Translator's Introduction

1. Overview

> Modern algebraic geometry has deservedly been considered for a long time as an exceedingly complex part of mathematics, drawing practically on every part to build up its concepts and methods and increasingly becoming an indispensable tool in many seemingly remote theories. It shares with number theory the distinction of having one of the longest and most intricate histories among all branches of our science, of having always attracted the efforts of the best mathematicians in each generation, and of still being one of the most active areas of research.
>
> Dieudonné (1972), p. 827.

> It seems to me that, in the spirit of the biogenetic law, the student who repeats in miniature the evolution of algebraic geometry will grasp the logic of the subject more clearly.
>
> Shafarevich (1994), p. vii.

Richard Dedekind and Heinrich Weber first worked together in 1874, as co-editors of Riemann's collected works. Weber was called into this project as a replacement for Clebsch, who had died unexpectedly of diptheria, and his expertise in mathematical physics complemented Dedekind's expertise in pure algebra and analysis. The fruit of this collaboration was their joint paper, Dedekind and Weber (1882), a ground-breaking contribution to the understanding and advancement of Riemann's ideas. *Theorie der algebraischen Functionen der einer Veränderlichen* (theory of algebraic functions of one variable) revolutionized algebraic geometry by introducing methods of algebraic number theory into the subject. This made possible the first rigorous proofs of theorems discovered with the help of physical intuition, and opened the way to an extension of algebraic-geometric concepts from the complex numbers to arbitrary fields.

In a sense, the paper is a sequel to Dedekind (1877), a long paper in which Dedekind expounded his theory of ideals and their applications to number theory. However, Dedekind and Weber give a self-contained exposition of their theory, which is at some points simpler than the ideal theory for algebraic numbers.

Like Dedekind (1877), the Dedekind-Weber paper starts with the concept of field, but this time it is a field of *functions*, the "algebraic functions of one variable." Following the example of number theory, they distinguish the ring of *integers* of this field, then the *primes*, and finally the *ideals*. As in number theory, it turns out that ideals are crucial to complete the analogy with the traditional arithmetic of integers. However, in the context of algebraic functions, ideals prove to be important in other ways, and indeed a more general idea that they call "polygons" is needed.

To show the value of these new ideas, Dedekind and Weber gave new proofs of two great theorems: *Abel's theorem* of Abel (1841) and the *Riemann-Roch theorem* of Riemann (1857) and Roch (1864). These theorems are as timely today as they were in 1882, but they require some introduction, which Dedekind and Weber do not supply. I would therefore like to present some historical background to these theorems, and to the theory of algebraic functions itself, with copious examples. In some ways, this introduction is a sequel to my introduction to Dedekind (1877), though I will recapitulate some points to keep it self-contained.

In preparing this material I have been greatly assisted by the first 35 pages of Dieudonné (1985), an extraordinarily rich and insightful account of the development of algebraic geometry up to the Dedekind-Weber paper, and Koch (1991), which places this development against the general background of 19th-century mathematics. Another helpful overview is the chapter by Geyer in Dedekind et al. (1981). As will become apparent, much of the algebra in modern algebraic geometry arose from problems in classical analysis, particularly the integral calculus. The first such result was the fundamental theorem of algebra, originally motivated by the desire to factorize polynomials for the purpose of integrating rational functions.

2. From Calculus to Abel's Theory of Algebraic Curves

> What a discovery is Abel's generalization of Euler's integral! I have never seen such a thing! But how can it be that this discovery, which could be the most important made in the mathematics of this century, and which was communicated to your Academy two years ago, has escaped the attention of you and your colleagues?
>
> Jacobi (1829) letter to Legendre, 14 March 1829.

When calculus was developed in the 17th century, the first really hard problems were problems of integration. This was especially true of the Leibniz approach, which sought integrals in "closed form," that is, in terms of functions from the small class known as "elementary." These are the algebraic functions, together with functions arising from them by composition with the exponential function and its relatives, the logarithm, circular functions, and inverse circular functions.

The only broad class of functions that can be integrated in Leibniz's sense are the rational functions, that is, the functions of the form $r(x) = p(x)/q(x)$, where p and q are polynomials. Any rational function can be integrated because the denominator $q(x)$ may be split into linear factors $(x - a)$, by the fundamental theorem of algebra, and the quotient $p(x)/q(x)$ may then be decomposed into partial fractions of the form $x^m/(x - a)^n$, which have rational integrals in all cases except

$$\int \frac{dx}{x-a} = \log(x-a) + \text{constant}.$$

Thus the integral of a rational function is itself a rational function, with the possible exception of some terms of the form $\log(x - a)$.

(In elementary calculus courses this simple picture is confused by the presence of partial fractions such as $1/(x^2 + 1)$, the integral of which is usually taken to be $\tan^{-1} x + \text{constant}$. However, we have

$$\frac{1}{x^2+1} = \frac{1/2i}{x-i} - \frac{1/2i}{x+i},$$

so we can also express $\int dx/(x^2+1)$ as a sum of logarithms, namely

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

This was first done, albeit with some confusion about the meaning of complex logarithms, by Johann Bernoulli (1702). Around 1800, when the fundamental theorem of algebra was finally proved, the meaning of complex numbers became better understood, and it became increasingly clear that they play an important role in the theory of integrals.)

When the rational functions are extended by as little as the square root function, the resulting integrals quickly fall outside the class of elementary functions. A famous example is the *lemniscatic* integral

$$\mathrm{sl}^{-1}(x) = \int_0^x \frac{dt}{\sqrt{1-t^4}},$$

so-called because it expresses the arc length of the lemniscate of Jakob Bernoulli (1694), shown in Figure 1.

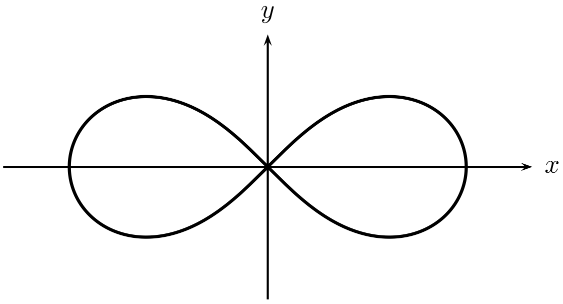

FIGURE 1. The lemniscate of Bernoulli

This curve has cartesian equation $(x^2+y^2)^2 = x^2-y^2$, and its arc length cannot be expressed in terms of elementary functions of x and y. However, Fagnano (1718) discovered that the lemniscatic integral satisfies an arc-length *doubling formula*

$$2\int_0^x \frac{dt}{\sqrt{1-t^4}} = \int_0^{2x\sqrt{1-x^4}/(1+x^4)} \frac{dt}{\sqrt{1-t^4}},$$

and Euler (1768) generalized Fagnano's formula to an arc-length *addition formula*

$$(*) \qquad \int_0^x \frac{dt}{\sqrt{1-t^4}} + \int_0^y \frac{dt}{\sqrt{1-t^4}} = \int_0^{(x\sqrt{1-y^4}+y\sqrt{1-x^4})/(1+x^2y^2)} \frac{dt}{\sqrt{1-t^4}}.$$

These results are analogous to properties of the *inverse sine integral*

$$\theta = \sin^{-1} x = \int_0^x \frac{dt}{\sqrt{1-t^2}},$$

which are derivable from basic properties of the sine and cosine functions. For example, the familiar angle-doubling formula

$$\sin 2\theta = 2\sin\theta\cos\theta = 2\sin\theta\sqrt{1-\sin^2\theta},$$

implies that

$$2\theta = \sin^{-1}(2\sin\theta\sqrt{1-\sin^2\theta}),$$

which gives the angle-doubling formula for integrals:

$$2\sin^{-1} x = 2\int_0^x \frac{dt}{\sqrt{1-t^2}} = \int_0^{2x\sqrt{1-x^2}} \frac{dt}{\sqrt{1-t^2}}.$$

And the familiar angle addition formula,

$$\begin{aligned}\sin(\theta+\varphi) &= \sin\theta\cos\varphi + \cos\theta\sin\varphi\\ &= \sin\theta\sqrt{1-\sin^2\varphi} + \sin\varphi\sqrt{1-\sin^2\theta},\end{aligned}$$

implies

$$\theta+\varphi = \sin^{-1}\left(\sin\theta\sqrt{1-\sin^2\varphi} + \sin\varphi\sqrt{1-\sin^2\theta}\right),$$

which gives the addition formula for arcsine integrals:

$$\int_0^x \frac{dt}{\sqrt{1-t^2}} + \int_0^y \frac{dt}{\sqrt{1-t^2}} = \int_0^{x\sqrt{1-y^2}+y\sqrt{1-x^2}} \frac{dt}{\sqrt{1-t^2}}.$$

Thus in both cases we find that a sum of two integrals, $\int_0^x f(t)\,dt + \int_0^y f(t)\,dt$, can be simplified to a single integral, $\int_0^z f(t)\,dt$, where z is an algebraic function of x and y.

It so happens that the integrand $1/\sqrt{1-t^2}$ of the inverse sine integral can be *rationalized* by the change of variable $t = 2s/(1+s^2)$—not so surprisingly, since the inverse sine is an elementary function—so we can eliminate the integral altogether in this case. However, in the case of the lemniscatic integral, reducing the sum of two integrals to one is the best we can do. The integrand $1/\sqrt{1-t^4}$ *cannot* be rationalized by a change of variable, and indeed Jakob Bernoulli (1704) made a remarkable attempt to prove this, using the theorem of Fermat that the equation $X^4 - Y^4 = Z^2$ has no solution in positive integers X, Y, Z. His attempt fell short, because it is not enough to know this theorem for integers X, Y, Z. But it is enough to know it for *polynomials* $X(t), Y(t), Z(t)$, and indeed polynomials behave enough like integers that Fermat's proof can be replayed for polynomials, though no one noticed this in Bernoulli's time.

Thus there is an essential difference between the ordinary sine function and the lemniscatic sine function, sl, defined as the inverse of the lemniscatic integral. Nevertheless there are enough similarities to enable the development of a theory of the lemniscatic sine function. This was begun by Gauss in 1796, and extended to a general theory of the so-called *elliptic functions* by Abel and Jacobi in the 1820s. Like the circular functions, the elliptic functions satisfy addition formulas and they are *periodic*, only more so. Just as the sine and cosine have *period* 2π, in the sense that

$$\sin(\theta+2\pi) = \sin\theta, \quad \cos(\theta+2\pi) = \cos\theta,$$

an elliptic function f has *two* periods ω_1, ω_2, in the sense that

$$f(z+\omega_1) = f(z), \quad f(z+\omega_2) = f(z).$$

The periods ω_1, ω_2 are complex numbers whose ratio is not real. For example, Gauss discovered that the two periods of sl are ϖ and $i\varpi$, where

$$\varpi = 2\int_0^1 \frac{dt}{\sqrt{1-t^4}}.$$

The double periodicity of elliptic functions was first explained by algebraic manipulation of integrals, but Riemann (1851) found a far more transparent geometric explanation (not unlike explaining the period 2π of sine and cosine by referring to the circle), which we will come to later.

