1. CIRCULAR FUNCTIONS

1. The cotangent as an infinite series. As a prelude to the study of elliptic functions (which are complex functions with two periods) it is useful to investigate the theory of circular (trigonometric) functions which are functions with a single period. If $f(z)$ is a function of period ω then $f(\omega z)$ has period 1 and so by a scale change one can restrict oneself to studying functions of period 1.

Euler was the first mathematician to develop the theory of circular functions systematically and to extend their definition to complex values of the argument. He established the central formulae of the theory and discovered the relation between the circular and the exponential functions given by

$$e^{ix} = \cos x + i \sin x, \qquad e^{i\pi} = -1$$

However it was only much later that a completely self contained theory of the circular functions would be created. Such a theory has several starting points. The usual way to do it is to begin with the power series expansions

$$e^z = 1 + z + \frac{z^2}{2!} + \ldots, \qquad e^{iz} = \cos z + i \sin z \qquad (z \in \mathbf{C})$$

In such a treatment the periodicity would not be manifest and will have to be established by some nontrivial argument.

Euler discovered two remarkable expressions of circular functions, one as an infinite product and the other as an infinite series. For the sine function he established the formula

$$\sin \pi x = \pi x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2}\right) = \pi x \prod_{-\infty}^{\infty} \left(1 + \frac{x}{n}\right) \tag{1}$$

By logarithmic differentiation one obtains

$$\pi \cot \pi x = PV \sum_{-\infty}^{\infty} \frac{1}{x+n} \tag{2}$$

where PV means *principal value,* namely, that the sum has to be interpreted as the limit

$$\lim_{N\to\infty}\sum_{n=N}^{N}\frac{1}{x+n}=\frac{1}{x}+\sum_{n=1}^{\infty}\Big[\frac{1}{x+n}+\frac{1}{x-n}\Big]$$

If we have already available to us the theory of circular functions we can give a direct proof of (2) instead of starting from the infinite product of Euler (Euler in fact had several proofs of both of these formulae). The method is to prove first that the RHS of (2) is meromorphic on $\mathbf{C}$ whose singularities are simple poles at the integers with residues 1, which has period 1, and which is bounded when $|z|\to\infty$ while $\Re(z)$ is bounded. These properties are obvious for the LHS, and so the difference between the two sides of (2) is an entire function which is bounded, hence a constant by Liouville's theorem. But it is immediate that the difference vanishes at $z=1/2$; hence the difference is 0.

It seems to have been **Eisenstein** who realized that one can use (2) as the basis of a theory of circular functions and then establish the theory of elliptic functions along similar but more complicated lines. This work of Eisenstein, long buried in a series of papers he wrote, was resurrected by **Weil** in recent years in his book *Elliptic Functions according to Eisenstein and Kronecker.* It is not our purpose to go in detail into the Weil treatment but just enough to see the architecture of such an independent development of the theory of circular functions based on (2). We shall see that it will be an excellent introduction to the conventional theory of elliptic functions due to **Weierstrass.**

2. The functions E_k. If one wants to build functions that have period 1 there are two very simple ways of doing it. In the first approach one starts with an arbitrary function g and defines the function S_g by

$$S_g(z)=\sum_{n=-\infty}^{\infty}g(z+n)$$

Changing z to $z+1$ can be realized as a change in the summation variable from n to $n+1$ and so the sum remains unaltered. The second method is to define the product P_g by

$$P_g(z)=\prod_{-\infty}^{\infty}g(z+n)$$

which has period 1 for the same reason. Of course the function g has to be chosen so that the series and products converge, or at least summable in some simple fashion. For the choice

$$g(z) = \frac{1}{z^k}$$

the summation method gives the functions E_k introduced by Eisenstein (although they are already in Euler and used by him to derive, among other things explicit expressions for the sums $\sum_{n\geq 1} n^{-2k}$):

$$E_k(z) = \sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^k} \qquad (k \geq 1) \tag{3}$$

For $k \geq 2$ the series converges absolutely, but for $k = 1$ we have to use the principal value. Following Weil we shall refer to taking the principal value as *Eisenstein summation.*

We take up the convergence when $k \geq 2$. Let D be any subset of the set of integers and consider

$$F_D(z) = \sum_{n\in D} (z+n)^{-k}$$

If z is in a compact set K that does not contain any element of D, we have $|z| \leq A$ for some constant A and so, for $|n| \geq 2A$ we have the estimate

$$|z+n|^k \geq (|n| - A)^k \geq (|n|/2)^k \qquad (z \in K)$$

so that the series

$$\sum_{n\in D} (z+n)^{-k}$$

converges normally on $\mathbf{C} \setminus D$. (We recall that a series $\sum_k u_k(z)$ is normally convergent on an open set U if for any compact set $K \subset U$ the series $\sum_k \sup_{z\in K} |u_k(z)|$ is convergent. If the u_k are holomorphic on U, then the sum $\sum_k u_k$ is holomorphic, and its derivatives can be computed by differentiating term by term. This is a standard result from complex analysis; the holomorphy of the sum is proved for instance by Morera's theorem, normal convergence being used to justify termwise integration; the possibility of formal differentiation is proved usually by Cauchy's formula.) If we take $D = \mathbf{Z}$ we see that $E_k(z)$ is holomorphic on $\mathbf{C} \setminus \mathbf{Z}$ and that

$$E_k' = -kE_{k+1} \tag{4}$$

If we take $D = \mathbf{Z} \setminus \{r, -r\}$, then we have

$$E_k(z) = (z+r)^{-k} + (z-r)^{-k} + \sum_{n \neq \pm r} (z+n)^{-k}$$

where the infinite sum on the right side is holomorphic also at $z = \pm r$. Hence we see that E_k has poles of order k at the points $z = \pm r$.

For $k = 1$ it is a question of taking principal value and so we write

$$E_1(z) = \frac{1}{z} + \sum_{n \geq 1} \Big[\frac{1}{z+n} + \frac{1}{z-n} \Big] = \frac{1}{z} + \sum_{n \geq 1} \frac{2z}{z^2 - n^2} \tag{5}$$

The exponent of n in the terms of the series is now 2 and so the earlier argument works again and shows that E_1 is holomorphic on $\mathbf{C} \setminus \mathbf{Z}$ and has simple poles at $z = n$ with residues 1. In particular, the differential equations (4) remain valid for all $k \geq 1$.

The periodicity of the E_k is immediate. For $k \geq 2$ it follows by rearranging the summation, a process that is justified by the absolute convergence. For $k = 1$ it follows from the fact that the terms of the series go to 0; indeed,

$$\sum_{-N}^{N} g(n+1) = \sum_{-N}^{N} g(n) - g(-N) + g(N+1)$$

and so, if $g(n) \to 0$ as $|n| \to \infty$, then

$$PV \sum_{-\infty}^{\infty} g(n+1) = PV \sum_{-\infty}^{\infty} g(n)$$

At this stage there are two courses that can be pursued. The first, which is what is usually done, is to establish the formula

$$E_1(z) = \pi \cot \pi z \tag{6}$$

and proceed to obtain a whole series of formulae. The second, and this is the direction we wish to go, is to use E_1 as the basis for a theory of the circular functions themselves. As was explained earlier, this is not just an academic or historical exercise; we wish to do it because of the insight

it offers into our real goal, namely the construction and study of doubly periodic functions.

3. Relations among the E_k. The differential equation for E_1. By expanding each term $(z+m)^{-k}$ as a power series in z we can obtain the local expansions of the E_k. The results are

$$E_1(z) = \frac{1}{z} - \sum_{r=1}^{\infty} \gamma_{2r} z^{2r-1}$$

$$E_k(z) = \frac{1}{z^k} + (-1)^k \sum_{r=1}^{\infty} \binom{2k-1}{k-1} \gamma_{2r} z^{2r-k}$$

Here

$$\gamma_{2r} = 2 \sum_{n+1}^{\infty} \frac{1}{n^{2r}}$$

and the binomial coefficients $\binom{2k-1}{k-1}$ are 0 when $2r < k$. To obtain non-linear relations among the E_k Eisenstein's method, elaborated in Weil's book, is to start with the algebraic identity

$$\frac{1}{pq} = \frac{1}{pr} + \frac{1}{qr} \qquad (r = p+q)$$

Differentiating with respect to p and q one obtains algebraic identities. For instance, differentiating once with respect to p and q we have

$$\frac{1}{p^2q^2} = \frac{1}{p^2r^2} + \frac{1}{q^2r^2} + \frac{2}{pr^3} + \frac{2}{qr^3}$$

Take now $p = z+m, q = \zeta+n-m$, apply Eisenstein summation with respect to m for fixed n and then sum with respect to n. The result is

$$2E_3(z+\zeta)\Big[E_1(z)+E_1(\zeta)\Big] = E_2(z)E_2(\zeta) - E_2(z)E_2(z+\zeta) - E_2(\zeta)E_2(z+\zeta)$$

Both sides are functions of ζ having a pole of order ≤ 2 at $\zeta = 0$ for fixed $z \notin \mathbf{Z}$, and so one can equate coefficients. Similarly, we can look at the local expansion at $w = 0$ for fixed $z \notin \mathbf{Z}$. In this manner we get relations among the $E_k(z)$. After some work this leads to the differential equation

$$E_1' = -(3\gamma_2 + E_1^2)$$

Write

$$V(z) = \theta^{-1} E_1(\theta^{-1} z) \qquad (\theta = \sqrt{3\gamma_2} \tag{7}$$

Then

$$V' = -(1 + V^2) \tag{8}$$

Here V is holomorphic away from the set $\mathbf{Z}\theta$ where it has simple poles with residue θ, and has period θ.

