem’, simply ‘Weierstrass theorem”, or (in Russia) ‘Sokhotskii’s theorem’.

The meaning of this theorem might be clearer in the following form.

Corollary 14.4 *Under the condition of Proposition* 14.3 *for any $c \in \mathbb{C}$ there exists a sequence $\{z_n\}_{n=1,2,\dots}$ of complex numbers such that $\lim_{n\to\infty} z_n = 0$ and $\lim_{n\to\infty} f(z_n) = c$.*

This follows from Proposition 14.3 immediately by taking $z_{\varepsilon,\delta}$ for $\varepsilon = \delta = \dfrac{1}{n}$ as z_n.

Proof (of Proposition* 14.3.*) If $f(z)$ were identically equal to c, $f(z)$ would not have an essential singularity. So, $f(z) \not\equiv c$. Put $\varphi(z) := \dfrac{1}{f(z) - c}$. Since $f(z)$ is meromorphic on $\{z \mid 0 < |z| < R\}$, $\varphi(z)$ is also meromorphic on the same place.

Suppose that for certain ε and δ there is no $z_{\varepsilon,\delta}$ which satisfies (14.4). Then

$$0 < |z| < \delta \Longrightarrow |\varphi(z)| \le \varepsilon^{-1}.$$

Therefore $\varphi(z)$ is bounded on this region. Hence we can extend $\varphi(z)$ holomorphically to $z = 0$ by Riemann’s theorem on removable singularities. This contradicts the assumption that $z = 0$ is an essential singularity of $f(z)$. □

[2] It was proved by C. Briot and C. Bouquet in 1859. (Corollaire II of Théorème IV in Ch.IV of the first part of [BB]. The statement is not correct, but the proof is correct.) They removed it from the second edition of the book (1875). According to the information in pp. 4–5 of [CL], many people published proofs: Weierstrass in his article, Zur Theorie der eindeutigen analytischen Functionen. *Abh. Königl. Akad. Wiss.* (1876); F. Casorati in his book on complex analysis, *Teorica delle funzioni di variabile complesse*, Pavia, 1868; J. W. Sokhotskii in his dissertation (St. Petersburg, 1873).

Using Proposition 14.3, we can prove the following more accurate theorem, which we shall use later. Because of the accuracy of the statement the proof is complicated.

Theorem 14.5 *Let $f(z)$ be a meromorphic function on $0 < |z| < R$ with an essential singularity at $z = 0$. For any complex number $c \in \mathbb{C}$ and any positive real number $\varepsilon > 0$ there exist $c' \in \mathbb{C}$ and a sequence $\{z_n\}_{n=1,2,\dots}$ such that $|c' - c| < \varepsilon$, $0 < |z_n| < R$ $(n = 1, 2, \dots)$, $\lim_{n\to\infty} z_n = 0$ and $f(z_n) = c'$.*

In Corollary 14.4 the sequence $f(z_n)$ *converges to* c, $\lim_{n\to\infty} f(z_n) = c$, while in Theorem 14.5 it is *always equal to* c', $f(z_n) = c'$.

Proof We inductively construct a sequence $\{\zeta_n\}_{n=1,2,\dots}$ in $\mathbb{C}$ and a sequence $\{\varepsilon_n\}_{n=1,2,\dots}$ $(\varepsilon_0 = \varepsilon)$ of positive real numbers satisfying the following conditions: We denote

- the punctured open disk $\left\{z \mid 0 < |z| < \dfrac{R}{2^n}\right\}$ by D_n,
- the closed disk $\{w \mid |w - f(\zeta_n)| \le \varepsilon_n\}$ by Δ_n,
- its interior $\{w \mid |w - f(\zeta_n)| < \varepsilon_n\}$ by Δ_n°,

and $\Delta_0 := \{w \mid |w - c| \le \varepsilon\}$, $\Delta_0^\circ := \{w \mid |w - c| < \varepsilon\}$. Then the conditions are

1. $\zeta_n \in D_{n-1}$, $f(\zeta_n) \in \Delta_{n-1}^\circ$,
2. $\Delta_n \subset f(D_{n-1} \setminus (\text{poles of } f)) \cap \Delta_{n-1}^\circ$, in particular $\Delta_n \subset \Delta_{n-1}$.
3. $\varepsilon_n \le 2^{-n}$.

We illustrate these complicated conditions in Fig. 14.1.

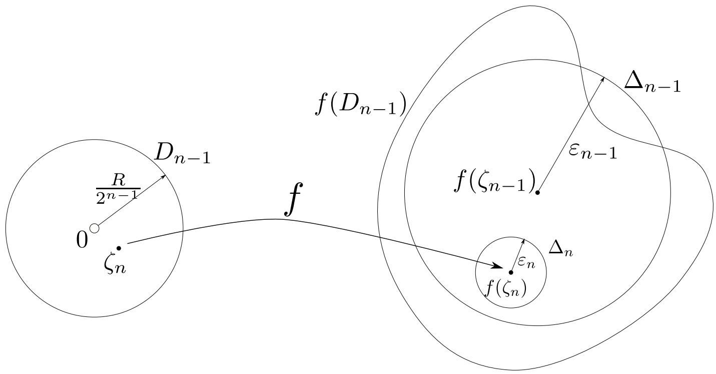

Fig. 14.1 Determine ζ_n and ε_n from ζ_{n-1} and ε_{n-1}.

Assume that ζ_{n-1} and ε_{n-1} are given. We determine ζ_n and ε_n as follows.

Condition (1) means that ζ_n is close to the origin (= the essential singularity) and that $f(\zeta_n)$ is close to the designated point ($f(\zeta_{n-1})$ for $n > 1$; c for $n = 1$). Therefore the existence of ζ_n follows directly from Proposition 14.3.

The existence of ε_n satisfying condition (2) is shown as follows: Since $D_{n-1} \setminus$ (poles of f) is an open set,[3] its image is also open by Theorem A.13. Hence its intersection with the open disk Δ°_{n-1} is also open. From condition (1) it follows that $f(\zeta_n) \in f(D_{n-1}) \cap \Delta^\circ_{n-1}$, which means that a sufficiently small neighbourhood of $f(\zeta_n)$ is contained in the open set $f(D_{n-1}) \cap \Delta^\circ_{n-1}$. Thus we can take ε_n smaller than 2^{-n} such that the closed disk Δ_n is contained in that small neighbourhood.

The centre $f(\zeta_n)$ of the disk Δ_n belongs to Δ°_N when $n \geq N$. Therefore, if $m, n \geq N$, $|f(\zeta_m) - f(\zeta_n)| < 2^{-N+1}$, which means that $\{f(\zeta_n)\}_{n=1,2,\dots}$ is a Cauchy sequence. As a Cauchy sequence converges, there exists the limit $c' := \lim_{n\to\infty} f(\zeta_n)$, which belongs to every Δ_n, since Δ_n is closed. In particular $c' \in \Delta_1 \subset \Delta^\circ_0$ and $|c' - c| < \varepsilon$.

For each n there exists a z_n such that $z_n \in D_n$, $f(z_n) = c'$, as $c' \in \Delta_{n+1} \subset f(D_n)$. These c' and $\{z_n\}_{n=1,2,\dots}$ are what we are looking for. □

Remark 14.6 In fact, a stronger statement holds:

Theorem 14.7 (Picard's great theorem)

If $f(z)$ is meromorphic on $D := \{z \mid 0 < |z| < R\}$ and has an essential singularity at $z = 0$, f takes any complex value on D with at most two exceptions.[4] *In other words, there is a finite set E, the number of elements of which is* 0 *or* 1 *or* 2, *such that $f(\{z \mid 0 < |z| < R\}) = (\mathbb{C} \cup \{\infty\}) \setminus E$.*

This theorem says that for almost all c we can take z_n such that $f(z_n) = c$, not taking an 'approximation c'' as in Theorem 14.5.

In Fig. 14.1 the image $f(D_{n-1})$ of the punctured disk D_{n-1} is 'almost' identical to $\mathbb{C}$, but, avoiding Picard's great theorem, we drew a 'usual open set'.

14.3 Proof of the Weierstrass–Phragmén Theorem

Now finally the algebraic addition formula

$$P(f(u), f(v), f(u+v)) = 0 \tag{14.5}$$

comes into play. In order to show Theorem 14.1, we have to prove that, if a meromorphic function satisfying (14.5) is not a rational function, it is either a rational function of $e^{\alpha u}$ for certain α or an elliptic function. A rational function of $e^{\alpha u}$ has period $\dfrac{2\pi i}{\alpha}$ and an elliptic function is also a periodic function. So, to begin with, let us show that a meromorphic function satisfying (14.5) which is *not* a rational function is a periodic function.

[3] Recall that the set of poles of f may have an accumulation point only at 0, which does not belong to D_{n-1}.

[4] This 'value' and 'exception' may be ∞. For example, if f is holomorphic on D, ∞ is one of the exceptional values and the statement becomes 'f takes any complex value with at most a single exception'.

We assume that $P(x,y,z)$ is a polynomial of degree N in z. According to Lemma 14.2, $u = \infty$ is an essential singularity of f. So, by applying Theorem 14.5 (with the coordinate change $z = u^{-1}$), for any $c \in \mathbb{C}$ and $\varepsilon > 0$ we can take a complex number $c' \in U_\varepsilon := \{z \mid |z - c| < \varepsilon\}$ and a sequence of complex numbers $\{u_n\}_{n=1,2,\dots}$ converging to ∞ such that $f(u_n) = c'$. What is important in this statement for our purpose is that 'there are at least $N+1$ distinct points, at which f takes the same value c''. So, for simplicity of notation, we denote c' by C and $N+1$ distinct points from the sequence $\{u_n\}_{n=1,2,\dots}$ by $a_0, \dots, a_N$:

$$f(a_n) = C \quad (n = 0, \dots, N), \qquad a_m \neq a_n \quad (m \neq n). \tag{14.6}$$

Lemma 14.8 *There exists a finite set E determined by the polynomial $P(x,y,z)$ such that, if $C \notin E$ and a meromorphic function $f(u)$ on $\mathbb{C}$ satisfies* (14.5) *and* (14.6), *then $a_{m_0} - a_{m_1}$ is a period of $f(u)$ for some m_0 and m_1 ($m_0 \neq m_1$).*

In the present situation we can take c and $\varepsilon > 0$ so that the ε-neighbourhood U_ε of c does not intersect with E. Hence Lemma 14.8 implies that $f(u)$ is a periodic function.

Proof (of Lemma* 14.8*) The claim of this lemma holds trivially for a constant function, so we assume that $f(u)$ is not constant. Actually, being a rational function, a constant function has already been excluded in the main context, but in this lemma it is not important whether f is rational or not. In the proof a constant function should be treated separately, so, just for simplicity, we exclude only a constant function.

As was discussed at the beginning of Section 14.1, the poles of f are countably many and distributed discretely, which allows us to choose $v_0 \in \mathbb{C}$ and its neighbourhood U_0 such that f is holomorphic in neighbourhoods U_0, $U_0 + a_0$, ..., $U_0 + a_N$ of $N+2$ points v_0, $v_0 + a_0$, ..., $v_0 + a_N$.

Fixing any point v in a neighbourhood U_0 of v_0 and substituting $u = a_n$ in the addition formula (14.5), we have

$$P(C, f(v), f(v + a_n)) = 0. \tag{14.7}$$

This means that all $f(v + a_n)$ $(n = 0, \dots, N)$ satisfy an equation

$$P(C, f(v), z) = 0 \tag{14.8}$$

for z.

"The equation $P(C, f(v), z) = 0$ of degree N in z has N roots with multiplicities, so there should be duplication in $f(v + a_n)$ $(n = 0, \dots, N)$." Is this really so? Is (14.8) really a *non-trivial equation*? To conclude that there is duplication in $\{f(v + a_0), \dots, f(v + a_N)\}$, we have to show that (14.8) *does* contain z. Let

$$P(x,y,z) = p_0(x,y) + p_1(x,y)z + \cdots + p_N(x,y)z^N \tag{14.9}$$

be the expansion of $P(x,y,z)$ in z. We want to show that at least one of the coefficients $p_1(x,y)$, ..., $p_N(x,y)$ does not vanish at $(x,y) = (C, f(v))$.

Here we may assume that $p_0(x,y)$ is a non-zero polynomial in x and y. In fact, if $p_0(x,y) = 0$ as a polynomial, $P(x,y,z)$ is factorised as

$$\begin{aligned}
P(x,y,z) &= z\tilde{P}(x,y,z), \\
\tilde{P}(x,y,z) &= p_1(x,y) + p_2(x,y)z + \cdots + p_N(x,y)z^{N-1}.
\end{aligned}$$

Hence the addition formula of f has the form $f(u+v)\,\tilde{P}(f(u), f(v), f(u+v)) = 0$. If we use variables $w = u+v$, $v = w-u$, this means that

$$f(w)\,\tilde{P}(f(u), f(w-u), f(w)) = 0$$

holds identically as a function of w. As $f(w)$ is not a constant function, it is not identically 0. Therefore by Theorem A.8 we have $\tilde{P}(f(u), f(w-u), f(w)) = 0$. In other words, we have an algebraic addition formula

$$\tilde{P}(f(u), f(v), f(u+v)) = 0.$$

If $p_1(x,y) = 0$ here, we can again simplify the addition formula to

$$\tilde{\tilde{P}}(f(u), f(v), f(u+v)) = 0,$$

where $\tilde{\tilde{P}}(x,y,z) = p_2(x,y) + p_3(x,y)z + \cdots + p_N(x,y)z^{N-2}$.

Since we can repeat this procedure until the constant term in z becomes non-zero, we may assume that $p_0(x,y)$ in (14.9) is a *non-vanishing* polynomial in x and y.

Having made this preparation, suppose $p_1(C, f(v)) = \cdots = p_N(C, f(v)) = 0$. Then (14.7) is simplified to

$$p_0(C, f(v)) = 0.$$

If the polynomial $p_0(x,y) = p_{00}(x) + p_{01}(x)y + \cdots + p_{0k}(x)y^k$ in two variables x and y contains y, the equation $p_0(C, f(v)) = 0$ means that "$y = f(v)$ is a root of $p_0(C,y) = 0$", which restricts the number of possible values of $f(v)$ to be finite. This cannot happen unless f is a constant function. Hence $p_0(x,y) = p_{00}(x)$ does not contain y and $x = C$ is a root of the equation $p_{00}(x) = 0$.

Since we are assuming that $p_0(x,y) = p_{00}(x) \neq 0$, the number of roots of $p_{00}(x) = 0$ is finite. 'The finite set E' in the condition of the lemma is the set of the roots of this equation. The condition $C \notin E$ implies $p_{00}(C) \neq 0$, which contradicts what we have shown above.

Thus we have proved that $P(C, f(v), z)$ is a non-constant polynomial in z of degree not more than N and that there are at most N solutions of the equation (14.8) in z.

As we have already mentioned, this shows that 'for any $v \in U_0$ there are duplicated values among the $N+1$ numbers $f(v+a_n)$ $(n = 0, \dots, N)$'. So, the product

$$F(v) := \prod_{0 \leqq m < n \leqq N} \bigl(f(v+a_m) - f(v+a_n)\bigr)$$

vanishes for any $v \in U_0$, as a certain factor vanishes. Applying Theorem A.8 to the holomorphic function $F(v)$, we have an identity

$$f(v + a_{m_0}) = f(v + a_{m_1})$$

for certain m_0 and m_1. A priori this holds for $v \in U_0$, but thanks to the identity theorem (Theorem A.7) it holds for arbitrary v. Replacing v with $u - a_{m_1}$, we have

$$f(u + a_{m_0} - a_{m_1}) = f(u)$$

identically, which means that $a_{m_0} - a_{m_1} (\neq 0)$ is a period of f.

This is the end of the proof of Lemma 14.8. □

Thus we have proved that $f(u)$ is a function with at least one period, say, Ω. Obviously $\tilde{f}(u) := f(\Omega u)$ has period 1 and the same algebraic addition formula (14.5) as f. If $\tilde{f}(u)$ is a rational function of $e^{2\pi i u}$, then f is a rational function of $e^{2\pi i \Omega^{-1} u}$ and if $\tilde{f}(u)$ is an elliptic function of u, then f is an elliptic function of u. Replacing f by $\tilde{f}$, we can assume that f has period 1 from the beginning. Our goal now is to show this f is a rational function of $e^{2\pi i u}$ or an elliptic function of u.

Let us use the periodicity of $f(u)$ to define a single valued function on $\mathbb{C} \setminus \{0\}$,

$$g(v) := f\left(\frac{\log v}{2\pi i}\right). \tag{14.10}$$

Of course the logarithm $\log v$ is a multi-valued function, but the multi-valuedness has the form, $\log v \mapsto \log v + 2\pi i \times$ (integer). Putting this in the definition (14.10) of g, we have $f\left(\dfrac{\log v}{2\pi i} + \text{(integer)}\right)$, which is equal to $f\left(\dfrac{\log v}{2\pi i}\right)$ by virtue of the periodicity of f. Hence the value of $g(v)$ is determined uniquely.

By the definition (14.10) $g(v)$ is a meromorphic function on $\mathbb{C} \setminus \{0\}$. If both $v = 0$ and $v = \infty$ are poles or removable singularities of $g(v)$, Lemma 14.2 implies that $g(v)$ is a rational function. This means that $f(u)$ is a rational function of $v = e^{2\pi i u}$, which proves half of our goal. It remains to show that, if either $v = 0$ or $v = \infty$ is an essential singularity, $f(u)$ has the second period.

Let us assume that $v = 0$ is an essential singularity of $g(v)$. (In the case when $v = \infty$ is an essential singularity, too, the following argument can be applied.) Then, similarly to (14.6), we can take β_n $(n = 0, \dots, N)$ such that

$$g(\beta_n) = C' \quad (n = 0, \dots, N), \qquad \beta_m \neq \beta_n \quad (m \neq n) \tag{14.11}$$

by Theorem 14.5. (We denote c' in Theorem 14.5 by C'.) Moreover we may assume that

$$|\beta_0| > |\beta_1| > \cdots > |\beta_N|, \tag{14.12}$$

because we take the complex numbers $\beta_0, \dots, \beta_N$ from a sequence $\{z_n\}$ which converges to the essential singularity $v = 0$.

In terms of the function $f(u)$, the condition (14.11) means

$$f(b_n) = C' \quad (n = 0, \dots, N), \qquad b_m \neq b_n \quad (m \neq n), \tag{14.13}$$

where $b_n := \dfrac{\log \beta_n}{2\pi i}$. The condition (14.12) means

$$\operatorname{Im} b_0 < \operatorname{Im} b_1 < \cdots < \operatorname{Im} b_N, \tag{14.14}$$

as $b_n = \dfrac{\log \beta_n}{2\pi i} = \dfrac{\arg \beta_n}{2\pi} - i \dfrac{\log |\beta_n|}{2\pi}$.

The condition (14.13) is essentially the same as (14.6). Hence we can apply Lemma 14.8 to show that $f(u)$ has a period of the form $b_{m_0} - b_{m_1}$. This period has non-zero imaginary part due to (14.14). Recall that the period which we obtained from (14.6) is 1 (by the scaling $u \mapsto \Omega u$). Therefore two periods, 1 and $b_{m_0} - b_{m_1}$, are linearly independent over $\mathbb{R}$. Therefore $f(u)$ is an elliptic function.

This is the end of the proof of Theorem 14.1. Thank you for your patience!

Remark 14.9 We proved the theorem for a single-valued meromorphic function, but Theorem 14.1 holds also for a multi-valued analytic function by replacing 'rational functions' by 'algebraic functions', 'rational functions of exponential functions' by 'algebraic functions of exponential functions' and 'elliptic functions' by 'algebraic functions of elliptic functions'. (See, for example, §50 of [Tkn].)

Chapter 15
Theta Functions

When we studied properties of elliptic functions in Section 11.2, we showed that 'a holomorphic elliptic function is constant'. This means that imposing both conditions, 'doubly periodic' and 'holomorphic', is too restrictive. When the condition 'holomorphic' is replaced by 'meromorphic', the fruitful theory of elliptic functions is developed, as we saw. In this chapter we loosen the condition 'doubly periodic'.

15.1 Definition of Theta Functions

Up to now we have used two general complex numbers $(\Omega_1, \Omega_2) \in \mathbb{C}^2$ linearly independent over $\mathbb{R}$ as periods of elliptic functions. Since we further use techniques like Fourier expansions, it is convenient to fix one of the periods to 1. For general periods (Ω_1, Ω_2) we can scale the independent variable as $u \mapsto \Omega_1 u$ to make the periods equal to $\left(1, \dfrac{\Omega_2}{\Omega_1}\right)$. If necessary, we replace $\dfrac{\Omega_2}{\Omega_1}$ by $-\dfrac{\Omega_2}{\Omega_1}$ and make the imaginary part of the second period positive without loss of generality: $\operatorname{Im} \dfrac{\Omega_2}{\Omega_1} > 0$. Hereafter we assume that the periods are 1 and τ ($\operatorname{Im} \tau > 0$) and we call $\Gamma = \mathbb{Z} + \mathbb{Z}\tau$ the period lattice. (What we are going to consider is *not* periodic but merely 'quasi'-periodic. So, strictly speaking, the words 'periods' and 'period lattice' are not correct. Well, let us relax and not be too strict.)

In order to escape from the spell of Liouville's theorem 'a holomorphic doubly-periodic function is constant', let us weaken the double periodicity to the following *(multiplicative) quasi-periodicity*

$$f(u+1) = f(u), \qquad f(u+\tau) = e^{au+b} f(u), \tag{15.1}$$

where a and b are fixed complex numbers.

This condition is an analogue of quasi-periodicity (12.23) of the sigma function $\sigma(u)$ considered in Section 12.3, but apparently stronger, as (15.1) requires period-

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_16

icity $f(u+1) = f(u)$ along one of the periods. In fact, this is not a serious difference. Let us suppose that a function $\tilde{f}(u)$ has the following quasi-peiodicity:

$$\tilde{f}(u+1) = e^{\alpha u+\beta}\tilde{f}(u), \qquad \tilde{f}(u+\tau) = e^{\gamma u+\delta}\tilde{f}(u), \tag{15.2}$$

instead of (15.1). (The sigma function with $\Omega_1 = 1$, $\Omega_2 = \tau$ is an example; according to (12.23) $e^{\alpha u+\beta} = -e^{\eta_1(u+\Omega_1/2)}$, $e^{\gamma u+\delta} = -e^{\eta_2(u+\Omega_2/2)}$, i.e., $\alpha = \eta_1$, $\beta = \pi i + \dfrac{\eta_1\Omega_1}{2}$, $\gamma = \eta_2$, $\delta = \pi i + \dfrac{\eta_2\Omega_2}{2}$.)

Then, it is easy to see that

$$f(u) := e^{-\mu u^2 - \nu u}\tilde{f}(u)$$

satisfies the condition (15.1), if the coefficients μ and ν are suitably chosen.

Exercise 15.1 Check this. In other words, find μ, ν such that $f(u) = e^{-\mu u^2 - \nu u}\tilde{f}(u)$ satisfies (15.1) and express a and b in terms of α, β, γ and δ. (Hint: Write down (15.1) in terms of $\tilde{f}(u)$, using (15.2).)

So, condition (15.1) is essentially the same as the more general condition (15.2), which is satisfied by $\sigma(u)$.

To estimate how many functions satisfy (15.1), let us use periodicity with respect to $u \mapsto u+1$ first. The basic of the most fundamental tool for handling periodic functions is the *Fourier expansion*,

$$f(u) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n u}. \tag{15.3}$$

Since $e^{2\pi i n u}$ ($n \in \mathbb{Z}$) has period 1 as a function of u, it is obvious that the right-hand side of (15.3) gives a periodic function, if it converges and really defines a function.

But the converse statement, "a periodic function has such an expansion", is not always true and the conditions for convergence and so on are rather complicated problems in real analysis.[1] Fortunately, we are considering only holomorphic functions on the whole complex plane. For such functions one can readily prove the existence of the Fourier expansion, using the theorem on the Laurent expansion.

The necessary technique has already been used in the previous chapter. As was explained after (14.10), a function $g(v) := f\left(\dfrac{\log v}{2\pi i}\right)$ constructed from $f(u)$ is a *single-valued* holomorphic function on $\mathbb{C} \setminus \{0\}$ because of the periodicity $f(u+1) = f(u)$ of f. The function $g(v)$ has an isolated singularity at $v = 0$, because $\log v$ is not defined at $v = 0$. So, it is expanded into a Laurent series as

$$g(v) = \sum_{n\in\mathbb{Z}} a_n v^n.$$

[1] One can refer any book on 'Fourier analysis'.

Setting $v = e^{2\pi i u}$, we obtain an expansion of $f(u)$ of the form (15.3).

Now we take the quasi-periodicity along the direction $u \mapsto u + \tau$ into account. Up to now, we have not imposed any restriction on the constants a and b which determine the quasi-periodicity, (15.1). But, if we choose a arbitrarily, in most cases only trivial functions remain. To see this, let us compute $f(u+1+\tau)$ in two ways. When we consider the argument as $u+1+\tau = (u+\tau)+1$, the periodicity and the quasi-periodicity imply

$$f(u+1+\tau) = f(u+\tau) = e^{au+b} f(u).$$

On the other hand, if we take the argument $u+1+\tau$ as $(u+1)+\tau$, then

$$f(u+1+\tau) = e^{a(u+1)+b} f(u+1) = e^{au+a+b} f(u).$$

Comparing two computations, we have

$$e^{au+b} f(u) = e^{au+a+b} f(u).$$

This holds for non-zero $f(u)$ only when $e^a = 1$, i.e., $a = 2\pi i k$ ($k \in \mathbb{Z}$). Thus we have to refine the quasi-periodicity condition (15.1) as follows:

$$f(u+1) = f(u), \qquad f(u+\tau) = e^{2\pi i k u + b} f(u). \tag{15.4}$$

The next step is to determine the coefficients a_n of the Fourier expansion (15.3) by means of the quasi-periodicity (15.4). Substituting the Fourier expansion into each side of the quasi-periodicity condition $f(u+\tau) = e^{2\pi i k u + b} f(u)$, we have

$$\begin{aligned} f(u+\tau) &= \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n \tau} e^{2\pi i n u}, \\ e^{2\pi i k u + b} f(u) &= \sum_{n\in\mathbb{Z}} a_n e^{2\pi i k u + b} e^{2\pi i n u} = \sum_{n\in\mathbb{Z}} e^b a_n \, e^{2\pi i (n+k) u}. \end{aligned}$$

The coefficients of $e^{2\pi i n u}$ should be equal: $a_n e^{2\pi i n \tau} = e^b a_{n-k}$. This yields a recurrence relation for a_n:

$$a_n = e^{-2\pi i n \tau + b} a_{n-k}. \tag{15.5}$$

The case $k = 0$ is easy, so we leave it to the reader as an exercise.

Exercise 15.2 Show that there exist $\alpha \in \mathbb{C}$ and $m \in \mathbb{Z}$ such that $f(u) = \alpha e^{2\pi i m u}$, when $k = 0$.