The theory of elliptic functions was the first great advance in integral calculus since the integration of rational functions. Nevertheless, this theory only scratched the surface of a huge and important world of calculus: the integrals of *algebraic functions*; that is, integrals of the form

$$\int g(s,t)\,dt, \quad \text{where } s \text{ satisfies a polynomial equation} \quad P(s,t) = 0.$$

The lemniscatic integral is $\int dt/s$, where $s^2 = 1 - t^4$, and the general theory of elliptic functions deals with the integrals $\int dt/s$ where s^2 equals a polynomial of degree 3 or 4 in t. But what can one say, for example, about the integral

$$\int_0^x \frac{dt}{\sqrt{1-t^6}} \,?$$

It turns out that this integral does *not* satisfy an addition formula

$$\int_0^x \frac{dt}{\sqrt{1-t^6}} + \int_0^y \frac{dt}{\sqrt{1-t^6}} = \int_0^z \frac{dt}{\sqrt{1-t^6}},$$

where z is an algebraic function of x and y. However, Abel (1841) discovered a wonderful substitute for an addition formula: *any sum of integrals,*

$$\int_0^{x_1} \frac{dt}{\sqrt{1-t^6}} + \cdots + \int_0^{x_m} \frac{dt}{\sqrt{1-t^6}}$$

is equal to the sum of two *integrals*

$$\int_0^{z_1} \frac{dt}{\sqrt{1-t^6}} + \int_0^{z_2} \frac{dt}{\sqrt{1-t^6}}, \quad \text{where } z_1, z_2 \text{ are algebraic functions of } x_1, \ldots, x_m,$$

plus some "trivial" algebraic and logarithmic terms.

This result is only an illustration of the amazingly general:

Abel's Theorem. *For any integral of the form $\int g(s,t)\,dt$, where g is a rational function and s and t are connected by a polynomial relation $P(s,t) = 0$, there is a number p such that any sum of integrals*

$$\int_0^{x_1} g(s,t)\,dt + \cdots + \int_0^{x_m} g(s,t)\,dt$$

equals a sum of at most p integrals

$$\int_0^{z_1} g(s,t)\,dt + \cdots + \int_0^{z_p} g(s,t)\,dt,$$

where $z_1, \ldots, z_p$ are algebraic functions of $x_1, \ldots, x_m$, plus terms that are either rational functions or their logarithms.

The number p depends only on the polynomial P. It was later called the *genus* of the curve defined by $P(s,t) = 0$, and it too found a natural geometric interpretation in Riemann (1851), as we will see in the next section. In particular, the curve $s^2 = 1 - t^4$ that yields the lemniscatic integral has genus 1, because any sum of lemniscatic integrals reduces to one integral by repeated application of the addition formula (*). More generally, any *elliptic curve*[1] $s^2 = q(t)$, where $q(t)$ is of degree 3 or 4 without repeated roots, is of genus 1, because there is an addition formula for the corresponding integral $\int dt/s$.

Finally, any curve $s = P(w)$, $t = Q(w)$ parameterized by rational functions P and Q is of genus zero, because the corresponding integral $\int g(s,t)\,dt$ is the integral of the rational function $g(P(w), Q(w))Q'(w)$.

Abel submitted his paper to Cauchy in 1826 but, due to inattention by the mathematicians of the Paris Academy, it was not published at the time. It was noticed by Jacobi, however, who in 1829 wrote the letter to Legendre quoted at the beginning of this section. Even the intervention of Jacobi failed to wake up the Academicians, and Abel's paper did not appear until 1841, long after Abel had died. There is a further excruciating twist to this story of neglected genius. The other mathematician notoriously ignored by the Paris Academy, Evariste Galois, also seems to have discovered Abel's theorem, independently of Abel, but some years later. It is mentioned in his letter to Auguste Chevalier, Galois (1846), written on the night before his death in 1832. He states the theorem without proof, but with some additional remarks that suggest that he already had some of the ideas developed by Riemann 20 years later.

3. Riemann's Theory of Algebraic Curves

> It is quite a paradox that in the work of this prodigious genius, out of which algebraic geometry emerges entirely regenerated, there is almost no mention of algebraic curve; it is from his theory of algebraic *functions* and their integrals that all of the birational geometry of the nineteenth and the beginning of the twentieth century issues.
>
> Dieudonné (1985), p. 18.

In the 1850s, two papers by Bernhard Riemann[2] completely changed the face of complex analysis and algebraic geometry. Riemann (1851) and Riemann (1857) viewed algebraic curves in a new way, as what we now call *Riemann surfaces.* In retrospect, this development seems unsurprising and even inevitable. Since around 1800, mathematicians had become used to the idea that the complex "line" $\mathbb{C}$ was geometrically a plane, so the idea that a complex "curve" should be some kind of surface was just over the horizon. Nevertheless, Riemann's description of these surfaces was received skeptically by most of his contemporaries. The underlying topological ideas, though very intuitive and persuasive, did not yet have a rigorous foundation. And, to make matters worse, Riemann made connections between topology and analysis by appealing to physics. Then, as now, this was considered mathematically dubious.

[1] The name "elliptic" became attached to the curves of genus 1 because the corresponding integrals ("elliptic integrals") include the integral for the arc length of the ellipse. Unfortunately, the ellipse itself has genus 0, and hence is *not* an elliptic curve.

[2] Page numbers in references to these papers in this Introduction refer to the original papers. However, many readers will find it more convenient to consult the English translation of Riemann's works, Riemann (2004). To make this easier to do, I also give section numbers, which are the same in the original papers and in the translation.

But if Riemann's proofs were not rigorous, his results were so stunning that they demanded explanation, and this became the task of later mathematicians, among them Dedekind and Weber.

Today, the necessary foundations of topology and analysis have been constructed, so we have the luxury of describing Riemann's ideas in informal terms similar to his own. I think that it is useful to do so, because some of the algebraic concepts devised by Dedekind and Weber are scarcely comprehensible if one has not seen the topological concepts they replace. In particular, I doubt that readers should be confronted with the "ramification ideal" before they have seen a picture of "ramification," or "branching." Such a picture was given in Neumann (1865), the first textbook on Riemann's theory (Figure 2).

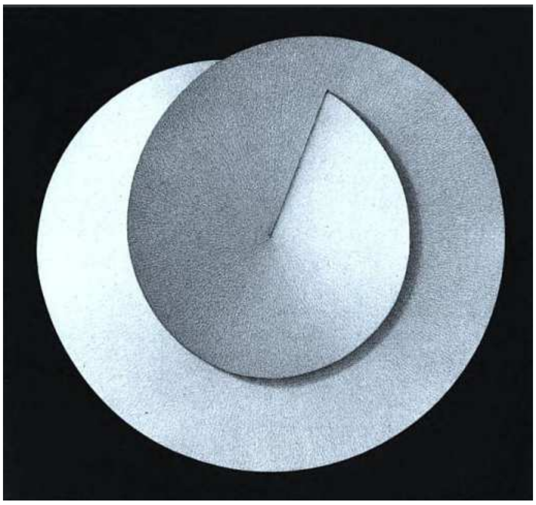

FIGURE 2. Neumann's picture of a branch point

This picture springs to mind when one attempts to visualize the curve $y^2 = x$ for *complex* variables x and y or, equivalently, the "two-valued function" $y = \pm\sqrt{x}$. Riemann imagined the two values $+\sqrt{x}$ and $-\sqrt{x}$ lying above x on a *two-sheeted covering* of the plane $\mathbb{C}$, as shown in Figure 3. Notice that, as x moves continuously once around a circle, the corresponding point $\sqrt{x}$ moves continuously around the lower sheet, then the upper sheet, of the two-sheeted covering, eventually taking the value $-\sqrt{x}$ that also lies above x. Thus the function $\sqrt{x}$ becomes "single-valued" on the covering surface.

The point $x = 0$ at which the two sheets fuse is called a *branch point* or *ramification point* of the covering, because one used to speak of the "branches of the multi-valued function"—in this case the two "branches" are $\sqrt{x}$ and $-\sqrt{x}$. The awkward feature of the picture—that the two sheets appear to pass through each other—is a result of representing the relation $y = x^2$ in three dimensions, one fewer than the four dimensions it really requires. One can visually add a fourth dimension, "shade of gray," to the sheets to avoid their meeting in the fourth dimension. This has actually happened in the Neumann picture, where one sheet is white where they appear to meet and the other is dark gray.

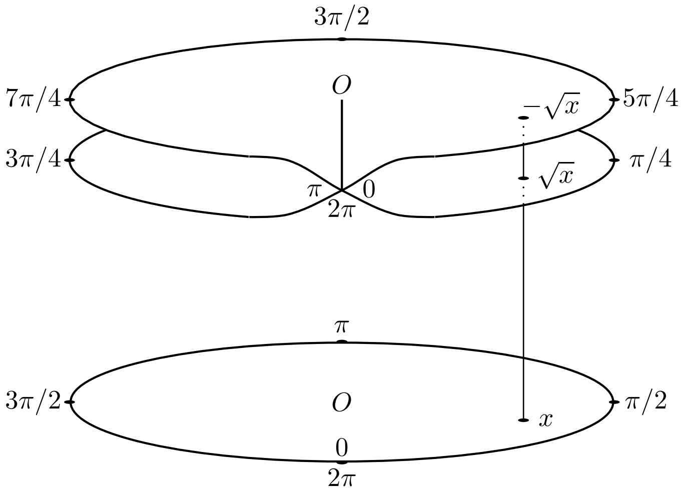

FIGURE 3. Branch point for the square root

Just as the curve $y^2 = x$ has a branch point of the two sheets at $x = 0$, the curve $y^n = x$ has a branch point of n sheets. An arbitrary algebraic curve $P(x, y) = 0$, where P is a polynomial of degree n, is an n-sheeted covering of $\mathbb{C}$ with a finite number of branch points. Since behavior of a curve at infinity is important, Riemann (1857) (Section 1, p. 117) extended $\mathbb{C}$ by a point ∞, and the resulting set $\mathbb{C} \cup \{\infty\}$ can be viewed as a *sphere* via the stereographic projection map shown in Figure 4. This idea is made explicit in Neumann (1865), p. 132. Under stereographic projection, each point $z \in \mathbb{C}$ corresponds to a point z' on the sphere other than the north pole N, and N itself naturally corresponds to ∞.

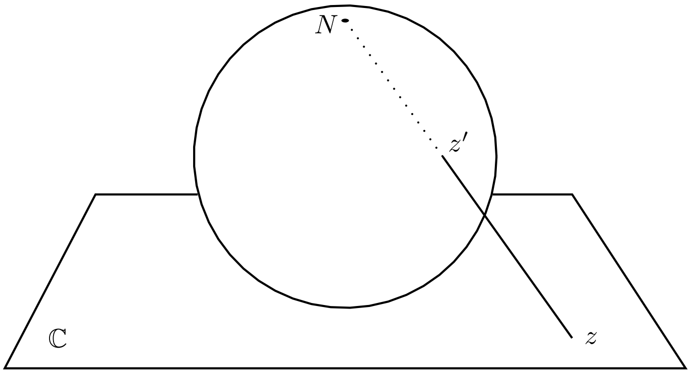

FIGURE 4. Stereographic projection of the sphere to $\mathbb{C} \cup \{\infty\}$

Corresponding to this completion of $\mathbb{C}$ to a sphere, we have a completion of each algebraic curve to a finite-sheeted covering of the sphere with finitely many branch points. Riemann realized that the covering surface S is topologically characterized by none other than Abel's number p, later dubbed the *genus* (or *Geschlecht* in German) by Clebsch (1865). Riemann described p topologically as half the number of closed cuts needed to make S simply connected (that is, such that any closed curve can be contracted to a point). In this case the resulting simply connected surface is a polygon with $4p$ sides. Möbius (1863) gave an even simpler interpretation of p, by showing that each Riemann surface is topologically equivalent to a member of the sequence of surfaces shown in Figure 5, namely, the one with p "holes."