Proceeding formally one can solve the differential equation as the variables V and z are separated. The equation (8) can be written as

$$dz = -\frac{dV}{1 + V^2}$$

giving the solution

$$z = -\int_0^Y \frac{dv}{1 + v^2}$$

However, as the integrand has singularities at $\pm i$, the function z is not single valued; the value of the integral depends on the path of integration. Nevertheless let us proceed formally. Writing

$$-2i\frac{1}{1 + v^2} = \frac{1}{v + i} - \frac{1}{v - i}$$

we have

$$z = \log(V + i) - \log(V - i) + k_1$$

where k_1 is a constant, so that

$$k^{2iz} = \frac{V + i}{V - i}$$

where k is a constant. This suggests the transformation

$$U = \frac{V + i}{V - i}$$

A simple calculation shows that (8) becomes

$$U' = 2iU$$

from which we conclude that

$$U = ke^{2iz}$$

for some constant k confirming our formal calculation. Now $E_1(1/2) = 0$ gives $V(\theta/2) = 0$ and so $U(\theta/2) = -1$, showing that $k = -e^{i\theta}$. Hence

$$V(z) = -\tan(z + \theta/2)$$

so that

$$E_1(z) = \theta \tan \theta(z + (1/2)) \qquad (e^{2i\theta} = 1)$$

One can show that θ is the *exact period* of e^{2iz}, namely that $e^{2iz} = 1$ if and only if $z = m\theta$ for some integer m. Hence $e^{i\theta} = -1$ and we have

$$E_1(z) = \theta \cot \theta z$$

Of course

$$\theta = \pi$$

but this is a matter of definition. Note that in this approach we automatically get Euler's formula

$$\sum_{n=1}^{\infty} = \frac{\pi^2}{6}$$

A little more work is needed to identify 2θ with the length of the unit circle.

2. DOUBLY PERIODIC FUNCTIONS

1. Lattices. If f is a meromorphic function on $\mathbf{C}$, the set P of its periods is an additive subgroup of $\mathbf{C}$; indeed, from $f(z+a) = f(z)$ and $f(z+b) = f(z)$ it is immediate that $f(z+ma+nb) = f(z)$ for all integers m, n. Somewhat less obvious is the fact that if f is not a constant, then P has to be a *discrete* subgroup of $\mathbf{C}$. Here if V is a real vector space and D is an additive subgroup of V, D is said to be *discrete* if it has no accumulation point in V. If P has an accumulation point c, we can find $a_n \in P, a_n \to c, a_n \neq a_{n+1}$, so that $f(z+b_n) = f(z), b_n = a_n - a_{n+1} \neq 0, b_n \to 0$. For a fixed z_0 this means that the point 0 is a zero of the function $g(w) = f(z_0+w) - f(z_0)$ which is not isolated, contradicting the assumption that f is not a constant.

The main result on discrete subgroups of vector spaces is the following.

Theorem 1. *Let V be a real vector space of dimension d and L a subgroup of V. Then L is discrete if and only if there is a basis $(e_i)_{1\leq i\leq d}$ of V and an integer r $(0 \leq r \leq d)$ such that*

$$L = \left\{ m_1 e_1 + \ldots + m_r e_r \;\middle|\; m_i \in \mathbf{Z} \right\}$$

Proof. We remark that a subgroup L of V is discrete if and only if 0 is an isolated point of V; this has been already observed above in the case of $\mathbf{C}$ and the argument is the same in the general case.

If L has the form described in the theorem, we can use the map

$$x_1 e_1 + \ldots + x_d e_d \longmapsto (x_1, \ldots, x_d)$$

to set up an isomorphism of V with $\mathbf{R}^d$ in such a manner that L goes over to $\mathbf{Z}^r \times (0)$ and the discreteness of L is obvious. The converse requires more effort and we use induction on d to prove it.

Since L is discrete, any compact set in V has only finitely many elements of L in it. We claim first that we can choose an element of smallest norm from $L \setminus (0)$, i.e., there is an element $e_1 \in L, e_1 \neq 0$ such that for any $x \in L \setminus (0)$ we have $||x|| \geq ||e_1||$. To see this, select some euclidean norm $||\cdot||$ for V (using an isomorphism of V with $\mathbf{R}^d$) and some $R > 0$ such that the open ball of radius R contains nonzero elements of L; such elements of L form a finite set and any element of minimum norm from this finite set can be chosen as e_1.

We claim first that $L \cap \mathbf{R}e_1 = \mathbf{Z}e_1$. If $x = re_1 \in L$ for a real number r, select an integer m such that $0 \leq s = r - m < 1$; then $y = se_1 = x - me_1 \in L$ and $||y|| < ||e_1||$, so that $y = 0$, i.e., $s = 0$. Hence $r = m$ as we claimed. This already proves that $L = \mathbf{Z}e_1$ if $d = \dim(V) = 1$.

Let us now assume that $d = \dim(V) > 1$ and that the theorem has been proved when the dimension is $< d$. Let $V_1 = V/\mathbf{R}e_1$ and L_1 the image of L under the natural map $V \longrightarrow V_1$. It is a question of showing that L_1 is discrete in V_1. For, suppose that we have done this. Then, by the induction hypothesis, we can find a basis $(e_2')_{2 \leq i \leq d}$ for V_1 and an integer $r \geq 1$ with $L = \oplus_{2 \leq i \leq r} \mathbf{Z}e_i'$. If e_i is an element of L that gets mapped onto e_i' it is immediate that $(e_i)_{1 \leq i \leq d}$ is a basis of V and $L = \oplus_{1 \leq i \leq r} \mathbf{Z}e_i$.

To prove that L_1 is discrete, let us suppose that this is not the case. The nondiscreteness of L_1 means that 0 is not an isolated point of L_1 and so we can find a sequence $x_n' \in L_1, x_n' \neq 0$ such that $x_n' \to 0$. If $x_n \in L$ maps to x_n', then there are $r_n \in \mathbf{R}$ such that $x_n - r_n e_1 \to 0$ in V. Writing $r_n = r_n' + k_n$ where $k_n \in \mathbf{Z}$ and $0 \leq r_n' < 1$ and $y_n = x_n - k_n e_1$ we have $y_n \in L$, $y_n - r_n' e_1 \to 0$ in V. By passing to a subsequence we may assume that $r_n' \to r$. Then $y_n \to re_1$ and so $y_n = re_1$ for all large n, a contradiction, since y_n lies above $x_n' \neq 0$.

We say that L is a *lattice* if $r = d$. This is the same as requiring that there is an isomorphism of V with $\mathbf{R}^d$ that carries L onto $\mathbf{Z}^d$. It is also the same as requiring that V/L is compact. Notice that

$$\mathbf{R}^d/\mathbf{Z}^d \simeq T^d, \qquad T = \text{ the unit circle}$$

so that $\mathbf{R}^d/\mathbf{Z}^d$ is a *torus* of dimension d. Notice that *this isomorphism is not only topological but also group theoretic.*

We apply these considerations to $V = \mathbf{C}$ regarded as a real vector space of dimension 2. The only discrete subgroups of $\mathbf{C}$ are, by the above theorem, as follows:

(i) the trivial subgroup (0)

(ii) The subgroup generated by an element $a \in \mathbf{C} \setminus (0)$

(iii) the lattice

$$L = \left\{ m_1\omega_1 + m_2\omega_2 \;\middle|\; m_i \in \mathbf{Z} \right\}$$

where ω_1, ω_2 are two nonzero elements such that $\omega_1/\omega_2 \notin \mathbf{R}$.

The functions with periods in a subgroup of type (ii) above are the singly periodic functions and they are the circular functions up to a scale change. The functions with periods from a lattice are the doubly periodic functions.