The case $k > 0$:

For any $n \in \mathbb{Z}$ there exist $m \in \mathbb{Z}$ and $n_0 \in \mathbb{Z}$, $0 \le n_0 < k$, such that $n = km + n_0$. For the moment we assume $n \ge 0$. (Then, $m \ge 0$.) Repeated use of the recurrence relation (15.5) gives

$$
\begin{aligned}
a_n &= e^{-2\pi i n\tau+b}\, a_{n-k} \\
&= e^{-2\pi i n\tau+b} e^{-2\pi i (n-k)\tau+b}\, a_{n-2k} = \cdots \\
&= e^{-2\pi i \left(n+(n-k)+\cdots+(k+n_0)\right)\tau+mb}\, a_{n_0} \\
&= e^{-\pi i m(m+1)k\tau-2\pi i m n_0\tau+mb}\, a_{n_0}.
\end{aligned} \tag{15.6}
$$

(The last line is due to the formula for arithmetic series.) It is easy to check that this equation holds also for $n < 0$, using the recurrence formula $a_n = e^{2\pi i n\tau+2\pi i k\tau-b} a_{n+k}$ obtained from (15.5). The equation (15.6) says, 'If we fix k coefficients $a_0, \dots, a_{k-1}$ (i.e., a_{n_0} in the above notation), the other coefficients are determined automatically.' "Thus we have obtained a holomorphic function $f(u)$ on $\mathbb{C}$ with k parameters." Really? We defined $f(u)$ as an infinite series. Without proof of convergence it is meaningless. Does a Fourier series with coefficients a_n determined by (15.6) express a holomorphic function?

Here let us call the function $g(v) = \sum_{n\in\mathbb{Z}} a_n v^n = f\left(\dfrac{\log v}{2\pi i}\right)$ back on stage, which was used in the proof of the existence of Fourier expansions of periodic holomorphic functions. This function is holomorphic on $\mathbb{C} \smallsetminus \{0\}$. By the formula for the coefficients of the Laurent expansion (derived from the Cauchy integral formula) we can express a_n as

$$
a_n = \frac{1}{2\pi i} \int_{|v|=R} \frac{g(v)}{v^{n+1}}\, dv.
$$

We can take any positive real number as R, since $g(v)$ is holomorphic everywhere on $\mathbb{C}$ except at $v = 0$. Let M_R be the maximum value of $|g(v)|$ on the circle with centre 0 and radius R: $M_R := \max_{|v|=R} |g(v)|$. We can estimate $|a_n|$ by the above integral expression, which is used quite often in this way in complex analysis:

$$
\begin{aligned}
|a_n| &\le \frac{1}{2\pi} \int_{|v|=R} \frac{|g(v)|}{|v|^{n+1}}\, |dv| \\
&\le \frac{1}{2\pi} \int_{|v|=R} \frac{M_R}{R^{n+1}}\, |dv| = \frac{M_R}{R^n} \xrightarrow{n\to+\infty} 0.
\end{aligned} \tag{15.7}
$$

Hence a_n converges to 0 when n tends to $+\infty$. On the other hand, it follows from the expression (15.6) that

$$
\begin{aligned}
|a_{km+n_0}| &= \left| e^{-\pi i m(m+1)k\tau-2\pi i m n_0\tau+mb}\, a_{n_0} \right| \\
&= e^{m(\operatorname{Re} b+\pi(k+2n_0)\operatorname{Im}\tau)+m^2\pi k \operatorname{Im}\tau} |a_{n_0}|.
\end{aligned}
$$

Here the dominant factor in this expression for $m \to \infty$ is $e^{m^2\pi k \operatorname{Im}\tau}$. Since $e^{\pi k \operatorname{Im}\tau} > 1$, the factor $e^{m^2\pi k \operatorname{Im}\tau}$ diverges very rapidly. As a result, if $a_{n_0} \neq 0$, a_{km+n_0} diverges when $m \to +\infty$, which contradicts $a_n \to 0$.

This implies that there is not a non-zero function $f(u)$ satisfying the conditions (15.4), when $k > 0$.

The case $k = -1$:
As in the case $k > 0$, the recurrence relation (15.5) implies

$$a_n = a_0 e^{\pi i n(n-1)\tau - nb}$$

for any $n \in \mathbb{Z}$. Thus $f(u)$ has the form

$$f(u) = a_0 \sum_{n \in \mathbb{Z}} e^{\pi i n^2 \tau + 2\pi i n(u - b/2\pi i - \tau/2)}.$$

The contents in the parentheses after $2\pi i n$ in the right-hand side are nothing more than a shift of the argument u. Replacing that part with u and the overall constant a_0 with 1, we define the following function.

Definition 15.3 The function

$$\theta(u, \tau) := \sum_{n \in \mathbb{Z}} e^{\pi i n^2 \tau + 2\pi i n u} \tag{15.8}$$

is called *Jacobi's elliptic theta function*.

Using this notation, the desired function $f(u)$ is expressed as

$$f(u) = a_0 \theta\left(u - \frac{b}{2\pi i} - \frac{\tau}{2}, \tau\right).$$

Uh, yes, if it really converges and expresses a holomorphic function. Doesn't this series diverge, as it did when $k > 0$? Fortunately, in this case everything goes well. The following lemma guarantees that the series (15.8) defines a holomorphic function.

Lemma 15.4 *The series* (15.8) *converges uniformly and absolutely on*

$$U_{C,\varepsilon} := \{(u, \tau) \mid |\operatorname{Im} u| \le C,\ \operatorname{Im} \tau \ge \varepsilon\}$$

for any positive C and ε.

Therefore by virtue of Weierstrass's theorem (Theorem A.9) $\theta(u, \tau)$ is

- holomorphic in u on the whole plane $\mathbb{C}$ for a fixed τ;
- holomorphic in τ on the upper half plane $\mathsf{H} := \{\tau \mid \operatorname{Im} \tau > 0\}$ for a fixed u.

In fact, for any compact subset $K \subset \mathbb{C}$ and a fixed τ there exist large $C > 0$ and small $\varepsilon > 0$ such that $K \times \{\tau\} \subset U_{C,\varepsilon}$. Hence from Lemma 15.4 it follows that the series converges uniformly and absolutely on K as a function of u. Thus the conditions of Theorem A.9 are satisfied and $\theta(u, \tau)$ is holomorphic in u. Holomorphicity in τ is proved similarly.

Proof (of Lemma* 15.4*) Each term of the series (15.8) is estimated from above as follows:

$$\left|e^{\pi i n^2\tau+2\pi i n u}\right| = e^{-\pi n^2 \operatorname{Im}\tau - 2\pi n \operatorname{Im} u}$$
$$\le e^{-\pi n^2\varepsilon + 2\pi|n|C}.$$

The exponent $-\pi n^2\varepsilon + 2\pi|n|C = -\pi|n|^2\varepsilon + 2\pi|n|C$ of the last exponential function in this inequality is a quadratic function of $|n|$ with negative coefficient $-\pi\varepsilon$ of the quadratic term. It becomes less than a linear function $-|n|$ of $|n|$ when $|n|$ is sufficiently large: $-\pi|n|^2\varepsilon + 2\pi|n|C < -|n|$. Hence, except for a finite number of n, we have an inequality $e^{-\pi n^2\varepsilon+2\pi|n|C} < e^{-|n|}$. The series with terms in the right-hand side of this inequality,

$$\sum_{n\in\mathbb{Z}} e^{-|n|} = 1 + 2\sum_{n=1}^{\infty} e^{-n}, \tag{15.9}$$

is a convergent series (1+ a convergent geometric series) and does not depend on the variable u. Since the absolute values of the terms of the series (15.8) are less than the terms of the series (15.9) (except for a finite number of terms), it converges uniformly and absolutely on $U_{C,\varepsilon}$ by Weierstrass's M-test. □

By our discussion and construction it is obvious that the theta function $\theta(u,\tau)$ has quasi-periodicity in u:

$$\begin{aligned} \theta(u+1,\tau) &= \theta(u,\tau), \\ \theta(u+\tau,\tau) &= e^{-\pi i\tau - 2\pi i u}\theta(u,\tau). \end{aligned} \tag{15.10}$$

For an arbitrary element $m + n\tau \in \Gamma$ ($m, n \in \mathbb{Z}$) of the period lattice, it transforms as

$$\theta(u+m+n\tau,\tau) = e^{-\pi i n^2\tau - 2\pi i n u}\,\theta(u,\tau),$$

which is immediately proved by (15.10).

We have found a function $f(u)$ satisfying (15.4) for $k = -1$. For the remaining cases, $k < -1$, a function $f(u)$ satisfying (15.4) is expressed in terms of $\theta(u,\tau)$. This is an exercise for the reader.

Exercise 15.5 Fix a positive integer k. Show that the space $\Theta_{k,b}$ of entire functions (functions holomorphic on the whole $\mathbb{C}$) with quasi-periodicity

$$f(u+1) = f(u), \quad f(u+\tau) = e^{-2\pi i k u + b} f(u)$$

is a k-dimensional linear space over $\mathbb{C}$ and construct its basis, using theta functions. (Hint: Repeat the discussion with the recurrence relation (15.5); all coefficients of the Fourier expansion are determined by $a_0, \dots, a_{k-1}$.)

In the construction of Jacobi's elliptic functions, for example, we shall use variants of theta functions, *theta functions with characteristics*. They are associated to real numbers $a, b \in \mathbb{R}$ called *characteristics* (usually $a, b \in \mathbb{Q}$ and, in this book, a and b are 0 or $\frac{1}{2}$) and defined by

$$\theta_{a,b}(u,\tau) := \sum_{n\in\mathbb{Z}} e^{\pi i (n+a)^2\tau + 2\pi i (n+a)(u+b)}. \tag{15.11}$$

(This is nothing more than shifting of the summation index n of the series by a and the variable u by b.)

In the same way as we did with the theta function $\theta(u,\tau)$ (without characteristics) we can show that the theta functions with characteristics is holomorphic on the whole $\mathbb{C}$ in u and on the upper half plane H in τ.

The following properties are directly derived from the definition (15.11):

- $\theta_{0,0}(u,\tau) = \theta(u,\tau)$.
- $\theta_{a,b+b'}(u,\tau) = \theta_{a,b}(u+b',\tau)$.
- $\theta_{a+a',b}(u,\tau) = e^{\pi i a'^2\tau + 2\pi i a'(u+b)}\theta_{a,b}(u+a'\tau,\tau)$.
- For integers p, q, $\theta_{a+p,b+q}(u,\tau) = e^{2\pi i a q}\theta_{a,b}(u,\tau)$.

Hereafter we use only the characteristics $a, b \in \left\{0, \frac{1}{2}\right\}$, so let us introduce a simplified notation,

$$\begin{aligned}\theta_{kl}(u,\tau) &:= \theta_{k/2,l/2}(u,\tau)\\ &= \sum_{n\in\mathbb{Z}} e^{\pi i\left(n+\frac{k}{2}\right)^2\tau + 2\pi i\left(n+\frac{k}{2}\right)\left(u+\frac{l}{2}\right)},\end{aligned} \tag{15.12}$$

where $k, l \in \{0, 1\}$. They are related to $\theta(u,\tau)$ defined by (15.8) as follows:

$$\begin{aligned}&\theta_{00}(u,\tau) = \theta(u,\tau), \qquad \theta_{01}(u,\tau) = \theta\left(u+\frac{1}{2},\tau\right),\\ &\theta_{10}(u,\tau) = e^{\pi i\tau/4+\pi i u}\theta\left(u+\frac{\tau}{2},\tau\right),\\ &\theta_{11}(u,\tau) = i e^{\pi i\tau/4+\pi i u}\theta\left(u+\frac{1+\tau}{2},\tau\right).\end{aligned} \tag{15.13}$$

Remark 15.6 The notations introduced above are those of D. Mumford, [M]. (Almost the same notations were introduced by H. Weber, (pp. 51–52 of [Wb]), in which the sign of $\theta_{11}(u,\tau)$ differs from ours.) Jacobi's notations $\theta_1(u,\tau)$, $\theta_2(u,\tau)$, $\theta_3(u,\tau)$ and $\theta_4(u,\tau)$ are probably used more often.[2] They correspond to the above introduced theta functions with characteristics, $-\theta_{11}\left(\frac{u}{\pi},\tau\right)$, $\theta_{10}\left(\frac{u}{\pi},\tau\right)$, $\theta_{00}\left(\frac{u}{\pi},\tau\right)$, $\theta_{01}\left(\frac{u}{\pi},\tau\right)$, respectively.

There are also other notations and other variants, so you should pay attention to definitions when you read the literature.

---

[2] Jacobi's original notations in [Ja2] are $\theta_j(u,q)$ $(j = 1,2,3,\ q = e^{\pi i\tau})$ and $\theta(u,q)$ instead of θ_4.

15.2 Properties of Theta Functions

Theta functions possess many good properties. In this section we consider properties of theta functions which are directly derived from the definitions. Most of the results in this section are for fixed $\tau \in \mathsf{H}$. So, for simplicity, we denote $\theta(u,\tau)$ by $\theta(u)$ except in Section 15.2.3, Exercise 15.10 and Exercise 15.11.

15.2.1 Quasi-periodicity

The following quasi-periodicities of $\theta_{kl}(u)$ are easily derived from those of $\theta(u)$, (15.10), and expressions of $\theta_{kl}(u)$ by $\theta(u)$, (15.13).

Quasi-periodicities with respect to $u \mapsto u+1$:

(15.14)
$$\begin{aligned} &\theta_{00}(u+1) = \theta_{00}(u), && \theta_{01}(u+1) = \theta_{01}(u), \\ &\theta_{10}(u+1) = -\theta_{10}(u), && \theta_{11}(u+1) = -\theta_{11}(u). \end{aligned}$$

Quasi-periodicities with respect to $u \mapsto u+\tau$:

(15.15)
$$\begin{aligned} \theta_{00}(u+\tau) &= e^{-\pi i\tau - 2\pi i u}\theta_{00}(u), \\ \theta_{01}(u+\tau) &= -e^{-\pi i\tau - 2\pi i u}\theta_{01}(u), \\ \theta_{10}(u+\tau) &= e^{-\pi i\tau - 2\pi i u}\theta_{10}(u), \\ \theta_{11}(u+\tau) &= -e^{-\pi i\tau - 2\pi i u}\theta_{11}(u). \end{aligned}$$

Transformation rules of characteristics (quasi-periodicities with respect to half periods):

Shifting u by half periods $\frac{1}{2}, \frac{\tau}{2}$, we obtain theta functions with different characteristics. To write down formulae, it is convenient to regard characteristics values 0 and 1 as elements of $\{0,1\} = \mathbb{Z}/2\mathbb{Z}$ so that '$1+1=0$'. In this notation, we have

(15.16)
$$\begin{aligned} \theta_{kl}\left(u+\frac{1}{2}\right) &= (-1)^{kl}\theta_{k,l+1}(u), \\ \theta_{kl}\left(u+\frac{\tau}{2}\right) &= (-i)^{l} e^{-\pi i\tau/4 - \pi i u}\theta_{k+1,l}(u). \end{aligned}$$

These formulae are directly checked by the definitions (15.11).

15.2.2 Parity

Substituting u with $-u$ in the definition of $\theta(u)$ and changing the summation index to $n' = -n$, we have

$$\theta(-u) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2\tau - 2\pi i n u}$$
$$= \sum_{n'\in\mathbb{Z}} e^{\pi i n'^2\tau + 2\pi i n' u} = \theta(u).$$

Hence $\theta(u)$ is an even function. Therefore from (15.16), (15.14) and (15.15) it follows that

$$\theta_{kl}(u) \text{ is } \begin{cases} \text{an even function} & (k,l) \neq (1,1), \\ \text{an odd function} & (k,l) = (1,1). \end{cases} \tag{15.17}$$

15.2.3 Heat equations

The uniform and absolute convergence of the series (15.12) on a compact subset of $\mathbb{C}$ is shown in the same way as Lemma 15.4. Weierstrass's theorem (Theorem A.9) guarantees termwise differentiability of such series. By differentiating each term of (15.12) twice by u or once by τ, we obtain

$$\frac{\partial^2}{\partial u^2} e^{\pi i\left(n+\frac{k}{2}\right)^2\tau + 2\pi i\left(n+\frac{k}{2}\right)\left(u+\frac{l}{2}\right)} = -4\pi^2\left(n+\frac{k}{2}\right)^2 e^{\pi i\left(n+\frac{k}{2}\right)^2\tau + 2\pi i\left(n+\frac{k}{2}\right)\left(u+\frac{l}{2}\right)},$$
$$\frac{\partial}{\partial \tau} e^{\pi i\left(n+\frac{k}{2}\right)^2\tau + 2\pi i\left(n+\frac{k}{2}\right)\left(u+\frac{l}{2}\right)} = \pi i\left(n+\frac{k}{2}\right)^2 e^{\pi i\left(n+\frac{k}{2}\right)^2\tau + 2\pi i\left(n+\frac{k}{2}\right)\left(u+\frac{l}{2}\right)}.$$

Therefore $\theta_{kl}(u,\tau)$ satisfies a second-order partial differential equation of the form

$$\frac{\partial}{\partial \tau}\theta_{kl}(u,\tau) = \frac{1}{4\pi i}\frac{\partial^2}{\partial u^2}\theta_{kl}(u,\tau). \tag{15.18}$$

Let us restrict the domain of variables (u,τ) to $\mathbb{R}\times i\mathbb{R}_{>0}$ ($\mathbb{R}_{>0} := \{t>0\} \subset \mathbb{R}$) and replace the variables by $x = u \in \mathbb{R}$ and $t = \dfrac{\tau}{i} \in \mathbb{R}_{>0}$. Then the equation (15.18) is rewritten in the real form,

$$\frac{\partial}{\partial t}\theta_{kl}(x,it) = \frac{1}{4\pi}\frac{\partial^2}{\partial x^2}\theta_{kl}(x,it).$$

This is the *heat equation* in physics.[3]

[3] In the above equation the space is one-dimensional and the thermal conductivity is fixed to a special value.

Theta functions play an important role in the theory of heat equations. (They are fundamental solutions under the periodic boundary condition.) We do not go into details here.

15.2.4 Zeros

Let us find zeros of theta functions. Quasi-periodicities (15.14) and (15.15) have the form 'shifting u by 1 or τ is equivalent to multiplying $\theta_{kl}(u)$ by non-zero factors'. Hence we obtain all zeros in $\mathbb{C}$ by shifting zeros in one period parallelogram with vertices a, $a+1$, $a+1+\tau$ and $a+\tau$ for some $a \in \mathbb{C}$.

First we prove the following lemma.

Lemma 15.7 *Each $\theta_{kl}(u)$ has one and only one zero in a period parallelogram. Here we assume that there are no zeros on the sides of the parallelogram.*

Proof As all theta functions with characteristics $\theta_{kl}(u)$ have the form (non-vanishing function) $\times\theta(u+\text{shift})$, their zeros are those of $\theta(u)$ with shifts. Hence it is enough to prove the lemma for $\theta(u)$.

We denote the period parallelogram under consideration by Π (Fig. 15.1) and assume that $\theta(u) \neq 0$ on its boundary.

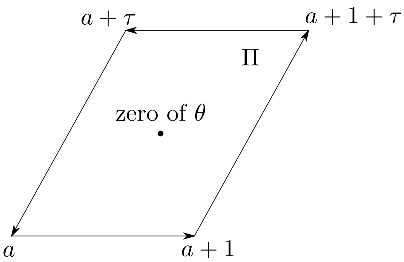

Fig. 15.1 A period parallelogram Π of the theta function.

Then, by the argument principle (Theorem A.10) the number of zeros of $\theta(u)$ in Π is equal to the integral

(15.19)
$$\frac{1}{2\pi i}\int_{\partial\Pi}\frac{\theta'(u)}{\theta(u)}\,du = \frac{1}{2\pi i}\int_{\partial\Pi}\frac{d}{du}\log\theta(u)\,du.$$

(The orientation of the boundary $\partial\Pi$ is as shown in Fig. 15.1.)

Let us compute the integral, decomposing it as

(15.20)
$$\int_{\partial\Pi} = \int_a^{a+1} + \int_{a+1}^{a+1+\tau} + \int_{a+1+\tau}^{a+\tau} + \int_{a+\tau}^{a}.$$

The idea is the same as the proof of Theorem 11.10: the integrals on the opposite sides 'almost' cancel. In fact, for the elliptic functions considered in Theorem 11.10 the integrals completely cancel, but in the case of theta functions they do not.

How does the integrand $\dfrac{d}{du}\log\theta(u)$ of (15.19) change by the shifts $u \mapsto u+1$, $u \mapsto u+\tau$ of the variable? The logarithms of the quasi-periodicity conditions (15.10) are

$$\begin{aligned}\log\theta(u+1) &= \log\theta(u),\\ \log\theta(u+\tau) &= \log\theta(u) - \pi i\tau - 2\pi i u.\end{aligned}$$

Their derivatives are

$$\begin{aligned}\frac{d}{du}\log\theta(u+1) &= \frac{d}{du}\log\theta(u),\\ \frac{d}{du}\log\theta(u+\tau) &= \frac{d}{du}\log\theta(u) - 2\pi i.\end{aligned} \tag{15.21}$$

Hence the number of zeros, (15.19), is computed as

$$\begin{aligned}\frac{1}{2\pi i}\int_{\partial\Pi}\frac{d}{du}\log\theta(u)\,du &= \frac{1}{2\pi i}\int_a^{a+1}\left(\frac{d}{du}\log\theta(u) - \frac{d}{du}\log\theta(u+\tau)\right)du\\ &\quad + \frac{1}{2\pi i}\int_{a+\tau}^{a}\left(\frac{d}{du}\log\theta(u) - \frac{d}{du}\log\theta(u+1)\right)du\\ &= \frac{1}{2\pi i}\int_a^{a+1}(2\pi i)\,du = 1.\end{aligned}$$

This proves the lemma. □

We have found the number of zeros of $\theta_{kl}(u)$. Their locations are also easily found. As $\theta_{11}(u)$ is an odd function, it vanishes at the origin: $\theta_{11}(0) = 0$. From this fact, quasi-periodicities (15.14), (15.15) and Lemma 15.7 it follows that (the set of zeros of $\theta_{11}(u)) = \Gamma = \mathbb{Z} + \mathbb{Z}\tau$.

The remaining theta functions $\theta_{kl}(u)$ are expressed in terms of $\theta_{11}(u)$ as

$$\begin{aligned}\theta_{10}(u) &= -\theta_{11}\left(u+\frac{1}{2}\right),\\ \theta_{01}(u) &= (\text{non-vanishing function}) \times \theta_{11}\left(u+\frac{\tau}{2}\right),\\ \theta_{00}(u) &= (\text{non-vanishing function}) \times \theta_{11}\left(u+\frac{1+\tau}{2}\right),\end{aligned}$$

due to (15.16). Hence their zeros are located as

$$
\begin{aligned}
&\theta_{00}(u) = 0 \Leftrightarrow u \in \Gamma + \frac{1+\tau}{2},\\
&\theta_{01}(u) = 0 \Leftrightarrow u \in \Gamma + \frac{\tau}{2},\\
&\theta_{10}(u) = 0 \Leftrightarrow u \in \Gamma + \frac{1}{2},\\
&\theta_{11}(u) = 0 \Leftrightarrow u \in \Gamma.
\end{aligned}
\tag{15.22}
$$

(See Fig. 15.2.) By Lemma 15.7 they are all zeros of the first order.

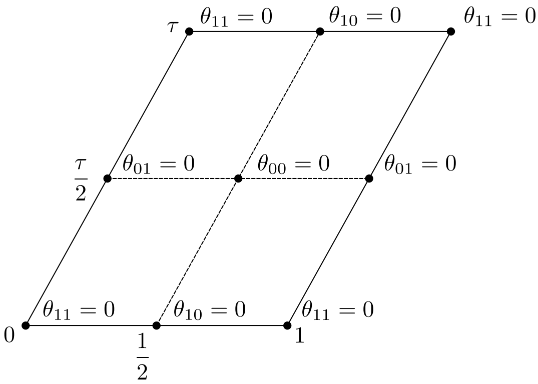

Fig. 15.2 Positions of zeros of the theta function with characteristics.

In Section 15.1 we introduced theta functions, saying ‘as there are no holomorphic doubly-periodic non-constant functions, we have to consider *quasi-periodic* functions’. But, theta functions would be dissatisfied with this description of them as a “compromise”. In fact, theta functions are fundamental building blocks of elliptic functions. Elliptic functions are written as ratios of products of theta functions, just as rational functions are written as ratios of polynomials, $R(x) = \dfrac{P(x)}{Q(x)}$. The exact statement is as follows:

Theorem 15.8 *Let a_i and b_i $(i = 1, \dots, N)$ be sets of complex numbers satisfying $\sum_{i=1}^{N} a_i = \sum_{i=1}^{N} b_i$ and c be an arbitrary complex number. The following function is an elliptic function with periods 1 and τ:*

$$
f(u) = c\, \frac{\theta_{11}(u-a_1)\cdots\theta_{11}(u-a_N)}{\theta_{11}(u-b_1)\cdots\theta_{11}(u-b_N)}. \tag{15.23}
$$

Conversely, any elliptic function with periods 1 and τ is expressed as above with suitable a_i and b_i $(i = 1, \dots, N)$.

The proof of Theorem 15.8 is an easy application of properties of theta functions and Liouville's theorems in Section 11.2. We leave the details to the reader as an exercise with detailed hints.

Exercise 15.9 Prove Theorem 15.8. (Hint: The first half is a consequence of direct computation using properties of θ_{11} (quasi-periodicity and positions of zeros). For the latter half we use Liouville's theorems as follows: according to Theorem 11.14 there are the same numbers of zeros and poles in a period parallelogram, which we name a_i and b_i ($i = 1,\dots,N$) respectively. Show that we may assume $\sum_{i=1}^N a_i = \sum_{i=1}^N b_i$ with due modification of a_i's and b_i's by periods $m + n\tau$ (Theorem 11.15). Using those a_i's and b_i's define the ratio of products of theta functions as above. Show that its ratio with the original elliptic function is a constant by Theorem 11.9.)

We can use other theta functions with characteristics instead of θ_{11}. The proof is similar.

Usually equalities among theta functions are proved by such comparison of quasi-periodicities, positions of zeros and positions of poles. As an example, try to show the following formulae, which express values of theta functions with τ doubled by means of values of theta functions with the original τ.

Exercise 15.10 Show the following formulae (*Landen's transformations*).

$$\theta_{01}(2u,2\tau) = \frac{\theta_{01}(0,2\tau)}{\theta_{01}(0,\tau)\theta_{00}(0,\tau)}\theta_{00}(u,\tau)\theta_{01}(u,\tau), \tag{15.24}$$

$$\theta_{11}(2u,2\tau) = \frac{\theta_{01}(0,2\tau)}{\theta_{01}(0,\tau)\theta_{00}(0,\tau)}\theta_{10}(u,\tau)\theta_{11}(u,\tau). \tag{15.25}$$

(Hint: Compare zeros and quasi-periodicities of both sides to show that their ratio is constant by Theorem 11.9. The proportional constant in the first equation (15.24) can be found by setting $u = 0$. The second equation (15.25) is obtained by shifting u to $u + \dfrac{\tau}{2}$.)