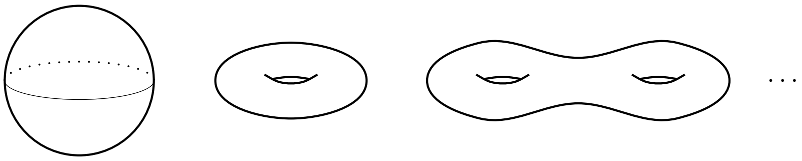

FIGURE 5. Riemann surfaces of genus 0, 1, 2, . . .

As an example, consider the elliptic curve

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

This curve is a two-sheeted cover of the sphere, with branch points like that shown in Figure 3 at $x = 0, 1, -1, \infty$. If we slit the sheets by cuts from 0 to ∞, and from 1 to -1, then, in order to obtain the branching, the edges of the cuts need to be identified so that the like-labeled edges shown in Figure 6 come together.

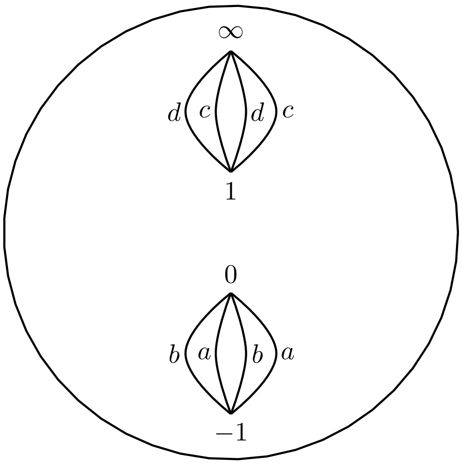

FIGURE 6. How edges are identified at branch points

But we can make a surface that is topologically the same by separating the two sheets before making the identifications, as shown in Figure 7.

The resulting surface is topologically a torus, shown in more familiar form in Figure 8. Thus Abel's number $p = 1$ agrees with the topological genus of the torus, because the torus has one "hole." (Notice that if 0 and ∞ are the only branch points, as is the case with $y^2 = x$, then the result of joining the two sheets is topologically a sphere, so the genus of $y^2 = x$ is zero.)

Moreover, as promised in the previous section, we can now see the reason for the two periods of elliptic functions associated with the curve $y^2 = x(x-1)(x+1)$. The periods are integrals over independent closed paths on the torus surface, such as the paths C_1 and C_2 shown in Figure 8.

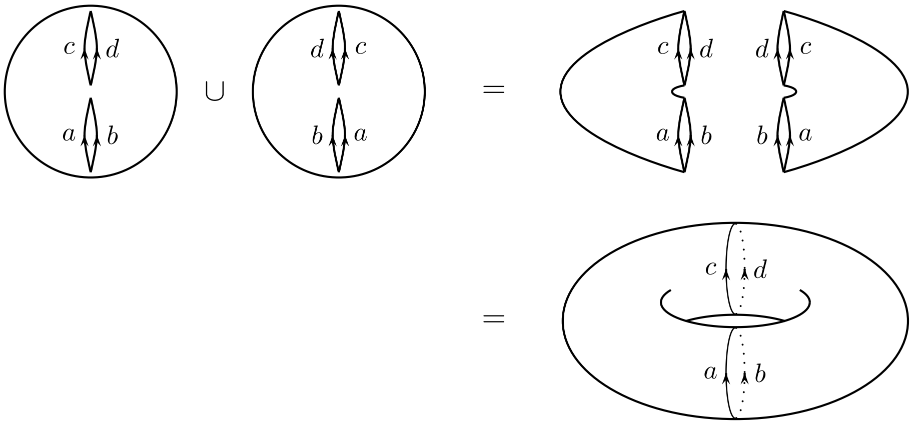

Figure 7. Identifying edges after separating the sheets

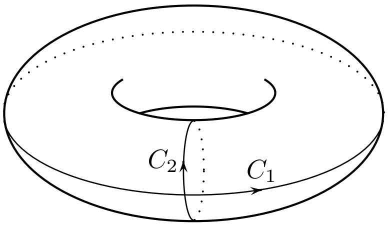

Figure 8. Independent closed paths on the torus

4. The Riemann-Hurwitz Formula

> For a simply connected surface, spread over a finite region of the z-plane, there is a relationship between the number of simple branch points and the number of windings of its boundary curve .... From this there results a relation, for a multiply connected surface, between these numbers and the number of transverse cuts needed to transform it into a simply connected surface. This relation, which does not depend on metric considerations and belongs to *analysis situs*, can be derived as follows ...
>
> Riemann (1857), Section 7, pp. 127–128.

As we saw in the previous section, Riemann viewed an algebraic curve as a surface covering the sphere $\mathbb{C} \cup \{\infty\}$. For a curve of degree n the covering is of n "sheets": that is, above each point of $\mathbb{C} \cup \{\infty\}$ there are n different points of the curve, with the exception of finitely many *ramification points* where two or more sheets come together. Riemann (1857) showed that the genus can be computed from the degree n and numbers e_P that give the number of sheets fused together at a ramification point P. Because of a generalization made by Hurwitz (1891), the method is now called the *Riemann-Hurwitz formula.*

The Riemann-Hurwitz formula is most easily explained in terms of the oldest manifestation of the genus concept, the *Euler characteristic*, which goes back to Euler (1752). Euler observed that, for any decomposition of the sphere into F faces, E edges, and V vertices one has $V - E + F = 2$. More generally, the number $V - E + F$ is invariant for any surface, and it equals $2 - 2p$, where p is the genus. The invariance of $V - E + F$ follows by observing that any two partitions of a surface have a common subdivision, obtained by superimposing one on the other. (This argument is merely plausible, since an edge in one partition may intersect edges of the other partition infinitely often, but Riemann used it in Riemann (1851), Section 6, p. 10.) And any subdivision may be obtained by a series of elementary subdivisions of the following two types:

(1) Subdividing an edge by a new vertex.
(2) Subdividing a face by a new edge connecting two of its vertices.

The first increases both V and E by 1, the second increases both E and F by 1, so neither changes $V - E + F$.

Thus $V - E + F$ is invariant and, by computing it for a standard subdivision of the genus p surface (such as the one shown for the torus in Figure 8), one finds that $V - E + F = 2 - 2p$. This quantity is now called the Euler characteristic, $\chi(\mathcal{S})$, of the surface $\mathcal{S}$ of genus p. It enables us to compute the genus from a decomposition of the surface into vertices, edges, and faces, which is particularly convenient when $\mathcal{S}$ is realized as a ramified covering of the sphere. If $\mathcal{S}$ were simply an n-sheeted covering of the sphere then we should have $\chi(\mathcal{S}) = 2n$, since 2 is the Euler characteristic of the sphere and $\mathcal{S}$ has n vertices over each vertex on the sphere, n edges over each edge, and n faces over each face.

However, when $n > 1$, the covering has ramification points, so we have to adjust the count of vertices. We assume that the ramification points are included among the vertices on the sphere, in which case the count of vertices on $\mathcal{S}$ must be reduced by $e_P - 1$ at each ramification point P. On the other hand, the numbers of edges and faces on $\mathcal{S}$ are still n times the corresponding numbers on the sphere, so we have

$$\chi(\mathcal{S}) = 2n - \sum_P (e_P - 1).$$

Thus the genus p of $\mathcal{S}$ depends only on the degree n of the curve and its ramification numbers e_P.

As an example, consider the 2-sheeted covering of the sphere with the four branch points $P = -1, 0, 1, \infty$ in the previous section. At each of these points $e_P = 2$, so the Riemann-Hurwitz formula gives $2 - 2p = 2 \cdot 2 - 4 = 0$. Hence $p = 1$, as obtained previously.

Since $\chi(\mathcal{S}) = V - E + F = 2 - 2p$, the above formula for $\chi(\mathcal{S})$ may be written as the following formula for genus:

$$p = \frac{1}{2}w - n + 1,$$

where $w = \sum_P (e_P - 1)$ is the sum of the ramification numbers. This formula is used as the *definition* of genus in §24 of the Dedekind-Weber paper.

The determination of genus by degree and ramification numbers is crucial to Dedekind and Weber's program of deriving Riemann's results by algebraic methods. Their algebraic definition of a "Riemann surface" $\mathcal{S}$ lacks geometric structure, so it is not meaningful to decompose $\mathcal{S}$ by vertices, edges, and faces (nor can one make "cuts" in it between the ramification points, as in Figure 6). However, it *is* possible to show that there are n distinct points of $\mathcal{S}$ over each point of the sphere $\mathbb{C} \cup \{\infty\}$, with the exception of finitely many ramification points P, and the ramification number e_P at each such point can be defined algebraically. Thus they can define genus in a way that agrees with Riemann's concept of genus, by means of the Riemann-Hurwitz formula.

5. Functions on Riemann Surfaces

> A system of like-branching algebraic functions and their integrals is the main object of our study; but instead of proceeding from expressions for these functions we define them by their discontinuities with the help of Dirichlet's principle.
>
> Riemann (1857), Section 1, p. 117.

> It certainly is somewhat daring to infer the existence of U from its hydrodynamic significance.
>
> Weyl (1964), p. 107.

With the insight gained from his view of algebraic curves as surfaces covering the plane $\mathbb{C}$ or the sphere $\mathbb{C} \cup \{\infty\}$, Riemann was able to generalize Cauchy's theory of integration on $\mathbb{C}$ to integration on arbitrary algebraic curves. To see where this may lead, let us first recall some of Cauchy's main results, and some of their immediate consequences. Cauchy developed his theory, in stages of increasing generality, between 1814 and 1831. For a very readable and insightful account of this period, see Smithies (1997).

The fundamental result is *Cauchy's theorem*, according to which

$$\int_C f(z)\,dz = 0,$$

where f is differentiable or *holomorphic* (and hence not infinite) on and inside the closed curve C in the complex plane $\mathbb{C}$. It follows that $\int_a^b f(z)\,dz$ does not depend on the path chosen between a and b, as long as the paths lie in a simply connected region of $\mathbb{C}$ where f does not become infinite.

Of course, it is a different story for functions that *do* become infinite, such as $f(z) = 1/z$ at $z = 0$. We find, for example, that

$$\int_C \frac{dz}{z} = 2\pi i,$$

where C is a clockwise path around the unit circle. Nevertheless, the value of the integral of $1/z$ around a closed path C depends only slightly on C—it is the same around any path C' to which C can be deformed without crossing the point $z = 0$.

This example is generalized in *Cauchy's integral formula* and *Cauchy's residue theorem.* The integral formula says that

$$f(a) = \frac{1}{2\pi i} \int_C \frac{f(z)\,dz}{z - a},$$

where C is a circle with center a in a region where f is differentiable, and it has the consequence that

$$f^{(n)}(a) = \frac{n!}{2\pi i} \int_C \frac{f(z)\,dz}{(z - a)^{n+1}},$$

so a differentiable complex function f in fact has derivatives of all orders. It also follows that f has a Taylor series expansion in the neighborhood of a point where f' exists, and from this we can conclude that if $f(a_k) = 0$ for a convergent sequence of points a_k then f is constantly zero. Another important consequence is the so-called *Liouville's theorem*: if f is differentiable and *bounded* on $\mathbb{C}$ then f is constant. (This theorem was discovered by Liouville in 1844 but first published by Cauchy (1844); see Lützen (1990), p. 124.)