For any lattice L in $\mathbf{C}$ let $\mathcal{M}(L)$ be the set of all meromorphic functions f on $\mathbf{C}$ which are periodic with respect to L, i.e., $f(z+\omega) = f(z)$ for all $\omega \in L$. It is obvious that $\mathcal{M}(L)$ is a *field* closed under d/dz, a *differential field.* The elements of $\mathcal{M}(L)$ for various lattices L are called *elliptic functions.*

The essential difference between the circular and elliptic functions is that one cannot go from one lattice to another by a scale change. To see this more clearly, let us say that two lattices L, L' are *equivalent*, $L \sim L'$, if $L' = \lambda L$ for some $\lambda \in \mathbf{C}$; in this case

$$f(z) \in \mathcal{M}(L') \Longleftrightarrow f(\lambda z) \in \mathcal{M}(L)$$

To see that this is a highly nontrivial relation, we proceed as follows. By a $\mathbf{Z}$–basis of a lattice L we mean a pair of elements ω_1, ω_2 of L such that $\omega_1/\omega_2 \notin \mathbf{R}$ and $L = \mathbf{Z}\omega_1 \oplus \mathbf{Z}\omega_2$. Multiplying L by ω_2^{-1} we see that $L \sim L_\tau$ where $L_\tau = \mathbf{Z}\tau \oplus \mathbf{Z}1$. Changing τ to $-\tau$ we may assume that $\Im(\tau) > 0$. Let $\mathcal{H}$ be the Poincaré upper half plane, i.e.,

$$\mathcal{H} = \left\{ \tau \in \mathbf{C} \;\middle|\; \Im(\tau) > 0 \right\}$$

Thus any lattice is equivalent to a lattice L_τ for some $\tau \in \mathcal{H}$. For the equivalences among the L_τ we have the following result.

Theorem 2. *For $\tau, \tau' \in \mathcal{H}$, we have*

$$L_\tau \sim L_{\tau'} \Longleftrightarrow \tau' = \frac{a\tau + b}{c\tau + d}$$

for some matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *where* a, b, c, d *are integers and* $ad - bc = 1$.

Proof. If L is a lattice with $\mathbf{Z}$–bases $(\omega_i), (\omega_i')$, we have, for suitable 2×2 matrices A, A' with integer entries,

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = A' \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}, \qquad \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = A \begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix}$$

Hence

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = A'A \begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix}, \qquad \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = AA' \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$$

It follows from the linear independence of the ω_i and ω_i' over $\mathbf{R}$ that $A'A = AA' = I$. This means that $\det(A) = \pm 1$ and $A' = A^{-1}$. We now take $L = L_\tau$ with $\mathbf{Z}$–basis $(\tau, 1)$. If $\lambda L_{\tau'} = L_\tau$, then the above remark leads to

$$\begin{pmatrix} \lambda\tau' \\ \lambda \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau \\ 1 \end{pmatrix}$$

where a, b, c, d are integers with $ad - bc = \pm 1$. But then

$$\tau' = \frac{a\tau + b}{c\tau + d}$$

and as

$$\Im(\tau') = |c\tau + d|^{-2}(ad - bc)\Im(\tau)$$

we must have $ad - bc = 1$. Conversely, if τ' and τ are related as described, we can find a $\lambda \neq 0$ such that

$$\begin{pmatrix} \lambda\tau' \\ \lambda \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} \lambda\tau' \\ \lambda \end{pmatrix}$$

Hence $\lambda L_{\tau'} = L_\tau$.

For any commutative ring R with unit let us write SL(2, R) for the group of 2×2 matrices with entries from R and determinant 1. Then the group SL(2, $\mathbf{R}$) acts on $\mathcal{H}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tau \longmapsto \frac{a\tau + b}{c\tau + d}$$

This is the group of *real fractional linear transformations,* also known as the *real Möbius group.* If we allow the numbers a, b, c, d to be complex, $\mathcal{H}$ is no longer invariant and we have to enlarge it to the extended complex plane $\mathbf{C} \cup \infty$. If we restrict the a, b, c, d to be integers, we get the *modular group.* Thus, $L_\tau \sim L_{\tau'}$ if and only if τ and τ' are in the *same orbit* under the modular group. Since the modular group is countable, it is clear that there are continuum many points in $\mathcal{H}$ which are mutually inequivalent under the modular group.

The integers act on the real line by translation and if we take the interval $I = [0, 1]$, then any real number can be translated by an integer to belong to I, and the only two elements in I that are equivalent under the action of $\mathbf{Z}$ are 0 and 1. One can ask if we can construct a subset of $\mathcal{H}$ that has a similar property with respect to the action of the modular group. The following theorem answers this question. We shall prove it later.

Theorem 3. *Let*

$$D = \left\{ \tau \in \mathcal{H} \;\middle|\; |\Re(\tau)| \leq \frac{1}{2}, |\tau| \geq 1 \right\}$$

Then D contains a representative from each orbit of the modular group acting on $\mathcal{H}$; and the only pairs (τ, τ') of points of D which lie in the same orbit under the modular group are

$$\Re(\tau) = -\frac{1}{2}, \tau' = \tau + 1, \qquad |\tau| = 1, \tau' = \frac{-1}{\tau}$$

In particular, two distinct points of the interior of D are never in the same orbit.

2. Theorems of Liouville. Let L be a lattice in $\mathbf{C}$ and let $\mathcal{M}(L)$ be the field of meromorphic functions on $\mathbf{C}$ that are periodic with respect to L. Then one can prove some general theorems about the zeros and poles of elements of $\mathcal{M}(L)$. For any nonconstant $f \in \mathcal{M}(L)$ let $Z(f)$ be the set of its zeros and $P(f)$ the set of its poles. Both $Z(f)$ and $P(f)$ are discrete sets and are invariant under translations by elements of L. For any $a \in \mathbf{C}$ let $m_a(f)$ be the order of f and Res $_a(f)$ the residue of f at a. A sum written as $\sum_{a \in \mathbf{C}/L}$ means that summation is over a complete set of elements that are mutually incongruent mod L.

Theorem 4. *Let f be a nonconstant meromorphic function on $\mathbf{C}$ with period lattice L. Then $Z(f)$ and $P(f)$ are finite sets mod L. Moreover we have the following.*

(i) f *has at least one pole*

(ii) $\sum_{a\in\mathbf{C}/L}$ *Res* $_a(f) = 0$

(iii) $Z(f)$ *and* $P(f)$ *have the same number of elements mod* L*, counting multiplicities. More precisely,*

$$\sum_{a\in\mathbf{C}/L} m_a(f) = 0$$

(iv) $\sum_{a\in\mathbf{C}/L} m_a{\cdot}a \equiv 0(\ mod\ L)$

Proof. First of all it is clear that any element of $\mathbf{C}$ is congruent mod L to an element of the parallelogram

$$\mathbf{P}_0 = \left\{ a_1\omega_1 + a_2\omega_2 \;\middle|\;) \le a_1, a_2 \le 1 \right\}$$

$\mathbf{P}_0$, and more generally, any translate $\mathbf{P}$ of $\mathbf{P}_0$ is called a *fundamental parallelogram (with respect to L).* Any $f \in \mathcal{M}(L)$ is determined completely once it is known on any fundamental parallelogram. Given $f \in \mathcal{M}(L)$ one can choose a fundamental parallelogram $\mathbf{P}$ such that the boundary $\partial\mathbf{P}$ of $\mathbf{P}$ does not contain any zero or pole of f; in this case, the zeros or poles of f within $\mathbf{P}$ is a complete set of representatives mod L for $Z(f)$ and $P(f)$. In particular, as there can be only finitely many zeros and poles within $\mathbf{P}$, it is clear that $Z(f)$ and $P(f)$ are finite sets mod L, and the summations in the assertions of the theorem can be taken to be sums over the finite sets of zeros and poles within $\mathbf{P}$. In what follows $\mathbf{P}$ is a fundamental parallelogram whose boundary does not contain any zero or pole of f.

Proof of (i): If f is holomorphic, it is bounded on $\mathbf{P}$, hence bounded on $\mathbf{C}$, hence a constant by Liouville's theorem.

Proof of (ii): Let $\mathbf{P}$ have vertices A, B, C, D and let the boundary be described in the order $ABCD$. Then

$$\oint_{ABCD} f(z)dz = 2\pi i(\text{sum of residues of } f \text{ inside } \mathbf{P})$$

On the other hand, the integrals over pairs of parallel sides of $\mathbf{P}$ cancel by periodicity of f, so that

$$\oint_{ABCD} f(z)dz = 0$$

This proves that the sum of the residues inside $\mathbf{P}$ must be 0.

Proof of (iii): Let $g = f'/f$; clearly $g \in \mathcal{M}(L)$. It is easy to see from the local expression for f at a zero or pole that the poles of g are precisely the points in $R = Z(f) \cup P(f)$, and the residue at a point $a \in R$ is $m_a(f)$ (write $f(z) = (z-a)^{m_a}h(z)$ where h is holomorphic at a and $h(a) \neq 0$; then $g(z) = (m_a/(z-a)) + k(z)$ where k is holomorphic at a). By (ii) applied to g we get (iii) immediately.