Another important formula of this kind is the following relation of the sigma function and $\theta_{11}(u,\tau)$. In fact, both of them have simple zeros at each point of Γ and are multiplicatively quasi-periodic. So it is natural to expect a relation between them.

Exercise 15.11 Let Ω_1 and Ω_2 be the periods defined in Section 12.3: They are linearly independent over $\mathbb{R}$ and $\operatorname{Im}\dfrac{\Omega_2}{\Omega_1} > 0$. Prove the relation

$$\sigma(u) = \frac{\Omega_1}{\theta_{11}'(0,\tau)} e^{-\Omega_1\eta_1 z^2/2}\,\theta_{11}(z,\tau), \qquad z = \frac{u}{\Omega_1},\ \tau = \frac{\Omega_2}{\Omega_1}, \tag{15.26}$$

between $\sigma(u)$ defined by (12.18) and $\theta_{11}(z,\tau)$ in the following way:

(i) Show that $\varphi(z) := \exp\left(-\frac{\Omega_1\eta_1 z^2}{2}\right)\sigma(\Omega_1 z)$ has the same quasi-periodicity (15.14), (15.15) as $\theta_{11}(z,\tau)$: $\varphi(z+1) = -\varphi(z)$, $\varphi(z+\tau) = -e^{-\pi i\tau - 2\pi i z}\varphi(u)$. (Hint: The first one is a direct consequence of (12.23). The second one is due to (12.23) and (12.16).)

(ii) Show that $C := \dfrac{\varphi(z)}{\theta_{11}(z,\tau)}$ is a constant. (Hint: Show that this ratio is a doubly periodic holomorphic function and apply Theorem 11.9.)

(iii) Find C explicitly. (Hint: Compute $\lim\limits_{z\to 0}\dfrac{\varphi(z)}{\theta_{11}(z,\tau)}$ by l'Hôpital's rule or by Taylor expansions of $\varphi(z)$ and $\theta_{11}(z,\tau)$ at $z = 0$.)

This relation (15.26) shows that the sigma function and $\theta_{11}(u,\tau)$ is almost the same. For example, we can replace $\theta_{11}(u)$ in Theorem 15.8 with $\sigma(u)$ as in the following corollary.

Corollary 15.12 *Let Ω_1 and Ω_2 be as in Exercise* 15.11*, a_i and b_i ($i = 1,\dots,N$) be sets of complex numbers satisfying $\sum_{i=1}^{N} a_i = \sum_{i=1}^{N} b_i$ and c be an arbitrary complex number. The following function is an elliptic function with periods Ω_1 and Ω_2:*

$$f(u) = c\,\frac{\sigma(u-a_1)\cdots\sigma(u-a_N)}{\sigma(u-b_1)\cdots\sigma(u-b_N)}. \tag{15.27}$$

Conversely, any elliptic function with periods Ω_1 and Ω_2 is expressed as above with suitable a_i and b_i ($i = 1,\dots,N$).

Exercise 15.13 Show Corollary 15.12. (Hint: Apply the same argument as in Exercise 15.9 or rewrite the expression (15.23) by (15.26).)

15.3 Jacobi's Theta Relations

Jacobi's theta relations[4] are a sort of addition theorem for theta functions.

Let us introduce the following matrix.

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \tag{15.28}$$

It is a symmetric matrix ${}^tA = A$ and its square is equal to the identity matrix: $A^2 = \mathrm{Id}_4$. Hence it is an orthogonal matrix, i.e., a matrix satisfying

$${}^tA\,A = (\text{identity matrix}). \tag{15.29}$$

Jacobi's theta relations express the sum of products of four theta functions of independent variables x_j ($j = 1, \dots, 4$) as a single product of theta functions of variables y_j ($j = 1, \dots, 4$) defined by

$$\mathbf{y} := A\mathbf{x} = \frac{1}{2}\begin{pmatrix} x_1 + x_2 + x_3 + x_4 \\ x_1 + x_2 - x_3 - x_4 \\ x_1 - x_2 + x_3 - x_4 \\ x_1 - x_2 - x_3 + x_4 \end{pmatrix}, \tag{15.30}$$

where $\mathbf{x} := \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ and $\mathbf{y} := \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$.

Theorem 15.14 (Jacobi's theta relation)

$$\prod_{j=1}^{4} \theta_{00}(x_j) + \prod_{j=1}^{4} \theta_{01}(x_j) + \prod_{j=1}^{4} \theta_{10}(x_j) + \prod_{j=1}^{4} \theta_{11}(x_j) = 2\prod_{j=1}^{4} \theta_{00}(y_j). \tag{J0}$$

Proof We denote the inner product[5] of four-dimensional complex vectors $\mathbf{a}, \mathbf{b} \in \mathbb{C}^4$ by $(\mathbf{a}, \mathbf{b}) = a_1 b_1 + \cdots + a_4 b_4$. We are going to use products over $j = 1, \dots, 4$, $\prod_{j=1}^{4}$, and sums over quadruplets of integers $\mathbf{m} \in \mathbb{Z}^4$, $\sum_{\mathbf{m} \in \mathbb{Z}^4}$, many times, so we denote them simply as $\prod$ and $\sum$ respectively in this proof.

[4] Those formulae are named "Riemann's theta relations" in [M] or references citing it. However they already appeared in Jacobi's work [Ja2]. Jacobi proved them in the same way as our proof here, while they are proved by means of quasi-periodicity in §21.22 of [WW].

[5] As we do not take the complex conjugate of one of the vectors, maybe you would prefer the name 'bilinear form'.

By the definitions (15.12) of theta functions with characteristics, each term of the left-hand side of (J0) is expanded as

$$\begin{aligned}
\prod \theta_{00}(x_j) &= \sum \exp\left(\pi i\tau(\mathbf{m},\mathbf{m}) + 2\pi i(\mathbf{m},\mathbf{x})\right),\\
\prod \theta_{01}(x_j) &= \sum \exp\left(\pi i\tau(\mathbf{m},\mathbf{m}) + 2\pi i(\mathbf{m},\mathbf{x}) + \pi i(m_1+\cdots+m_4)\right),\\
\prod \theta_{10}(x_j) &= \sum \exp\left(\pi i\tau(\mathbf{m}',\mathbf{m}') + 2\pi i(\mathbf{m}',\mathbf{x})\right),\\
\prod \theta_{11}(x_j) &= \sum \exp\left(\pi i\tau(\mathbf{m}',\mathbf{m}') + 2\pi i(\mathbf{m}',\mathbf{x}) + \pi i(m_1'+\cdots+m_4')\right),
\end{aligned}$$

where $\mathbf{m}' = (m_i')_{i=1,\dots,4}$, $m_i' = m_i + \dfrac{1}{2}$.

When they are summed up as in (J0),

- terms in which $m_1+\cdots+m_4$ or $m_1'+\cdots+m_4'$ is odd cancel, as $\exp(\pi i \times (\text{odd number})) = -1$.
- terms in which $m_1+\cdots+m_4$ or $m_1'+\cdots+m_4'$ is even are doubled, as $\exp(\pi i \times (\text{even number})) = 1$.

Hence,

$$\text{(15.31)} \qquad \text{Left-hand side of (J0)} = 2{\sum}' \exp\left(\pi i\tau(\mathbf{m},\mathbf{m}) + 2\pi i(\mathbf{m},\mathbf{x})\right),$$

where the summation $\sum'$ means sum over vectors $\mathbf{m}$ in $\dfrac{1}{2}\mathbb{Z}$ satisfying the following conditions (i) or (ii):

(i) $m_j \in \mathbb{Z}$ for all j and $m_1+\cdots+m_4$ is even. (Such terms arise from θ_{00} and θ_{01}.)

(ii) $m_j \in \mathbb{Z}+\dfrac{1}{2}$ for all j and $m_1+\cdots+m_4$ is even. (Such terms arise from θ_{10} and θ_{11}.)

Note that $\mathbf{n} := A\mathbf{m}$ satisfies the following:

- $(\mathbf{m},\mathbf{m}) = (\mathbf{n},\mathbf{n})$, $(\mathbf{m},\mathbf{x}) = (\mathbf{n},\mathbf{y})$.
- $\mathbf{m}$ satisfies (i) or (ii) if and only if $\mathbf{n} \in \mathbb{Z}^4$.

(The former follows from the orthogonality (15.29) of A. The latter can be shown by direct computation from the explicit form (15.28) of A.)

Thus, rewriting the sum (15.31) as a sum over $\mathbf{n} \in \mathbb{Z}^4$, we have

$$\begin{aligned}
{\sum}' \exp\left(\pi i\tau(\mathbf{m},\mathbf{m}) + 2\pi i(\mathbf{m},\mathbf{x})\right) &= \sum_{\mathbf{n}\in\mathbb{Z}^4} \exp\left(\pi i\tau(\mathbf{n},\mathbf{n}) + 2\pi i(\mathbf{n},\mathbf{y})\right)\\
&= \prod_{j=1}^{4} \theta_{00}(y_j).
\end{aligned}$$

This equation together with (15.31) proves (J0). □

There are dozens of variants of the theta relation, among which we will use the following three.

Corollary 15.15 *Variables* y_j $(j = 1,\dots,4)$ *are defined by* (15.30) *as in Theorem* 15.14. *Then,*

$$\text{(J1)} \qquad \prod \theta_{00}(x_j) - \prod \theta_{01}(x_j) - \prod \theta_{10}(x_j) + \prod \theta_{11}(x_j) = 2\prod \theta_{11}(y_j).$$

$$\text{(J2)} \qquad \prod \theta_{00}(x_j) + \prod \theta_{01}(x_j) - \prod \theta_{10}(x_j) - \prod \theta_{11}(x_j) = 2\prod \theta_{01}(y_j).$$

$$\text{(J3)} \qquad \begin{aligned} &\theta_{00}(x_1)\theta_{01}(x_2)\theta_{10}(x_3)\theta_{11}(x_4) + \theta_{01}(x_1)\theta_{00}(x_2)\theta_{11}(x_3)\theta_{10}(x_4) \\ &+ \theta_{10}(x_1)\theta_{11}(x_2)\theta_{00}(x_3)\theta_{01}(x_4) + \theta_{11}(x_1)\theta_{10}(x_2)\theta_{01}(x_3)\theta_{00}(x_4) \\ &= 2\theta_{11}(y_1)\theta_{10}(y_2)\theta_{01}(y_3)\theta_{00}(y_4). \end{aligned}$$

Here $\prod$ *denotes the product over* $j = 1,2,3,4$.

Proof By shift of x_1 to $x_1 + 1 + \tau$ in (J0) each theta function in the left-hand side becomes $\theta_{kl}(x_1 + 1 + \tau) = (-1)^{k+l} e^{-\pi i \tau - 2\pi i x_1} \theta_{kl}(x_1)$ by virtue of quasi-periodicities (15.14) and (15.15). In the right-hand side y_j is shifted to $y_j + \dfrac{1+\tau}{2}$, which transforms $\theta_{00}(y_j)$ to $\theta_{11}(y_j)$ with an exponential factor by the transformation rule (15.16) of characteristics. Thus we obtain (J1).

The equation (J2) is obtained similarly by changing x_1 to $x_1 + 1$ in (J0).

The proof of (J3) is almost the same with following shifts of variables:

$$x_1 \mapsto x_1, \quad x_2 \mapsto x_2 + \frac{1}{2}, \quad x_3 \mapsto x_3 + \frac{\tau}{2}, \quad x_4 \mapsto x_4 + \frac{1+\tau}{2}.$$

Theta relations resemble addition theorems in the sense that they algebraically relate values of theta functions at sums of variables with their values at the original variables. In fact, we can make them look more like 'addition formulae' by specialising variables. There are many variants of addition theorems, but here we show only those which we shall use later.

For simplicity we use the notation $\theta_{kl} := \theta_{kl}(0)$. (In this notation $\theta_{11} = 0$ because of (15.22).)

Corollary 15.16 (Addition theorems of theta functions)

$$\text{(A1)} \qquad \begin{aligned} \theta_{00}(x+u)\theta_{00}(x-u)\theta_{00}^2 &= \theta_{00}(x)^2\theta_{00}(u)^2 + \theta_{11}(x)^2\theta_{11}(u)^2 \\ &= \theta_{01}(x)^2\theta_{01}(u)^2 + \theta_{10}(x)^2\theta_{10}(u)^2. \end{aligned}$$

$$\text{(A2)} \qquad \theta_{01}(x+u)\theta_{01}(x-u)\theta_{01}^2 = \theta_{01}(x)^2\theta_{01}(u)^2 - \theta_{11}(x)^2\theta_{11}(u)^2.$$

$$\text{(A3)} \qquad \begin{aligned} \theta_{11}(x+u)\theta_{01}(x-u)\theta_{10}\theta_{00} &= \theta_{00}(x)\theta_{10}(x)\theta_{01}(u)\theta_{11}(u) \\ &\quad + \theta_{01}(x)\theta_{11}(x)\theta_{00}(u)\theta_{10}(u). \end{aligned}$$

In particular, an important formula

$$\text{(A4)} \qquad \theta_{00}^4 = \theta_{01}^4 + \theta_{10}^4$$

follows from (A1), $x = u = 0$.

Proof Specialisation of variables x_j in the theta relation (J1) as $x_1 = x_2 = x, x_3 = x_4 = u$ changes the y_j's to $y_1 = x+u$, $y_2 = x-u$, $y_3 = y_4 = 0$ according to the definition (15.30). Therefore from (J1) it follows that

$$\begin{aligned}&\theta_{00}(x)^2\theta_{00}(u)^2 - \theta_{01}(x)^2\theta_{01}(u)^2 - \theta_{10}(x)^2\theta_{10}(u)^2 + \theta_{11}(x)^2\theta_{11}(u)^2 \\ &= 2\theta_{11}(x+u)\theta_{11}(x-u)\theta_{11}^2.\end{aligned}$$

As we have mentioned above, the right-hand side = 0 because $\theta_{11} = 0$. This shows the second equality in (A1).

Applying the same specialisation to (J0), we have

$$\begin{aligned}\theta_{00}(x)^2\theta_{00}(u)^2 + \theta_{01}(x)^2\theta_{01}(u)^2 + \theta_{10}(x)^2\theta_{10}(u)^2 + \theta_{11}(x)^2\theta_{11}(u)^2 \\ = 2\theta_{00}(x+u)\theta_{00}(x-u)\theta_{00}^2.\end{aligned} \tag{15.32}$$

The second equality in (A1), which we proved above, implies that the left-hand side of (15.32) becomes $2(\theta_{00}(x)^2\theta_{00}(u)^2 + \theta_{11}(x)^2\theta_{11}(u)^2)$. On the other hand, its right-hand side is twice the left-most side of (A1). Thus the first equality in (A1) has been proved.

The other addition formulae, (A2) and (A3), are derived similarly from specialisation of (J2) and (J3) respectively. □

Exercise 15.17 Complete the proofs of (A2) and (A3) omitted in the above proof. (Hint: (A2) follows from (J2) by the same specialisation as in the proof of (A1). To prove (A3), apply specialisations $x_1 = x_3 = x$, $x_2 = x_4 = u$ and $x_1 = -x_3 = x$, $x_2 = -x_4 = u$ to (J3) and combine the results.)

15.4 Jacobi's Derivative Formula

Using one of the theta relations and the heat equation proved in the previous sections, we can show an important relation called *Jacobi's derivative formula*.[6]

According to the results (15.22), the values of theta functions at $u = 0$, $\theta_{kl} = \theta_{kl}(0,\tau)$ (*theta constants* or *theta zero-values*) are non-zero for $(k,l) \neq (1,1)$. When (k,l) is $(1,1)$, $\theta_{11} = 0$ but the derivative $\theta_{11}' := \left.\dfrac{\partial}{\partial u}\right|_{u=0} \theta_{11}(u,\tau)$ does not vanish, because $u = 0$ is the first order zero of $\theta_{11}(u)$. These values are connected by the following formula.

Theorem 15.18 (Jacobi's derivative formula)

$$\theta_{11}' = -\pi\theta_{00}\theta_{01}\theta_{10}. \tag{15.33}$$

[6] Equation (28) in §4 of [Ja2].

Proof The specialisation $x_1 = x$, $x_2 = x_3 = x_4 = 0$ of (J3) leaves only one term in the left-hand side because terms with $\theta_{11}(x_2)$, $\theta_{11}(x_3)$ or $\theta_{11}(x_4)$ vanish by $\theta_{11}(0) = 0$:

$$\theta_{11}(x)\,\theta_{10}\,\theta_{01}\,\theta_{00} = 2\,\theta_{11}\left(\frac{x}{2}\right)\theta_{10}\left(\frac{x}{2}\right)\theta_{01}\left(\frac{x}{2}\right)\theta_{00}\left(\frac{x}{2}\right). \tag{15.34}$$

Let us expand this equation around $x = 0$, using the Taylor expansions of $\theta_{kl}(x)$ and $\theta_{11}(x)$,

$$\begin{aligned}\theta_{kl}(x) &= \theta_{kl} + \frac{\theta''_{kl}}{2}x^2 + O(x^4) \quad ((k,l) \neq (1,1)),\\ \theta_{11}(x) &= \theta'_{11}x + \frac{\theta'''_{11}}{6}x^3 + O(x^5).\end{aligned}$$

Here we use abbreviations

$$\theta''_{kl} = \left.\frac{\partial^2}{\partial u^2}\right|_{u=0} \theta_{kl}(u,\tau), \qquad \theta'''_{11} = \left.\frac{\partial^3}{\partial u^3}\right|_{u=0} \theta_{11}(u,\tau),$$

together with θ_{kl}, θ'_{11}. Note that in these Taylor expansions only even degree terms remain for $\theta_{kl}(x)$ $((k,l) \neq (1,1))$, while only odd degree terms remain for $\theta_{11}(x)$, because of their parity (15.17).

Comparing coefficients of x^3 in the Taylor expansion of (15.34) thus obtained, we have relations:

$$\frac{1}{6}\theta'''_{11}\theta_{10}\theta_{01}\theta_{00} = \frac{1}{24}\theta'''_{11}\theta_{10}\theta_{01}\theta_{00} + \frac{1}{8}\theta'_{11}(\theta''_{10}\theta_{01}\theta_{00} + \theta_{10}\theta''_{01}\theta_{00} + \theta_{10}\theta_{01}\theta''_{00}).$$

By dividing both sides by $\theta_{00}\theta_{01}\theta_{10}\theta'_{11}(\neq 0)$ we can simplify this relation as

$$\frac{\theta'''_{11}}{\theta'_{11}} - \frac{\theta''_{00}}{\theta_{00}} - \frac{\theta''_{01}}{\theta_{01}} - \frac{\theta''_{10}}{\theta_{10}} = 0.$$

The Taylor expansions of the (complex) heat equations (15.18) give $\theta''_{kl} = 4\pi i \dfrac{\partial}{\partial \tau}\theta_{kl}$ $((k,l) \neq (1,1))$ and $\theta'''_{11} = 4\pi i \dfrac{\partial}{\partial \tau}\theta'_{11}$. Substituting them into the above relation, we obtain

$$0 = \frac{\dfrac{\partial}{\partial \tau}\theta'_{11}}{\theta'_{11}} - \frac{\dfrac{\partial}{\partial \tau}\theta_{00}}{\theta_{00}} - \frac{\dfrac{\partial}{\partial \tau}\theta_{01}}{\theta_{01}} - \frac{\dfrac{\partial}{\partial \tau}\theta_{10}}{\theta_{10}} = \frac{\partial}{\partial \tau}\left(\log \frac{\theta'_{11}}{\theta_{00}\,\theta_{01}\,\theta_{10}}\right).$$

This shows that (the logarithm of) $\dfrac{\theta'_{11}}{\theta_{00}\,\theta_{01}\,\theta_{10}}$ does not depend on τ.

In order to compute the value of this constant, let us restrict τ to the imaginary axis $\tau = it$ $(t \in \mathbb{R}_{>0})$ and take the limit $t \longrightarrow +\infty$. In other words, replacing the variable $q = e^{\pi i \tau} = e^{-\pi t}$, we take the limit $q \longrightarrow 0$.

To find this limit, let us expand θ_{kl} and θ'_{11} as a series in q. Substituting $u = 0$ into the Fourier series (15.12) (and its derivative), we have

$$\begin{aligned}
\theta_{00} &= \sum e^{\pi i n^2 \tau} = 1 + O(q), \\
\theta_{01} &= \sum e^{\pi i n^2 \tau + \pi i n} = 1 + O(q), \\
\theta_{10} &= \sum e^{\pi i (n+\frac{1}{2})^2 \tau} = 2q^{1/4} + O(q), \\
\theta'_{11} &= \sum \pi i (2n+1) e^{\pi i (n+\frac{1}{2})^2 \tau + \pi i (n+\frac{1}{2})} \\
&= -2\pi q^{1/4} + O(q).
\end{aligned} \tag{15.35}$$

Therefore the constant which we are looking for is

$$\frac{\theta'_{11}}{\theta_{00}\,\theta_{01}\,\theta_{10}} = \lim_{q\to 0} \frac{-2\pi q^{1/4} + O(q)}{2q^{1/4} + O(q)} = -\pi.$$

This proves Jacobi's derivative formula (15.33). □

15.5 Modular Transformations of Theta Functions

At the beginning of Section 15.1, mainly for reasons of convenience, we fixed the periods to $(1,\tau)$ ($\operatorname{Im}\tau > 0$) instead of general periods (Ω_1,Ω_2) by scaling $u \mapsto \Omega_1 u$. If we use general periods (Ω_1,Ω_2) as they are, not normalising, what 'theta functions' do we get?

For example, let us consider a function

$$\theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right) := \theta\left(\frac{u}{\Omega_1}, \frac{\Omega_2}{\Omega_1}\right) \tag{15.36}$$

defined by the above mentioned scaling. (For the moment we use only the theta function $\theta(u,\tau) = \theta_{00}(u,\tau)$ without characteristics.) Here Ω_1 and Ω_2 are arbitrary non-zero complex numbers which satisfy $\operatorname{Im}\dfrac{\Omega_2}{\Omega_1} > 0$. The last condition is required for substitution in the theta function. Since the argument of $\dfrac{\Omega_2}{\Omega_1}$ is equal to $\arg\Omega_2 - \arg\Omega_1$, this condition means that the vector corresponding to Ω_2 in the complex plane $\mathbb{C} \cong \mathbb{R}^2$ is in the left side of the vector corresponding to Ω_1 (Fig. 15.3).

It is easy to see from the quasi-periodicities of the theta function (15.14), (15.15) that the function defined by (15.36) has the following quasi-periodicities:

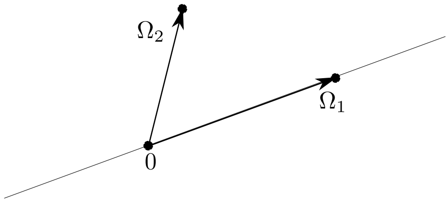

Fig. 15.3 $\operatorname{Im}\dfrac{\Omega_2}{\Omega_1} > 0$.

$$\theta\left(u+\Omega_1 \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right) = \theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right),$$
$$\theta\left(u+\Omega_2 \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right) = e^{-\pi i \frac{\Omega_2}{\Omega_1} - 2\pi i \frac{u}{\Omega_1}}\,\theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right).$$

"OK, then the general version is almost the same, easy stuff." It might seem so, but there is an important problem behind it. Let us consider an elliptic curve as a quotient of the plane $\mathbb{C}$ by the lattice $\Gamma = \mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2$ (the image of the Abel–Jacobi map): $\mathcal{E}_{\Omega_1,\Omega_2} := \mathbb{C}/\mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2$. In this construction the essential ingredient is the lattice Γ. If other periods (Ω_1', Ω_2') define the same lattice Γ, we obtain one and the same elliptic curve:

$$\Gamma = \mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2 = \mathbb{Z}\Omega_1' + \mathbb{Z}\Omega_2' \implies \mathcal{E}_{\Omega_1,\Omega_2} = \mathcal{E}_{\Omega_1',\Omega_2'}.$$

Therefore this elliptic curve has two 'theta functions' $\theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right)$ and $\theta\left(u \,\middle|\, \begin{matrix}\Omega_2'\\ \Omega_1'\end{matrix}\right)$ on it. How are they related to each other?

As the simplest and most important example, let us consider the case

$$\Omega_1' = \Omega_2, \qquad \Omega_2' = -\Omega_1.$$

(See Fig. 15.4.) The minus sign in the second equation is required by the condition $\operatorname{Im}\dfrac{\Omega_2'}{\Omega_1'} > 0$.

Obviously the pair (Ω_1', Ω_2') generates the same lattice as (Ω_1, Ω_2): $\mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2 = \mathbb{Z}\Omega_1' + \mathbb{Z}\Omega_2'$. Since $\dfrac{\Omega_2'}{\Omega_1'} = -\dfrac{\Omega_1}{\Omega_2}$, $\dfrac{u}{\Omega_1'} = \dfrac{u}{\Omega_1} \times \dfrac{\Omega_1}{\Omega_2}$, the function $\theta\left(u \,\middle|\, \begin{matrix}\Omega_2'\\ \Omega_1'\end{matrix}\right)$ is expressed as

$$\theta\left(u \,\middle|\, \begin{matrix}\Omega_2'\\ \Omega_1'\end{matrix}\right) = \theta\left(\frac{u/\Omega_1}{\tau}, -\frac{1}{\tau}\right),$$

where $\tau = \dfrac{\Omega_2}{\Omega_1}$. Hence, to compare $\theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right)$ and $\theta\left(u \,\middle|\, \begin{matrix}\Omega_2'\\ \Omega_1'\end{matrix}\right)$, it is sufficient to compare $\theta(u,\tau)$ and $\tilde{\theta}(u,\tau) := \theta\left(\dfrac{u}{\tau}, -\dfrac{1}{\tau}\right)$.

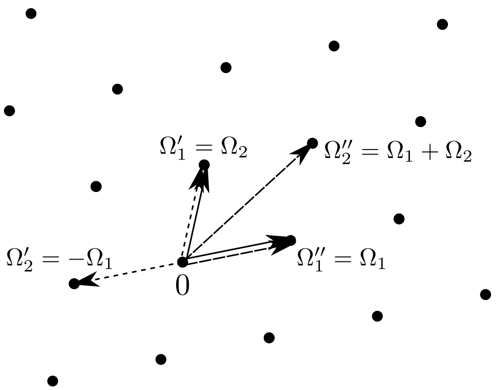

Fig. 15.4 Bases of Γ, $(\Omega_1', \Omega_2') = (\Omega_2, -\Omega_1)$, $(\Omega_1'', \Omega_2'') = (\Omega_1, \Omega_1 + \Omega_2)$.

As is claimed at the end of Section 15.2, we can identify theta functions by quasi-periodicities and zeros. It follows from (15.15) and (15.14) that $\tilde{\theta}(u,\tau)$ is quasi-periodic as

(15.37)
$$\begin{aligned}\tilde{\theta}(u+1,\tau) &= e^{\frac{\pi i}{\tau}+\frac{2\pi i u}{\tau}}\tilde{\theta}(u,\tau),\\ \tilde{\theta}(u+\tau,\tau) &= \tilde{\theta}(u,\tau).\end{aligned}$$

Exercise 15.19 Check the quasi-periodicity (15.37). (Hint: The first one follows from (15.15). Find a formula expressing $\theta(u-\tau,\tau)$ in terms of $\theta(u,\tau)$ and then replace τ by $-\dfrac{1}{\tau}$.)