The residue theorem generalizes the property of the function $1/z$ to an arbitrary *meromorphic* function f—one that is differentiable except at isolated *poles* $z = a$, in the neighborhood of which

$$f(z) = c_{-h}(z-a)^{-h} + \cdots + c_{-1}(z-a)^{-1} + c_0 + c_1(z-a) + c_2(z-a)^2 + \cdots .$$

The integer h is called the *multiplicity* or *order* of the pole $z = a$, and the coefficient c_{-1} is called the *residue* of f at $z = a$. For such a function,

$$\int_C f(z)\,dz = 2\pi i c_{-1},$$

where C is a clockwise path around $z = a$, small enough not to touch or enclose any other pole of f. Thus the value of the integral depends only on the residue. More generally, if C is a path running clockwise around poles of f at which the residues are $r_1, \ldots, r_m$, then

$$\int_C f(z)\,dz = 2\pi i(r_1 + \cdots + r_m).$$

With these theorems we can find *all the functions f on $\mathbb{C} \cup \{\infty\}$ that are meromorphic. They are precisely the rational functions.* A rational function

$$f(z) = \frac{p(z)}{q(z)}, \quad \text{where } p \text{ and } q \text{ are polynomials,}$$

is differentiable except at its finitely many poles where $q(z) = 0$, so a rational function is certainly meromorphic.

Conversely, if $f(z)$ is meromorphic on $\mathbb{C} \cup \{\infty\}$ and f is not the constant zero, then it can have only finitely many zeros and poles. (If not, there are infinitely many zeros or poles in a neighborhood of ∞. Infinitely many zeros force f to be the constant zero, and infinitely many poles destroy differentiability near infinity.) Now, multiplying $f(z)$ by a rational function $g(z)$ that cancels its zeros and poles, we obtain a function $h(z) = f(z)g(z)$ that is finite and nonzero everywhere except possibly at $z = \infty$. In any case, either $h(z)$ or $1/h(z)$ is bounded, hence constant by Liouville's theorem. Thus $f(z)$ is a constant multiple of the rational function $1/g(z)$, hence rational itself.

Now a rational function $f(z) = p(z)/q(z)$ is determined, up to a constant multiple, by its zeros and poles and their orders. Assuming any common factors of $p(z)$ and $q(z)$ have been canceled, the zeros of $p(z)$ give zeros of $f(z)$ and the zeros of $q(z)$ give poles of $f(z)$. In addition, if $\deg(p) \neq \deg(q)$, there will be a zero or pole at ∞ according as $\deg(p) < \deg(q)$ or $\deg(p) > \deg(q)$, making the total order of zeros equal to the total order of poles.

To summarize: *a meromorphic function on $\mathbb{C} \cup \{\infty\}$ is determined, up to a constant multiple, by a finite set of zeros and poles, with associated orders. A function with given zeros and poles actually exists (namely, a rational function) provided that the total order of zeros equals the total order of poles.*

Thus Riemann's program of "defining a function by its discontinuities" is easily carried out on $\mathbb{C}\cup\{\infty\}$, a Riemann surface of genus zero. We need only understand the "discontinuities" to be the zeros and poles. Implicitly, Cauchy had the result, though perhaps not the means of expressing it in terms of functions on a surface. However, extending this result to *all* surfaces of genus zero is already a significant problem. It was solved by Riemann (1851) via the *Riemann mapping theorem*, a difficult result that he proved by appealing to what he called the *Dirichlet principle.* The Dirichlet principle is a powerful method for proving the existence of functions with given properties (such as specified zeros and poles) on surfaces, but it stems from physical intuition about the flow of electricity and was not proved in a form suitable for Riemann's applications until after 1900.

It should be stressed that this remarkably simple view of meromorphic functions on $\mathbb{C}\cup\{\infty\}$ is made possible by the point ∞. We would now say that the role of ∞ is to make the surface *compact.* Compactness ensures that any infinite set of points has a limit, which makes possible the above argument that a meromorphic function on $\mathbb{C}\cup\{\infty\}$ is rational. The meromorphic functions on $\mathbb{C}$ are a much less manageable class, since they include functions that are not even algebraic, such as e^z. The Riemann (1857) view of an algebraic curve as a surface covering the sphere led him to an equally simple view of the meromorphic functions on such a curve—one that would be the starting point of the Dedekind-Weber theory—they are *algebraic functions*, of degree bounded by the degree of the curve.

To see why, suppose that f is s meromorphic function on a Riemann surface $\mathcal{S}$, so f has only finitely many poles on $\mathcal{S}$. If $\mathcal{S}$ is of degree n then there are n points $x_1,\ldots,x_n$ (not necessarily distinct) over each $x\in\mathbb{C}\cup\{\infty\}$. Consider the values $f(x_1),\ldots,f(x_n)$ of f at these points. Then the "multi-valued function" $y=f(x)$ satisfies

$$(y-f(x_1))\cdots(y-f(x_n))=0.$$

Expanding the left side, we get

$$(*)\qquad y^n+a_{n-1}y^{n-1}+\cdots+a_1y+a_0=0,$$

where

$$\begin{aligned} a_0&=(-1)^n f(x_1)\cdots f(x_n),\\ &\ \vdots\\ a_{n-1}&=-(f(x_1)+\cdots+f(x_n)) \end{aligned}$$

are the elementary symmetric functions of $f(x_1),\ldots,f(x_n)$. Because of their symmetry, $a_0,\ldots a_{n-1}$ are well-defined meromorphic functions of $x\in\mathbb{C}\cup\{\infty\}$. So, $y=f(x)$ satisfies a polynomial equation (*) of degree n whose coefficients are rational functions of x, and this means that $f(x)$ is an algebraic function of x, of degree at most n. This is essentially the argument of Riemann (1857), Section 5, p. 123.

On a surface of genus $p>0$ it could likewise be proved that a meromorphic function is determined, up to a constant multiple, by its zeros and poles, and that the total order of zeros equals the total order of poles. However, it could also be proved that there are further *constraints on existence*, arising from closed paths that do not bound a piece of the surface and the corresponding periods of integrals. For example, on a surface of genus 1 there is no function with one pole of order 1 and one zero of order 1. This raises the problem of finding functions, with given zeros and poles, that satisfy the known constraints.

For genus $p = 1$, constraints were found by Abel (1827), and he was also able to prove that any finite set of zeros and poles satisfying these constraints could be realized by a meromorphic function, that is, by an elliptic function. He did this with the help of "expressions for these functions," which he and Jacobi had already raised to a fine art. "Expressions" for functions on surfaces of higher genus were much less developed in Riemann's time, so he actually made a virtue of a necessity by appealing to Dirichlet's principle in order to prove their existence. The trouble was, Weierstrass (1870) showed that the Dirichlet principle *fails* in certain cases, so Riemann's methods were under suspicion for some decades.

Riemann (1857) considered functions on a surface of genus p with simple poles (that is, poles of order 1) at r given points. These functions form a vector space (over the field $\mathbb{C}$) whose dimension l Riemann proved to satisfy

$$l \geq r - p + 1$$

(Riemann's inequality). Riemann did not use the language of vector spaces and dimensions,[3] which did not yet exist; he said that the functions have l "arbitrary constants." The inequality generalizes to the case where the poles have multiplicities $d_1, \ldots, d_r$, in which case

$$l \geq (d_1 + \cdots + d_r) - p + 1.$$

Riemann's student Roch (1865) turned the inequality into an equality by interpreting the difference

$$l - [(d_1 + \cdots + d_r) - p + 1]$$

as the dimension of a space of certain functions, today called the *canonical class.*

In the special case of meromorphic functions on $\mathbb{C} \cup \{\infty\}$ we can see that

$$l = (d_1 + \cdots + d_r) - p + 1,$$

because in this case $p = 0$ and the meromorphic functions with r poles are of the form

$$f(z) = \frac{p(z)}{k(z - p_1)^{d_1} \cdots (z - p_r)^{d_r}},$$

if ∞ is not one of the poles. In this case $p(z)$ can be any polynomial of degree at most $d_1 + \cdots + d_r$ (the degree of $q(z)$), and the space of such functions indeed has dimension $(d_1 + \cdots + d_r) + 1$, because there are $(d_1 + \cdots + d_r) + 1$ arbitrary constants in the definition of a polynomial of degree $(d_1 + \cdots + d_r)$. If ∞ *is* a pole, then its order is the difference $\deg(p) - \deg(q)$, and it is easily checked that $l = (d_1 + \cdots + d_r) + 1$ in this case also.

We will explain in Section 9 how Dedekind and Weber overhauled these ideas so as to avoid assuming the Dirichlet principle, and thereby transformed Riemann-Roch into a theorem of algebra. There were attempts to prove the Riemann-Roch theorem without appealing to analysis and topology before Dedekind and Weber, but these proofs were not completely general, as Dedekind and Weber indicate in the first few sentences of their paper. They point out that these previous attempts were

[3] Indeed, the concept is still struggling to emerge in Dedekind and Weber's paper. They use the term *Schaar* for what we call a vector space, but they reprove the basic vector space properties for each new *Schaar* that comes up. Dedekind formally described the vector space properties of a *Schaar* for the first time on pp. 467–468 of Dedekind (1894).

made "under certain restrictive assumptions about the singularities of the functions under consideration." This was the case, in particular, for the proof in Brill and Noether (1874), where the term "Riemann-Roch theorem" is first employed. For more details see Gray (1998). In any case, before discussing Dedekind and Weber, we should say a few words on the later development of Riemann's ideas, and their eventual vindication.

6. Later Development of Analysis on Riemann Surfaces

> Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry.
>
> From Poincaré's essay "Mathematical creation" in Poincaré (1918).

As explained in the previous section, Riemann viewed algebraic curves as surfaces, so that "multi-valued functions" such as $\sqrt{z}$ became single-valued and the genus p had a simple topological meaning. He also interpreted meromorphic functions in terms of *flows of electricity*, brought about by applying the poles of a battery to points of a surface covered by an infinitely thin layer of conducting material. (In fact, this is apparently where the word "pole" comes from.) An account of Riemann's theory in these frankly unrigorous, but intuitively helpful, terms was given by Klein (1882).

Klein's book (appearing in the same year as the Dedekind and Weber paper) made no advance towards a proof of Riemann's theorems, since the crucial Dirichlet principle was still assumed without proof, and the concept of surface remained vague. However, it did inspire other mathematicians to rigorize the questionable parts of Riemann's work. Hilbert (1904) proved a "Dirichlet principle" strong enough for Riemann's needs, and Weyl (1913) completed the theory with a precise definition of Riemann surfaces. In fact, Weyl's definition was soon imitated in all parts of complex analysis and differential geometry, where it was useful to have a concept of a *manifold with differentiable structure*, and where one wants to decide which structures are isomorphic.

This was not the only direction in which the concept of Riemann surface developed. Another was in *generalizing the theory of elliptic functions to higher genus.* To explain what there is to generalize, I will outline the basic facts about elliptic functions and their relation to Riemann surfaces.

Thanks to the work of Abel and Jacobi, elliptic functions were already well known in the 1820s as doubly-periodic functions on $\mathbb{C}$. As we have already mentioned in Section 3, the Riemann surface concept explains the two periods as integrals around certain closed curves on the torus, such as C_1 and C_2 in Figure 8. However, it is also possible to exhibit double periodicity in a *formula* for an elliptic function. The simplest possible formula is one due to Eisenstein (1847):

$$f(z)=\sum_{m,n\in\mathbb{Z}}\frac{1}{(z+m\omega_1+n\omega_2)^2}.$$

Assuming that this series is meaningful (which it is, if the summation is interpreted properly), then it clearly remains the same if z is replaced by $z+\omega_1$ or $z+\omega_2$.