Proof of (iv): We take now $g(z) = zf'(z)/f(z)$. g is meromorphic but is not in $\mathcal{M}(L)$ because of the factor z. Nevertheless we can proceed as before but with greater care in evaluating the integral of g over $\partial\mathbf{P}$. We need the result which we formulate as a lemma below. Let $\mathbf{P} - \mathbf{P}_0 + a$ so that the vertices of $\mathbf{P}$ are $a, a+\omega_2, a+\omega_1+\omega_2, a+\omega_1$. We have

$$\oint_b^{b+\omega_1} z\frac{f'(z)}{f(z)}dz = \oint_{a+\omega_2}^{a+\omega_1+\omega_2} (u-\omega_2)\frac{f'(u)}{f(u)}du$$

so that

$$\oint_{BC} z\frac{f'(z)}{f(z)}dz - \oint_{AD} z\frac{f'(z)}{f(z)}dz = \omega_2 \oint_{BC} \frac{f'(u)}{f(u)}du$$

Similarly

$$\oint_{AB} z\frac{f'(z)}{f(z)}dz - \oint_{DC} z\frac{f'(z)}{f(z)}dz = -\omega_1 \oint_{DC} \frac{f'(u)}{f(u)}du$$

Now the residue of g at a point a is $m_a \cdot a$. Hence

$$2\pi i \sum_{a \in \mathbf{C}/L} m_a \cdot a = \oint_{\partial\mathbf{P}} g(z)dz = -\omega_1 \oint_{DC} \frac{f'(u)}{f(u)}du + \omega_2 \oint_{BC} \frac{f'(u)}{f(u)}du$$

So to prove (iv) it is enough to show that

$$\frac{1}{2\pi i}\oint_{BC} \frac{f'(z)}{f(z)}dz, \qquad \frac{1}{2\pi i}\oint_{DC} \frac{f'(z)}{f(z)}dz$$

are both integers. Now, as f is nowhere 0 on the line segments BC and DC, we can find small rectangles containing these line segments on which f has no zero. The required conclusion then follows from the following lemma.

Lemma. *Let U be a simply connected domain in $\mathbf{C}$ and $p : [a, b] \longrightarrow U$ a continuous map. Suppose that f has no zero in U and $f(p(a)) = f(p(b))$. Then*

$$\frac{1}{2\pi i} \oint_p \frac{f'}{f} dz \in \mathbf{Z}$$

Proof. Since U is simply connected, we can find a holomorphic u on U such that $u' = f'/f$. Locally on U we can define $\log f$ and hence for such a choice we have $(\log f)' = f'/f$. Thus $u = \log f +$ constant locally, and so $e^u = \text{constant} f$ locally, hence globally on U. But then, as

$$\oint_p \frac{f'}{f} dz = u(p(b)) - u(p(a))$$

and

$$e^{u(p(b))-u(p(a))} = \frac{f(p(a))}{f(p(b))} = 1$$

we see that

$$\frac{1}{2\pi i} \oint \frac{f'}{f} dz$$

must be an integer.

3. THE WEIERSTRASS FUNCTION

1. The function $\wp(z)$. We begin with

Lemma 1. *Let L be a lattice in a real euclidean vector space of dimension d. Then the series*

$$\sum_{0\neq\ell\in L} \frac{1}{||\ell||^k} < \infty$$

if and only if $k > d$.

Proof. Let $\alpha_1, \ldots, \alpha_d$ form a $\mathbf{Z}$–basis for L. Let

$$Q(x) = ||x_1e_1 + \ldots + x_de_d||^2 \qquad (x = (x_1, \ldots, x_d) \in \mathbf{R}^d)$$

We wish to determine when the seies

$$\sum_{x\in\mathbf{Z}^d\setminus\{0\}} \frac{1}{Q(x)^{k/2}} < \infty$$

Now Q is a positive definite quadratic form in the x_j and so its matrix is symmetric and has only positive eigenvalues. Since this matrix can be diagonalised by an orthogonal matrix, it is clear that if m and M are its least and greatest eigenvalues,

$$m(\sum_i x_i^2) \leq Q(x) \leq M(\sum_i x_i^2) \qquad (x = (x_1, \ldots, x_d) \in \mathbf{R}^d)$$

Hence it is a question of determining when

$$\sum_{x\in\mathbf{Z}^d\setminus\{0\}} (x_1^2 + \ldots + x_d^2)^{-k/2} < \infty$$

But for $x \in \mathbf{Z}^d \setminus \{0\}$,

$$\frac{1}{d}\Big(\sum_i |x_i|\Big)^2 \leq \sum_i x_i^2 \leq d\Big(\sum_i |x_i|\Big)^2$$

and hence it comes to determining when

$$\sum_{x \in \mathbf{Z}^d \setminus \{0\}} (|x_1| + \ldots + |x_d|)^{-k} < \infty$$

We may clearly restrict the sum to integers $x_i \geq 0$. The number of $x \in \mathbf{Z}^d$ with $x_i \geq 0$ and $\sum_i |x_i| = r$ is $\sim Cr^{d-1}$ as $r \to \infty$ for some constant $C > 0$, and so it comes down to when

$$\sum_{r \geq 1} \frac{r^{d-1}}{r^k} < \infty$$

and this is clearly the case if and only if $k > d$.

We are now interested in constructing meromorphic functions over $\mathbf{C}$ with period lattice L. Let ω_1, ω_2 be a $\mathbf{Z}$–basis of L. Following the method discussed in the case of circular functions (it is precisely to motivate this construction that we treated the circular case earlier) let us define

$$W_k(z) = \sum_{\omega \in L} \frac{1}{(z+\omega)^k} \qquad (k \geq 3) \tag{1}$$

For any $A > 0$ and $|z| \leq A$, we have, for $\omega \in L$ with $|\omega| \geq 2A$,

$$|z+\omega|^k \geq (|\omega| - |z|)^k \geq (|\omega|/2)^k$$

and so the series

$$\sum_{\omega \in L, |\omega| > 2A} \frac{1}{(z+\omega)^k}$$

converges normally on $\mathbf{C} \setminus \{\omega \in L | |\omega| \geq 2A\}$ when $k \geq 3$ by Lemma 1. Hence, for any subset D of L the function

$$\sum_{\omega \in D} \frac{1}{(z+\omega)^k}$$

is holomorphic away from D and its derivatives can be calculated by termwise differentiation. This shows first of all, taking $D = L$, that W_k is holomorphic on $\mathbf{C} \setminus L$; and secondly, taking $D = L \setminus \{\omega_0\}$ where $\omega_0 \in L$, that

$$W_k(z) = \frac{1}{(z+\omega_0)^k} + H(z)$$

where H is holomorphic in a small neighborhood of ω_0. Hence W_k is meromorphic on $\mathbf{C}$, with singularities only at the elements of L where it has poles of order $-k$, the principal part at ω_0 being

$$\frac{1}{(z+\omega_0)^k}$$

Moreover we have

$$\frac{dW_k}{dz} = -kW_{k+1} \qquad (k \geq 3) \tag{2}$$

We cannot take $k = 2$ in the above argument as the series will not converge normally since the series

$$\sum_{\omega \in L \setminus \{0\}} \frac{1}{\omega^2}$$

is not convergent. However if we subtract this infinite sum, we will obtain a renormalized sum that will define an elliptic function. This is the famous Weierstrass $\wp$–function.

Theorem 2. *The function*

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in L \setminus \{0\}} \Big(\frac{1}{(z+\omega)2} - \frac{1}{\omega^2} \Big)$$

is well defined, even, and meromorphic on $\mathbf{C}$ *with period lattice* L*. Its degree is* 2*, and it has double poles at* $\omega \in L$ *with principal parts* $(z+\omega)^{-2}$ *and no other singularities.*

Proof. We have

$$\Big(\frac{1}{(z+\omega)2} - \frac{1}{\omega^2} \Big) = \frac{2\omega z - z^2}{\omega^2 (z+\omega)^2}$$

For $|z| \leq A$ and $|\omega| \geq 2A$ where $A > 1$ is fixed, we have

$$|(z+\omega)^2| \geq (|\omega| - |z|)^2 \geq (|\omega|/2)^2$$

so that

$$\Big| \frac{2\omega z - z^2}{\omega^2 (z+\omega)^2} \Big| \geq B|\omega|^{-3}$$

for some constant $B > 0$ independent of ω. This proves the normal convergence. This defines $\wp$ as a meromorphic function with singulatiries at the points of L and nowhere else, the points of L being double poles with principal parts $(z+\omega)^{-2}$. Sice the sum over $L \setminus \{0\}$ is unchanged when we change z to $-z$ and ω to $-\omega$, it follows that $\wp$ is an even function.