The quasi-periodicities (15.37) of $\tilde{\theta}(u,\tau)$ are different from those of $\theta(u,\tau)$ or theta functions with characteristics, (15.14) and (15.15), in that the shift by 1 induces multiplication by a rather complicated exponential factor, while $\tilde{\theta}(u,\tau)$ is periodic with respect to τ. It might seem a new kind of quasi-periodicity, but, in fact, a small modification is enough to identify $\tilde{\theta}(u,\tau)$ with $\theta(u,\tau)$. In fact, the idea is the same as the reduction of the Weierstrass sigma function $\sigma(u)$ to the theta function discussed at the beginning of Section 15.1. (See Exercise 15.1.)

Let us take a quadratic form $Q(u) = \dfrac{\pi i u^2}{\tau}$. It satisfies

(15.38)
$$\begin{aligned}Q(u+1)-Q(u) &= \frac{2\pi i u}{\tau}+\frac{\pi i}{\tau},\\ Q(u+\tau)-Q(u) &= 2\pi i u + \pi i \tau.\end{aligned}$$

Hence the function

(15.39)
$$\tilde{\tilde{\theta}}(u,\tau) := e^{-Q(u)}\tilde{\theta}(u,\tau)$$

satisfies

(15.40)
$$\begin{aligned}\tilde{\tilde{\theta}}(u+1,\tau) &= \tilde{\tilde{\theta}}(u,\tau),\\ \tilde{\tilde{\theta}}(u+\tau,\tau) &= e^{-\pi i\tau-2\pi iu}\tilde{\tilde{\theta}}(u,\tau)\end{aligned},$$

by virtue of (15.37) and (15.38), which exactly coincide with quasi-periodicities (15.14), (15.15) of $\theta(u,\tau)$.

On the other hand, as the exponential function never vanishes, the zeros of $\tilde{\tilde{\theta}}(u,\tau)$ coincide with those of $\tilde{\theta}(u,\tau) = \theta\left(\frac{u}{\tau},-\frac{1}{\tau}\right)$. Therefore it follows from (15.22) that

$$\begin{aligned}\theta\left(\frac{u}{\tau},-\frac{1}{\tau}\right) = 0 &\Longleftrightarrow \frac{u}{\tau} \in \frac{1+(-\tau^{-1})}{2} + \mathbb{Z} + \mathbb{Z}\left(-\frac{1}{\tau}\right)\\ &\Longleftrightarrow u \in \frac{1+\tau}{2} + \mathbb{Z} + \mathbb{Z}\tau,\end{aligned}$$

which means that the zeros of $\tilde{\tilde{\theta}}(u,\tau)$ and those of $\theta(u,\tau)$ are the same. Hence the ratio $\dfrac{\tilde{\tilde{\theta}}(u,\tau)}{\theta(u,\tau)}$ of these functions is a doubly-periodic function without poles, which is a constant not depending on u by Theorem 11.9: $\tilde{\tilde{\theta}}(u,\tau) = A(\tau)\,\theta(u,\tau)$. (Note that the constant $A(\tau)$ *does* depend on τ.) In the original notation, we have

(15.41)
$$\theta\left(\frac{u}{\tau},-\frac{1}{\tau}\right) = A(\tau)e^{\frac{\pi iu^2}{\tau}}\,\theta(u,\tau).$$

To find the value of $A = A(\tau)$ explicitly, we need similar transformation formulae for theta functions with characteristics. The formula (15.41) implies

$$\begin{aligned}\theta_{00}\left(\frac{u}{\tau},-\frac{1}{\tau}\right) &= Ae^{\frac{\pi iu^2}{\tau}}\,\theta_{00}(u,\tau),\\ \theta_{01}\left(\frac{u}{\tau},-\frac{1}{\tau}\right) &= Ae^{\frac{\pi iu^2}{\tau}}\,\theta_{10}(u,\tau),\\ \theta_{10}\left(\frac{u}{\tau},-\frac{1}{\tau}\right) &= Ae^{\frac{\pi iu^2}{\tau}}\,\theta_{01}(u,\tau),\\ \theta_{11}\left(\frac{u}{\tau},-\frac{1}{\tau}\right) &= -iAe^{\frac{\pi iu^2}{\tau}}\,\theta_{11}(u,\tau)\end{aligned}$$

by (15.16). (Note that characteristics 01 and 10 get interchanged by the transformation in the second and third rows.) Setting u to 0 in the first three, we obtain

(15.42)
$$\tilde{\theta}_{00} = A\theta_{00},\ \tilde{\theta}_{01} = A\theta_{10},\ \tilde{\theta}_{10} = A\theta_{01}.$$

Here we use notations $\theta_{kl} = \theta_{kl}(0,\tau)$, $\tilde{\theta}_{kl} = \theta_{kl}\left(0,-\frac{1}{\tau}\right)$. The fourth formula gives $0 = 0$ by setting $u = 0$, so let us substitute $u = 0$ after differentiating it. In the same notations as above, the result is

$$\frac{1}{\tau}\tilde{\theta}'_{11} = -i\tau A\,\theta'_{11}.$$

Rewriting this equation by Jacobi's derivative formula (15.33), and the transformation formulae (15.42), we obtain $A^2 = -i\tau$, i.e.,

$$A(\tau) = \sqrt{-i\tau} \text{ or } -\sqrt{-i\tau}.$$

In order to fix the sign (more precisely, the argument) of this square root, let us find the sign of $A(\tau)$ on the imaginary axis, $\tau = it$, $t > 0$. Substituting $u = 0$, $\tau = it$ or $u = 0$, $\tau = -\dfrac{1}{it}$ into the definition of the theta function (15.12), we can compute θ_{00} and $\tilde{\theta}_{00}$ as follows:

$$\theta_{00} = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t}, \qquad \tilde{\theta}_{00} = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t^{-1}}.$$

They are series with positive real terms and thus give positive real numbers. Hence the ratio $A(\tau) = A(it)$ of θ_{00} and $\tilde{\theta}_{00}$ is a positive real number. Therefore $A(it) = \sqrt{t}$.

Being the ratio of two holomorphic functions θ_{00} and $\tilde{\theta}_{00}$ of τ on the upper half plane ((15.42)), $A(\tau)$ is also a holomorphic function on the upper half plane, which allows us to extend $A(it) = \sqrt{t}$ analytically on the whole upper half plane as

$$A(\tau) = \sqrt{t}\, e^{i\left(\frac{\varphi}{2} - \frac{\pi}{4}\right)},$$

where $\tau = t\,e^{i\varphi}$ ($t = |\tau|$, $\varphi = \arg\tau$, $0 < \varphi < \pi$). We denote this $A(\tau)$ simply by $\sqrt{-i\tau}$.

Thus we have proved *Jacobi's imaginary transformation*, or the *modular transformation formula* for $\tau \mapsto -\dfrac{1}{\tau}$,

$$\begin{aligned}
\theta_{00}\left(\frac{u}{\tau}, -\frac{1}{\tau}\right) &= \sqrt{-i\tau}\, e^{\frac{\pi i u^2}{\tau}}\, \theta_{00}(u,\tau),\\
\theta_{01}\left(\frac{u}{\tau}, -\frac{1}{\tau}\right) &= \sqrt{-i\tau}\, e^{\frac{\pi i u^2}{\tau}}\, \theta_{10}(u,\tau),\\
\theta_{10}\left(\frac{u}{\tau}, -\frac{1}{\tau}\right) &= \sqrt{-i\tau}\, e^{\frac{\pi i u^2}{\tau}}\, \theta_{01}(u,\tau),\\
\theta_{11}\left(\frac{u}{\tau}, -\frac{1}{\tau}\right) &= -i\sqrt{-i\tau}\, e^{\frac{\pi i u^2}{\tau}}\, \theta_{11}(u,\tau).
\end{aligned} \tag{15.43}$$

Let us return to the starting point of the discussion. These equations connect $\theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right)$ and $\theta\left(u \,\middle|\, \begin{matrix}-\Omega_1\\ \Omega_2\end{matrix}\right)$ as

$$\theta\left(u \,\middle|\, \begin{matrix}-\Omega_1\\ \Omega_2\end{matrix}\right) = \sqrt{-i\frac{\Omega_2}{\Omega_1}}\, e^{\frac{\pi i u^2}{\Omega_1\Omega_2}}\, \theta\left(u \,\middle|\, \begin{matrix}\Omega_2\\ \Omega_1\end{matrix}\right).$$

Change of bases of Γ, $(\Omega_1, \Omega_2) \to (\Omega_2, -\Omega_1)$, modifies the theta function by an exponential factor and does not change it too much.

There are infinitely many pairs (Ω_1, Ω_2) which generate the same lattice Γ: $\mathbb{Z}\Omega_1 + \mathbb{Z}\Omega_2 = \mathbb{Z}\Omega_1' + \mathbb{Z}\Omega_2'$ ($\mathrm{Im}\,\dfrac{\Omega_2}{\Omega_1}$ and $\mathrm{Im}\,\dfrac{\Omega_2'}{\Omega_1'}$ are positive) if and only if there exists an integer matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $(a, b, c, d \in \mathbb{Z})$ with determinant 1 such that

$$\begin{pmatrix} \Omega_2' \\ \Omega_1' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \Omega_2 \\ \Omega_1 \end{pmatrix}.$$

There is a transformation formula (the *modular transformation formula*) like (15.43) for any such matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Jacobi's imaginary transformation is for the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Later in Section 17.2 we will use the following modular transformation for $\tau \mapsto \tau + 1$, which corresponds to $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (or $(\Omega_1'', \Omega_2'') = (\Omega_1, \Omega_1 + \Omega_2)$ in Fig. 15.4), in addition to Jacobi's imaginary transformation.

Exercise 15.20 Prove the following formula.

$$\begin{aligned} \theta_{00}(u, \tau + 1) &= \theta_{01}(u, \tau), \\ \theta_{01}(u, \tau + 1) &= \theta_{00}(u, \tau), \\ \theta_{10}(u, \tau + 1) &= e^{\frac{\pi i}{4}} \theta_{10}(u, \tau), \\ \theta_{11}(u, \tau + 1) &= e^{\frac{\pi i}{4}} \theta_{11}(u, \tau). \end{aligned} \tag{15.44}$$

(Hint: Substitute $\tau + 1$ for τ in the definition (15.12).)

Remark 15.21 Combining this formula (15.44) and Jacobi's imaginary transformation (15.43), we can construct a modular transformation formula for any integer matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $(a, b, c, d \in \mathbb{Z}$, determinant $= 1)$.

We have discussed most of the important properties of theta functions, which we will use later. There are still many good properties and formulae but in the next chapter as the last example we will only discuss infinite product factorisations of theta functions, which are important in applications.

Chapter 16
Infinite Product Factorisation of Theta Functions

When we introduced elliptic integrals and elliptic functions, we showed how they were applied. Now it should be theta functions' turn. But, although theta functions are used in many branches of mathematics and physics, many prerequisites are required to understand those applications, which makes it difficult to explain them in this book. Here, instead, we prove infinite product factorisations of theta functions often used in applications.

16.1 Infinite Product of Functions

In contrast to the theory of infinite sums, i.e., the theory of series, which is one of the musts in calculus courses, infinite products are not always discussed in detail in university courses. So, we review the theory of infinite products briefly here.

Let us begin with a simple case. A polynomial with complex coefficients is defined by a sum,

$$P(z) = a_0 + a_1 z + \cdots + a_n z^n, \qquad a_i \in \mathbb{C},$$

and, thanks to the *fundamental theorem of algebra*[1], it can be always expressed as a product of linear forms with complex coefficients:

$$P(z) = a_n (z - \alpha_1) \cdots (z - \alpha_n). \tag{16.1}$$

Needless to say, a factor $z - \alpha_i$ $(i = 1, \ldots, n)$ corresponds to a zero $z = \alpha_i$ of a function $P(z)$.

What about a function with infinitely many zeros? For example, Euler proved ([E], Chapter 9, §156) that $\sin z$ is factorised as

[1] If $n \geq 1$, then there exists a complex number $z_0 \in \mathbb{C}$ satisfying $P(z_0) = 0$. In spite of its name, it is an analytic theorem in nature.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_17

$$\sin z = z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2\pi^2}\right). \tag{16.2}$$

Like the factorisation (16.1) of a polynomial, factors in the right-hand side corresponds to zeros of $\sin z$, $z = 0$ and $z = \pm n\pi$ ($n = 1, 2, \dots$).

However, you may have noticed that the expression for a polynomial and the expression for $\sin z$ are a little different. In the case of a polynomial a factor for a zero α_i has a simple form $z - \alpha_i$, while for $\sin z$ two zeros $n\pi$ and $-n\pi$ are paired and the corresponding factor has the form $1 - \dfrac{z^2}{n^2\pi^2}$, not the simplest form $(z - n\pi)(z + n\pi)$. Why can't we use the form "(constant) $\times \prod\limits_{n\in\mathbb{Z}} (z - n\pi)$"?

This is the difference caused by the finiteness/infiniteness of the number of zeros. In the product $\prod\limits_{n\in\mathbb{Z}} (z - n\pi)$ the absolute value of a factor $z - n\pi$ grows unboundedly when $|n|$ grows, and the product diverges. In order for an infinite product $\prod\limits_{n} a_n$ of complex numbers a_n to converge, the sequence $\{a_n\}$ should converge to 1. Therefore we multiply '1 − (something)' in the expression (16.2), where 'something' converges to 0 as $|n|$ grows.

Definition 16.1 Let $\{a_n\}_{n=1,2,\dots}$ be a sequence of non-zero complex numbers. The infinite product $\prod\limits_{n=1}^{\infty} a_n$ is said to *converge to P* if the sequence $\{p_n\}_{n=1,2,\dots}$ of *partial products* $p_n = \prod\limits_{k=1}^{n} a_k$ converges to a *non-zero* complex number P: $P = \prod\limits_{n=1}^{\infty} a_n = \lim\limits_{n\to\infty} p_n$.

If there are a finite number of zeros among the a_n's, $\prod\limits_{n=1}^{\infty} a_n$ is said to converge to 0 if the infinite product of $\{a_n\}_{n=1,2,\dots}$ except for zero factors converges in the above sense.

The notions of absolute convergence and uniform convergence, which are important for series, also play important roles, but their definitions are not obtained by simple replacement of word 'sum' to 'product'.

Lemma 16.2 *(i) An infinite product* $\prod\limits_{n=1}^{\infty} (1 + u_n)$ *(*$\lim\limits_{n\to\infty} u_n = 0$*) is said to converge* absolutely *when* $\prod\limits_{n=1}^{\infty} (1 + |u_n|)$ *converges. This is equivalent to convergence of the series* $\sum\limits_{n=1}^{\infty} |u_n|$*. When* $\prod\limits_{n=1}^{\infty} (1 + u_n)$ *converges absolutely, it converges and the value does not depend on the order of factors.*

(ii) Assume that each $u_n(z)$ is a function on a subset D of $\mathbb{C}$. When the sequence of functions $\left\{p_n(z) := \prod_{k=1}^{n}(1+u_k(z))\right\}_{n=1,2,\dots}$ *converges uniformly on D, the infinite product* $\prod_{n=1}^{\infty}(1+u_n(z))$ *is said to converge* uniformly *on D. If each $u_n(z)$ is a holomorphic function on a domain D, the product* $\prod_{n=1}^{\infty}(1+u_n(z))$ *is holomorphic on D.*

(iii) If there exists a sequence $\{M_N\}_{n=1,2,\dots}$ of positive real numbers satisfying the following two conditions, the infinite product $\prod_{n=1}^{\infty}(1+u_n(z))$ *converges absolutely and uniformly on D:*

1. $|u_n(z)| \le M_n$ *for any n and $z \in D$.*
2. $\sum_{n=1}^{\infty} M_n < \infty$.

For the proofs we refer to §2.2 of Chapter 5 in [Ah].

Using the above facts, we can show that the right-hand side of the factorisation (16.2) really gives a holomorphic function.

Exercise 16.3 Using Lemma 16.2, show that the infinite product (16.2) converges absolutely and uniformly on any compact subset of $\mathbb{C}$. Why shouldn't we express this factorisation as follows?

$$\sin z \text{``}=\text{''} z \prod_{n=1}^{\infty}\left(1-\frac{z}{n\pi}\right)\prod_{n=1}^{\infty}\left(1+\frac{z}{n\pi}\right).$$

Remark 16.4 In general, a holomorphic function on the whole plane $\mathbb{C}$ can be factorised as an infinite product of linear forms, possibly modified by exponential functions. For example, instead of the false factorisation in Exercise 16.3 we can factorise $\sin z$ as

$$\sin z = z \prod_{n=1}^{\infty}\left(1-\frac{z}{n\pi}\right)e^{z/n\pi}\prod_{n=1}^{\infty}\left(1+\frac{z}{n\pi}\right)e^{-z/n\pi}.$$

Another example is the factorisation (12.19) of the sigma function,

$$\sigma(u) = u \prod_{\Omega\in\Gamma\smallsetminus\{0\}}\left(1-\frac{u}{\Omega}\right)\exp\left(\frac{u}{\Omega}+\frac{u^2}{2\Omega^2}\right).$$

See §2.3 of Chapter 5 in [Ah] for details on such factorisations.

16.2 Infinite Product Factorisation of Theta Functions

Moving away from general theory, let us factorise theta functions as such infinite products, since we know their zeros. Below we use the variable $z = e^{\pi i u}$, as the expression of the factorisation of $\theta_{kl}(u,\tau)$ ($k,l \in \{0,1\}$) is simpler in that variable.[2]

The zeros of theta functions are of first order and located as described in (15.22). In terms of variables $z = e^{\pi i u}$, $q = e^{\pi i \tau}$, this is rewritten as

$$
\begin{aligned}
\theta_{00}(u) = 0 &\Longleftrightarrow z = \pm i q^{n+\frac{1}{2}},\\
\theta_{01}(u) = 0 &\Longleftrightarrow z = \pm q^{n+\frac{1}{2}},\\
\theta_{10}(u) = 0 &\Longleftrightarrow z = \pm i q^{n},\\
\theta_{11}(u) = 0 &\Longleftrightarrow z = \pm q^{n}.
\end{aligned}
\tag{16.3}
$$

Here n is an arbitrary integer.

Let us begin with $\theta(u) = \theta_{00}(u,\tau)$. As shown in (16.3), the zeros of $\theta(u)$ are $\pm i q^{n+\frac{1}{2}}$ in terms of z. Therefore from the analogy with $\sin z$, one might infer that $\theta(u)$ would be a constant multiple of

$$
\prod_{n\in\mathbb{Z}} \left(1 - \frac{z}{i q^{n+\frac{1}{2}}}\right)\left(1 + \frac{z}{i q^{n+\frac{1}{2}}}\right) = \prod_{n\in\mathbb{Z}} \left(1 + \frac{z^2}{q^{2n+1}}\right).
$$

However, this product diverges, as $\dfrac{z^2}{q^{2n+1}}$ grows when $n \to +\infty$. (Note that $|q| < 1$, as $\operatorname{Im}\tau > 0$.)

A puzzle presents itself: how can we modify this product so that it is convergent? The answer is as follows. It is sufficient to modify the part $n \geq 0$ as

$$
\left(1 - \frac{z^{-1}}{\left(i q^{n+\frac{1}{2}}\right)^{-1}}\right)\left(1 + \frac{z^{-1}}{\left(i q^{n+\frac{1}{2}}\right)^{-1}}\right) = 1 + \frac{z^{-2}}{q^{-2n-1}}.
$$

Thus a new guess is

$$
\begin{aligned}
\theta(u) &= (\text{constant}) \times \prod_{n\geq 0} (1 + q^{2n+1} z^{-2}) \prod_{n\leq -1} (1 + q^{-2n-1} z^2)\\
&= (\text{constant}) \times \prod_{n=1}^{\infty} (1 + q^{2n-1} z^{-2}) \prod_{n=1}^{\infty} (1 + q^{2n-1} z^2).
\end{aligned}
\tag{16.4}
$$

[2] The functions $\theta_{00}(u,\tau)$ and $\theta_{01}(u,\tau)$ are factorised even more simply with the variable $\tilde{z} = e^{2\pi i u}$. If you find the following discussion too messy, it is a good exercise to rewrite it with $\tilde{z}$.

First of all, we have to verify that this infinite product converges and defines a holomorphic function. This is a direct consequence of Lemma 16.2 (iii), but, as we omitted its proof, let us discuss this example in detail.

Lemma 16.5 *(i) The infinite product* $\prod_{n=1}^{\infty}(1+q^{2n-1}z^2)$ *converges uniformly as a function of* z *on any bounded closed set (= any compact set)* $K \subset \mathbb{C}$.

(ii) The infinite product $\prod_{n=1}^{\infty}(1+q^{2n-1}z^{-2})$ *converges uniformly as a function of* z *on any bounded closed set* $K \subset \mathbb{C} \setminus \{0\}$.

(iii) The infinite product $\prod_{n=1}^{\infty}(1+q^{2n-1}z^2)(1+q^{2n-1}z^{-2})$ *converges uniformly as a function of* z *on any bounded closed set* $K \subset \mathbb{C} \setminus \{0\}$. *Its value is equal to the product of the infinite products in (i) and (ii).*

Proof The statement (iii) follows from (i), (ii) and properties of uniformly convergent limits (cf. Exercise 16.6 (ii)). The statement (ii) is reduced to (i) by the change of variable $z = w^{-1}$.

Proof of (i): Let us show that the infinite product $\prod_{n=1}^{\infty}(1+q^{2n-1}z^2)$ converges uniformly on a closed disk $D_R := \{z \mid |z| \le R\}$ for any $R > 0$. (This is sufficient, as any bounded set is contained in some D_R.) Since we can omit a finite number of factors, when we show convergence, we take a large natural number n_0 such that

$$|q^{2n_0-1}| < \frac{1}{2R^2}, \tag{16.5}$$

(recall that $|q| < 1$), and consider the convergence of a sequence of functions,

$$p_n(z) := \prod_{k=n_0}^{n}(1+q^{2k-1}z^2), \qquad n \ge n_0. \tag{16.6}$$

Inequality (16.5) implies $|q^{2k-1}| < \dfrac{|q|^{2(k-n_0)}}{2R^2}$ for $k \ge n_0$, from which follows the estimate,

$$|q^{2k-1}z^2| < \frac{|q|^{2(k-n_0)}}{2} < \frac{1}{2} \tag{16.7}$$

for $z \in D_R$. Using the Taylor expansion[3] $\log(1+w) = \sum_{n=1}^{\infty}(-1)^{n-1}\dfrac{w^n}{n}$ for $|w| < 1$ and the triangle inequality, we have

---

[3] Exactly speaking, we *fix* a value (a branch) of the multi-valued function $\log(1+w)$ by this Taylor expansion.

$$
\begin{aligned}
|\log(1+q^{2k-1}z^2)| &\le \sum_{n=1}^{\infty} \left| \frac{(q^{2k-1}z^2)^n}{n} \right| \\
&\le \sum_{n=1}^{\infty} \left|q^{2k-1}z^2\right|^n = \frac{\left|q^{2k-1}z^2\right|}{1-\left|q^{2k-1}z^2\right|}.
\end{aligned}
$$

The estimate (16.7) implies $\dfrac{1}{1-\left|q^{2k-1}z^2\right|} < 2$. Applying this to the right-most side of the above inequality and using (16.7) once more, we obtain

$$
|\log(1+q^{2k-1}z^2)| < 2|q|^{2(k-n_0)}. \tag{16.8}
$$

On the other hand, as

$$
\sum_{k=n_0}^{\infty} 2|q|^{2(k-n_0)} = \frac{2}{1-|q|^2} < +\infty,
$$

we can apply Weierstrass's M-test to the series

$$
\sum_{k=n_0}^{\infty} \log(1+q^{2k-1}z^2) \tag{16.9}
$$

thanks to (16.8). Thus we have proved the uniform convergence of (16.9) as a series of functions of z on D_R. This means that a sequence of functions

$$
\log p_n(z) = \sum_{k=n_0}^{n} \log(1+q^{2k-1}z^2)
$$

uniformly converges to the function (16.9) on D_R. It only remains to exponentiate this. It is a good exercise in elementary calculus to check that exponential functions of a uniformly converging sequence of functions also converge uniformly (Exercise 16.6 (iii)). □

Exercise 16.6 (i) Let $\{f_n(z)\}_{n=1,2,\dots}$ be a sequence of continuous functions on a compact set K, which uniformly converges to $f(z)$ on K. Show that there exists a positive real number M such that $|f(z)| < M$ and $|f_n(z)| < M$ for any $n = 1, 2, \dots$ and $z \in K$. (Hint: The limit of a uniform convergent sequence of continuous functions is continuous. A continuous function is bounded on a compact set. If n is large enough, $|f(z)|$ and $|f_n(z)|$ do not differ much. So, first show the statement for $n \ge N$, where N is a large natural number, and then extend the result to all n.)

(ii) Let $\{f_n(z)\}_{n=1,2,\dots}$ and $\{g_n(z)\}_{n=1,2,\dots}$ be sequences of continuous functions on a compact set K, which uniformly converge to $f(z)$ and $g(z)$ on K respectively.

Show that the sequence of products $\{f_n(z)\,g_n(z)\}_{n=1,2,\dots}$ converges uniformly to $f(z)\,g(z)$. (Hint: Estimate $|f(z)\,g(z) - f_n(z)\,g_n(z)|$, using (i).)

(iii) Let $\{f_n(z)\}_{n=1,2,\dots}$ be a sequence of complex-valued continuous functions on a compact set K, which uniformly converges to $f(z)$ on K, and $F(w)$ be a continuous function on $\mathbb{C}$.

Show that $\{F(f_n(z))\}_{n=1,2,\dots}$ converges uniformly to $F(f(z))$ on K. (Hint: The result of (i) allows us to assume that the image of any f_n belongs to the same closed disk $D_M = \{w \mid |w| \le M\}$. The function F is *uniformly continuous* on this disk.)

According to Lemma 16.5, the infinite product (16.4) converges uniformly on any compact set in $\mathbb{C} \smallsetminus \{0\}$. Hence by virtue of Weierstrass's double series theorem (Theorem A.9 (i)) this infinite product defines a holomorphic function of z on $\mathbb{C} \smallsetminus \{0\}$. Let us show that this is really a constant multiple of $\theta(u)$ and determine the constant. There are many proofs, among which we first apply the method which honestly expands the partial product

$$f_N(z) = \prod_{n=1}^{N} (1 + q^{2n-1} z^{-2})(1 + q^{2n-1} z^2) \tag{16.10}$$

into sums and compares with the definition of the theta function.