Thus f has two periods, ω_1 and ω_2. Almost as simple, and more standard, is the *pe-function* of Weierstrass (1863):

$$\wp(z) = \frac{1}{z^2} + \sum_{(m,n)\neq(0,0)} \left(\frac{1}{(z+m\omega_1+n\omega_2)^2} - \frac{1}{(m\omega_1+n\omega_2)^2} \right).$$

The Weierstrass function $\wp(z)$ is not so obviously periodic (one first proves this for $\wp'$), but it is easier to work with because its series is uniformly convergent except at the double poles where $z = m\omega_1 + n\omega_2$. For example, Weierstrass was able to show, by simple series manipulations, that

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

where g_2 and g_3 are certain constants depending on ω_1 and ω_2. Thus the functions $x = \wp(z)$ and $y = \wp'(z)$ parameterize the curve

$$y^2 = 4x^3 - g_2x - g_3,$$

to which any curve of genus 1 happens to be birationally equivalent (for suitable choice of g_2 and g_3).

Any function with periods ω_1 and ω_2, such as $\wp$, has values that repeat in each parallelogram in the tessellation of the plane $\mathbb{C}$ shown in Figure 9. We assume that ω_1 and ω_2 lie in different directions from O, so that their integer combinations $m\omega_1 + n\omega_2$ form a *lattice* of parallelograms. The points marked by stars, which are "equivalent modulo the lattice," form a set on which all values of the function are the same.

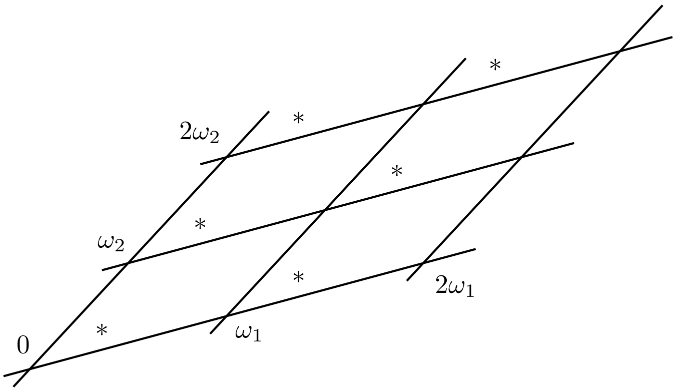

FIGURE 9. Lattice-equivalent points

Thus $\wp$ can be viewed as a function on the surface whose "points" are the classes of lattice-equivalent points

$$z + \omega_1\mathbb{Z} + \omega_2\mathbb{Z} = \{z + m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}.$$

We call the surface the *quotient* $\mathbb{C}/\Gamma$ of the plane $\mathbb{C}$ by the group Γ of translations $z \mapsto z + m\omega_1 + n\omega_2$ for $m, n \in \mathbb{Z}$. This Riemann surface, not surprisingly, is a torus, obtained by identifying (or "pasting") opposite sides of the parallelogram with vertices $0, \omega_1, \omega_2, \omega_1 + \omega_2$, as shown in Figure 10. Thus, we can also arrive at meromorphic functions on the torus by starting with suitably *periodic* functions in the plane $\mathbb{C}$. In the years 1880–1882 (when the Dedekind-Weber paper was awaiting

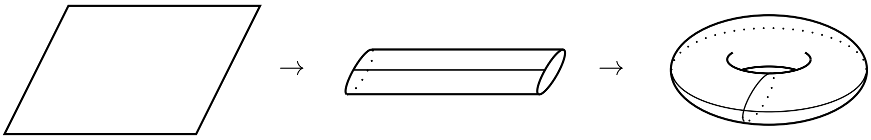

Figure 10. Construction of a torus by pasting

publication) a remarkable approach to meromorphic functions on surfaces of genus $p > 1$ was discovered through an exploration of *non-Euclidean periodicity*, on the half-plane $\{z : \mathrm{Im}(z) > 0\}$ or the disk $\{z : |z| < 1\}$.

Isolated examples of functions with striking periodic behavior were discovered before 1880, but their periodicity did not have a context or a name. The first picture of this new kind of periodicity to appear in print was given by Schwarz (1872), and it exhibits the periodicity of a function now known as a Schwarz triangle function. The periodicity is indicated in Figure 11 via a tessellation of the disk by curvilinear triangles, in each of which the function repeats its values.

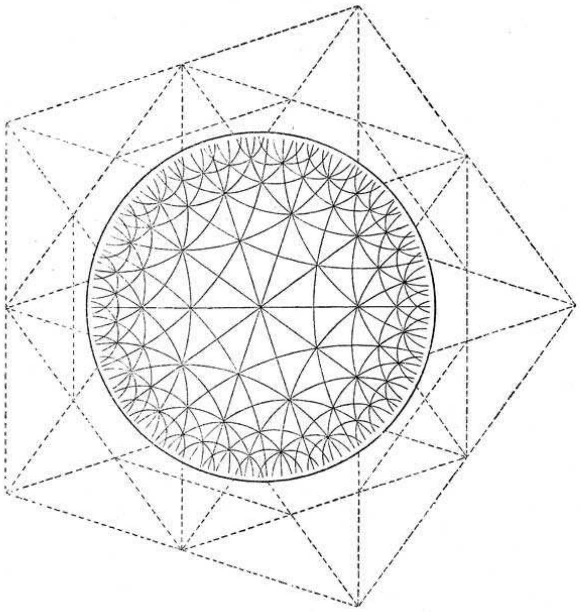

Figure 11. The Schwarz tessellation

The triangles are of course not congruent in the euclidean sense, but in 1880 Poincaré realized that they are *congruent in the sense of the non-Euclidean geometry* of Bolyai and Lobachevsky. Following this discovery, Poincaré (1882) brought non-Euclidean geometry into mainstream mathematics, and unveiled a new view of functions on Riemann surfaces—as functions with non-Euclidean periodicity.

The relationship of the disk to a Riemann surface $\mathcal{S}$ of genus $p > 1$ is like that of the plane to the torus: $\mathcal{S}$ is a quotient of the disk by a discrete group of (non-Euclidean) translations, and the disk is tessellated by congruent copies of a polygon obtained by suitably cutting the surface. The simplest case is the surface of genus 2, which can be cut along the curves shown in Figure 12 to form an octagon whose corner angles sum to 2π.

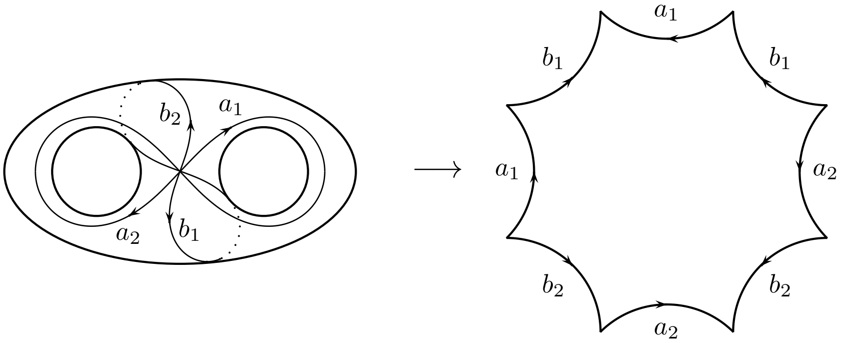

FIGURE 12. Dissection of genus 2 surface

The natural home of such an octagon is the non-Euclidean plane, which can be modeled by the unit disk with circular arcs orthogonal to the boundary circle as "lines." In particular, we can arrange eight such arcs so as to form an octagon with corner angles $\pi/4$, and (non-Euclidean) translations of this octagon fill the whole disk as shown in Figure 13.

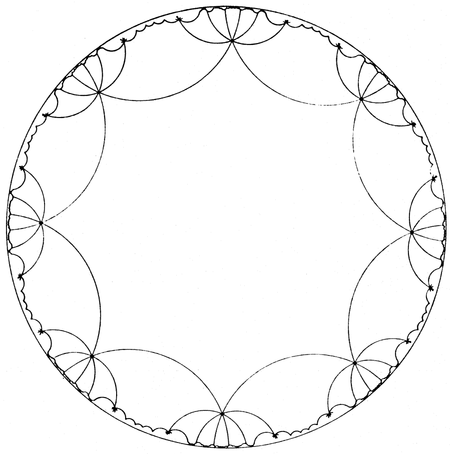

FIGURE 13. Tessellation of the disk by octagons

We can then recover the surface $\mathcal{S}$ of genus 2 as the quotient of the disk by the group Γ of all translations that map the tessellation onto itself. Finally, meromorphic functions on $\mathcal{S}$ correspond to meromorphic functions f on the disk with periodicity given by Γ, that is, functions with the property that

$$f(z) = f(g(z)), \quad \text{for any translation } g \in \Gamma.$$

Thus we are reduced to the problem of constructing meromorphic functions f on the disk with periodicity given by Γ. This problem was solved by Poincaré (1883), using insights from the theory of elliptic functions. The solution builds all elements $g \in \Gamma$ into f, though not so simply as Eisenstein (1847) built Euclidean translations into the definition of a doubly-periodic function. We omit the details.

Poincaré's construction of functions with non-Euclidean periodicity (*Fuchsian functions*, or *automorphic functions* as they became known) was a spectacular extension of the theory of elliptic functions to surfaces of higher genus. At the same time, it showed that elliptic functions and the torus are special, being associated with Euclidean geometry, and that the general Riemann surface should be viewed as non-Euclidean. However, like Riemann, Poincaré assumed an unproved theorem—the so-called *uniformization theorem*—to give his results their most general form.

Just as Riemann needed an assumption (the Dirichlet principle) to prove that any surface of genus 0 is isomorphic to the sphere $\mathbb{C} \cup \{\infty\}$, Poincaré needed to assume that any Riemann surface of genus $p > 1$ is isomorphic to a quotient of the disk by a discrete group of non-Euclidean translations. The uniformization theorem encompasses both these theorems by stating that any *simply connected* Riemann surface is isomorphic to either the sphere $\mathbb{C} \cup \{\infty\}$, the Euclidean plane $\mathbb{C}$, or the unit disk $\{z : |z| < 1\}$. Uniformization also extends the known parameterization theorems for algebraic curves of genus 0 (parameterization by rational functions) and genus 1 (parameterization by elliptic functions), because it implies that any algebraic curve of genus > 1 may be parameterized by automorphic functions.

Poincaré's work was clearly a brilliant extension of Riemann's ideas, but it remained under a cloud for some time because of doubts about its rigor. Indeed his papers of 1882 and 1883 were not published by the leading journals of the time; instead, they played a big part in establishing the new journal *Acta Mathematica*.[4] The doubts were eventually dispelled with proofs of the uniformization theorem by Poincaré (1907) and Koebe (1907), and by the rigorous theory of Riemann surfaces developed by Weyl (1913). Thus, it took more than 50 years for the analytic theory of algebraic curves to be fully accepted, and perhaps this worked to the advantage of algebra and algebraic geometry. The algebraic restructuring of Riemann's concepts by Dedekind and Weber may not have taken place had an analytic alternative been available, and we might thereby have missed a radically new insight into the nature of algebraic curves. The first textbook development of the algebraic theory was Hensel and Landsberg (1902), and then only with a reversion to certain ideas from analysis, such as infinite series expansions. For an English version of the Hensel-Landsberg approach, with some simplifications, see Bliss (1933).

Even today, it is a little strange to think that number theory inspired the first complete and rigorous proof of the Riemann-Roch theorem. Nevertheless, if one looks at the proof of the theorem in any modern book one will see that it involves things called "divisors." A divisor is nothing but a finite set of points with attached integer multiplicities—like a set of zeros and poles—yet it sounds as if the concept comes from number theory. In the next sections we explain how this happened.