The periodicity needs a little more care. The normal convergence allows us as usual to differentiate termwise and so we get

$$\wp' = -2W_3$$

So $\wp'$ has period lattice L. So, if $\mu \in L$, the function $\wp(z+\mu) - \wp(z)$ has vanishing derivative and so must be a constant, say $c(\mu)$:

$$\wp(z+\mu) - \wp(z) = c(\mu)$$

If $\mu/2 \notin L$, we can take $z = -\mu/2$ in the equation above and get, remembering that $\wp$ is even, we get

$$c(\mu) = \wp(\mu/2) - \wp(-\mu/2) = 0$$

Taking $\mu = \omega_1, \omega_2$ we see that ω_1 and ω_2 are periods of $\wp$, and so $\wp$ has period lattice L. Finally, the degree of $\wp$ is 2 since it has only one pole mod L, namely 0, and that has order -2.

2. The properties of the $\wp$–function. We shall now discuss some basic properties of the function $\wp$.

Theorem 3. *We have the following.*

- *(i) $\wp'$ has exactly three zeros mod L, namely $z = \omega_1/2, \omega_2/2, \omega_3/2$ where $\omega_3 = \omega_1 + \omega_2$. These are all simple.*
- *(ii) For any $a \in \mathbf{C} \setminus (1/2)L$, $\wp(z) - \wp(a)$ has exactly two zeros mod L, namely at $z = a$ and $z = -a$, and these are simple.*
- *(iii) If $a \in (1/2)L \setminus L$, i.e., if $a \equiv \omega_i (i = 1, 2, 3)$, then $\wp(z) - \wp(a)$ has exactly one zero mod L, namely $z = a$, and it is a double zero.*

Proof. Since $\omega_i/2 \equiv -\omega_i/2 \bmod L$, we have, as $\wp'$ is an odd function, $-\wp'(\omega_i/2) = \wp'(-\omega_i/2) = \wp'(\omega_i/2)$ and so $\wp'(\omega_i/2) = 0$. Since $\wp'$ has a pole of order 3 at $z = 0$ and no ther singulatities mod L, the degree of

$\wp'$ is 3. As we have found 3 distinct zeros, it must be that these are all simple and there are no other zeros. This proves (i).

Let now $a \notin (1/2)L$, i.e., $a \not\equiv -a \bmod L$. For such a, $\wp(z) - \wp(a)$ has a zero at $z = a$, hence by evenness a zero at $z = -a$ which is distinct from $z = a \bmod L$. Since the degree of $\wp(z) - \wp(a)$ is 2, $z = \pm a$ are both simple zeros and there are no other zeros. This proves (ii).

If $a \equiv -a \bmod L$, i.e., $a \equiv \omega/2$, then $\wp(z) - \wp(a)$ must have a double zero at $z = a$ since $\wp'(a) = 0$. Again there is no other zero because the degree of $\wp(z) - \wp(a)$ is 2.

Theorem 4. *We have the following.*

(i) The field of even elements of $\mathcal{M}(L)$ is generated by $\wp$, i.e., if f is in $\mathcal{M}(L)$ and is even, there is a rational function k of a single variable such that $f = k(\wp)$.

(ii) The field $\mathbf{C}(\wp, \wp')$ of all rational functions of $\wp$ and $\wp'$ is precisely $\mathcal{M}(L)$.

Proof. Let f be even. Let f be an even element of $\mathcal{M}(L)$ and non constant. We first observe that if $a \in \mathbf{C}$ with $a \notin (1/2)L$ and $z = a$ is a zero (resp. pole), then $z = -a$ is also a zero (resp. pole) and of the same order. For, if m is the order of $z = a$, then $f(z) = (z - a)^m g(z)$ for z near $z = a$ where g is holomorphic at $z = a$ and $g(a) \neq 0$. Then $f(z) = (-1)^m (z + a)^m h(z)$ where $h(z) = g(-z)$ is holomorphic at $z = -a$ and $h(-a) = g(a) \neq 0$. Next we observe that if $a \in (1/2)L \setminus L$ and $z = a$ is a zero (resp. pole) of f, its order is even. For, we again write $f(z) = (z - a)^m g(z)$ where g is holomorphic at $z = a$ and $g(a) \neq 0$. If $F(z) = f(z + a)$, then $F(-z) = f(-z + a) = f(z - a) = f(z + a) = F(z)$ so that F is even. Clearly the order of F at $z = 0$ is even, and this order is m. If we write $k(a)$ for $m(a)$ or $(1/2)m(a)$ according as $a \not\equiv -a$ or $a \equiv -a$, then $f(z)$ and $(\wp(z) - \wp(a))^{k(a)}$ has the same order of zero at $z = \pm a$. A similar result is valid when we deal with a pole at $z = a$.

This said, let $a_1, \ldots, a_r$ be the zeros of f with the following properties: no a_j is in L, $a_j \not\equiv \pm a_k$ if $j \neq k$, and if $z = a$ is a zero of f not in L, there is some j such that $a \equiv \pm a_j$. It is then immediate that

$$h(z) = \frac{\prod_{j=1}^{j=r} (\wp(z) - \wp(a_j))^{k(a_j)}}{\prod_{k=1}^{k=s} (\wp(z) - \wp(b_k))^{k(b_k)}}$$

has the same zeros and poles wiht the same orders as f on $\mathbf{C} \setminus L$, and hence without this restriction also, in view of Liouville's theorems. So f is a constant multiple of h. This proves (i).

To prove (ii) let f be odd. Then $f\wp'$ is even and so lies in $\mathbf{C}(\wp)$. But then $f \in \mathbf{C}(\wp, \wp')$.

We shall now show that $\wp$ and $\wp'$ are not algebraically independent, i.e., there is a polynomial relation between $\wp$ and $\wp'$. To describe this relation, let us define

$$G_k = G_k(L) = \sum_{\omega \in L \setminus \{0\}} \frac{1}{\omega^{2k}} \qquad (k \geq 2)$$

Note that the corresponding sums of odd powers are 0. The G_k are analogous to the lattice sums

$$\sum_{n \in \mathbf{Z} \setminus \{0\}} \frac{1}{n^{2k}}$$

whose computation dates back to Euler. The G_k however are not constants but depend on the lattice. We have

$$G_k(\lambda L) = \lambda^{-2k} G_k(L) \qquad (\lambda \in \mathbf{C} \setminus \{0\})$$

Theorem 5. *Let*

$$g_2 = 60G_2, \qquad g_3 = 140G_3$$

Then we have

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3$$

Proof. This will be done by looking at the local expansion of $\wp$ and $\wp'$ at $z = 0$. We have

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in L \setminus \{0\}} \Big(\frac{1}{(\omega - z)^2} - \frac{1}{\omega^2} \Big)$$

Since $(1-t)^{-2} = 1 + 2t + 3t^2 + \ldots + (r+1)t^r + \ldots$ for $|t| < 1$, we have,

$$\frac{1}{(\omega - z)^2} = \frac{1}{\omega^2} \frac{1}{(1 - \frac{z}{\omega})^2} = \sum_{r \geq 0} \frac{(r+1)z^r}{\omega^{r+2}} \qquad (|z| < \alpha)$$

where α is the minimum value of $\omega|$ as ω varies over $L \setminus \{0\}$. Substituting this in the expression for $\wp$ and rearranging the sum we get the local expansion for $\wp$. This can be justified of course by checking that the double series converges in the indicated region for z if everything is replaced by its absolute value (see problems below). So we have

$$\wp(z) = \frac{1}{z^2} + \sum_{k \geq 1} (2k+1) G_{k+1} z^{2k}$$

We see that the G_k are the coefficients of the local expansion of $\wp$.

We now differentiate and multiply this local expansion to get the following:

$$\begin{aligned}
\wp &= \frac{1}{z^2} + 3G_2 z^2 + 5G_3 z^4 + \dots \\
\wp' &= \frac{-2}{z^3} + 6G_2 z + 20G_3 z^3 + \dots \\
\wp'^2 &= \frac{4}{z^6} - \frac{24G_2}{z^2} - 80G34 + \dots \\
\wp^3 &= \frac{1}{z^6} + \frac{9G_2}{z^2} + 15G_4 + \dots
\end{aligned}$$

A simple calculation then shows that

$$\wp'^2 - 4\wp^3 + 60G_2\wp + 140G_3$$

has no pole at $z = 0$ and no constant term, and so, being in $\mathcal{M}(L)$, must be a constant, and thus 0. This finishes the proof.