The expansion of $f_N(z)$ has the form

$$\begin{aligned} f_N(z) &= \sum_{k=0}^{2N} a_{k-N}^{(N)} z^{2k-2N} \\ &= a_{-N}^{(N)} z^{-2N} + \cdots + a_{-1}^{(N)} z^{-2} + a_0^{(N)} + a_1^{(N)} z^2 + \cdots + a_N^{(N)} z^{2N}, \end{aligned} \tag{16.11}$$

in which only even powers of z appear. Moreover, as the definition (16.10) is symmetric with respect to interchange of z and z^{-1}, the coefficients of z^{2k} and z^{-2k} are equal: $a_{-k}^{(N)} = a_k^{(N)}$. The top coefficient $a_N^{(N)}$ can be easily found, since the term $a_N^{(N)} z^{2N}$ in (16.11) is the product of $q^{2n-1} z^2$ in the factors $(1 + q^{2n-1} z^2)$, $n = 1, \dots, N$:

$$a_N^{(N)} = \prod_{n=1}^{N} q^{2n-1} = q^{1+3+\cdots+(2N-1)} = q^{N^2}. \tag{16.12}$$

It is not so easy to find other coefficients directly, but regarding $f_N(z)$ as their *generating function*[4], we can obtain the explicit form.

Substituting qz for z in $f_N(z)$, we obtain

[4] For a sequence $\{a_n\}$ a series (or a finite sum) $f(z) = \sum a_n z^n$ is called a *generating function* of the sequence $\{a_n\}$. Applying various operations (not only arithmetic operations but also derivation, difference and q-difference operators, which we shall discuss below), we can study various properties of the sequence $\{a_n\}$. There are other types of generating functions like an exponential-type generating function $f(z) = \sum \frac{a_n}{n!} z^n$ or a Dirichlet-type generating function $f(s) = \sum \frac{a_n}{n^s}$.

$$\begin{aligned} f_N(qz) &= \prod_{n=1}^{N} (1+q^{2n-1}q^{-2}z^{-2})(1+q^{2n-1}q^2z^2) \\ &= \frac{(1+q^{-1}z^{-2})(1+q^{2N+1}z^2)}{(1+q^{2N-1}z^{-2})(1+qz^2)} f_N(z). \end{aligned} \tag{16.13}$$

A relation of functions $f(z)$, $f(qz)$, $f(q^2z)$, ... is called a *q-difference equation*[5] of f. In this terminology (16.13) is a q-difference equation of $f_N(z)$. We can rewrite it as

$$(qz^2+q^{2N}) f_N(qz) = (1+q^{2N+1}z^2) f_N(z),$$

because $1+qz^2 = qz^2(1+q^{-1}z^{-2})$. Expanding $f_N(qz)$ and $f_N(z)$ by (16.11) and comparing the coefficients of z^{2k-2N} $(1 \le k \le 2N)$, we obtain relations among coefficients $a_k^{(N)}$,

$$a_{k-N-1}^{(N)} q^{2k-2N-1} + a_{k-N}^{(N)} q^{2k} = a_{k-N}^{(N)} + a_{k-N-1}^{(N)} q^{2N+1}.$$

(The coefficients of z^{-2N} and z^{2N+2} are trivially equal.) Thus we have found a recurrence relation

$$a_{k-1}^{(N)} = \frac{1-q^{2k+2N}}{q^{2k-1}(1-q^{2N-2k+2})} a_k^{(N)} \tag{16.14}$$

of $a_k^{(N)}$ for k, $-N+1 \le k \le N$. Recall that we know the explicit expression (16.12) of $a_N^{(N)}$. Therefore the recurrence relation (16.14) gives the explicit expression of $a_k^{(N)}$ as

$$\begin{aligned} a_k^{(N)} &= a_N^{(N)} \prod_{l=k+1}^{N} \frac{1-q^{2l+2N}}{q^{2l-1}(1-q^{2N-2l+2})} \\ &= q^{k^2} \frac{\prod_{l=k+1}^{N} (1-q^{2(l+N)})}{\prod_{l=1}^{N-k} (1-q^{2l})} = q^{k^2} \frac{\prod_{l=1}^{2N} (1-q^{2l})}{\prod_{l=1}^{k+N} (1-q^{2l}) \prod_{l=1}^{N-k} (1-q^{2l})}. \end{aligned} \tag{16.15}$$

We already know by Lemma 16.5 that the limit $(N \to \infty)$ of the sequence of partial products $f_N(z)$ defined by (16.10), i.e., the infinite product

$$f_\infty(z) := \prod_{n=1}^{\infty} (1+q^{2n-1}z^{-2}) \prod_{n=1}^{\infty} (1+q^{2n-1}z^2), \tag{16.16}$$

[5] On the other hand, a *difference equation* is a relation of $f(z)$, $f(z+h)$, $f(x+2h)$, ... for a constant h.

is a holomorphic function on $\mathbb{C} \setminus \{0\}$. The coefficients of the Laurent expansion of $f_\infty(z)$ around $z = 0$,

$$f_\infty(z) = \sum_{k\in\mathbb{Z}} a_k z^{2k}, \tag{16.17}$$

are the limits of the coefficients $a_k^{(N)}$ of the expansion (16.11) of the partial product $f_N(z)$. Therefore, taking the limit of (16.15), we have

$$\begin{aligned} a_k &= \lim_{N\to\infty} a_k^{(N)} \\ &= q^{k^2} \frac{\prod_{l=1}^{\infty}(1-q^{2l})}{\prod_{l=1}^{\infty}(1-q^{2l}) \prod_{l=1}^{\infty}(1-q^{2l})} = \frac{q^{k^2}}{\prod_{l=1}^{\infty}(1-q^{2l})}, \end{aligned} \tag{16.18}$$

and thus,

$$f_\infty(z) = \frac{\sum_{k\in\mathbb{Z}} q^{k^2} z^{2k}}{\prod_{l=1}^{\infty}(1-q^{2l})}. \tag{16.19}$$

Substitutions $z = e^{\pi i u}$ and $q = e^{\pi i \tau}$ turn the numerator of this right-hand side into a function of u and τ,

$$\sum_{k\in\mathbb{Z}} q^{k^2} z^{2k} = \sum_{k\in\mathbb{Z}} e^{\pi i k^2 \tau + 2\pi i k u},$$

which is nothing but the definition (15.8) of $\theta(u,\tau)$ in Section 15.1. So, finally, we have obtained the infinite product factorisation of $\theta(u,\tau)$,

(16.20)

$$\begin{aligned} \theta(u,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n}) \prod_{n=1}^{\infty}(1+q^{2n-1}z^2) \prod_{n=1}^{\infty}(1+q^{2n-1}z^{-2}) \\ &= \prod_{n=1}^{\infty}(1-e^{2n\pi i\tau}) \prod_{n=1}^{\infty}(1+e^{(2n-1)\pi i\tau+2\pi i u}) \prod_{n=1}^{\infty}(1+q^{(2n-1)\pi i\tau-2\pi i u}) \end{aligned}$$

from (16.19) and (16.16).

In the research area called ‘q-analysis’, which studies q-difference equations and so on, such infinite products appear quite frequently and the following notation is often used:

$$(a;p)_\infty := \prod_{n=0}^{\infty}(1-ap^n).$$

Using this notation, the infinite product expression of the θ-function is written as

$$\theta(u,\tau) = (q^2;q^2)_\infty\,(-qz^2;q^2)_\infty\,(-qz^{-2};q^2)_\infty. \tag{16.21}$$

An infinite product factorisation of theta functions with characteristics, $\theta_{kl}(u,\tau)$, can be found immediately by substituting (16.20) into (15.13):

$$\begin{aligned}
\theta_{00}(u,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+q^{2n-1}z^2)\prod_{n=1}^{\infty}(1+q^{2n-1}z^{-2}),\\
\theta_{01}(u,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1-q^{2n-1}z^2)\prod_{n=1}^{\infty}(1-q^{2n-1}z^{-2}),\\
\theta_{10}(u,\tau) &= q^{1/4}(z+z^{-1})\prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+q^{2n}z^2)\prod_{n=1}^{\infty}(1+q^{2n}z^{-2}),\\
\theta_{11}(u,\tau) &= iq^{1/4}(z-z^{-1})\prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1-q^{2n}z^2)\prod_{n=1}^{\infty}(1-q^{2n}z^{-2}).
\end{aligned} \tag{16.22}$$

Or, we can also express them in the following way:

$$\begin{aligned}
\theta_{00}(u,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+2q^{2n-1}\cos(2\pi u)+q^{4n-2}),\\
\theta_{01}(u,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1-2q^{2n-1}\cos(2\pi u)+q^{4n-2}),\\
\theta_{10}(u,\tau) &= 2q^{1/4}\cos(\pi u)\prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+2q^{2n}\cos(2\pi u)+q^{4n}),\\
\theta_{11}(u,\tau) &= -2q^{1/4}\sin(\pi u)\prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1-2q^{2n}\cos(2\pi u)+q^{4n}).
\end{aligned} \tag{16.23}$$

Putting $u=0$ in the above formulae, we obtain impressive infinite product expressions for theta zero-values, which are often useful.

$$\begin{aligned}
\theta_{00}(0,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+q^{2n-1})^2,\\
\theta_{01}(0,\tau) &= \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1-q^{2n-1})^2,\\
\theta_{10}(0,\tau) &= 2q^{1/4}\prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+q^{2n})^2.
\end{aligned} \tag{16.24}$$

Exercise 16.7 Verify the infinite product factorisations (16.22) and (16.23). How are they written with the notation $(a;q^2)_\infty$?

According to the 'dogma' at the end of Section 15.2, 'Formulae on theta functions can be proved once we know quasi-periodicity and positions of zeros', the infinite product factorisations should be proved in this way. In fact it is not very complicated, so we leave it to the reader.[6] The proof makes it clear that the form of the infinite product exactly corresponds to the quasi-periodicity of theta functions.

Exercise 16.8 Similarly to (16.16), we define a function $\varphi(u,\tau)$ by

$$\varphi(u,\tau) := \prod_{n=1}^{\infty}(1+q^{2n-1}e^{-2\pi iu})\prod_{n=1}^{\infty}(1+q^{2n-1}e^{2\pi iu}), \tag{16.25}$$

where $q = e^{\pi i\tau}$. According to Lemma 16.5, this function is holomorphic on the whole plane, $\mathbb{C}$ ($z = e^{\pi iu}$).

(i) Show that the quasi-periodicities of $\varphi(u,\tau)$ with respect to shifts $u \mapsto u+1$ and $u \mapsto u+\tau$ coincide with those of $\theta_{00}(u,\tau)$. (Hint: The proof for the shift $u \mapsto u+1$ is easy. For the shift by τ note that the indices of the infinite product get shifted by one.)

(ii) Show that there is a complex number $c(\tau)$ depending on τ such that $\theta_{00}(u,\tau) = c(\tau)\,\varphi(u,\tau)$.

(iii) Using the result of (ii), find infinite product factorisations of $\theta_{01}(u,\tau)$, $\theta_{10}(u,\tau)$ and $\theta_{11}(u,\tau)$. At this stage factorisations still contain $c(\tau)$ as an undetermined function.

(iv) Find $\theta_{00}(0,\tau)$, $\theta_{01}(0,\tau)$, $\theta_{10}(0,\tau)$ and $\theta_{11}'(0,\tau) = \left.\dfrac{d}{du}\right|_{u=0}\theta_{11}(u,\tau)$, using the result of (iii). (Hint: Rewriting the factorisation of $\theta_{11}(u,\tau)$ in the form $\theta_{11}(u,\tau) = (e^{\pi iu} - e^{-\pi iu})\psi(u,\tau)$, we have $\theta_{11}'(0,\tau) = 2\pi i\,\psi(0)$.)

(v) Substitute the results of (iv) into Jacobi's derivative formula (15.33) and show $c(\tau)^2 = \left(\prod_{n=1}^{\infty}(1-q^{2n})\right)^2$.

(vi) It remains to determine the sign of $c(\tau)$, which can be found by observing the behaviour for $\operatorname{Im}\tau \to \infty$, i.e., the behaviour around $q = 0$. Comparing the result of (iv) and the expansion of $\theta_{00} = \theta_{00}(0,\tau)$ into a series in q which was used in the proof of Theorem 15.18, show $c(\tau) = \prod_{n=1}^{\infty}(1-q^{2n})$.

In the above proof Jacobi's derivative formula is used, but, conversely, we can prove that formula using the infinite product factorisations.

Exercise 16.9 (i) Show

$$\prod_{n=1}^{\infty}(1+q^{2n-1})\prod_{n=1}^{\infty}(1-q^{2n-1})\prod_{n=1}^{\infty}(1+q^{2n}) = 1. \tag{16.26}$$

[6] The strategy suggested in Exercise 16.8 is cited from [M], Chapter I, §14. Another proof is in Chapter 5, §5.4 of [Ji].

(Hint: Multiply the left-hand side by $\prod_{n=1}^{\infty}(1-q^{2n})$ and use $(1-q^{2n-1})(1+q^{2n-1}) = 1-q^{4n-2}$, $(1+q^{2n})(1-q^{2n}) = 1-q^{4n}$. Recall that an even number is expressed either as $4n-2$ or as $4n$.)

(ii) Using the infinite product factorisation of $\theta_{11}(u,\tau)$, show that $\theta'_{11}(0,\tau) = -2\pi q^{1/4}\prod_{n=1}^{\infty}(1-q^{2n})^3$. (Hint: See the hint of Exercise 16.8 (iv).)

(iii) Show Jacobi's derivative formula, using the results of (i) and (ii).

For example, substituting the definition of the theta function into the left-hand side of (16.20) and expressing the result in the variables $q = e^{\pi i\tau}$ and $z = e^{\pi i u}$, we have an equality,

$$\sum_{n\in\mathbb{Z}} q^{n^2} z^{2n} = \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1+q^{2n-1}z^2)\prod_{n=1}^{\infty}(1+q^{2n-1}z^{-2}),$$

of 'infinite sum = infinite product' type. If we further put $z = 1$, we obtain

$$\sum_{n\in\mathbb{Z}} q^{n^2} = \prod_{n=1}^{\infty}(1-q^{2n})\left(\prod_{n=1}^{\infty}(1+q^{2n-1})\right)^2,$$

which looks very simple but miraculous and hard to guess its proof.[7]

There are many formulae of this type, among which the most well known is the following *pentagonal number theorem*,

$$\sum_{n\in\mathbb{Z}}(-1)^n q^{3n^2+n} = \prod_{n=1}^{\infty}(1-q^{2n}), \tag{16.27}$$

which Euler found in 1741 and proved in 1750.[8] Many proofs are known. For example, a combinatorial one is in §3.5, Chapter 3 and §5.4, Chapter 5 of [AE]. Here we sketch two proofs by infinite product factorisation of theta functions and leave the (not very complicated) details to the reader as an exercise.

The first one[9] is by the infinite product factorisation of $\theta_{11}(u,\tau)$.

Exercise 16.10 (i) Show *Jacobi's triple product identity*

$$\sum_{n\in\mathbb{Z}}(-1)^n q^{n(n-1)} z^{2n} = \prod_{n=1}^{\infty}(1-q^{2n})\prod_{n=1}^{\infty}(1-q^{2n-2}z^2)\prod_{n=1}^{\infty}(1-q^{2n}z^{-2}), \tag{16.28}$$

using the infinite product factorisation (16.22) of $\theta_{11}(u,\tau)$.

---

[7] It is recommended to compute products of several factors in the right-hand side to appreciate how terms *miraculously* cancel.

[8] For a detailed history, see Chapter 7 of [No].

[9] Corollary 5.21 in §5.4, Chapter 5 of [Ji].

(ii) Putting $z = q^{2/3}$ in (i), prove the pentagonal number theorem (16.27). (Hint: In order to reproduce the formula (16.27) exactly as it is, it is necessary to change q.)

The next proof[10] uses a theta function with characteristics different from 0 or $\pm\dfrac{1}{2}$.

Exercise 16.11 (i) Express $\theta_{1/6,1/2}(0,3\tau)$ in terms of $\theta_{00}\left(\dfrac{1+\tau}{2},3\tau\right)$, using the general transformation rule of characteristics (cf. Section 15.1). Then express it as an infinite product, using (16.20).

(ii) Expand $\theta_{1/6,1/2}(0,3\tau)$ by the definition of a theta function as a series and express it as a power series of $q = e^{\pi i\tau}$.

(iii) Comparing (i) and (ii), prove the pentagonal number theorem (16.27).

The following problem will be used later. The infinite products are used at the end of (ii).

Exercise 16.12 According to Exercise 15.5, the space $\Theta := \Theta_{2,-2\pi i\tau}$, i.e., the space of entire functions satisfying

$$f(u+1) = f(u), \quad f(u+\tau) = e^{-4\pi iu-2\pi i\tau} f(u),$$

is a two-dimensional complex linear space.

(i) Show that $\theta_{00}(u,\tau)^2$ and $\theta_{01}(u,\tau)^2$ belong to Θ and that they are linearly independent. (Hint: To show the linear independence it is sufficient to show that they have different zeros. Why?)

(ii) Show that $\theta_{00}(2u,2\tau)$ and $\theta_{10}(2u,2\tau)$ are also elements of Θ and that they are expanded by the basis $\{\theta_{00}(u,\tau)^2, \theta_{01}(u,\tau)^2\}$ as

$$\theta_{00}(2u,2\tau) = \frac{1}{2\theta_{00}(0,2\tau)}(\theta_{00}(u,\tau)^2 + \theta_{01}(u,\tau)^2), \tag{16.29}$$

$$\theta_{10}(2u,2\tau) = \frac{1}{2\theta_{00}(0,2\tau)}(\theta_{00}(u,\tau)^2 - \theta_{01}(u,\tau)^2). \tag{16.30}$$

(Hint: To find the coefficients in the expansion (16.29) put $u = \dfrac{\tau}{2}$ or $u = \dfrac{1+\tau}{2}$. Make them ratios of theta zero-values, using transformation rules of characteristics (15.16), and apply the infinite product expansions (16.24).)

Remark 16.13 Infinite product factorisations of theta functions, which we introduced in this chapter, appear in various fields of mathematics and physics. To name a few,

- in number theory theta zero values $\theta_{ab}(0)$ play an important role (for example, the generating function of the number of representations of natural numbers as a sum of squares);

[10] §14, Chapter I of [M].

- in the representation theory of infinite-dimensional Lie algebras theta functions appear as characters of infinite-dimensional representations;
- in string theory (elementary particle physics) or the theory of solvable lattice models (statistical mechanics) theta functions appear as partition functions of models.

In these examples, theta functions are generating functions[11] of important quantities (number of solutions of equations of integers, dimensions of so-called weight spaces of representations, certain physical quantities and so on). Generating functions are a priori defined by sums (series) and their expression as infinite products reveal deep properties.

In the next chapter we construct Jacobi's elliptic functions (sn, cn, dn) from theta functions and relate them with the definitions as 'inverse functions of elliptic integrals'.

[11] See footnote 4 in this chapter (p. 281).

Chapter 17
Complex Jacobian Elliptic Functions

In the previous two chapters we loosened the 'double periodicity' condition for elliptic functions to the 'quasi-periodicity' condition and defined theta functions which are holomorphic everywhere on the complex plane. They satisfy many good properties, in particular, Jacobi's relations, addition theorems and modular transformations. In this chapter we redefine Jacobi's elliptic functions sn, cn and dn as complex functions by means of theta functions.

17.1 Definition of Jacobi's Elliptic Functions as Complex Functions

In Section 4.1 and in Section 11.1 we defined Jacobi's elliptic function $\mathrm{sn}(u,k)$ as the inverse function of the incomplete elliptic integral of the first kind,

$$u = \int_0^{\mathrm{sn}(u,k)} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}. \tag{17.1}$$

On the other hand, Theorem 15.8 (Exercise 15.9) claims that any elliptic function is expressed as a ratio of theta functions.

Then the natural question which follows from this context is:

How are Jacobi's elliptic functions expressed as ratios of theta functions?

It is an interesting exercise to find expressions of sn, cn and dn in terms of theta functions by examining their properties, but here we give the answer first and then prove that those ratios are really the same functions as those introduced in terms of the elliptic integral.

We fix τ ($\mathrm{Im}\,\tau > 0$) and write $\theta_{kl}(u,\tau)$ as $\theta_{kl}(u)$, unless we have to specify τ. We also use the abbreviation $\theta_{kl} := \theta_{kl}(0)$ as before. In these notations Jacobi's elliptic functions are defined as follows:

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7_18

$$\mathrm{sn}(u) = -\frac{\theta_{00}}{\theta_{10}} \frac{\theta_{11}(v)}{\theta_{01}(v)}, \tag{17.2}$$

$$\mathrm{cn}(u) = \frac{\theta_{01}}{\theta_{10}} \frac{\theta_{10}(v)}{\theta_{01}(v)}, \tag{17.3}$$

$$\mathrm{dn}(u) = \frac{\theta_{01}}{\theta_{00}} \frac{\theta_{00}(v)}{\theta_{01}(v)}, \tag{17.4}$$

where $v := \dfrac{u}{\pi\theta_{00}^2}$.

We are going to show that they are the same as those defined before. The first thing which we have to check is that they are really 'elliptic functions':

- $\mathrm{sn}(u)$, $\mathrm{cn}(u)$, $\mathrm{dn}(u)$ are meromorphic functions on $\mathbb{C}$.
- $\mathrm{sn}(u)$ defined by (17.2) has periods $2\pi\,\theta_{00}^2$ and $\pi\tau\,\theta_{00}^2$.
- $\mathrm{cn}(u)$ defined by (17.3) has periods $2\pi\,\theta_{00}^2$ and $\pi(1+\tau)\,\theta_{00}^2$.
- $\mathrm{dn}(u)$ defined by (17.4) has periods $\pi\,\theta_{00}^2$ and $2\pi\tau\,\theta_{00}^2$.

Meromorphicity is a direct consequence of the fact that theta functions are holomorphic on the whole plane $\mathbb{C}$. Periodicities can be shown by quasi-periodicities (15.14) and (15.15) of theta functions, which we leave to the reader as an exercise.

Exercise 17.1 Verify that the functions defined by (17.2), (17.3) and (17.4) have the periods shown above. (Hint: Do not forget that we use the variable $v = \dfrac{u}{\pi\theta_{00}^2}$!)

Thus we have shown that $\mathrm{sn}(u)$, $\mathrm{cn}(u)$ and $\mathrm{dn}(u)$ defined above are elliptic functions.

From the positions (15.22) of zeros of theta functions and definitions (17.2), (17.3), (17.4) it immediately follows that

$$\mathrm{sn}(0) = 0, \qquad \mathrm{cn}(0) = \mathrm{dn}(0) = 1, \tag{17.5}$$

and

$$\begin{aligned} &\mathrm{sn}(u) = 0 \Leftrightarrow u \in \tilde{\Gamma}, \\ &\mathrm{cn}(u) = 0 \Leftrightarrow u \in \frac{\pi\theta_{00}^2}{2} + \tilde{\Gamma}, \\ &\mathrm{dn}(u) = 0 \Leftrightarrow u \in \frac{(1+\tau)\pi\theta_{00}^2}{2} + \tilde{\Gamma}, \end{aligned} \tag{17.6}$$

where $\tilde{\Gamma} := \mathbb{Z}\pi\theta_{00}^2 + \mathbb{Z}\tau\pi\theta_{00}^2$. (Note that this $\tilde{\Gamma}$ is *not* the period lattice of sn, cn, dn.) The positions of poles are common to sn, cn and dn:

$$\{\text{poles of sn (or cn or dn)}\} = \frac{\tau\pi\theta_{00}^2}{2} + \tilde{\Gamma}. \tag{17.7}$$

All poles are simple. It is easy to see that each period parallelogram contains exactly two poles, which means that Jacobi's elliptic functions are of order two.

The parities,

$$\operatorname{sn}(u): \text{ an odd function}, \quad \operatorname{cn}(u), \operatorname{dn}(u): \text{ even functions}, \tag{17.8}$$

also follow directly from the parities of theta functions (15.17) and definitions (17.2), (17.3) and (17.4).

The following properties were definitions of $\operatorname{cn}(u)$ and $\operatorname{dn}(u)$, when they were defined over $\mathbb{R}$. In the present context these formulae are reduced to the addition theorems of theta functions.

Lemma 17.2

$$\operatorname{sn}^2 u + \operatorname{cn}^2 u = 1, \qquad k^2 \operatorname{sn}^2 u + \operatorname{dn}^2 u = 1.$$

Here we define the modulus *k by*

$$k = k(\tau) := \frac{\theta_{10}^2}{\theta_{00}^2}. \tag{17.9}$$

Proof By the definitions (17.2) and (17.3) the left-hand side of the first equation is equal to

$$\operatorname{sn}^2 u + \operatorname{cn}^2 u = \frac{\theta_{00}^2\, \theta_{11}(v)^2 + \theta_{01}^2\, \theta_{10}(v)^2}{\theta_{10}^2\, \theta_{01}(v)^2}. \tag{17.10}$$

Let us put $x \mapsto v$ and $u \mapsto \dfrac{1+\tau}{2}$ into the last half of the addition theorem (A1) of Corollary 15.16,

$$\theta_{00}(x+u)\, \theta_{00}(x-u)\, \theta_{00}^2 = \theta_{01}(x)^2\, \theta_{01}(u)^2 + \theta_{10}(x)^2\, \theta_{10}(u)^2.$$

Rewriting the obtained formula by the transformation rules (15.16) of the characteristics of theta functions and removing exponential factors common to both sides, we obtain

$$\theta_{11}(v)^2\, \theta_{00}^2 = \theta_{01}(v)^2\, \theta_{10}^2 - \theta_{10}(v)^2\, \theta_{01}^2.$$

From this and (17.10) the equation $\operatorname{sn}^2 u + \operatorname{cn}^2 u = 1$ follows.

The next formula is obtained from the first half of the addition theorem (A1) of Corollary 15.16,

$$\theta_{00}(x+u)\, \theta_{00}(x-u)\, \theta_{00}^2 = \theta_{00}(x)^2\, \theta_{00}(u)^2 + \theta_{11}(x)^2\, \theta_{11}(u)^2.$$

Substituting $x \mapsto v$, $u \mapsto \dfrac{1}{2}$ into this formula and using the transformation rules (15.16) of the characteristics again, we obtain

$$\theta_{01}(v)^2\, \theta_{00}^2 = \theta_{00}(v)^2\, \theta_{01}^2 + \theta_{11}(v)^2\, \theta_{10}^2.$$

Dividing both sides by $\theta_{01}(v)^2\, \theta_{00}^2$, we have

$$1 = \frac{\theta_{10}^4}{\theta_{00}^4} \frac{\theta_{00}^2 \theta_{11}(v)^2}{\theta_{10}^2 \theta_{01}(v)^2} + \frac{\theta_{01}^2 \theta_{00}(v)^2}{\theta_{00}^2 \theta_{01}(v)^2},$$

which is equivalent to $k^2 \operatorname{sn}^2(u) + \operatorname{dn}^2(u) = 1$. □

The key to the relation with an elliptic integral is the following derivative formula.