[4] Looking back to his founding of *Acta Mathematica*, Mittag-Leffler (1923) recalled:

> Kronecker, for example, expressed to me via a mutual friend his regret that the journal seemed bound to fail, without help, through publishing a work so incomplete, unripe, and obscure.

7. Origins of Algebraic Number Theory

> It is greatly to be lamented that this virtue of the real numbers [i.e., the rational integers] to be decomposable into prime factors, always the same ones for a given number, does not also belong to the complex numbers; were this the case, the whole theory, which is still laboring under such difficulties, could easily be brought to its conclusion. For this reason, the complex numbers we have been considering seem imperfect, and one may well ask whether one ought to look for another kind which would preserve the analogy with the real numbers ...
>
> Translation by Weil (1975) from Kummer (1844).

The concept of divisibility has been fundamental in number theory for more than 2000 years. Euclid's *Elements*, Book VII, introduces the Euclidean algorithm in Proposition 1, and shows that it yields the greatest common divisor of two numbers in Proposition 2. Eventually Euclid deduces, in Proposition 30, that if a prime p divides a product ab then p divides a or p divides b. This prime divisor property easily yields what we now call the *fundamental theorem of arithmetic*: each natural number has unique prime factorization (up to the order of factors).

Unique prime factorization was first explicitly stated as a theorem in Gauss (1801), but before then it was frequently assumed without comment. Indeed, unique prime factorization was sometimes assumed for numbers seemingly far from the natural numbers. One of the first instances was the spectacular proof of Euler (1770), p. 401, that the only natural number solution of

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

is $x = 5$, $y = 3$, a result that had been claimed by Fermat. To prove this result, Euler factorized the right side of the equation, obtaining

$$y^3 = (x + \sqrt{-2})(x - \sqrt{-2}),$$

and then worked with numbers of the form $a + b\sqrt{-2}$ (for integers a and b) as if they were ordinary integers. In particular, he argued that $x + \sqrt{-2}$ and $x - \sqrt{-2}$ are relatively prime and therefore (apparently assuming unique prime factorization) that each is a cube, because their product is a cube. But if

$$x + \sqrt{-2} = (a + b\sqrt{-2})^3$$

it follows that

$$x = a^3 - 6ab^2 = a(a^2 - 6b^2), \quad 1 = 3a^2b - 2b^3 = b(3a^2 - 2).$$

The only integer solutions of the latter equation are $a = \pm 1$, $b = 1$, of which only $a = -1$, $b = 1$ give a natural number solution, namely $x = 5$.

This is magnificent, but is it number theory?

Evidently, we need to rebuild the theory of divisibility and primes for new types of numbers, such as those of the form $a + b\sqrt{-2}$. The first to take steps in this direction was Gauss (1832), who worked out the basic theory of what we now call the *Gaussian integers*, $\mathbb{Z}[i]$. These are the numbers of the form $a + bi$, where a and b are integers and $i = \sqrt{-1}$.

We now sketch the divisibility theory of $\mathbb{Z}[i]$, since it admits a visual interpretation that will be helpful in other cases. The only result we really need to prove is the *division property* (sometimes called the "division algorithm"): *if* α *and* $\beta \neq 0$

are Gaussian integers, then there are Gaussian integers μ and ρ such that

$$\alpha = \mu\beta + \rho, \quad \textit{where} \quad |\rho| < |\beta|.$$

If we can prove this, then there is a Euclidean algorithm, and unique prime factorization in $\mathbb{Z}[i]$ follows as it does for ordinary integers.

Well, for any Gaussian integer $\beta \neq 0$, consider the multiples $\mu\beta$ of β by all the Gaussian integers μ. These are the integer combinations of the perpendicular vectors β and $i\beta$, which form squares of side length $|\beta|$, as shown in Figure 14.

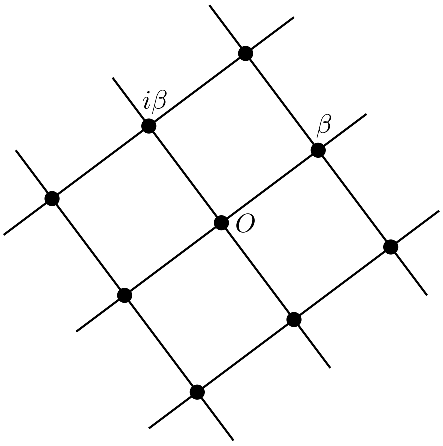

FIGURE 14. Multiples of β in $\mathbb{Z}[i]$

The Gaussian integer α falls in one of these squares, so if $\mu\beta$ is the nearest corner we have

$$|\rho| < |\beta|, \quad \text{where} \quad \rho = \alpha - \mu\beta,$$

because the distance of a point in a square from the nearest corner is less than the side length. This establishes the division property, and hence unique prime factorization in $\mathbb{Z}[i]$.

The argument is similar for the set $\mathbb{Z}[\sqrt{-2}]$ of numbers of the form $a + b\sqrt{-2}$ considered by Euler. The multiples $\mu\beta$ of such a number β form a grid of rectangles, with sides of length $|\beta|$ and $|\beta|\sqrt{2}$, but it remains true that the distance of a point α from the nearest corner is less than $|\beta|$. Thus the division property holds for $\mathbb{Z}[\sqrt{-2}]$, hence a Euclidean algorithm and unique prime factorization, so the main assumption of Euler's proof is valid. (The other assumption, about relative primality, is also easy to justify.)

Many results about ordinary integers are most easily proved by passing to "quadratic integers," such as $\mathbb{Z}[i]$, and appealing to unique prime factorization. For example, consider the prime 797, which is the sum of two squares $26^2 + 11^2$. Is this sum of squares unique? Yes, because if $797 = a^2 + b^2$ we have $797 = (a+bi)(a-bi)$, which is a *prime* factorization because $|a + bi|^2 = |a - bi|^2$ is the ordinary prime $a^2 + b^2 = 797$ (so neither $a+bi$ nor $a-bi$ is a product of smaller Gaussian integers). Then, since Gaussian prime factorization is unique, so is the decomposition of 797 into a sum of squares.

Unfortunately, there is trouble not far ahead. If we wish to prove theorems about natural numbers of the form a^2+5b^2 by using the factorization

$$a^2+5b^2=(a+b\sqrt{-5})(a-b\sqrt{-5})$$

we cannot assume unique prime factorization, because it *fails* for the set $\mathbb{Z}[\sqrt{-5}]$ of numbers of the form $a+b\sqrt{-5}$. The classical example, from Dedekind (1877), is

$$6=2\cdot 3=(1+\sqrt{-5})(1-\sqrt{-5}).$$

Each of the numbers 2, 3, $1+\sqrt{-5}$, $1-\sqrt{-5}$ is "prime," in the sense that none of them is a product of smaller numbers in $\mathbb{Z}[\sqrt{-5}]$, so we have nonunique prime factorization in $\mathbb{Z}[\sqrt{-5}]$. Gauss seemed to be aware that unique prime factorization could fail in cases like this; his theory of quadratic forms in Gauss (1801) seems designed to avoid such problems, sometimes at considerable expense.

However the first to note the phenomenon in print, and to declare that something should be done about it, was Kummer (1844) (see the quote at the beginning of this section). In a bold attempt to rescue unique prime factorization, Kummer introduced what he called "ideal numbers." The name was apparently motivated by the "ideal" elements that were then becoming accepted in geometry, such as points at infinity, but the *concept* drew inspiration from Jacobi's work on number theory in the 1830s, as has been shown by Lemmermeyer (2009). In Kummer's work, "ideal numbers" arise in a rather complicated way, and we will instead follow Dedekind (1877) and explain how they work in $\mathbb{Z}[\sqrt{-5}]$.

To reconcile the the factorizations

$$6=2\cdot 3=(1+\sqrt{-5})(1-\sqrt{-5})$$

with unique prime factorization, we have to believe that 2, 3, $1+\sqrt{-5}$, $1-\sqrt{-5}$ are *not* primes, after all. They must somehow split into smaller, "ideal," factors—but what are these ideal factors? A candidate that comes to mind is the greatest common divisor of 2 and $1+\sqrt{-5}$. We cannot put our finger on any such number but, by borrowing an idea from classical number theory, we can describe the *set of multiples* of a greatest common divisor. In classical number theory, the multiples of $\gcd(a,b)$ are all the numbers of the form $ma+nb$. In $\mathbb{Z}[\sqrt{-5}]$, there is no actual number that can serve as $\gcd(2,1+\sqrt{-5})$, but the set

$$\{2\mu+(1+\sqrt{-5})\nu:\mu,\nu\in\mathbb{Z}[\sqrt{-5}]\}=\{2m+(1+\sqrt{-5})n:m,n\in\mathbb{Z}\}$$

is perfectly concrete, and it can serve as the set of "multiples of the ideal number $\gcd(2,1+\sqrt{-5})$."

Figure 15 shows what this set looks like. Its members are the black dots on the grid of points in $\mathbb{Z}[\sqrt{-5}]$. Notice that these black dots do *not* form a grid of rectangles, as they would if they were multiples of an actual number in $\mathbb{Z}[\sqrt{-5}]$—because the multiples of a number α are simply the rectangular grid $\mathbb{Z}[\sqrt{-5}]$, magnified by $|\alpha|$ and rotated through the argument of α.

The number $\gcd(2,1+\sqrt{-5})$ is "ideal," but the set of its multiples is perfectly real, and visible. When we admit "ideal numbers," via their sets of multiples, it turns out that there are "ideal primes," and that each number in $\mathbb{Z}[\sqrt{-5}]$ factorizes uniquely into ideal primes. This is how Kummer recovered unique prime factorization, and hence preserved the analogy with the ordinary integers.

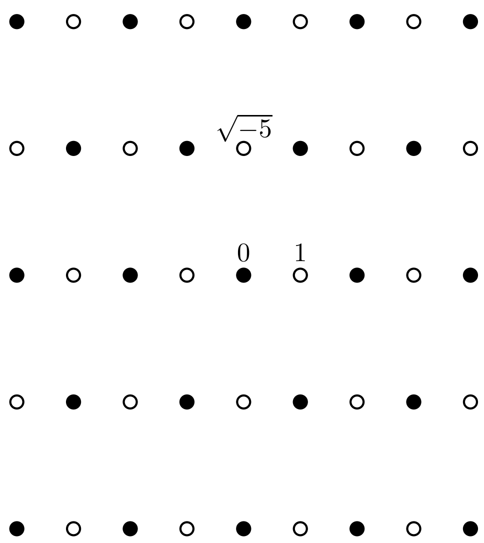

FIGURE 15. Multiples of the ideal number $\gcd(2, 1+\sqrt{-5})$ in $\mathbb{Z}[\sqrt{-5}]$

8. Dedekind's Theory of Algebraic Integers

> The great success of Kummer's researches in the domain of circle division allows us to suppose that the same laws hold in *all* numerical domains .... I did not achieve the general theory ... until I abandoned the old formal approach and replaced it by another; a fundamentally simpler conception focussed directly on the goal. In the latter approach I need no concept more novel than that of Kummer's *ideal numbers*, and it is sufficient to consider a *system of actual numbers* that I call an *ideal*.
>
> Dedekind (1877), pp. 56–57.

Kummer was interested in a particular class of numbers, now known as *cyclotomic integers*, most famously for their application to Fermat's last theorem. He developed his theory of "ideal numbers" only for the cylotomic integers, and it was left to Dedekind to develop a general theory of algebraic integers, and what we now know as the theory of ideals. Dedekind's first account of this theory appeared in Dedekind (1871), an appendix to Dirichlet's *Vorlesungen über Zahlentheorie*, which Dedekind edited. This did not attract as much interest as Dedekind hoped, and he produced a more down-to-earth exposition of the theory in Dedekind (1877). I have drawn on the latter version here.