4. COMPLEX TORI

1. Complex manifolds of dimension 1. In elementary complex function theory one knows that it is conceptually worthwhile to introduce the extended complex plane $\mathbf{C} \cup (\infty)$, which is topologically the sphere S^2 of real dimension 2. The rational functions are then precisely the meromorphic functions on $\mathbf{C} \cup (\infty)$. The starting point of Riemann's more profound theory of complex functions is the recognition that one has to consider other spaces than the sphere on which one can make sense of analytic functions. From this point of view the elliptic functions appear as meromorphic functions on a torus in the same way as the rational functions are meromorphic functions on $\mathbf{C} \cup (\infty)$. The basic notion here is that of a *complex manifold of dimension* 1, or a *Riemann surface.* This was the central concept in Riemann's epoch–making theory; however its true scope was not completely understood immediately because in Riemann's work the Riemann surfaces seemed to be only a device to study multi–valued functions. It was Klein who recognized that Riemann surfaces have to be treated as objects of independent study, and that they are the true domains on which complex function theory should be studied. Klein did not complete his program, and in particular did not completely clarify the notion of a Riemann surface. This was left to Weyl who clearly understood what the conceprual basis of the theory of Riemann surfaces should be and codified it in his famous book *Die Idee der Riemannschen Fläsche*. It was in Weyl's book that the notion of what is meant by a complex manifold of dimension 1 first found its clearest formulation.

We shall give a brief treatment of the notion of a complex manifold. For our purposes we need only the cases when the dmension of the manifold is either 1 or 2. We start with the case of dimension 1. To say that X is a complex manifold of dimension 1 we should have the following data.

(i) X should be a Hausdorff topological space; we assume that X is connected and satisfies the second axiom of countability.

(ii) X should be covered by open sets X_i and for each i there should be a homomorphism z_i of X_i with an open set D_i of the complex plane $\mathbf{C}$; z_i is called the *local coordinate* on X_i.

(iii) If the indices i, j are such that $X_{ij} = X_i \cap X_j$ is nonempty, and if D_{ij} and D_{ji} are the images under z_i and z_j respectively of X_{ij}, then the function

$$f_{ij} : D_{ij} \longrightarrow D_{ji}$$

defined by

$$z_j = f_{ij} \circ z_i^{-1}$$

is holomorphic. By interchanging i and j we see that f_{ij} is invertible (as a map) and f_{ji} is the inverse function to f_{ij}.

With this data it becomes possible to speak, for any open set $Y \subset X$, of holomorphic functions on Y: a function $f(Y \longrightarrow \mathbf{C})$ is holomorphic if for any index i, the restriction of f to $Y \cap X_i$ is of the form $f_i \circ z_i$ where f_i is a holomorphic function on $z_i(Y \cap X_i)$.

The above notion of holomorphic function has all the properties that are familiar to us from elementary function theory. The essential new ingredient is that unlike the case of $\mathbf{C}$, there is no distinguished analytic function that can serve as a local coordinate everywhere. Indeed, this is not true even for the extended complex plane $\mathbf{C} \cup (\infty)$ as we shall see below. However even for $\mathbf{C}$, at a given point p of a complex manifold X, there are many possible *local coordinates.* Let z be a local coordinate at p with $z(p) = 0$; then, a function ζ, defined and holomorphic around p, is of the form

$$\zeta(q) = Z(z(q)) \qquad (q \in U)$$

where U is an open neighborhood of p and Z is an analytic function on the complex plane, defined in a neighborhood of the origin. ζa local coordinate if and only if

$$\Big(\frac{\partial Z}{\partial t}\Big)_{t=0} \neq 0$$

The necessity of the condition is clear because from $t = t(Z), Z = Z(t)$ we get $(\partial t/\partial Z)(\partial Z/\partial t) = 1$ which implies that $(\partial Z/\partial t)_{t=0} \neq 0$. The sufficiency for example can be established by using power series. If $c = (\partial Z/\partial t)_{t=0} \neq 0$, we can replace Z by $c^{-1}Z$ and assume that

$$Z = t + a_2 t^2 + \ldots + a_n t^n + \ldots$$

where the power seires converges on some disk $|t| < \rho$. To invert this equation let us write

$$t = Z + b_2 Z^2 + \ldots + b_n Z^n + \ldots$$

and it is a question of showing that this defines the b_n uniquely and the series above converges in a disk $|Z| < \delta$. Substituting for t in this power series we get a formal powewr series in Z all of whose powers higher than 1 must have vanishing coefficients. For the coefficient of Z^2 we have $b_2 + a_2$ and so

$$b_2 + a_2 = 0$$

which determines b_2 as $-a_2$. Equating the coefficient of Z^n to 0 we get an equation

$$b_n + P_{n,2}a_2 + \ldots + P_{n,n-1}a_{n-1} + a_n = 0$$

where the $P_{n,j}$ are *universal polynomials in* $b_2, \ldots, b_{n-1}$ *with nonnegative integer coefficients.* This shows that the b_n for $n \geq 2$ are uniquely determined by the a's; moreover it can be shown by the so called *method of majorants* that the power series $z + b_2 z^2 + \ldots + b_n z^n + \ldots$ has a positive radius of convergence (see problem #1).

Let us now discuss a few examples.

C: Here $X = \mathbf{C}$ and z is a coordinate at all points. The same applies to any open subset of **C**.

$X = \mathbf{C} \cup (\infty)$: The neighborhoods of ∞ are the sets $\{|z| > A\} \cup (\infty)$. X is covered by $X_0 = \mathbf{C}$ and $X_\infty = X \setminus (0)$. On X_0 the coordinate is z; on X_∞ the coordinate is t which is defined as 0 at ∞ and z^{-1} everywhere else. On the overlap $X_0 \cap X_\infty = \mathbf{C} \setminus (0)$, the relation between z and t is given by

$$t = \frac{1}{z}, \qquad z = \frac{1}{t}$$

The complex manifold thus defined is *compact.* For, if (U_i) is a covering of X by open sets, ∞ must be in one of them, say U_{i_0}, so that for some $A > 0$ we must have $\{|z| > A\} \subset U_{i_0}$; then we can find $i_1, \ldots, i_k$ such that $\{|z| \leq A\} \subset U_{i_1} \cup \ldots \cup U_{i_k}$, so that X is covered by U_{i_r}, $(0 \leq r \leq k)$. As one knows from elementary fumction theory, the rational functions are precisely the meromorphic functions on X.

The complex structure on X can also be defined using stereographic projection. We start with $X = S^2$, the unit sphere in $\mathbf{R}^3$. Let N be the north pole, the point $(0, 0, 1)$. Then we can project $X \setminus (N)$ onto **C** bijectively by sending the point P to the unique point on the x_1x_2–plane

in which the line joining P and N meets this plane. If $(x_1(P), x_2(P))$ are the coordinates of this point, we put

$$z(P) = x_1(P) + ix_2(P)$$

We use z as a local coordinate on $X \setminus (N)$. We replace N by S, the south pole with coordinates $(0, 0, -1)$, and define the local coordinate t on $X \setminus (S)$ by

$$t(P) = t_1(P) - it_2(P)$$

where $(t_1(P), t_2(P))$ are the coordinates of the unique point of the x_1x_2–plane in which the line joining S and P meets it (note the change in sign). It is an easy verification that if P is different from both N and S, we have

$$t(P_= \frac{1}{z(P)}, \qquad z(P) = \frac{1}{t(P)}$$

$X = \mathbf{CP}_1$: This is the *complex projective line.* It is defined as the set of lines through the origin in $\mathbf{C}^2$. Any point p of $\mathbf{C}^2 \setminus (0, 0)$ lies on a unique line and so we have a surjective map

$$\pi : \mathbf{C}^2 \setminus (0, 0) \longrightarrow \mathbf{CP}_1, \ \pi(z_1, z_2) := [(z_1, z_2)], \text{ the line containing } (z_1, z_2)$$

We give $\mathbf{CP}_1$ the quotient topology; a subset U of $\mathbf{CP}_1$ is open if and only if $\pi^{-1}(U)$ is open in $\mathbf{C}^2 \setminus (0, 0)$. It is easy to verify that $\mathbf{CP}_1$ is Hausdorff, connected, and satisfies the second axionm of countability. Since every line contains a point (z_1, z_2) with $|z_1|^2 + |z_2|^2 = 1$, we see that $\mathbf{CP}_1$ is the image under π of the unit sphere in $\mathbf{C}^2$ and so $\mathbf{CP}_1$ is compact. If the line in $\mathbf{C}^2$ does not lie in the plane $z_2 = 0$, its points have $z_2 \neq 0$ and so any such line contains a unique point of the form $(z_1, 1)$; thus we have a bijection

$$[(z, 1)] \leftrightarrow z$$

from the open set X_2 of lines not on the plane $z_2 = 0$; (this set is the image under π of the open set $\{z_2 \neq 0\}$ in $\mathbf{C}^2 \setminus (0, 0)$, hence open as π is an open map) to $\mathbf{C}$; we use z as a coordinate on X_2. There is only one line outside X_2, namely the line whose points are $(a, 0)$. We call this ∞. Thus we see that $\mathbf{CP}_1$ is obtained from $\mathbf{C}$ by adding a point at infinity. Let X_1 be the set of lines that are not in the plane $z_1 = 0$; we have a bijection

$$[(1, t)] \leftrightarrow t$$

and so we can use t as a coordinate on X_1. X is the union of X_1 and X_2; on $X_1 \cap X_2$ we have

$$[(z_1, z_2)] = [(z_1/z_2, 1)] = [(1, z_2/z_1)]$$

so that the coordinates z and t are related by

$$t = \frac{1}{z}, \qquad z = \frac{1}{t}$$

So X is just the extended complex plane defined earlier.