Lemma 17.3

$$\frac{d}{du} \operatorname{sn}(u) = \operatorname{cn}(u) \operatorname{dn}(u). \tag{17.11}$$

Proof Let us differentiate the definition (17.2) by u. As $v = \dfrac{u}{\pi \theta_{00}^2}$, we have

$$\begin{aligned} \frac{d}{du} \operatorname{sn}(u) &= \frac{dv}{du} \frac{d}{dv} \left(-\frac{\theta_{00}}{\theta_{10}} \frac{\theta_{11}(v)}{\theta_{01}(v)} \right) \\ &= -\frac{1}{\pi \theta_{00} \theta_{10}} \frac{\theta_{11}'(v)\theta_{01}(v) - \theta_{11}(v)\theta_{01}'(v)}{\theta_{01}(v)^2}. \end{aligned} \tag{17.12}$$

In order to rewrite the numerator, we need the coefficient of u^1 in the Taylor expansion of the addition formula (A3),

$$\begin{aligned} &\theta_{11}(x+u)\theta_{01}(x-u)\theta_{10}\theta_{00} \\ &\qquad = \theta_{00}(x)\theta_{10}(x)\theta_{01}(u)\theta_{11}(u) + \theta_{01}(x)\theta_{11}(x)\theta_{00}(u)\theta_{10}(u), \end{aligned}$$

of Corollary 15.16 around $u = 0$. Being an even function of u, the expansion of the second term in the right-hand side of (A3) does not have a term of order u^1. Since $\theta_{11}(0) = 0$, we obtain

$$\big(\theta_{11}'(x)\,\theta_{01}(x) - \theta_{11}(x)\,\theta_{01}'(x)\big)\theta_{10}\,\theta_{00} = \theta_{00}(x)\,\theta_{10}(x)\,\theta_{01}\,\theta_{11}'$$

from the terms of order u^1. Changing the variable x in this formula to v and applying it to (17.12), we can prove

$$\begin{aligned} \frac{d}{du} \operatorname{sn}(u) &= -\frac{1}{\pi \theta_{00} \theta_{10}} \frac{\theta_{00}(v)\,\theta_{10}(v)\,\theta_{01}\,\theta_{11}'}{\theta_{00}\,\theta_{10}\,\theta_{01}(v)^2} \\ &= \frac{\theta_{01}^2}{\theta_{00}\theta_{10}} \frac{\theta_{00}(v)\,\theta_{10}(v)}{\theta_{01}(v)^2} = \operatorname{cn}(u) \operatorname{dn}(u). \end{aligned}$$

We used Jacobi's derivative formula (15.33) when we moved to the second line. □

When we discussed Jacobi's elliptic functions over $\mathbb{R}$, the derivative formula (17.11) was a consequence of the definition 'sn(u) is the inverse function of the incomplete elliptic integral of the first kind' and the derivation formula of the inverse function. Tracing back that discussion, we can show that the function $\mathrm{sn}(u)$ defined by (17.2) as a ratio of theta functions is the 'inverse function' of the incomplete elliptic integral of the first kind. Let us have a look at the storyline in the slightly rough discussion below.

The equation (17.11) can be rewritten solely in terms of $\mathrm{sn}(u)$ as

$$\frac{d}{du}\,\mathrm{sn}(u) = \sqrt{(1-\mathrm{sn}^2(u))(1-k^2\,\mathrm{sn}^2(u))}$$

with the help of Lemma 17.2. This means that $\mathrm{sn}(u)$ is a solution of a differential equation

$$\frac{dz}{du} = \sqrt{(1-z^2)(1-k^2z^2)}. \tag{17.13}$$

Hence the inverse function $u = u(z)$ of $z = \mathrm{sn}(u)$ has the derivative

$$\begin{aligned}\frac{du}{dz} = \left(\frac{dz}{du}\right)^{-1} &= \frac{1}{\sqrt{(1-\mathrm{sn}^2(u))(1-k^2\,\mathrm{sn}^2(u))}}\\ &= \frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}}\end{aligned}$$

by the derivation formula of the inverse function. Integration of this formula from 0 to $\mathrm{sn}(u)$ by z gives

$$u = \int_0^{\mathrm{sn}(u)} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}. \tag{17.14}$$

(Here we used $u(0) = 0$, which follows from $\mathrm{sn}(0) = 0$.) Thus the definitions of $\mathrm{sn}(u)$ in the real case (17.1) and in the complex case (17.2) are connected.

Why did we describe the above discussion as 'slightly rough'? There are two reasons. One is that we do not fix the sign of the root in the differential equation (17.13). Those who have read this book up to here would have responded "If we have a problem of signs, we need a Riemann surface." immediately. Indeed, as we take the root, we must regard z in the equation (17.13) as the coordinate of a point (z,w) on the Riemann surface $\mathcal{R} = \{(z,w) \mid w^2 = (1-z^2)(1-k^2z^2)\}$. Another coordinate w corresponds to $\dfrac{dz}{du} = \mathrm{sn}'(u)$ as is shown from the equation (17.13). On the other hand, $\mathrm{sn}'(0) = 1$ follows from (17.12) ($u = 0$, i.e., $v = 0$) and Jacobi's derivative formula (15.33). Therefore we can assume that the differential equation (17.13) is a priori considered in a neighbourhood of the point $(z,w) = (0,1)$ of $\mathcal{R}$.

"Uh-oh, only in a neighbourhood?" Yes, at least we have to start from a small neighbourhood. This is another 'rough' part of the above discussion. For an arbitrary integration contour the value of the integral (17.14) cannot be determined uniquely even when terminal points 0 and $\mathrm{sn}(u)$ are fixed, because the periods of the Abelian differential of the first kind are non-zero (cf. Lemma 10.2). After all, a periodic function like $\mathrm{sn}(u)$ cannot have an 'inverse function' on the whole plane $\mathbb{C}$. The relation (17.14) holds only when both the domain of definition for u and the image of $z = \mathrm{sn}(u)$ are restricted to a small domain, exactly speaking, to a neighbourhood of $u = 0$ and to a neighbourhood of $(z, w) = (0, 1)$ on the Riemann surface $\mathcal{R}$. Once this relation between neighbourhoods is established, we can extend it analytically to $\mathbb{C}$ for u and justify (17.14) by suitably extending the integration contour on the elliptic curve.

Let us take this opportunity to clarify yet another 'rough argument', which we have deliberately not mentioned so far. We have denoted the function sn by $\mathrm{sn}(u)$ since we defined it by (17.2). But in the real case we used the notation $\mathrm{sn}(u, k)$. The reader might think "Is this a problem? Isn't it nothing more than the abbreviation?" Is that really so? Let us review the definition (17.2). There appears τ in it, but not k. The modulus k is defined by (17.9) separately. This means that the function defined by (17.2) should be denoted as '$\mathrm{sn}(u, \tau)$' in principle.

"Then, is the notation $\mathrm{sn}(u, k)$ false?!" Do not panic. The variable u and '$z = \mathrm{sn}(u, \tau)$' are connected by the relation (17.14). This relation contains k but not τ explicitly. Therefore, if τ_1 and τ_2 give the same k, $k(\tau_1) = k(\tau_2)$, this relation implies '$\mathrm{sn}(u, \tau_1) = \mathrm{sn}(u, \tau_2)$' as '$\mathrm{sn}(u, \tau)$' is the inverse function of the same elliptic integral. In other words, the solution of the differential equation (17.13), or

$$\left(\frac{dz}{du}\right)^2 = (1 - z^2)(1 - k^2 z^2),$$

with the initial value $z(0) = 0$ is unique and depends on τ only through $k = k(\tau)$. Hence the function '$\mathrm{sn}(u, \tau)$' can be regarded as a function of u and k and the notation '$\mathrm{sn}(u, k)$' is justified.

17.2 Inversion from k^2 to τ

In the previous section we defined Jacobi's sn-function as a ratio of theta functions $\theta_{kl}(u, \tau)$, which depends on $k = k(\tau) = \dfrac{\theta_{10}^2}{\theta_{00}^2}$ rather than on τ. In fact, when we considered sn as a real function in Chapter 4 and when we applied it to physics in Chapter 5, the parameter used was not τ but the modulus k.

So, a natural question arises:

Does τ exist for a given modulus k?

This is the problem of existence of the solution τ of the equation

$$k = \frac{\theta_{10}(0, \tau)^2}{\theta_{00}(0, \tau)^2}$$

for a given k. Or, since the modulus k appears only in the form k^2, we have only to find τ which satisfies the equation

$$\mu = \frac{\theta_{10}(0,\tau)^4}{\theta_{00}(0,\tau)^4} \tag{17.15}$$

for a given number $\mu = k^2$.

In this section we prove the following proposition, which claims that the above equation is almost always solvable. (See also §21.7 of [WW].)

Proposition 17.4 *The equation* (17.15) *has a solution τ in the upper half plane* H *if and only if $\mu \neq 0$,* 1.

Remark 17.5 There are infinitely many solutions $\tau \in \mathsf{H}$ for each $\mu \neq 0$, 1. We shall remark on this again at the end of this section.

The right-hand side of (17.15),

$$\lambda(\tau) := \frac{\theta_{10}(0,\tau)^4}{\theta_{00}(0,\tau)^4}, \tag{17.16}$$

is called the *lambda function* or the *modular lambda function*. We define the *complementary lambda function* by

$$\tilde{\lambda}(\tau) := \frac{\theta_{01}(0,\tau)^4}{\theta_{00}(0,\tau)^4}, \tag{17.17}$$

which is the square of the *complementary modulus*,

$$k'(\tau) := \frac{\theta_{01}(0,\tau)^2}{\theta_{00}(0,\tau)^2}. \tag{17.18}$$

In the real case we defined the complementary modulus by $k' = \sqrt{1-k^2}$. (See Theorem 3.2.) In the complex case, this formula cannot determine the sign, so we define k' by (17.18). Indeed,

$$k(\tau)^2 + k'(\tau)^2 = 1, \tag{17.19}$$

namely,

$$\lambda(\tau) + \tilde{\lambda}(\tau) = 1, \tag{17.20}$$

follows immediately from $\theta_{10}^4 + \theta_{01}^4 = \theta_{00}^4$ ((A4) in Corollary 15.16).

It is obvious that $\mu = \lambda(\tau)$ cannot be 0, because $\theta_{10}(0,\tau) \neq 0$. (Recall that $\theta_{10}(u,\tau)$ has a zero in the set $\frac{1}{2} + \mathbb{Z} + \mathbb{Z}\tau$, (15.22).) Similarly, since $\theta_{01}(0,\tau) \neq 0$, $\tilde{\lambda}(\tau)$ cannot be 0, from which $\lambda(\tau) \neq 1$ follows because of (17.20). This exclusion of $\lambda = 0$ and $\lambda = 1$, or, equivalently, $k = 0$ and $k = \pm 1$ is quite natural, since the quartic polynomial

$\varphi(z) = (1-z^2)(1-k^2z^2)$ has multiple roots if $k=0$ or ± 1. In such cases, as we saw in Chapter 2, integrals with $\sqrt{\varphi(z)}$ are not elliptic integrals. Moreover the equation $w^2 = \varphi(z)$ does not define an elliptic curve, as the two slits $[\pm 1, \pm 1/k]$ in Fig. 8.4 degenerate or merge to one slit.

Let us prove the existence of a solution of (17.15), assuming that μ is neither 0 nor 1. For that purpose we need several properties of $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ (or, equivalently, properties of $k(\tau)$ and $k'(\tau)$).

Lemma 17.6

(17.21)
$$k(\tau) = 4q^{1/2}\frac{\prod_{n=1}^{\infty}(1+q^{2n})^4}{\prod_{n=1}^{\infty}(1+q^{2n-1})^4}, \qquad \lambda(\tau) = 16q\frac{\prod_{n=1}^{\infty}(1+q^{2n})^8}{\prod_{n=1}^{\infty}(1+q^{2n-1})^8},$$
$$k'(\tau) = \frac{\prod_{n=1}^{\infty}(1-q^{2n-1})^4}{\prod_{n=1}^{\infty}(1+q^{2n-1})^4}, \qquad \tilde{\lambda}(\tau) = \frac{\prod_{n=1}^{\infty}(1-q^{2n-1})^8}{\prod_{n=1}^{\infty}(1+q^{2n-1})^8},$$

where $q = e^{\pi i \tau}$.

This is a direct consequence of the infinite product factorisations of theta functions, (16.22), and the definitions, (17.16), (17.17).

Lemma 17.7 (Modular properties)

(i) When τ is shifted by ± 1, $k(\tau)$ and $k'(\tau)$ transform as

(17.22)
$$k(\tau\pm 1) = \pm i\frac{k(\tau)}{k'(\tau)}, \qquad k'(\tau\pm 1) = \frac{1}{k'(\tau)}.$$

Correspondingly, by this transformation $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ transform as

(17.23)
$$\lambda(\tau\pm 1) = -\frac{\lambda(\tau)}{\tilde{\lambda}(\tau)} = \frac{\lambda(\tau)}{\lambda(\tau)-1}, \qquad \tilde{\lambda}(\tau\pm 1) = \frac{1}{\tilde{\lambda}(\tau)}.$$

(ii) When τ is shifted by 2, $k(\tau)$ changes the sign and $k'(\tau)$ does not change:

(17.24)
$$k(\tau+2) = -k(\tau), \qquad k'(\tau+2) = k'(\tau).$$

Hence $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ are preserved by this transformation:

(17.25)
$$\lambda(\tau+2) = \lambda(\tau), \qquad \tilde{\lambda}(\tau+2) = \tilde{\lambda}(\tau).$$

(iii) By the transformation $\tau \mapsto -\dfrac{1}{\tau}$, $k(\tau)$ and $k'(\tau)$ are interchanged:

$$k\left(-\frac{1}{\tau}\right) = k'(\tau), \qquad k'\left(-\frac{1}{\tau}\right) = k(\tau). \tag{17.26}$$

Correspondingly, by this transformation $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ are interchanged:

$$\lambda\left(-\frac{1}{\tau}\right) = \tilde{\lambda}(\tau), \qquad \tilde{\lambda}\left(-\frac{1}{\tau}\right) = \lambda(\tau). \tag{17.27}$$

(iv) When τ is changed to $\dfrac{-\tau}{2\tau-1}$, $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ do not change:

$$\lambda\left(\frac{-\tau}{2\tau-1}\right) = \lambda(\tau), \qquad \tilde{\lambda}\left(\frac{-\tau}{2\tau-1}\right) = \tilde{\lambda}(\tau). \tag{17.28}$$

The second statement (ii) follows immediately from (i). The fourth statement (iv) follows from (ii) and (iii), because $\tau \mapsto \dfrac{-\tau}{2\tau-1}$ is a composition of the following transformations:

$$\tau \mapsto \tau_1 := -\frac{1}{\tau} \mapsto \tau_2 := \tau_1 + 2 \mapsto \tau_3 := -\frac{1}{\tau_2}.$$

The other statements are consequences of modular properties of theta functions in Section 15.5.

Exercise 17.8 Check (i), (iii) and (iv) of Lemma 17.7. (Hint: Use (15.44) for (i) and (15.43) for (iii).)

Remark 17.9 As was mentioned at the end of Section 15.5, the shift $\tau \mapsto \tau + 1$ corresponds to a matrix $T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and the transformation $\tau \mapsto -\dfrac{1}{\tau}$ corresponds to a matrix $S := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The above lemma shows that $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ are invariant under modular transformations corresponding to $T^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ ((ii) of Lemma 17.7) and $ST^2S = \begin{pmatrix} -1 & 0 \\ 2 & -1 \end{pmatrix}$ ((iv) of Lemma 17.7).

It is known that T^2 and ST^2S generate a subgroup of $SL(2,\mathbb{Z})$,

$$\Gamma(2) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod 2 \right\}, \tag{17.29}$$

which is called the *principal congruence subgroup of level* 2. Therefore, $\lambda(\tau)$ and $\tilde{\lambda}(\tau)$ are invariant under a modular transformation belonging to $\Gamma(2)$: for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(2)$,

$$\lambda\left(\frac{a\tau+b}{c\tau+d}\right) = \lambda(\tau), \qquad \tilde{\lambda}\left(\frac{a\tau+b}{c\tau+d}\right) = \tilde{\lambda}(\tau),$$

and called *modular functions of level* 2 or *modular forms of weight* 0 *and level* 2. For modular functions and modular forms we refer, for example, to §9 of Chapter 1 of [M], Chapter 4 of the 2nd part or §4, Chapter 7 of the 3rd part of [HC], or Chapter 7 of [Se].

Let us return to the equation (17.15), $\mu = \lambda(\tau)$. We prove the existence of τ for (I) $0 < \mu < 1$, (II) $\mu < 0$, (III) $\mu > 1$ and (IV) $\mu \notin \mathbb{R}$ separately. (We have already excluded the cases $\mu = 0$ and $\mu = 1$.)

(I) Case $0 < \mu < 1$.

When $\tau = it$ ($t \in \mathbb{R}_{>0}$) goes up from 0 to $i\infty$ along the imaginary axis, $q = e^{\pi i \tau} = e^{-\pi t}$ moves from 1 to 0 in the real interval $(0,1)$ and, correspondingly, each factor $\left(\dfrac{1-q^{2n-1}}{1+q^{2n-1}}\right)^8$ of the factorisation (17.21) of $\tilde{\lambda}(\tau)$ increases from 0 to 1 monotonically. Therefore a continuous function $\tilde{\lambda}(it)$ is monotonically increasing in t and attains the value $1-\mu \in (0,1)$ by the intermediate value theorem. Hence there exists a $t \in (0,+\infty)$ such that $\lambda(it) = 1 - \tilde{\lambda}(it) = \mu$.

(II) Case $\mu < 0$.

The argument is similar to (I), but we seek τ not on the imaginary axis but on the half lines $\{\pm 1 + it \mid t \in (0,+\infty)\}$. By the formula (17.23), $\tilde{\lambda}(\pm 1 + it)$ is equal to $\dfrac{1}{\tilde{\lambda}(it)}$, which monotonically decreases from $+\infty$ to 1 as t moves from 0 to $+\infty$ because $\tilde{\lambda}(it)$ increases from 0 to 1.

Therefore the intermediate value theorem assures us that there exists a $t \in (0,+\infty)$ such that $\tilde{\lambda}(\pm 1 + it) = 1 - \mu$ for any $\mu \in (-\infty, 0)$, and thus the existence of a solution t of the equation $\lambda(\pm 1 + it) = \mu$ has been proved.

(III) Case $\mu > 1$.

In this case τ can be found on a semicircle, $\left\{\tau \,\middle|\, \operatorname{Im}\tau > 0,\ \left|\tau - \dfrac{1}{2}\right| = \dfrac{1}{2}\right\}$. This semicircle corresponds to the half line $\{\tilde{\tau} = -1 + it \mid t \in (0,+\infty)\}$ by $\tilde{\tau} = -\dfrac{1}{\tau}$. As we have shown in case (II), there exists a $\tilde{\tau}$ on this half line, which satisfies $\tilde{\lambda}(\tilde{\tau}) = \mu$ for any $\mu > 1$. Because of the property (17.27) $\tau = -\dfrac{1}{\tilde{\tau}}$ is a solution of $\lambda(\tau) = \mu$. (Similarly, we can find another solution on the semicircle $\left\{\tau \,\middle|\, \operatorname{Im}\tau > 0,\ \left|\tau + \dfrac{1}{2}\right| = \dfrac{1}{2}\right\}$.)

(IV) Case $\mu \notin \mathbb{R}$.

In this case we shall show that there is one solution of the equation (17.15), $\mu = \lambda(\tau)$, in the shaded region of Fig. 17.1.

For that purpose, we compute the number of zeros of $\lambda(\tau) - \mu$ in the bounded domain $ABCDD'C'B'A'$ in Fig. 17.1 by the argument principle (Theorem A.10), i.e., by the integral

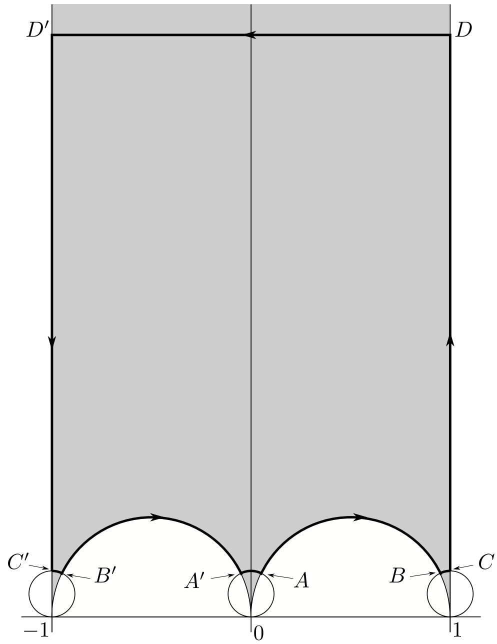

Fig. 17.1 Integration contour for (17.30).

$$\text{(17.30)}\quad \frac{1}{2\pi i}\int_{ABCDD'C'B'A'} \frac{\dfrac{d}{d\tau}(\lambda(\tau)-\mu)}{\lambda(\tau)-\mu}\,d\tau = \frac{1}{2\pi i}\int_{ABCDD'C'B'A'} d\log(\lambda(\tau)-\mu).$$

In Fig. 17.1 the segment DD' is a part of a horizontal line $\operatorname{Im}\tau = M$ for $M > 0$. The three small circles tangent to the real axis have the same radius $\dfrac{i}{2M}$. The two semicircles are those considered in case (III) and have diameters $[-1,0]$ and $[0,1]$ respectively. When M goes to infinity, the bounded domain $ABCDD'C'B'A'$ eventually covers the shaded region.

The following geometric properties, which can be checked by direct computation, are important in the computations below.

Exercise 17.10 (i) Show that the arc $A'A$ corresponds to the segment DD' by the transformation $\tau \mapsto -\dfrac{1}{\tau}$.

(ii) Show that the arc AB (respectively, $B'A'$) corresponds to the segment $D'C'$ (respectively, CD) by the transformation $\tau \mapsto -\dfrac{1}{\tau}$.

(iii) Show that the arc AB corresponds to the arc $B'A'$ by the transformation $\tau \mapsto \dfrac{-\tau}{2\tau-1}$. (Hint: This fractional linear transformation is the one which appeared in Lemma 17.7. As in its proof (Exercise 17.8), decompose it into three parts and use (ii).)

The integrals along the segments CD and $D'C'$ in (17.30) cancel, because these segments correspond to each other by $\tau \mapsto \tau+2$ and $\lambda(\tau)$ is invariant under this transformation, (17.25):

$$\begin{aligned}
&\left(\int_{CD} + \int_{D'C'}\right) d\log(\lambda(\tau)-\mu) \\
&= \int_{C'D'} d\log(\lambda(\tau+2)-\mu) + \int_{D'C'} d\log(\lambda(\tau)-\mu) \\
&= \int_{C'D'} d\log(\lambda(\tau)-\mu) - \int_{C'D'} d\log(\lambda(\tau)-\mu) = 0.
\end{aligned} \tag{17.31}$$

Similarly, the integrals along the arcs AB and $B'A'$ cancel because of the invariance (17.28). (See also Exercise 17.10 (iii).):

$$\begin{aligned}
&\left(\int_{AB} + \int_{B'A'}\right) d\log(\lambda(\tau)-\mu) \\
&= \int_{AB} d\log(\lambda(\tau)-\mu) + \int_{BA} d\log\left(\lambda\left(\frac{-\tau}{2\tau-1}\right)-\mu\right) \\
&= \int_{AB} d\log(\lambda(\tau)-\mu) - \int_{AB} d\log(\lambda(\tau)-\mu) = 0.
\end{aligned} \tag{17.32}$$

The integral (17.30) on the segment DD' is rewritten as

$$\int_{DD'} d(\log(\lambda(\tau)-\mu)) = -\int_{-1}^{1} \frac{1}{\lambda(t+iM)-\mu} \frac{d\lambda}{d\tau}(t+iM)\,dt \tag{17.33}$$

where the variable τ is replaced by $t+iM$ ($t \in [-1,1]$). As $q = e^{\pi i\tau} = e^{-M}e^{\pi it}$, it follows from the infinite product expression (17.21) that $\lambda(t+iM)$ converges to 0 uniformly on $[-1,1]$ as $M \to \infty$: $\lim_{M\to\infty} \lambda(t+iM) = 0$. Hence,

$$\lim_{M\to\infty} \frac{1}{\lambda(t+iM)-\mu} = -\frac{1}{\mu} \tag{17.34}$$

uniformly on $[-1,1]$.

Taking the logarithm of (17.21) and differentiating, we have

$$\frac{1}{\lambda(\tau)} \frac{d\lambda}{d\tau}(\tau) = \pi i + 8\pi i q \left(\sum_{n=1}^{\infty} \frac{2nq^{2n-1}}{1+q^{2n}} - \sum_{n=1}^{\infty} \frac{(2n-1)q^{2n-2}}{1+q^{2n-1}}\right),$$

from which it follows that

$$\lim_{M\to\infty} \frac{d\lambda}{d\tau}(t+iM) = 0 \tag{17.35}$$

uniformly on $[-1,1]$. From these estimates we can take the limit of (17.33):

$$\lim_{M\to\infty}\int_{DD'} d(\log(\lambda(\tau)-\mu)) = 0. \tag{17.36}$$

The integral (17.30) on the arc $A'A$ is transformed to an integral on the segment DD' by the change of variables $\tau \mapsto -\dfrac{1}{\tau}$, using (17.27) and Exercise 17.10 (i):

$$\int_{A'A} d(\log\lambda(\tau)-\mu) = -\int_{-1}^{1} \frac{1}{\tilde{\lambda}(t+iM)-\mu}\,\frac{d\tilde{\lambda}}{d\tau}(t+iM)\,dt. \tag{17.37}$$

Similarly to (17.34) and (17.35) we can show that

$$\lim_{M\to\infty}\frac{1}{\tilde{\lambda}(t+iM)-\mu} = \frac{1}{1-\mu}, \qquad \lim_{M\to\infty}\frac{d\tilde{\lambda}}{d\tau}(t+iM) = 0$$

uniformly on $[-1,1]$, by using the infinite product expression (17.21) of $\tilde{\lambda}(\tau)$. Hence the limit of (17.37) becomes 0:

$$\lim_{M\to\infty}\int_{A'A} d(\log(\lambda(\tau)-\mu)) = 0. \tag{17.38}$$

The remaining part of the integral (17.30) is the integral on the arcs BC and $C'B'$. First shifting $\tau \mapsto \tau' := \tau \pm 1$ and then using the change of variable $\tau' \mapsto \tilde{\tau} := -\dfrac{1}{\tau'}$, we can reduce the integral on BC and $C'B'$ to an integral on the segment DD' because of Exercise 17.10 (i):

$$\begin{aligned}
&\left(\int_{BC}+\int_{C'B'}\right) d\log(\lambda(\tau)-\mu) \\
&= \int_{DD'} d\log\left(-\frac{\tilde{\lambda}(\tilde{\tau})}{\lambda(\tilde{\tau})}-\mu\right) = \int_{DD'} \frac{1}{-\dfrac{\tilde{\lambda}(\tilde{\tau})}{\lambda(\tilde{\tau})}-\mu}\,\frac{d}{d\tilde{\tau}}\left(-\frac{\tilde{\lambda}(\tilde{\tau})}{\lambda(\tilde{\tau})}-\mu\right)d\tilde{\tau} \\
&\qquad = \int_{DD'} \frac{1}{1+\mu\dfrac{\lambda(\tilde{\tau})}{\tilde{\lambda}(\tilde{\tau})}}\,\frac{d}{d\tilde{\tau}}\log\left(-\frac{\tilde{\lambda}(\tilde{\tau})}{\lambda(\tilde{\tau})}\right)d\tilde{\tau}.
\end{aligned} \tag{17.39}$$

Here we used formulae (17.23) and (17.27) to rewrite the integrand.