From a handful of examples such as $\mathbb{Z}[i]$, $\mathbb{Z}[\sqrt{-2}]$, and Kummer's cyclotomic integers, Dedekind (1871) synthesized a general concept of *algebraic integer.* Then, by properly situating algebraic integers in the broader context of algebraic *numbers*, he was able to define ideal prime numbers and prove the uniqueness of ideal prime factorization. In outline, his train of thought was the following.

Begin with the *algebraic numbers.* A number α is algebraic if it is the root of an equation

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0, \quad \text{where} \quad a_0, a_1, \ldots, a_n \in \mathbb{Z}.$$

Among the algebraic numbers, those satisfying equations of this form in which $a_n = 1$ (*monic* equations) are called *algebraic integers.* It can be proved (not quite trivially) that the algebraic numbers are closed under the operations $+, -, \times, \div$ (by a nonzero number), and hence they form a field, and that the algebraic integers are closed under the operations $+, -, \times$, and hence they form a ring. The concept of algebraic integer is compatible with the ordinary concept of integer because any rational solution of a monic equation is an ordinary integer. This fact was already mentioned by Gauss (1801), article 11.

For these reasons, and others that will appear below, the concept of algebraic integer is a good generalization of the ordinary integer concept. This is not completely obvious, because there are algebraic integers that do not "look integral." One might think that the cube root of 1,

$$\zeta_3 = \frac{-1+\sqrt{-3}}{2},$$

should not be regarded as integer because of its "fractional" appearance. Yet it is a root of the monic equation, $x^2 + x + 1 = 0$. And in fact it is better to work in the ring

$$\mathbb{Z}[\zeta_3] = \{a + b\zeta_3 : a, b \in \mathbb{Z}\}$$

than in the ring

$$\mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\},$$

because $\mathbb{Z}[\zeta_3]$ has unique prime factorization and $\mathbb{Z}[\sqrt{-3}]$ does not.

However, it is *not* a good idea to work with the ring of all algebraic integers, because it cannot possibly have unique prime factorization. This is because any algebraic integer α has a factorization

$$\alpha = \sqrt{\alpha}\sqrt{\alpha},$$

and $\sqrt{\alpha}$ is also an algebraic integer. Dedekind (1871) saw that the way to avoid this problem is to work in an *algebraic number field F of finite degree over* $\mathbb{Q}$, and with the ring of algebraic integers in F. Any such field has the form[5]

$$\mathbb{Q}(\alpha) = \{p(\alpha)/q(\alpha) : p, q \text{ polynomials with integer coefficients}\},$$

for some algebraic number α. The degree of α (that is, the degree of the minimal polynomial satisfied by α) is called the *degree of* $F = \mathbb{Q}(\alpha)$, $\deg(F)$.

Each algebraic integer β in F has a minimal polynomial

$$x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0,$$

where $n \leq \deg(F)$, and $b_0, \ldots, b_{n-1}$ are ordinary integers. The ordinary integer $(-1)^n b_0$ is called the *norm* of β, $N(\beta)$. $N(\beta)$ is the product of β with all the other roots of the minimal polynomial, called the *conjugates* of β. For example,

[5] Dedekind actually defines a field F of finite degree n over $\mathbb{Q}$ more generally, as a field of dimension n as a vector space over $\mathbb{Q}$. But it is true that there is always a *primitive element* α such that $F = \mathbb{Q}(\alpha)$. This result, essentially due to Galois, also applies to the function fields that appear later, so we will also suppose each of them to be generated by a primitive element.

a Gaussian integer $\beta = a + bi$ has one conjugate, the ordinary complex conjugate $\overline{\beta} = a - bi$, and

$$N(\beta) = (a + bi)(a - bi) = a^2 + b^2 = |\beta|^2.$$

The familiar property of absolute value, $|\alpha\beta| = |\alpha||\beta|$, generalizes to the *multiplicative property of norm*:

$$N(\alpha\beta) = N(\alpha)N(\beta).$$

It follows from the multiplicative property that, if $\gamma = \alpha\beta$, then we have $N(\gamma) = N(\alpha)N(\beta)$. This reduces certain questions about divisibility of algebraic integers to questions about divisibility of ordinary integers. For example, if α divides γ, then $N(\alpha)$ divides $N(\gamma)$. It follows that the process of factorizing an algebraic integer γ must eventually terminate with a factorization

$$\gamma = \alpha_1\alpha_2 \cdots \alpha_k,$$

where $N(\alpha_1), N(\alpha_2), \ldots, N(\alpha_k)$ are ordinary primes.

Thus any integer γ of F has a factorization into integers $\alpha_1, \alpha_2, \ldots, \alpha_n$ that are *irreducible* in the sense that α_i is not the product of integers of smaller norm.

We know from the example

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

in $\mathbb{Z}[\sqrt{-5}]$, which is the ring of integers of $\mathbb{Q}(\sqrt{-5})$, that factorization into irreducibles is not always unique. But we also know that unique prime factorization can be recovered, in this case, by further splitting into "ideal numbers." Dedekind generalized this recovery program with his concept of *ideals.* An *ideal* $\mathfrak{a}$ is a set of integers of a field F with the following closure properties:

1. If $\alpha \in \mathfrak{a}$ and $\alpha' \in \mathfrak{a}$ then $\alpha + \alpha' \in \mathfrak{a}$ and $\alpha - \alpha' \in \mathfrak{a}$.
2. If $\alpha \in \mathfrak{a}$ and μ is any integer of F, then $\mu\alpha \in \mathfrak{a}$.

These properties capture the idea of a "set of integer multiples" of an integer, actual or ideal.

If α is an integer of F then its set of multiples

$$(\alpha) = \{\mu\alpha : \mu \text{ an integer of } F\}$$

certainly has the properties of an ideal. We call (α) the *principal ideal* generated by α. The set in $\mathbb{Z}[\sqrt{-5}]$,

$$\{2\mu + (1 + \sqrt{-5})\nu : \mu, \nu \in \mathbb{Z}[\sqrt{-5}]\} = \{2m + (1 + \sqrt{-5})n : m, n \in \mathbb{Z}\},$$

also has the required closure properties, so it is also an ideal. But, as we have seen, this ideal is *not* the set of integer multiples of an actual integer in $\mathbb{Q}(\sqrt{-5})$—it is a *nonprincipal* ideal.

If $\mathfrak{a}$ and $\mathfrak{b}$ are ideals, Dedekind (1871) defines their *product* by

$$\mathfrak{ab} = \{\alpha_1\beta_1 + \cdots + \alpha_m\beta_m : \alpha_1, \ldots, \alpha_m \in \mathfrak{a} \text{ and } \beta_1, \ldots, \beta_m \in \mathfrak{b}\}.$$

This definition agrees with the natural product $(\alpha)(\beta) = (\alpha\beta)$ when $\mathfrak{a} = (\alpha)$ and $\mathfrak{b} = (\beta)$ are principal ideals. Finally, Dedekind (1871) says that $\mathfrak{b}$ *divides* $\mathfrak{a}$ if $\mathfrak{b} \supseteq \mathfrak{a}$ ("to divide is to contain"). This also agrees with the natural concept of division for principal ideals. For example, 2 divides 6, so naturally (2) divides (6), and indeed $(2) \supseteq (6)$ because

$$(2) = \{0, \pm 2, \pm 4, \pm 6, \ldots\} \supseteq \{0, \pm 6, \pm 12, \ldots\} = (6).$$

Up to this point, the theory is a straightforward extension of the usual divisibility concept to ideals, if one views ideals as sets of multiples. But, as Dedekind (1877), §23, points out:

> It is very easy to see (§22,1) that any product of $\mathfrak{a}$ by an ideal $\mathfrak{b}$ is divisible by $\mathfrak{a}$, but it is by no means easy to show the converse, that each ideal divisible by $\mathfrak{a}$ is the product of $\mathfrak{a}$ by an ideal $\mathfrak{b}$.

Eventually, Dedekind overcomes the difficulty by generalizing the theory of divisibility to more general subsystems of integers of F called *orders*, and unique prime ideal factorization follows.

We will not go into further detail, because the details are somewhat different in the case of algebraic function fields, as Dedekind and Weber point out. We only wish to emphasize how important it was to be aware of the notion of ideal. The theory of algebraic functions would not have been discovered without this awareness of ideals, or something similar.

There is in fact an alternative to the concept of ideal, namely, the concept of *divisor* developed by Kronecker. For the full story, see Edwards (1980). Kronecker's work overlaps with Dedekind's and in fact his major paper, Kronecker (1882), appears in the same issue of the journal containing the Dedekind-Weber paper (probably not by accident,[6] since Kronecker was an editor of the journal). Kronecker's paper surpasses the Dedekind-Weber paper in some ways, by considering functions of several variables, but it falls short on specifics and applications, such as proofs of the Abel and Riemann-Roch theorems. For this reason, and its greater readability, the Dedekind-Weber paper has been far more influential. Nonetheless, we must give credit to Kronecker for his divisor concept. The word "divisor" is the one that has passed into algebraic geometry today, in place of what Dedekind and Weber called a "polygon." In the next section we explain how the concept of divisibility made its way from algebraic number theory to algebraic geometry.

9. Number Fields and Function Fields

> Dedekind and Weber propose to give algebraic proofs of all of Riemann's algebraic theorems. But their remarkable originality (which in all the history of algebraic geometry is only scarcely surpassed by that of Riemann) leads them to introduce a series of ideas that will become fundamental in the modern era.
>
> Dieudonné (1985), p. 29.

If we had to explain the nature of algebraic number theory in a nutshell, we could say that it is the result of generalizing the theory of $\mathbb{Z} \subseteq \mathbb{Q}$ (the "rational integers") to the integers in an extension field $\mathbb{Q}(\alpha)$ of $\mathbb{Q}$, where α is an algebraic number. In particular, one tries to generalize the theory of divisibility and primes, with a view to retaining, or recovering, unique prime factorization. The search for suitable "primes" leads to the discovery of the key concept of ideal.

The 19th-century theory of algebraic functions begins with the analogous theory of polynomials in the field $\mathbb{C}(x)$ of rational functions of x with complex coefficients, and it similarly studies extensions of $\mathbb{C}(x)$. So first we should see to what extent the polynomials behave like "integers" in the field of rational functions.

[6] Indeed, Kronecker delayed publication of the Dedekind-Weber paper for over a year, until his paper could appear at the same time. See Edwards (1980), p. 370.

Certainly, the polynomials form a ring. Also, it has been known since Stevin (1585) that they have a division property and hence they admit a Euclidean algorithm. The division property takes the following form: if $f(x)$ and $g(x) \neq 0$ are polynomials, then there are polynomials $q(x)$ and $r(x)$ ("quotient" and "remainder") such that

$$f(x) = g(x)q(x) + r(x), \quad \text{where} \quad \deg(r) < \deg(g).$$

Thus the remainder $r(x)$ is "smaller" than $g(x)$ when measured by its degree. The Euclidean algorithm therefore terminates, and yields the gcd of polynomials $f(x)$ and $g(x)$, in a number of steps bounded by the degree of g. It follows by the usual steps that unique prime factorization holds for the ring $\mathbb{C}[x]$ of polynomials in $\mathbb{C}(x)$. To be precise, prime factorization is unique up to nonzero constant factors. Moreover, we can say exactly what the prime polynomials are: by the fundamental theorem of algebra, they are the linear polynomials $x - c$, for complex numbers c.