$X = \mathbf{C}/L$: We now come to the example that is most important for us, namely the *torus.* Let L be a lattice in $\mathbf{C}$. Since $\mathbf{C}$ and L are additive abelian groups we can define the group $X = \mathbf{C}/L$. The topology of X is the quotient topology with respect to the natural map

$$\pi : \mathbf{C} \longrightarrow \mathbf{C}/L$$

Since X is the image under π of any fundamental parallelogram we see that X is compact. If $\{\omega_1, \omega_2\}$ is a $\mathbf{Z}$–basis of L, we have an isomorphism $(\mathbf{C}, L)$ with $(\mathbf{R}^2, \mathbf{Z}^2)$ so that

$$X \simeq \mathbf{R}^2/\mathbf{Z}^2 \simeq \mathbf{R}/\mathbf{Z} \times \mathbf{R}/\mathbf{Z}$$

Thus X is a torus in the topological sense. Actually this isomorphism is C^∞. The local coordinates on X are defined naturally. If x is any point of X and a is a point of $\mathbf{C}$ above x (so that $x = a + L$), we can find a small disk $D_a = \{|z - a| < \delta\}$ around a such that the map π is one to one on D_a so that we can use $z - a$ as a local coordinate on $\pi(D_a)$. This defines $\mathbf{C}/L$ as a compact complex manifold of dimension 1.

Any function on X or on an open subset of X can be lifted to a function on $\mathbf{C}$ or an open subset of $\mathbf{C}$ that is invariant under translations by elements of L, and conversely any such function arises from a function on X. Hence the elements of $\mathcal{M}(L)$ can be viewed as meromorphic functions on X. Thus we have a clear analogy with the case of $\mathbf{CP}_1$. However, function theory on $\mathbf{C}/L$ differs in mnay ways from function theory on $\mathbf{CP}_1$. For instance we know that in constructing rational functions with given zeros and poles there are no restrictions on the locations of the zeros and poles, but this is no longer true on $\mathbf{C}/L$; indeed, we have seen that

if $a_1, \ldots, a_r$ are the zeros and $\mathbf{B_1}, \ldots, \mathbf{b_r}$ are the poles for a meromorphic function on $\mathbf{C}/L$, counting multiplicities, then

$$a_1 + \ldots + a_r + b_1 + \ldots + b_r = 0$$

in $\mathbf{C}/L$. It is a deep result that we shall prove later that if this condition is satisfied, then there is a meromorphic function on $\mathbf{C}/L$ which ahs precisely these zeros and poles (Abel's theorem).

Recall that in defining a complex manifold we started with a topological space. It is a little easier to start with a manifold of class C^1 which is defined in the same way as a complex manifold, except that the transition between different local coordinates is only required to be continuously differentiable. Then one can define the structure of a complex manifold by selecting a subset of allowable local coordinate systems with the property that the transitions are holomorphic (the possibility of being able to make this selection is an assumption). The question, first studied by Riemann, is whether there is more than one way of making this further choice of local coordinates, i.e., whether there is more than one complex structure compatible with the structure of X as a C^1–manifold. It turns out that this is not so for S^2. The complex structure defined on S^2 above is the only one upto isomorphism.

Theorem 1. *Let X be a complex manifold that is C^1–ismorphic to S^2. Then X is complex analytically isomorphic to $\mathbf{CP}_1$.*

The case of the torus is more interesting. For any lattics L in $\mathbf{C}$ we have seen that $\mathbf{C}/L$ is C^∞–isomorphic to the standard torus $\mathbf{R}^2/\mathbf{Z}^2$ so that for any two lattices L, L' in $\mathbf{C}$ we have a C^∞–isomorphism

$$\mathbf{C}?L \simeq \mathbf{C}/L'$$

In fact, if $(\omega_i), (\omega_i')$ are $\mathbf{Z}$–bases for L, L', then the map $T : x_1\omega_1 + x_2\omega_2 \longmapsto x_1\omega_1' + x_2\omega_2'$ is a *real linear* isomorphism of $\mathbf{C}$ onto itself, and $t(L) = L'$. Then t induces a real analytic isomorphism

$$t^\sim : x + L \longrightarrow t(x)x + L'$$

But in general we cannot choose a *complex analytic* isomorphism of $\mathbf{C}/L$ with $\mathbf{C}/L'$. In fgactw e have the following theorem.

Theorem 2. *Let L, L' be lattices in $\mathbf{C}$.a complex manifold that is C^1–ismorphic to S^2. Then $\mathbf{C}/L$ and $\mathbf{C}/L'$ are complex analytically isomorphic if and only if $L \simeq L'$,i.e., there is a number $c \in \mathbf{C} \setminus (0)$ such that $L' = cL$. Moreover, if X is a complex analytic manifold whose underlying smooth manifold is a torus, then X is complex analytically isomorphic to $\mathbf{C}/L$ for some lattice L in $\mathbf{C}$.*

We have already seen that lattices viewed up to complex multiplication are parametrized by $\mathcal{H}/\mathrm{SL}(2,\mathbf{Z})$. So there are continuum many complex structures on a given smooth torus. A similar result is also valid for *annuli.* For any $0 < r < R$ let $A_{r,R}$ be the annulus $\{r < |z| < R\}$ with the usual structure as a complex manifold.

Theorem 3. *Let X be a complex manifold that is C^1–ismorphic to an annulus. Then X is complex analytically isomorphic to some $A_{r,R}$. Moreover $A_{r,R}$ and $A_{r',R'}$ are complex analytically isomorphic if and only if $r/R = r'/R'$.*

The proofs of these theorems are not trivial and we shall not discuss them at this stage.

5. COMPLEX TORI AS NONSINGULAR PLANE CUBIC CURVES

1. Complex manifolds of dimension 2. The definition of a complex manifold of dimension 2 is exactly the same as for a manifold of dimension 1 except that there are two local coordinates instead of one. To say that X is a complex manifold of dimension 2 we should have the following data.

(i) X should be a Hausdorff topological space, connected and satisfying the second axiom of countability.

(ii) X should be covered by open sets X_i and for each i there should be a homomorphism $h_i = (z_{i1}, z_{i2})$ of X_i with an open set D_i of the complex space $\mathbf{C}^2$; h_i is called the *local coordinate system* on X_i.

(iii) If the indices i, j are such that $X_{ij} = X_i \cap X_j$ is nonempty, and if D_{ij} and D_{ji} are the images under h_i and h_j respectively of X_{ij}, then the function

$$f_{ij} : D_{ij} \longrightarrow D_{ji}$$

defined by

$$f_{ij} = h_j \circ h_i^{-1}$$

is a holomorphic function of two variables. By interchanging i and j we see that f_{ij} is invertible (as a map) and f_{ji} is the inverse map to f_{ij}.

The space $\mathbf{C}^2$ or any open connected subset of it is a complex manifold of dimension 2 according to this definition with (z_1, z_2) serving as a system of coordinates everywhere. The *complex projective plane* $\mathbf{CP}^2$ is an example of a *compact* complex manifold of dimension 2 which is very important for us. It is defined as the set of all complex lines through the origin of the 3–dimensional vector space $\mathbf{C}^3$. We have a map

$$\pi : \mathbf{C}^3 \setminus (0) \longrightarrow \mathbf{CP}^2, \qquad \pi((a_1, a_2, a_3)) = [(a_1, a_2, a_3)]$$

where $[(a_1, a_2, a_3)]$ is the line containing (a_1, a_2, a_3). If ℓ is a line in $\mathbf{C}^3$ viewed as a point of $\mathbf{CP}^2$ and if (a_1, a_2, a_3) is a nonzero point of ℓ, we call

$(1_1, a_2, a_3)$ the *homogeneous coordinates of* ℓ*;* if $k \neq 0$, then (ka_1, ka_2, ka_3) are also the homogeneous coordinates of ℓ. We give to $\mathbf{CP}^2$ the quotient topology: a subset of $\mathbf{CP}^2$ is open if and only if its preimage under π is open in $\mathbf{C}^3 \setminus (0)$. It is easy to show that π is an open map, that $\mathbf{CP}^2$ is connected, Hausdorff, and second countable. Although a function f on $\mathbf{C}^3$ does not make sense on $\mathbf{CP}^2$, it will do so if it is homogeneous of degree 0; more generally, if f is homogeneous of some degree d, the set of points where f is nonzero is the preimage of a subset of $\mathbf{CP}^2$, and this set will be defined by abuse of language as the subset of $\mathbf{CP}^2$ where f is nonzero. Similarly, we can speak of the subset of $\mathbf{CP}^2$ where f is zero. If $X_i (i = 1, 2, 3)$ are the sets where $x_i \neq 0$, then each X_i is open and

$$\mathbf{CP}^2 = X_1 \cup X_2 \cup X_3$$

On each X_i we have coordinates that take X_i to $\mathbf{C}^2$; thus the coordinates on X_1 are $z_1 = x_2/x_1, z_2 = x_3/x_1$. The transformations between the overlaps are holomorphic. For instance, if (z_1, z_2) and (w_1, w_2) are the coordinates of a point in $X_1 \cap X_2$, then

$$w_1 = \frac{1}{z_1},\ w_2 = \frac{z_2}{z_1}, \qquad z_1 = \frac{1}{w_1},\ z_2 = \frac{w_2}{w_1}$$

We thus have a complex manifold of dimension 2. It can be viewed as being obtained by adding to $\mathbf{C}^2$ *a line at infinity,* with $\mathbf{C}^2$ identified with any one of the X_i.