With the help of the infinite product factorisation

$$\frac{\lambda(\tau)}{\tilde{\lambda}(\tau)} = 16q\,\frac{\displaystyle\prod_{n=1}^{\infty}(1+q^{2n})^8}{\displaystyle\prod_{n=1}^{\infty}(1-q^{2n-1})^8},$$

which is obtained from (17.21), we can show that

$$\lim_{M\to\infty} \frac{1}{1+\mu\dfrac{\lambda(t+iM)}{\tilde\lambda(t+iM)}} = 1, \qquad \lim_{M\to\infty} \frac{d}{d\tilde\tau} \log\left(-\frac{\tilde\lambda(\tilde\tau)}{\lambda(\tilde\tau)}\right)\bigg|_{\tilde\tau=t+iM} = -\pi i,$$

uniformly. Therefore the limit of the integral (17.39) is

$$\lim_{M\to\infty}\left(\int_{BC} + \int_{C'B'}\right) d\log(\lambda(\tau)-\mu) = 2\pi i. \tag{17.40}$$

Thus, putting the results (17.31), (17.32), (17.36), (17.38) and (17.40) together, we obtain the number of zeros of $\lambda(\tau)-\mu$ by (17.30):

$$\lim_{M\to\infty} \frac{1}{2\pi i}\int_{ABCDD'C'B'A'} d\log(\lambda(\tau)-\mu) = 1.$$

This is the end of the proof of Proposition 17.4. □

Remark 17.11 We can construct $\lambda(\tau)$ geometrically by the reflection principle (cf. Section 9.2). See §4, Chapter 7 of the 3rd part of [HC] or §5.3b of [Ko]. In these references the function is defined on a disk, so we need to compose a fractional linear transformation which maps the upper half plane to the disk.

Remark 17.12 Since the lambda function is invariant under the action of the principal congruence subgroup $\Gamma(2)$ (cf. Remark 17.9), if τ satisfies $\lambda(\tau)=\mu$, $\dfrac{a\tau+b}{c\tau+d}$ also satisfies $\lambda\left(\dfrac{a\tau+b}{c\tau+d}\right)=\mu$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(2)$. The result of Proposition 17.4 can be interpreted as follows: The lambda function $\lambda(\tau)$ gives a bijection between the quotient space $\mathsf{H}/\Gamma(2)$ and $\mathbb{C}\setminus\{0,1\}$. The shaded region in Fig. 17.1 is called the *fundamental domain* of $\Gamma(2)$.

In fact Gauss knew this fundamental domain and drew a figure essentially equivalent to Fig. 17.1 in 1827, which is shown in p. 477, Werke Bd.3 and in p. 105, Werke Bd.8, [G]. The meaning of this figure was not understood when the third volume of his complete works was published in 1866 and only in 1900 did Fricke explain its meaning in the eighth volume. For its history we refer to pp. 46–49 (Chapter 1, Reine Mathematik, Sachliche Ausführungen, Modulformen und Modulfunktionen) of [Kl], p. 27 of [Tkg] and p. 121 of [Ka].

17.3 Properties of $\mathrm{sn}(u,k)$

Now let us return to Jacobi's elliptic functions, in particular, to $\mathrm{sn}(u,k)$. In Section 4.2 we saw the following properties of Jacobi's elliptic function $\mathrm{sn}(u,k)$ defined as a real function:

- it is periodic with period $4K(k)$.
- in the limits $k \to 0$ and $k \to 1$ the elliptic function $\mathrm{sn}(u,k)$ becomes $\sin u$ and $\tanh u$ respectively.
- it satisfies an addition theorem.

There we derived the limits of sn from the limits of the elliptic integral for $k \to 0,1$. Then we showed the addition theorem, using the derivation formula (17.11), and applied the addition theorem to show $\mathrm{sn}(u+4K) = \mathrm{sn}(u)$. Of course these properties can be proved from the definition (17.2) by theta functions.

17.3.1 The addition theorem of sn

Replacing the variables in the addition formulae of theta functions, (A3), (A2) in Corollary 15.16, as $x \mapsto \dfrac{u}{\pi\theta_{00}^2}$ and $u \mapsto \dfrac{v}{\pi\theta_{00}^2}$, taking the ratio and multiplying by $-\dfrac{\theta_{01}^2}{\theta_{10}^2}$, we obtain from the left-hand side,

$$-\frac{\theta_{01}^2}{\theta_{10}^2} \times (\text{ratio of LHSs}) = -\frac{\theta_{00}}{\theta_{10}} \frac{\theta_{11}\left(\dfrac{u+v}{\pi\theta_{00}^2}\right)}{\theta_{01}\left(\dfrac{u+v}{\pi\theta_{00}^2}\right)} = \mathrm{sn}(u+v),$$

and from the right hand side,

$$\begin{aligned}
&-\frac{\theta_{01}^2}{\theta_{10}^2} \times (\text{ratio of RHSs}) \\
&= -\frac{\theta_{01}^2}{\theta_{10}^2} \frac{\begin{array}{c}\theta_{01}\left(\dfrac{u}{\pi\theta_{00}^2}\right)\theta_{11}\left(\dfrac{u}{\pi\theta_{00}^2}\right)\theta_{00}\left(\dfrac{v}{\pi\theta_{00}^2}\right)\theta_{10}\left(\dfrac{v}{\pi\theta_{00}^2}\right) \\ +\theta_{00}\left(\dfrac{u}{\pi\theta_{00}^2}\right)\theta_{10}\left(\dfrac{u}{\pi\theta_{00}^2}\right)\theta_{01}\left(\dfrac{v}{\pi\theta_{00}^2}\right)\theta_{11}\left(\dfrac{v}{\pi\theta_{00}^2}\right)\end{array}}{\theta_{01}\left(\dfrac{u}{\pi\theta_{00}^2}\right)^2\theta_{01}\left(\dfrac{v}{\pi\theta_{00}^2}\right)^2 - \theta_{11}\left(\dfrac{u}{\pi\theta_{00}^2}\right)^2\theta_{11}\left(\dfrac{v}{\pi\theta_{00}^2}\right)^2} \\
&= \frac{\mathrm{sn}(u)\,\mathrm{cn}(v)\,\mathrm{dn}(v) + \mathrm{sn}(v)\,\mathrm{cn}(u)\,\mathrm{dn}(u)}{1-k^2\,\mathrm{sn}(u)^2\,\mathrm{sn}(v)^2}.
\end{aligned}$$

Thus, putting them together, we have proved the addition theorem of $\mathrm{sn}(u)$,

$$\operatorname{sn}(u+v) = \frac{\operatorname{sn}(u)\operatorname{cn}(v)\operatorname{dn}(v) + \operatorname{sn}(v)\operatorname{cn}(u)\operatorname{dn}(u)}{1 - k^2 \operatorname{sn}(u)^2 \operatorname{sn}(v)^2}.$$

□

17.3.2 Limit as $k \to 0$

We want to take the limit as $k = \dfrac{\theta_{10}^2}{\theta_{00}^2} \to 0$ of theta functions. What limit of τ corresponds to this limit? As we saw in the proof of Jacobi's derivative formula (Theorem 15.18) θ_{00} and θ_{10} behave as

$$(17.41)\qquad \begin{aligned} \theta_{00} &= \sum_{n\in\mathbb{Z}} e^{\pi i n^2 \tau} = 1 + O(q), \\ \theta_{10} &= \sum e^{\pi i \left(n+\frac{1}{2}\right)^2 \tau} = 2q^{\frac{1}{4}} + O(q^{\frac{9}{4}}) \end{aligned}$$

in the neighbourhood of $q = 0$, where $q = e^{\pi i \tau}$. Hence, when q tends to 0, $k = k(\tau)$ behaves as

$$k(\tau) = \frac{4q^{\frac{1}{2}} + O(q^{\frac{5}{2}})}{1 + O(q)} \xrightarrow{q\to 0} 0.$$

Or, we can use the infinite product factorisation (17.21) of $k(\tau)$, from which the same asymptotic behaviour $k(\tau) \xrightarrow{q\to 0} 0$ follows.

Since $|q| = e^{-\pi \operatorname{Im}\tau}$, the limit $q \to 0$ is equivalent to $\operatorname{Im}\tau \to +\infty$. So, we have to find the limit of $\operatorname{sn}(u,k)$ defined by (17.2) for $\operatorname{Im}\tau \to +\infty$ or $q \to 0$.

By setting $q = e^{\pi i \tau}$ in the definition (15.12) of theta functions, the numerator and the denominator in the definition (17.2) of sn are expressed as series in q:

$$(17.42)\qquad \begin{aligned} \theta_{11}(v) &= q^{\frac{1}{4}}(ie^{\pi i v} - ie^{-\pi i v}) + q^{\frac{9}{4}}(-ie^{3\pi i v} + ie^{-3\pi i v}) + \cdots \\ &= -2q^{\frac{1}{4}}\sin(\pi v) + O(q^{\frac{9}{4}}), \\ \theta_{01}(v) &= 1 - q(e^{2\pi i v} + e^{-2\pi i v}) + \cdots \\ &= 1 + O(q). \end{aligned}$$

Substituting q-expansions (17.41) and (17.42) into (17.2), we obtain

$$(17.43)\qquad \operatorname{sn}(u) = \sin(\pi v) + O(q) \xrightarrow{q\to 0} \sin(u).$$

(Note that $v = \dfrac{u}{\pi\theta_{00}^2} \to \dfrac{u}{\pi}$.)

In the same way we can show

$$(17.44)\qquad \operatorname{cn}(u) \xrightarrow{q\to 0} \cos(u), \qquad \operatorname{dn}(u) \xrightarrow{q\to 0} 1.$$

(They can be derived also as consequences of (17.43) and the relations in Lemma 17.2.)

Exercise 17.13 Prove the limits (17.44) by using the limits of theta functions.

17.3.3 Jacobi's imaginary transformation and the limit as $k \to 1$

The limit as $k \to 1$ is equivalent to the limit as $k' \to 0$ of the complementary modulus. We have shown $\lim\limits_{k'\to 0} \mathrm{sn}(u,k') = \sin(u)$, $\lim\limits_{k'\to 0} \mathrm{cn}(u,k') = \cos(u)$, $\lim\limits_{k'\to 0} \mathrm{dn}(u,k') = 1$. So, in order to find the limits $\lim\limits_{k\to 1} \mathrm{sn}(u,k)$, $\lim\limits_{k\to 1} \mathrm{cn}(u,k)$ and $\lim\limits_{k\to 1} \mathrm{dn}(u,k)$, let us express $\mathrm{sn}(u,k)$, $\mathrm{cn}(u,k)$ and $\mathrm{dn}(u,k)$ in terms of $\mathrm{sn}(u,k')$, $\mathrm{cn}(u,k')$ and $\mathrm{dn}(u,k')$ or vice versa.

As was shown in Lemma 17.7 (iii), k and k' are interchanged by the transformation $\tau \mapsto -\dfrac{1}{\tau}$. We apply this transformation to the theta functions in (17.2), (17.3), (17.4) and use Jacobi's imaginary transformation formulae (15.43). For example, sn transforms as

$$\begin{aligned}
\mathrm{sn}(u,k') &= -\frac{\theta_{00}\left(0,-\dfrac{1}{\tau}\right)}{\theta_{01}\left(0,-\dfrac{1}{\tau}\right)} \frac{\theta_{11}\left(v',-\dfrac{1}{\tau}\right)}{\theta_{10}\left(v',-\dfrac{1}{\tau}\right)} \\
&= -\frac{\sqrt{-i\tau}\theta_{00}(0,\tau)}{\sqrt{-i\tau}\theta_{01}(0,\tau)} \frac{-i\sqrt{-i\tau}e^{\pi i(\tau v')^2/\tau}\theta_{11}(\tau v',\tau)}{\sqrt{-i\tau}e^{\pi i(\tau v')^2/\tau}\theta_{10}(\tau v',\tau)} = i\frac{\theta_{00}}{\theta_{01}}\frac{\theta_{11}(\tau v')}{\theta_{10}(\tau v')},
\end{aligned}$$

where $v' = \dfrac{u}{\pi\theta_{00}^2\left(0,-\dfrac{1}{\tau}\right)}$. Of course, we have to rewrite this theta function with $-\dfrac{1}{\tau}$ in v' in terms of a theta function with τ:

$$\tau v' = \frac{\tau u}{\pi\,\theta_{00}\left(0,-\dfrac{1}{\tau}\right)^2} = \frac{\tau u}{\pi\,(-i\tau)\,\theta_{00}^2(0,\tau)} = \frac{iu}{\pi\,\theta_{00}^2(0,\tau)} = iv.$$

Summarising, we have

$$\mathrm{sn}(u,k') = i\frac{\theta_{00}}{\theta_{01}}\frac{\theta_{11}(iv)}{\theta_{10}(iv)},$$

which is rewritten as

$$\mathrm{sn}(u,k') = -i\frac{\mathrm{sn}(iu,k)}{\mathrm{cn}(iu,k)}, \tag{17.45}$$

by the definitions (17.2), (17.3). As in the case of theta functions, this formula is called *Jacobi's imaginary transformation* of sn.

Exercise 17.14 Prove Jacobi's imaginary transformations of cn and dn:

$$\mathrm{cn}(u,k') = \frac{1}{\mathrm{cn}(iu,k)}, \quad \mathrm{dn}(u,k') = \frac{\mathrm{dn}(iu,k)}{\mathrm{cn}(iu,k)}.$$

Using (17.43) and (17.44), we can take the limit as $k' \to 1$, which is equivalent to $k \to 0$, of the formula (17.45):

$$\lim_{k'\to 1} \mathrm{sn}(u,k') = \lim_{k\to 0} -i\frac{\mathrm{sn}(iu,k)}{\mathrm{cn}(iu,k)} = -i\frac{\sin(iu)}{\cos(iu)} = \frac{-e^{-u}+e^{u}}{e^{-u}+e^{u}} = \tanh u.$$

Thus we have proved the first formula in Exercise 4.4 in terms of theta functions.

Exercise 17.15 Prove $\mathrm{cn}(u,k), \mathrm{dn}(u,k) \xrightarrow{k\to 1} \dfrac{1}{\cosh u}$ in the same way as above.

17.3.4 The periods of sn are $4K(k)$ and $2iK'(k)$

In this subsection we assume that τ is a purely imaginary number, $\tau = it$ ($t > 0$). Then $q = e^{\pi i\tau} = e^{-\pi t}$ satisfies $0 < q < 1$ and both θ_{00} and θ_{10} are power series in q as in (17.41). Similarly θ_{01} is expressed as

$$\theta_{01} = \sum_{n\in\mathbb{Z}} (-1)^n q^{n^2}.$$

All of these three numbers are real and, because of the relation $\theta_{10}^4 + \theta_{01}^4 = \theta_{00}^4$ ((A4) in Corollary 15.16), inequalities $0 < \theta_{10}^2 < \theta_{00}^2$ and $0 < \theta_{01}^2 < \theta_{00}^2$ hold. They imply

$$0 < k < 1, \qquad 0 < k' < 1,$$

because of the definitions (17.9) and (17.18).[1] This is the condition which we used up to Chapter 5.

We saw in Section 17.1 that $\mathrm{sn}(u,k)$ has periods $2\pi\theta_{00}^2$ and $\pi\tau\theta_{00}^2$, while we know that the real function $\mathrm{sn}(u,k)$ has period $4K(k)$, where $K(k)$ is the complete elliptic integral

$$K(k) = \int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}. \tag{17.46}$$

Moreover in Proposition 8.3 we computed the A- and the B-periods of the elliptic integral

$$u(x) = \int_0^x \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}$$

[1] In the proof of Proposition 17.4, case (I), we showed these inequalities (for λ and $\tilde{\lambda}$) by infinite product factorisations.

and obtained $\Omega_A = 4K(k)$, $\Omega_B = 2iK'(k)$ ($K'(k) := K(k')$). As sn is the inverse function of this elliptic integral in the sense of Example 11.7 (or Example 11.4), sn has periods $4K(k)$ and $2iK'(k)$.

So the period lattice of sn has two bases, $(4K(k), 2iK'(k))$ and $(2\pi\theta_{00}^2, \pi\tau\theta_{00}^2)$:

$$4\mathbb{Z}K(k) + 2\mathbb{Z}iK'(k) = 2\mathbb{Z}\pi\theta_{00}^2 + \mathbb{Z}\pi\tau\theta_{00}^2,$$

which means that there exist integers m_1, m_2, n_1, n_2, such that

$$\begin{aligned} 4K(k) &= 2m_1\pi\theta_{00}^2 + n_1\pi\tau\theta_{00}^2, \\ 2iK'(k) &= 2m_2\pi\theta_{00}^2 + n_2\pi\tau\theta_{00}^2. \end{aligned} \tag{17.47}$$

These relations can also be derived and even refined by observing values of sn and cn at $u = K(k)$. Comparing the definition (17.46) of $K(k)$ and the integral relation (17.14), we have $\mathrm{sn}(K(k),k) = 1$. Hence, $\mathrm{cn}(K(k),k) = 0$ follows from the relation $\mathrm{sn}^2 u + \mathrm{cn}^2 u = 1$ (Lemma 17.2). We know the positions of the zeros of $\mathrm{cn}(u)$ as shown in (17.6), from which it follows that there exist integers $\tilde{m}_1$, $\tilde{n}_1$ such that

$$K(k) = \frac{\pi\theta_{00}^2}{2} + \tilde{m}_1\pi\theta_{00}^2 + \tilde{n}_1\tau\pi\theta_{00}^2, \text{ i.e., } 4K(k) = 2(2\tilde{m}_1+1)\pi\theta_{00}^2 + 4\tilde{n}_1\tau\pi\theta_{00}^2.$$

This recovers the first equation of (17.47) with $m_1 = 2\tilde{m}_1 + 1$ and $n_1 = 4\tilde{n}_1$.

For $K'(k) = K(k')$, since k and k' are interchanged by $\tau \mapsto -\dfrac{1}{\tau}$ (Lemma 17.7 (iii)), we have the same relation with θ_{00} replaced by $\tilde{\theta}_{00} := \theta_{00}\left(0, -\dfrac{1}{\tau}\right)$: there exist integers $\tilde{m}_2$ and $\tilde{n}_2$ such that

$$K'(k) = \frac{2\tilde{n}_2+1}{2}\pi\tilde{\theta}_{00}^2 + \tilde{m}_2\left(-\frac{1}{\tau}\right)\pi\tilde{\theta}_{00}^2.$$

Applying the modular transformation (15.43), we have $\tilde{\theta}_{00}^2 = -i\tau\theta_{00}^2$. By this formula the above equation is rewritten as

$$2iK'(k) = (2\tilde{n}_2+1)\pi\tau\theta_{00}^2 - 2\tilde{m}_2\pi\theta_{00}^2,$$

which recovers the second equation of (17.47) with $m_2 = -\tilde{m}_2$ and $n_2 = 2\tilde{n}_2 + 1$.

Actually, simpler relations hold as follows.

Theorem 17.16

$$K(k) = \frac{\pi}{2}\theta_{00}^2, \qquad K'(k) = \frac{\pi\tau}{2i}\theta_{00}^2. \tag{17.48}$$

This means that the periods $(2\pi\theta_{00}^2, \pi\tau\theta_{00}^2)$ of sn defined by theta functions are equal to $(4K(k), 2iK'(k))$.

Proof As $0 < k < 1$ and $0 < k' < 1$, the complete elliptic integrals $K(k)$ and $K'(k) = K(k')$ defined by (17.46) are positive real numbers. Note that $2\pi\theta_{00}^2$ in the right-hand sides of the equations in (17.47) is a positive real number, while $\pi\tau\theta_{00}^2 = i\pi t\theta_{00}^2$ is a purely imaginary number with a positive imaginary part. Therefore $n_1 = m_2 = 0$ and m_1 and n_2 are positive integers. Thus the relations (17.47) are simplified as

$$K(k) = \frac{m_1\pi}{2}\theta_{00}^2, \qquad K'(k) = \frac{n_2\pi\tau}{2i}\theta_{00}^2, \tag{17.49}$$

where m_1 and n_2 are positive integers.

Let us regard the incomplete elliptic integral

$$u(x) := \int_0^x \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}$$

as a function of $x \in [0,1]$. We saw in Section 4.1 that $u(x)$ satisfies the following properties:

- $u(0) = 0$, $u(1) = K(k)$.
- $u(x)$ is a strictly increasing continuous function. (This is because the integrand is positive.)

Suppose $m_1 > 1$ in (17.49). Then

$$u(0) = 0 < \frac{\pi}{2}\theta_{00}^2 = \frac{K(k)}{m_1} < K(k) = u(1),$$

and by the intermediate value theorem there exists a $c \in (0,1)$ such that

$$\frac{\pi}{2}\theta_{00}^2 = u(c) = \int_0^c \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}.$$

According to (17.14), this means

$$\operatorname{sn}\left(\frac{\pi}{2}\theta_{00}^2\right) = c < 1.$$

However, putting $u = \dfrac{\pi}{2}\theta_{00}^2$ in the definition (17.2) of sn, we have

$$\operatorname{sn}\left(\frac{\pi}{2}\theta_{00}^2\right) = -\frac{\theta_{00}}{\theta_{10}}\frac{\theta_{11}\left(\frac{1}{2}\right)}{\theta_{01}\left(\frac{1}{2}\right)} = -\frac{\theta_{00}}{\theta_{10}}\frac{(-\theta_{10})}{\theta_{00}} = 1.$$

(Here we used the transformation rules of characteristics, (15.16).) This is a contradiction. The first equation in (17.48) has been proved.

By replacing τ with $-\dfrac{1}{\tau}$, k becomes k' (Lemma 17.7 (iii)) and θ_{00} becomes $\sqrt{-i\tau}\theta_{00}$, (15.43). Hence the first equation in (17.48) is rewritten as

$$K(k') = \frac{\pi}{2}\,(-i\tau)\,\theta_{00}^2 = \frac{\pi\tau}{2i}\,\theta_{00}^2,$$

which proves the second equation in (17.48). □

The relations (17.48) allow us to continue K and K' analytically as analytic functions of τ, but as analytic functions of k they become multi-valued functions because of the modular invariance with respect to $\Gamma(2)$ discussed in Section 17.2 (cf. §21.712 of [WW]).

17.4 The Arithmetic-Geometric Mean Revisited

In the previous section we expressed the complete elliptic integral of the first kind $K(k)$ in terms of a theta function. Since theta functions satisfy many good formulae, we can expect that this expression might be useful in studying $K(k)$. Indeed it *is*. Recall that we encountered $K(k)$ in Section 3.1 when we studied the arithmetic-geometric mean. We proved that the arithmetic-geometric mean is expressed by the complete elliptic integral of the first kind by change of variables of integrals. This is a strategy which Gauss took in 1818 (p. 352, Werke Bd.III, [G][2]). As a matter of fact he proved the theorem earlier (1799), essentially using theta functions. See the chapter of 'Zur Theorie des arithmetisch-geometrischen Mittels', Analysis, Werke Bd.X Abt1, [G]. We follow exposition in §9 of [Ka]. For simplicity we assume that $\tau = it$ $(t > 0)$ is purely imaginary.

The key is the following.

Lemma 17.17

$$\theta_{00}(0,2\tau)^2 = \frac{1}{2}\left(\theta_{00}(0,\tau)^2 + \theta_{01}(0,\tau)^2\right), \tag{17.50}$$

$$\theta_{01}(0,2\tau)^2 = \theta_{00}(0,\tau)\,\theta_{01}(0,\tau). \tag{17.51}$$

Proof The first formula (17.50) is a direct consequence of (16.29).

The second formula (17.51) follows from the infinite product expressions of theta zero-values, (16.24):

$$\begin{aligned}
&\theta_{01}(0,2\tau)^2 \\
&= \prod_{n=1}^{\infty}(1-q^{4n})^2 \prod_{n=1}^{\infty}(1-q^{4n-2})^4 \\
&= \prod_{n=1}^{\infty}(1-q^{2n})^2 \prod_{n=1}^{\infty}(1+q^{2n})^2 \prod_{n=1}^{\infty}(1+q^{2n-1})^4 \prod_{n=1}^{\infty}(1-q^{2n-1})^4
\end{aligned}$$

[2] Gauss used a different change of variables from ours.

$$= \prod_{n=1}^{\infty} (1-q^{2n})^2 \prod_{n=1}^{\infty} (1+q^{2n-1})^2 \prod_{n=1}^{\infty} (1-q^{2n-1})^2,$$

due to the formula (16.26). On the other hand the last expression is equal to $\theta_{00}(0,\tau)\,\theta_{01}(0,\tau)$ again by (16.24). □

This lemma says that $\theta_{00}(0,2\tau)^2$ is the arithmetic mean of $\theta_{00}(0,\tau)^2$ and $\theta_{01}(0,\tau)^2$, while $\theta_{01}(0,2\tau)^2$ is the geometric mean of $\theta_{00}(0,\tau)^2$ and $\theta_{01}(0,\tau)^2$.

Therefore, if we can find μ and τ such that

$$a = \mu\,\theta_{00}(0,\tau)^2, \qquad b = \mu\,\theta_{01}(0,\tau)^2,$$

we can express the sequence (a_n, b_n) $(n = 0, 1, 2, \dots)$ defined by $a_0 := a$, $b_0 := b$ and (3.1),

$$a_{n+1} := \frac{a_n + b_n}{2}, \qquad b_{n+1} := \sqrt{a_n b_n},$$

explicitly as follows:

$$a_n = \mu\,\theta_{00}(0,2^n\tau)^2, \qquad b_n = \mu\,\theta_{01}(0,2^n\tau)^2. \tag{17.52}$$

Are there really such μ and τ?

Yes, that is what we proved in Section 17.2. Let us take two positive real numbers a and b, $a > b > 0$. (We exclude the trivial case $a = b$.) Then $0 < \dfrac{b}{a} < 1$, so there exists $\tau = it$ $(t > 0)$ such that

$$k'(\tau) := \frac{\theta_{01}(0,\tau)^2}{\theta_{00}(0,\tau)^2} = \frac{b}{a},$$

as we showed in the proof of Proposition 17.4, Case (I). (Recall that $\tilde{\lambda}(\tau) = (k'(\tau))^2$.) Then μ defined by

$$\mu := \sqrt{\frac{ab}{\theta_{00}(0,\tau)^2\,\theta_{01}(0,\tau)^2}}$$

satisfies

$$a = a_0 = \mu\,\theta_{00}(0,\tau)^2, \qquad b = b_0 = \mu\,\theta_{01}(0,\tau)^2.$$

Now we have an expression (17.52) for the sequences $\{a_n\}$ and $\{b_n\}$. The arithmetic-geometric mean is the common limit of these sequences: $M(a,b) = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.