Now the field $\mathbb{C}(x)$ of rational functions is, as we saw in Section 4, the field of meromorphic functions on the sphere $\mathbb{C} \cup \{\infty\}$. The primes $x - c$ of this field correspond to the points c of $\mathbb{C}$. So, in a sense, we can recover the points of the surface on which the field $\mathbb{C}(x)$ "lives" from the primes of the field itself. (Admittedly, we are still missing the point ∞, but presumably we can squeeze it out of $\mathbb{C}(x)$ somehow.)

This seemingly retrograde idea—defining a Riemann surface from the functions on it—is one of the most original and prescient ideas of the Dedekind-Weber paper. But before we can explain how it works, we must say more about function fields, in order to show how a function field may be plausibly related to a Riemann surface in the first place. As mentioned in Section 4, the connection arises from Riemann's discovery that the meromorphic functions on a Riemann surface are algebraic, that is, solutions of polynomial equations. It is also clear that the set of all meromorphic functions on a given Riemann surface forms a field, because it is obviously closed under the operations of $+$, $-$, $\times$, and $\div$ (by a nonzero function).

Now, if we pursue that idea of constructing algebraic function fields in analogy with algebraic number fields, then we should extend $\mathbb{C}(x)$ to $\mathbb{C}(x)(y)$, where y (a "primitive element" for the field) satisfies an equation of the form

$$a_n y^n + a_{n-1} y^{n-1} + \cdots + a_1 y + a_0 = 0, \quad \text{where} \quad a_0, a_1, \ldots, a_n \in \mathbb{C}(x).$$

The elements of $\mathbb{C}(x)$ are rational functions, that is, quotients of polynomials. By multiplying by their common denominator we can therefore rewrite the equation satisfied by y in the form

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

where $f(x, y)$ is a polynomial. In other words, y is defined by the equation of an algebraic curve, and hence of a Riemann surface. The function field generated by y consists of the rational functions in y,

$$\begin{aligned}\mathbb{C}(x)(y) &= \{p(y)/q(y) : p, q \text{ polynomials with coefficients in } \mathbb{C}(x)\}\\ &= \{\text{rational functions in } x, y \text{ such that } f(x, y) = 0\}\\ &= \mathbb{C}(x, y)/(f(x, y)),\end{aligned}$$

which is called the field of rational functions *on* the curve $f(x, y) = 0$, or on the Riemann surface $f(x, y) = 0$. Thus a finite-degree extension of $\mathbb{C}(x)$ is naturally viewed as the field of rational functions on a Riemann surface.

Exactly why the field Σ rational functions in x, y such that $f(x, y) = 0$ should be called "the rational functions on the curve $f(x, y) = 0$" is seldom explained. One of the few authors who bothers to do so is Walker (1950), p. 132 (by "a point of f" he means a point of the curve $f(x, y) = 0$):

> If $g_1(x, y)/h_1(x, y)$ and $g_2(x, y)/h_2(x, y)$ are rational functions ... and (a, b) is any point of f at which these functions have values, then these values will be equal if $g_1(\xi, \eta)/h_1(\xi, \eta) = g_2(\xi, \eta)/h_2(\xi, \eta)$. Conversely, if this last equality does not hold, then $f(x, y)$ does not divide $g_1(x, y)h_2(x, y) - g_2(x, y)h_1(x, y)$, and there are points of f at which the two rational functions assume different values. In other words, as far as the points of f are concerned the rational functions behave like elements of Σ.

We now pause to look at some examples of finite-degree extensions of $\mathbb{C}(x)$.

A genus 0 example. The rational functions on $x^2 + y^2 = 1$.

This curve is just the complex version of the circle, which is well known to have a parameterization by rational functions:

$$x = \frac{2t}{1+t^2}, \quad y = \frac{1-t^2}{1+t^2}.$$

(This result is equivalent to Euclid's parameterization of Pythagorean triples, given in a lemma following Prop. 28 in Book X of the *Elements.*)

A rational function of x and y is therefore a rational function of t, so our function field in this case is a subfield of $\mathbb{C}(t)$. In fact our field is *exactly* $\mathbb{C}(t)$, because

$$\frac{x}{1+y} = t,$$

so any rational function of t is a rational function of x and y.

Thus, the field of rational functions on $x^2 + y^2 = 1$ is the same as the field of rational functions on $\mathbb{C} \cup \{\infty\}$. If we hope to recover each Riemann surface from the field of rational functions on it, the Riemann surface $x^2 + y^2 = 1$ therefore needs to be the "same" as the sphere $\mathbb{C} \cup \{\infty\}$ in some sense. The equations above show that $x^2 + y^2 = 1$ and $\mathbb{C} \cup \{\infty\}$ are indeed the "same" in the following strong sense: they are *birationally equivalent*; that is, there is a one-to-one correspondence $(x, y) \leftrightarrow t$ that is rational in both directions. □

This notion of equivalence was discovered by Riemann (1851), who showed a much stronger result, already alluded to in Section 4. Assuming the Dirichlet principle, he showed that any Riemann surface of genus zero is birationally equivalent to the sphere $\mathbb{C} \cup \{\infty\}$. A simpler fact about birational equivalence is that *algebraic curves $f(x, y) = 0$ and $g(x, y) = 0$ are birationally equivalent if and only if their function fields are isomorphic*, which is an easy generalization of the example above.

A genus 1 example. The rational functions on $y^2 = 1 - x^4$.

The Riemann surface $y^2 = 1 - x^4$ is *not* birationally equivalent to the sphere, because it has no parameterization $x = f(t)$, $y = g(t)$ by rational functions f and g. This is the result foreseen by Jakob Bernoulli (1704) that we mentioned in Section 2. It is worth saying a little more about it now, because it nicely underlines the analogy between polynomials and integers.

Suppose on the contrary that there are rational functions $x(t)$ and $y(t)$ with

$$y(t)^2 = 1 - x(t)^4.$$

Writing $x(t) = p(t)/r(t)$ and $y(t) = q(t)/r(t)$ as quotients of polynomials with a common denominator we get the equation

$$q(t)^2 r(t)^2 = r(t)^4 - p(t)^4,$$

and hence

$$s(t)^2 = r(t)^4 - p(t)^4,$$

for certain polynomials $p(t)$, $r(t)$, and $s(t)$.

Now, as Jakob Bernoulli noticed, there is a theorem of Fermat that the equation $s^2 = r^4 - p^4$ is impossible for *nonzero integers* p, r, s. This does not rule out the equation for nontrivial *polynomials* p, r, s, but Bernoulli was nevertheless on the right track. One has only to imitate Fermat's argument, using polynomials in place of integers. The two domains are sufficiently similar that the arguments carry over.

For example, the starting point of Fermat's argument is essentially the fact mentioned in the previous example: if X, Y are rational numbers such that

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

then we can write

$$X = \frac{2T}{1+T^2}, \quad Y = \frac{1-T^2}{1+T^2}, \quad \text{for the rational number } T = \frac{X}{1+Y}.$$

In our case we have rational functions $X(t), Y(t)$ such that $X(t)^2 + Y(t)^2 = 1$, and exactly the same calculation shows that

$$X(t) = \frac{2T(t)}{1+T(t)^2},\ Y(t) = \frac{1-T(t)^2}{1+T(t)^2},\ \text{for the rational } \textit{function } T(t) = \frac{X(t)}{1+Y(t)}.$$

The rest of the argument may be similarly rewritten, using divisibility of polynomials in place of divisibility of numbers where appropriate.

Thus the polynomial version of Fermat's argument shows that the Riemann surface $y^2 = 1 - x^4$ is not birationally equivalent to the sphere. It follows, by the remark after the previous example, that the field of rational functions on $y^2 = 1 - x^4$ is not isomorphic to $\mathbb{C}(x)$.[7]

The function field of $y^2 = 1 - x^4$ therefore reflects the difference between this genus 1 curve and the genus 0 curve $\mathbb{C} \cup \{\infty\}$. In fact, it reflects more than the difference in genus, because the genus 1 curves actually fall into infinitely many

[7] The function field of $y^2 = 1 - x^4$ in fact equals $\mathbb{C}(\mathrm{sl}(t), \mathrm{sl}'(t))$, where sl is the lemniscatic sine function defined, as in Section 2, by

$$u = \mathrm{sl}^{-1}(x) = \int_0^x \frac{dt}{\sqrt{1-x^4}}.$$

This is because the curve $y^2 = 1 - x^4$ has the parameterization

$$x = \mathrm{sl}(u), \quad y = \mathrm{sl}'(u).$$

The definition of sl^{-1} of course gives $x = \mathrm{sl}(u)$, and differentiation gives

$$\frac{du}{dx} = \frac{1}{\sqrt{1-x^4}} = \frac{1}{y},$$

whence

$$\mathrm{sl}'(u) = \frac{dx}{du} = y.$$

Thus the rational functions on $y^2 = 1 - x^4$ are generated from the elliptic functions sl and sl'.

birational equivalence classes. This follows from a theorem of Salmon (1851), if one suitably reinterprets Salmon's result. He stated his result in terms of projective equivalence, which happens to be the same as birational equivalence, and also as a result about curves of degree 3 rather than curves of genus 1. The curve $y^2 = 1-x^4$ is not itself of degree 3, of course, but it is birationally equivalent to the cubic curve

$$Y^2 = 4X^3 - 6X^2 + 4X - 1$$

under the substitution

$$X = \frac{1}{1-x}, \quad Y = \frac{y}{(1-x)^2}. \qquad \square$$

The correspondence between birational equivalence classes of curves and isomorphism classes of their function fields is now taken for granted—so much so that certain theorems that were discovered as results about birational equivalence are now stated (without comment) as theorems about function fields. A well-known example is the so-called *Lüroth's theorem.* The original statement of the theorem, in Lüroth (1875), is that any curve parameterized by rational functions can be parameterized *bijectively* by rational functions (with the exception of finitely many points, for example, if the curve has self-intersections). A typical modern statement of the theorem is that any subfield of $\mathbb{C}(z)$ containing more than $\mathbb{C}$ is isomorphic to $\mathbb{C}(z)$.

The equivalence of these two statements is not obvious, though not very hard to prove either. For an elementary proof that the modern form of the theorem implies the original form, see Shafarevich (1994), pp. 9–10.

10. Algebraic Functions and Riemann Surfaces

> These prime ideals correspond to the linear factors in the theory of polynomials. On this basis one attains a completely precise and general definition of a "point of a Riemann surface," i.e., a complete system of numerical values that can be consistently attached to the functions of the field.
>
> Dedekind and Weber (1882), Introduction.

The basic idea of Dedekind and Weber is very natural: a point a on a Riemann surface $\mathcal{S}$ gives a value $f(a)$ to each rational function f on $\mathcal{S}$, and the values given to different functions f and g are *consistent* in the sense that

- The value given to a constant function c is c.
- The value given to $f+g$ is $f(a)+g(a)$.
- The value given to $f-g$ is $f(a)-g(a)$.
- The value given to $f \times g$ is $f(a) \times g(a)$.
- The value given to $f \div g$ is $f(a) \div g(a)$.

When the values include ∞, as they necessarily do for rational functions, one has to make conventions such as $1/\infty = 0$ and $1/0 = \infty$, but this is not a serious problem. The basic question is: does each consistent assignment of values to the rational functions on a surface $\mathcal{S}$ arise from a unique point of $\mathcal{S}$? If so, we can reverse the original idea: namely, start with a function field F, and *define* points of a surface $\mathcal{S}_F$ for which F is the field of rational functions. Just say: a point $P \in \mathcal{S}_F$ *is* an assignment of values (from $\mathbb{C} \cup \{\infty\}$) to the functions in F that satisfies the consistency conditions above.