When n tends to ∞, $2^n\tau = i2^n t$ diverges to $i\infty$ and consequently $q_n := \exp(\pi i 2^n \tau)$ tends to 0. Hence by the expansions (15.35) or by the infinite product expressions (16.24),

$$\lim_{n\to\infty} \theta_{00}(0,2^n\tau) = \lim_{n\to\infty} \theta_{01}(0,2^n\tau) = 1.$$

Thus we have an expression for the arithmetic-geometric mean in terms of a theta function:

$$M(a,b) = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \mu = \frac{a}{\theta_{00}(0,\tau)^2}. \tag{17.53}$$

In particular, if $a = 1$, then $b = \dfrac{b}{a} = k'$ and

$$M(1,k') = \frac{1}{\theta_{00}(0,\tau)^2},$$

which is equal to $\dfrac{\pi}{2K(k)}$ because of (17.48). This completes another proof of Theorem 3.2. (Though k and k' are interchanged, it is not essential.)

We close our journey around the 'fairyland of mathematics' (cf. the Preface) with this theorem, one of the starting points of the theory of elliptic integrals and elliptic functions. The author wishes the reader a pleasant further trip!

Appendix A
Theorems in Analysis and Complex Analysis

This appendix collects together theorems in real and complex analysis which are used in the main text. Most proofs are omitted or sketchy. For details, we refer to textbooks in calculus and complex analysis.

A.1 Convergence Theorems of Integrals

Most of the material collected in this section is cited from [Su], but the details can be found in any calculus textbooks.

The following exchangeability of limit (or differentiation) and integral for the Riemann integral is well known.

Proposition A.1 *Let $I = [a, b]$ be a closed interval and T be a subset of $\mathbb{R}$. If a family $\{f_t(x)\}_{t\in T}$ of continuous functions on I with a parameter t in T converges to $f(x)$ uniformly when $t \longrightarrow t_0 \in \bar{T}$ (the closure of T), then the integral of $f_t(x)$ on I converges to the integral of $f(x)$:*

$$\lim_{t\to t_0} \int_a^b f_t(x)\, dx = \int_a^b \lim_{t\to t_0} f_t(x)\, dx.$$

When T is unbounded from above (respectively, from below), we consider $t \longrightarrow +\infty$ (respectively, $t \longrightarrow -\infty$).

Proposition A.2 *Let $I = [a, b]$ be a closed interval, J be an interval in $\mathbb{R}$ and $f(x,t)$ be a continuous function on $I \times J$.*

Then $F(t) := \displaystyle\int_a^b f(x,t)\, dx$ is a continuous function on J.

Moreover, if $\dfrac{\partial f}{\partial t}(x,t)$ exists and is continuous on $I \times J$, then $F(t)$ is continuously differentiable on J and

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7

$$\frac{dF}{dt}(t) = \int_a^b \frac{\partial f}{\partial t}(x,t)\,dx.$$

The same is true for improper integrals under a condition on convergence.

Proposition A.3 *Let $I = [a,b)$ and J be intervals in $\mathbb{R}$ and $f(x,t)$ be a continuous function on $I \times J$. ($f(x,t)$ might diverge when $x \to b$.) Assume that the integral of $f(x,t)$ on $[a,u] \subset I$,*

$$F_u(t) := \int_a^u f(x,t)\,dx,$$

converges to the improper integral

$$F(t) := \int_a^b f(x,t)\,dx$$

uniformly on any compact set K in J when u tends to b, which means that

$$\lim_{u \to b-0} \left(\sup_{t \in K} |F(t) - F_u(t)| \right) = 0.$$

Then,
(i) $F(t)$ is continuous on J.
(ii) For any bounded closed interval $[c,d] \in J$,

$$\int_c^d F(t)\,dt = \int_a^b \left(\int_c^d f(x,t)\,dt \right) dx.$$

Proof (i) For any $u \in [a,b)$ $F_u(t)$ is continuous on J due to Proposition A.2. Since $F_u(t)$ converges to $F(t)$ uniformly, $F(t)$ is also continuous.

(ii) For proper integrals it is well known that the order of integrals can be exchanged:

$$\int_c^d F_u(t)\,dt = \int_a^u \left(\int_c^d f(x,t)\,dt \right) dx.$$

It follows from the uniformity of the convergence $F_u(t) \rightrightarrows F(t)$ that the left-hand side converges to $\int_c^d F(t)\,dt$ when $u \to b$, which proves statement (ii). □

Proposition A.4 *Let I and J be the same as in Proposition* A.3 *and $f(x,t)$ be a continuous function on $I \times J$ satisfying the following conditions:*

- *For any $t \in J$ the improper integral $F(t) := \int_a^b f(x,t)\,dx$ converges.*
- *$\dfrac{\partial f}{\partial t}(x,t)$ exists and is continuous on $I \times J$.*
- *The improper integral $G(t) := \int_a^b \dfrac{\partial f}{\partial t}(x,t)\,dx$ converges uniformly on any compact set in J.*

Then, $F(t)$ *is continuously differentiable and* $\dfrac{dF}{dt}(t) = G(t)$ *on* I.

Proof The function $G(t)$ is continuous by Proposition A.3 (i). Therefore for any closed interval $[c,s] \subset I$, applying Proposition A.3 (ii), we have

$$\begin{aligned}\int_c^s G(t)\,dt &= \int_a^b \left(\int_c^s \frac{\partial f}{\partial t}(x,t)\,dt\right) dx \\ &= \int_a^b (f(x,s) - f(x,c))\,dx = F(s) - F(c).\end{aligned}$$

By differentiating this with respect to s we obtain $G(s) = F'(s)$. □

The uniform convergence of improper integrals can be verified by the Weierstrass M-test.

Proposition A.5 *Let* I *and* J *be the same as in Proposition* A.3 *and* $f(x,t)$ *be a function on* $I \times J$ *satisfying the following:*

- *There exists a function* $M(x)$ *on* I *such that* $|f(x,t)| \le M(x)$ *for any* $(x,t) \in I \times J$.
- *The improper integral* $\displaystyle\int_a^b M(x)\,dx$ *converges.*

Then the improper integral $F(t) := \displaystyle\int_a^b f(x,t)\,dx$ *converges uniformly on* J.

Proof Let us prove that for any $\varepsilon > 0$ there exists a $c \in [a,b)$ such that

$$\sup_{t \in J} |F_u(t) - F_v(t)| \le \varepsilon$$

for any $u, v \in (c,b)$, where $F_u(t) := \displaystyle\int_a^u f(x,t)\,dx$.

By the assumption on $M(x)$ there exists a $c \in [a,b)$ such that

$$|M_u - M_v| < \varepsilon$$

for any $u, v \in (c,b)$, where $M_u := \displaystyle\int_a^u M(x)\,dx$. Using the estimate of $|f(x,t)|$ by $M(x)$, we have

$$\begin{aligned}|F_u(t) - F_v(t)| &= \left|\int_a^u f(x,t)\,dx - \int_a^v f(x,t)\,dx\right| = \left|\int_v^u f(x,t)\,dx\right| \\ &\le \left|\int_v^u M(x)\,dx\right| = |M_u - M_v| < \varepsilon\end{aligned}$$

for any $t \in J$. This proves the proposition. □

A.2 Several Facts in Complex Analysis

A.2.1 Integral depending on a parameter

(For details we refer to, for example, Chapter 4, §2 in [Ah].)

The following lemma is quite useful.

Lemma A.6 *Let $\varphi(z,w)$ be a function on $D \times \Omega \subset \mathbb{C}^2$ which is continuous as a function of both variables and analytic as a function of $z \in D$. If C is a piecewise smooth curve in Ω with finite length, then*

$$F(z) := \int_C \varphi(z,w)\, dw$$

is analytic in z and

$$F'(z) = \int_C \frac{\partial \varphi}{\partial z}(z,w)\, dw.$$

One can prove this directly, or by using the Cauchy integral formula $\varphi(z,w) = \dfrac{1}{2\pi i}\displaystyle\int_\gamma \varphi(\zeta,w)\, d\zeta$ and changing the order of integrations by ζ and by w.

A.2.2 Consequences of the Taylor expansion

Let $f(z)$ and $g(z)$ be holomorphic function on a domain (= connected open set) $D \subset \mathbb{C}$. The following two theorems are consequences of the Taylor expansion.

Theorem A.7 (Identity theorem)

If there is a subset S of D such that

- *S has an accumulation point in D;*
- *f and g coincide on S,*

then $f(z) = g(z)$ for any $z \in D$.

Typically, for the subset S, we use an open subset of D or a sequence $\{z_n\}_{n=1,2,\dots}$ which converges to a point $z_0 \in D$.

Proof (of Theorem* A.7*) Let D_∞ be the subset of D consisting of points at which all Taylor coefficients of f and g coincide:

$$D_\infty := \{z \in D \mid f^{(m)}(z) = g^{(m)}(z) \text{ for all } m = 0,1,2,\dots.\}.$$

We want to show $D_\infty = D$, which is a consequence of the following three facts because of the connectedness of D:

- D_∞ is closed.

- D_∞ is open.
- $D_\infty \neq \emptyset$.

Note that $D_\infty = \bigcap_{m=0}^{\infty} D_m$, where $D_m := \{z \in D \mid f^{(m)}(z) = g^{(m)}(z)\}$. As $f^{(m)}(z) - g^{(m)}(z)$ is a continuous function, D_m = (the inverse image of a point set $\{0\}$ by $f^{(m)} - g^{(m)}$) is closed. Since the intersection of closed sets is closed, D_∞ is closed.

To show that D_∞ is open, it is sufficient to show that for each point $z_0 \in D_\infty$ there exists a neighbourhood of z_0 which is contained in D_∞. Since f and g have the same Taylor expansion at z_0, their Taylor expansions have the same convergence radius ρ_0. Consequently f and g coincide on the open disk $\Delta_0 := \{z \mid |z - z_0| < \rho_0\}$. We can take $\Delta_0 \cap D$ as the desired neighbourhood of z_0.

Lastly we show $D_\infty \neq \emptyset$, using the existence of the subset S. We denote an accumulating point of S by $z_0 \in D$ and take a sequence $\{z_n\}_{n=1,2,\dots} \subset S$ which converges to z_0: $\lim_{n\to\infty} z_n = z_0$. By the assumption on S, $f(z_n) = g(z_n)$ holds for all n. Since $h(z) := f(z) - g(z)$ is continuous, $h(z_0) = \lim_{n\to\infty} \bigl(f(z_n) - g(z_n)\bigr) = 0$. Hence the 0-th coefficient a_0 of the Taylor expansion of h at z_0,

$$h(z) = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \cdots,$$

vanishes: $a_0 = 0$. Inductively applying the same argument to

$$h_m(z) := \frac{h(z)}{(z - z_0)^m} = a_m + a_{m+1}(z - z_0) + a_{m+2}(z - z_0)^2 + \cdots,$$

which satisfies $h_m(z_n) = 0$ for all n, we obtain $a_m = 0$. Thus we have proved that all the Taylor coefficients of h at z_0 are 0, i.e., $z_0 \in D_\infty$. □

Theorem A.8 (Integral domain)

If the product $f(z)\,g(z)$ is identically 0 on D, either $f(z)$ or $g(z)$ is identically 0. In the terminology of algebra, the ring of holomorphic functions on D is an integral domain.

This is not true for C^∞-functions. For example, take the C^∞-function defined by $f(x) = 0$ $(x \leq 0)$ and e^{-1/x^2} $(x > 0)$. Both $f(x)$ and $f(-x)$ are not identically 0 on $\mathbb{R}$, but their product $f(x)\,f(-x)$ vanishes identically.

Proof (of Theorem* A.8*) Assume that both $f(z)$ and $g(z)$ are not identically 0 on D and let

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n, \qquad g(z) = \sum_{n=0}^{\infty} b_n(z - z_0)^n$$

be their Taylor expansions around a point $z_0 \in D$. There are non-negative integers M and N such that

$$a_i = 0 \ (i = 0, \dots, M-1), \quad a_M \neq 0,$$
$$b_j = 0 \ (j = 0, \dots, N-1), \quad b_N \neq 0.$$

(If all coefficients of the Taylor expansion are 0, the function is identically 0 in a neighbourhood of z_0 and by virtue of the identity theorem (Theorem A.7) it is 0 everywhere in D.)

Therefore the Taylor expansion of the product $f(z)\,g(z)$ starts from the term of degree $M+N$:

$$f(z)\,g(z) = a_M b_N\, z^{M+N} + (\text{terms of degree not less than } M+N+1).$$

As the term of degree $M+N$ remains, $f(z)\,g(z)$ cannot be identically 0. This proves the theorem. □

A.2.3 Weierstrass's double series theorem

(For details we refer to, for example, §1.1 of Chapter 5 in [Ah].)

For a sequence $\{f_n(z)\}_{n=1,2,\dots}$ or a series $\sum_{n=1}^{\infty} f_n(z)$ of holomorphic functions $f_n(z)$ the following theorems are fundamental.

Theorem A.9 (Weierstrass's double series theorem)

Let $f_n(z)$ $(n = 1,2,\dots)$ be a holomorphic functions on a domain D.

(i) If $f_n(z)$ converges to a function $f(z)$ uniformly on any compact subset of D, $f(z)$ is holomorphic on D. Moreover the sequence of derivatives $f_n'(z)$ converges to $f'(z)$ uniformly on any compact subset of D.

(ii) If the series $\sum_{n=1}^{\infty} f_n(z)$ converges uniformly on any compact subset of D, it is a holomorphic function on D and

$$\frac{d}{dz}\left(\sum_{n=1}^{\infty} f_n(z)\right) = \sum_{n=1}^{\infty} \frac{df_n}{dz}(z).$$

This is a corollary of Cauchy's integral formula. If we expand $f_n(z)$ in (ii) into the Taylor series $f_n(z) = \sum_{m=1}^{\infty} a_{m,n}(z-a)^m$, the statement of the theorem says that we can change the order of the sums by m and by n, from which comes the name of the theorem.

A.2.4 The argument principle and its generalisation

(For details we refer to, for example, Chapter 4, §3.3 in [Ah].)

Let $g(w)$ be a meromorphic function in a domain (= a connected open set) $D \subset \mathbb{C}$, which is not a constant function 0: $g(w) \not\equiv 0$. Let us take a Jordan closed curve (= a simple closed curve = a closed curve without self intersection) C in D with positive (counter-clockwise) orientation. By the Jordan curve theorem C divides $\mathbb{C}$ into its interior and exterior. We assume that C is rectifiable (and hence the contour integral along C is defined), that the interior of C is included in D and that g does not have zeros or poles along C.

We define non-negative integers $N(g;C)$ and $P(g;C)$ as follows:

$$N(g;C) := \text{the number of zeros of } g \text{ inside } C \text{ with multiplicities,}$$
$$P(g;C) := \text{the number of poles of } g \text{ inside } C \text{ with multiplicities.}$$

For example, if $g(w)$ has a Laurent expansion of the form

$$g(w) = c_k(w-w_0)^k + c_{k+1}(w-w_0)^{k+1} + \cdots, \qquad c_k \neq 0 \tag{A.1}$$

at w_0 and C is a small circle around w_0, inside of which there are neither zeros nor poles of g except w_0, then

$$N(g;C) = \begin{cases} k & (k \geq 0), \\ 0 & (k < 0), \end{cases} \qquad P(g;C) = \begin{cases} 0 & (k \geq 0), \\ -k & (k < 0). \end{cases}$$

Theorem A.10 (Argument principle)

$$\frac{1}{2\pi i}\int_C \frac{g'(w)}{g(w)}\,dw = N(g;C) - P(g;C). \tag{A.2}$$

The proof is an easy application of the residue theorem to $\dfrac{g'(w)}{g(w)}$. In fact, $\dfrac{g'(w)}{g(w)} = \dfrac{k}{w-w_0} +$ (holomorphic function), if $g(w)$ has a Laurent expansion of the form (A.1), and the left-hand side of (A.2) is the sum of residues $\operatorname{Res}_{w=w_0} \dfrac{g'(w)}{g(w)}dw = k$ at the zeros and poles of g inside C.

The following theorem can be proved in the same way.

Theorem A.11 (Generalised argument principle)

If $\varphi(w)$ is a holomorphic function in D,

$$\frac{1}{2\pi i}\int_C \varphi(w)\frac{g'(w)}{g(w)}\,dw = \sum_{a:\text{ zero of } g \text{ inside } C} \varphi(a) - \sum_{b:\text{ pole of } g \text{ inside } C} \varphi(b), \tag{A.3}$$

where the zeros and poles in the right-hand side are counted with multiplicities.

In the simplest case, (A.3) gives an integral expression of the inverse function.

Corollary A.12 *Assume $g : D \to \tilde{D}$ is a holomorphic bijection and that a closed disk $\{w \mid |w-c| \le r\}$ is entirely contained in D. Then the inverse function $g^{-1} : \tilde{D} \to D$ is expressed by the integral*

$$g^{-1}(\zeta) = \frac{1}{2\pi i} \int_{|w-c|=r} w \frac{g'(w)}{g(w)-\zeta} \, dw \tag{A.4}$$

in a sufficiently small neighbourhood of $g(c)$. The inverse function $g^{-1}(\zeta)$ is a holomorphic function of ζ.

The proof is an application of Theorem A.11 to $g(w)-\zeta$ (as $g(w)$ in the theorem) and w (as $\varphi(w)$). The last statement is due to Lemma A.6.

It is noteworthy that the inverse function of a holomorphic function is always holomorphic, in contrast to real analysis. (The inverse function $\sqrt[3]{y}$ of a differentiable function $y = x^3$ $(x, y \in \mathbb{R})$ is not differentiable at $y = 0$.)

Another important application of the argument principle is the following.

Theorem A.13 *If $f(z)$ is holomorphic and not a constant function, then f maps an open set onto an open set.*

It is sufficient to show that, if $w_0 = f(z_0)$ belongs to the image of f, then a sufficiently small neighbourhood of w_0 is contained in the image of f. In other words, if w is sufficiently close to w_0, equation $f(z) = w$, or $F(z; w) := f(z) - w = 0$ has a solution. Let us take a small circle C with its centre at z_0. We want to show $N(F(\cdot; w), C) \ge 1$. At $w = w_0$ this is true because $w_0 = f(z_0)$. By the argument principle (A.2) we have an integral representation of $N(F(\cdot; w), C)$,

$$N(F(\cdot; w), C) = \frac{1}{2\pi i} \int_C \frac{f'(\zeta)}{f(\zeta) - w} d\zeta,$$

which is continuous as a function of w, if w stays close to w_0. The number of zeros $N(F(\cdot; w), C)$ is, of course, an integer, so it is a locally constant function in a neighbourhood of w_0. Thus $N(F(\cdot; w), C) \ge 1$ follows from $N(F(\cdot; w_0), C) \ge 1$.

References

[AE] Andrews, G. E. and Eriksson, K., *Integer partitions*, Cambridge University Press, 2004.

[Ah] Ahlfors, L. V., *Complex analysis. An introduction to the theory of analytic functions of one complex variable*, Third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., 1978.

[Ay] Ayoub, R., Leminiscate and Fagnano's Contributions to Elliptic Integrals, *Archive for History of Exact Sciences* **29** (1984), 131–149. http://www.jstor.org/stable/41133708

[B] Bellman, R., *A Brief Introduction to Theta Functions*, Holt, Rinehart and Winston, 1961; Dover reprint, 2013.

[BB] Briot, C. and Bouquet, C., *Théorie des fonctions doublement périodiques et, en particulier, des fonctions elliptiques*, Mallet-Bachelier, 1859.

[C] Conrad, B., *Integration in elementary terms*, https://math.stanford.edu/~conrad/papers/elemint.pdf (viewed on 17 May 2022)

[CL] Collingwood, E. F. and Lohwater, A. J., *The theory of cluster sets*, Cambridge University Press, 1966.

[CLO] Cox, D. A., Little, J. and O'Shea, D., *Ideals, Varieties, and Algorithms — An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Fourth edition, Undergraduate Texts in Mathematics, Springer, 2015.

[E] Euler, L., *Introductio in Analysin Infinitorum*, Tomus Primus, 1748.

[FK] Farkas, H. M. and Kra, I., *Riemann Surfaces*, Springer-Verlag, 1980.

[G] Gauss, C. F., *Werke* III, Königlichen Gesellschaft der Wissenschaften, 1866; *Werke* VIII, Königlichen Gesellschaft der Wissenschaften, 1900; *Werke* X Abteilung 1, Königlichen Gesellschaft der Wissenschaften, 1917.

[Ha] Hatcher, A., *Algebraic Topology*, Cambridge University Press, 2001.

[Hi] Hitotsumatsu, S., *Shotou kansuu no suuchi keisan (Numerical computation of elementary functions)* (in Japanese), Kyoiku Shuppan, 1974.

[HC] Hurwitz, A. and Courant, R., *Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen* (in German), die Grundlehren der mathematischen Wissenschaften **3**, Springer-Verlag, 1929.

[INU] Iitaka, S., Namikawa, Y. and Ueno, K., *Descartes no seishin to daisuu kika (Spirit of Descartes and algebraic geometry)* (in Japanese), Nippon Hyoron sha, 1993.

[Ja1] Jacobi, C. G. J., Considerationes generales de transcendentibus Abelianis, *Crelle's Journal für die reine und angewandte Mathematik* **9**, 394–403; *Gesammelte Werke* **2**, 5–16.

[Ja2] Jacobi, C. G. J., Theorie der elliptischen Functionen aus den Eigenschaften der Thetareihen abgeleitet, *Nach einer Vorlesung Jacobis in dessen Auftrag aufgearbeitet von C. W. Borchardt*, *Gesammelte Werke* **1**, 497–538.

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7

[Ji] Jimbo, M., *Fukuso kansuu nyuumon (Introduction to complex functions)* (in Japanese), Iwanami Shoten, 2003.

[Ka] Kawada, Y., *Gauss no daen kansuu-ron (Gauss and Elliptic Functions)* (in Japanese), Sophia Kokyuroku in Mathematics **24**, Sophia University, Department of Mathematics, 1986.

[Kl] Klein, F., *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert* (in German), Springer, 1926; *Development of Mathematics in the 19th Century (Lie Groups)* (English translation), Math Sci Press, 1979. https://gdz.sub.uni-goettingen.de/id/PPN375425993

[Ko] Kodaira, K., *Fukuso kaiseki (Complex analysis)*, Second edition, (in Japanese) Iwanami Shoten, 1983; *Complex analysis* (English translation) Cambridge Studies in Advanced Mathematics, **107**, 2007.

[Li] Liouville, J., Leçons sur les fonctions doublement périodiques faites en 1847 par M. J. Liouville, *Journal für die reine und angewandte Mathematik* **88** (1880), 277–310. (Lecture notes taken by C. W. Borchardt in French.)

[Lv] Lvovski, S., *Principles of Complex Analysis*, Moscow Lectures, Springer, 2020.

[M] Mumford, D., *Tata Lectures on Theta* I, Progress in Mathematics **28**, Birkhäuser Boston, Inc., 1983.

[Na] Natanzon, S. M., *Complex Analysis, Riemann Surfaces and Integrable Systems*, Moscow Lectures, Springer, 2019.

[Ne] Newman, D. J., A simplified version of the fast algorithms of Brent and Salamin. *Math. Comp.* **44** (1985), no. 169, 207–210.

[No] Noumi, M., *Euler ni manabu (Learning from Euler)* (in Japanese), Nippon Hyoron sha, 2013.

[P] Phragmén, E., Sur un théorème concernant le fonctions elliptiques, *Acta Math.* **7** (1883), no. 1, 33–42.

[PS] Prasolov, V. and Solovyev, Y., *Elliptic functions and algebraic equations* (in Russian), Factorial, 1997; *Elliptic functions and elliptic integrals* (English translation), Translations of Mathematical Monographs **170**, American Mathematical Society, 1997.

[Ri] Riemann, B., *Bernard Riemann's Gesammelte mathematische Werke und wissenschaftlicher Nachlass*, Teuber, 1876.

[Ro] Rosenlicht, M., Integration in finite terms, *American Math. Monthly* **79** (1972), 963–972.

[Se] Serre, J.-P., *Cours d'Arithmétique* (in French), Presses Universitaires de France, 1970; *A Course in Arithmetic* (English translation), Graduate texts in mathematics **7**, Springer-Verlag, 1973.

[Sk1] Sklyanin, E. K., Some Algebraic Structures Connected with the Yang-Baxter Equation, *Funkts. analiz i ego Prilozh.* **16-4** 27–34 (1982) (in Russian); *Funct. Anal. Appl.* **16** 263–270 (1983) (English translation)

[Sk2] Sklyanin, E. K., Some Algebraic Structures Connected with the Yang-Baxter Equation. Representations of Quantum Algebras, *Funkts. analiz i ego Prilozh.* **17-4** 34–48 (1983) (in Russian); *Funct. Anal. Appl.* **17** 273–284 (1984) (English translation)

[Su] Sugiura, M., *Kaiseki nyuumon I (Introduction to analysis I)* (in Japanese), University of Tokyo Press, 1980.

[ST] Singer, I. M., Thorpe, J. A., *Lecture Notes on Elementary Topology and Geometry*, Springer, 1967.

[Tkb] Takebe, T., *Daen sekibun to daen kansuu — otogi no kuni no aruki-kata (Elliptic integrals and elliptic functions — a tour guide of the fairyland)* (in Japanese), Nippon Hyoron sha, 2019.

[Tkg] Takagi, T., *Kinsei Suugaku Shidan (Historical episodes in modern mathematics)*, Kyoritsu Shuppan, 1933; Iwanami Shoten, 1995 (paperback).

[Tkn] Takenouchi, T., *Daen kansuu ron (Theory of elliptic functions)* (in Japanese), Iwanami Shoten, 1936.

[Tks] Takase, M., *Abel — daen kansuu-ron he no michi (Abel — a way to the theory of elliptic functions)* (in Japanese), Gendai Suugaku sha, 2016.

[To1] Toda, M., *Koushi soliton no hakken (Discovery of lattice solitons)* (in Japanese), Nihon Butsuri Gakkai Kaishi (Monthly Magazine of the Physical Society of Japan) **51-3**, (1996). https://www.jps.or.jp/books/50thkinen/50th_03/004.html

[To2] Toda, M., *Daen kansuu nyuumon (Introduction to elliptic functions)* (in Japanese), Nippon Hyoron sha, 2001.

[Tu] Tu, L. W., *An Introduction to Manifolds*, Springer, 2011.

[U] Umemura, H., *Daen kansuu-ron — daen kyokusen no kaiseki-gaku (Theory of elliptic functions — Analysis on elliptic curves)* (in Japanese), University of Tokyo Press, 2000.

[Wb] Weber, H., *Elliptische Functionen und algebraische Zahlen, Akademische Vorlesungen von H. Weber*, Vieweg, 1891.

[Wi] Weierstrass, K., *Mathematische Werke* II, Mayer & Müller, 1895.

[Wy] Weyl, H., *Die Idee der Riemannschen Fläche* (in German), 3te Aufl., Teubner, 1955; *The concept of a Riemann surface* (English translation), Addison-Wesley, 1964.

[WW] Whittaker, E. T. and Watson, G. N., *A course of modern analysis, the fourth edition*, Cambridge University Press, 1927.

Index

T. Takebe, *Elliptic Integrals and Elliptic Functions*, Moscow Lectures 9,
https://doi.org/10.1007/978-3-031-30265-7