In a general complex manifold X, if p is a point and z_1, z_2 are local coordinates at p, then for a pair w_1, w_2 of functions defined and holomorphic around p, the necessary and sufficient condition that w_1, w_2 be also a system of local coordinates at p is that the jacobian determinant between the w_i and z_i be nonzero at p:

$$\Big(\frac{\partial(w_1, w_2)}{\partial(z_1, z_2)}\Big)_P \neq 0$$

The manifold $\mathbf{CP}^2$ is the stage on which the geometric theory of plane curves takes place. Originally curves were studied only on affine space $\mathbf{C}^2$. But it was soon realized that the theory takes a simpler and more harmonious form if points at infinity were added and the curves treated as curves in the projective plane. Topologically, adding points at infinity is equivalent to compactifying the space and it is clear on topological grounds

that such a process will result in a simplification and streamlining of the theory.

2. Nonsingular cubic curves. We are interested in *cubic* curves in $\mathbf{CP}^2$. As a preliminary to this topic let us consider a curve defined by an equation

$$F(X, Y, W) = 0$$

where F is a homogeneous polynomial of degree N; the above equation is then said to define a *curve of degree* N and we denote th curve by C_F. If P is a point of C_F with homogeneous coordinates (p_1, p_2, p_3), P is said to be a *nonsingular point* of C_F if one of the three numbers

$$(\partial F/\partial X)_P,\ (\partial F/\partial Y)_P\ (\partial F/\partial W)_P$$

is nonzero. The geometrical meaning of this condition is easy to understand. Suppose that

$$(\partial F/\partial X)_P \neq 0$$

Then first of all either p_2 or p_3 is different from zero. For, otherwise, P is the point $(1, 0, 0)$, and the condition that P lies on C_F means that the term X^N does not enter F; but then $\partial F/\partial X$ has to vanish at P. If p_3 (for instance) is different from 0, we may take $W = 1$ and consider C_F in the neighborhood of P as the affine curve

$$F(X, Y, 1) = 0$$

where $(\partial F/\partial X)(P) \neq 0$. If we write $U = F, V = Y$, then

$$\frac{\partial(U, V)}{\partial(X, Y)} = \frac{\partial F}{\partial X} \neq 0$$

at P and so (U, V) is a system of local coordinates at P. In this coordinate system, the curve appears locally as $U = 0$ and so has the structure of a complex manifold of dimension 1. If every point of C_F is nonsingular, C_F is called a *nonsingular* or *smooth* curve. A nonsingular curve can thus be viewed as a compact complex manifold of dimension 1.

Consider now a curve in affine $\mathbf{C}^2$ with equation

$$Y^2 = P(X), \qquad P(X) = a_0X^3 + a_1X^2 + a_2X + a_3\ (a_0 \neq 0) \qquad (1)$$

For a point (X_0, Y_0) of the curve to be singular the condition is

$$Y_0 = 0,\ P(X_0),\ P'(X_0) = 0$$

so that if P has no multiple roots, the affine curve is nonsingular. Let us now examine if the curve is nonsingular in the projective plane when P has no multiple roots. In this case, in the projective plane the equation of the curve is

$$Y^2W = P(X, W), \qquad P(X, W) = a_0X^3 + a_1X^2W + a_2XW^2 + a_3W^3 \quad (1')$$

Since the degree of P is 3, the coefficient of X^3 in P is nonzero and so the points at infinity of the curve are given by $X^3 = 0$. In other words, there is exactly one point at infinity on the curve, namely $(0, 1, 0)$. To analyse the curve near $(0, 1, 0)$ we take $Y = 1$, so that the point $(0, 1, 0)$ reduces to $(0, 0)$ and the curve to the affine curve in the (X, W)–plane

$$W = a_0X^3 + a_1X^2W + a_2XW^2 + a_3W^3$$

It is immediate that $(0, 0)$ is a nonsingular point. Thus the projective cubic curve (1') is nonsingular if and only if P has 3 distinct roots.

Let us take the cubic curve in the form

$$Y^2 = 4X^3 - a_2X - a_3 \tag{2}$$

The nonsingularity is equivalent to the fact that the polynomial on the RHS has 3 distinct roots. Let us write down the condition for this. This condition is equivalent to saying that the polynomial and its derivative have no root in common. The derivative is $12X^2 - a_2$ whose roots are $x = \pm a_2/2\sqrt{3}$. So the condition is $x(4x^2 - a_2) \neq \pm a_3$, i.e., $x^2(4x^2 - a_2)^2 \neq a_3^2$. This simplifies to

$$\Delta := a_2^3 - 27a_3^2 \neq 0 \tag{3}$$

3. The isomorphism of complex tori with nonsingular plane curves. Recall that if L is a lattice in $\mathbf{C}$, the Weierstrass function

$$\wp = \wp_L$$

satisfies the equation

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3 = 0 \tag{4}$$

where

$$g_2 = g_2(L) = 60 \sum_{0 \neq \omega \in L} \omega^{-4}, \qquad g_3 = g_3(L) = 140 \sum_{0 \neq \omega \in L} \omega^{-6} \tag{5}$$

Theorem 1. *We have the factorization*

$$4X^3 - g_2 X - g_3 = 4(X - e_1)(X - e_2)(X - e_3) \qquad e_i = \wp(\omega_i/2)$$

where ω_1, ω_2 is a $\mathbf{Z}$–basis for L and $\omega_3 = \omega_1 + \omega_2$. Moreover the e_i are distinct. In particular,

$$\Delta(L) := g_2^3 - 27 g_3^2 \neq 0 \tag{6}$$

Proof. Since $\wp$ is of order 3, the function $4\wp^3 - g_2\wp - g_3$ is of order 6 and 0 is its sole singularity, a pole of order 6. Now $\wp'$ vanishes at $\omega_i/2$ and as the order of $\wp'$ is 3, these are simple zeros of $\wp'$. So $z = \omega_i/2$ is a double zero for $\wp - e_i$, and the e_i are distinct as $\wp$ has order 2. Hence

$$\wp'^2 = 4(\wp - e_1)(\wp - e_2)(\wp - e_3)$$

So

$$4\wp^3 - g_2\wp - g_3 = 4(\wp - e_1)(\wp - e_2)(\wp - e_3)$$

showing that

$$4X^3 - g_2 X - g_3 = 4(X - e_1)(X - e_2)(X - e_3)$$

The condition (6) is immediate from the criterion (3). This finishes the proof.

The point $(\wp(z), \wp'(z))$ is thus on the cubic curve

$$Y^2 = P(X), \qquad P(X) = 4X^3 - g_2 X - g_3$$

If Γ_L is the corresponding projective curve, the map

$$F : z + L \longmapsto [(\wp(z), \wp'(z), 1)] \tag{7}$$

is well defined and holomorphic from $\mathbf{C}/L \setminus (0)$ to Γ_L.

Theorem 2. *The map F extends holomorphically to all of $\mathbf{C}/L$ and is a complex analytic isomorphism of $\mathbf{C}/L$ with the nonsingular cubic curve Γ_L.*

Proof. The map F is clearly holomorphic from $\mathbf{C}/L \setminus (0)$ into $\mathbf{CP}^2$ and goes into Γ_L. So it is a holomorphic map into Γ_L. Close to 0 we can also write it as

$$F(z) = [(\wp(z)/\wp'(z)), 1, (1/\wp'(z))]$$

which is of the form

$$F(z) = [(az + \ldots, 1, bz^3 + \ldots)]$$

This shows that if we extend F to 0 by setting

$$F(0) = [(0, 1, 0)]$$

then F is holomorphic also at $z = 0$ and takes $z = 0$ to $[(0, 1, 0)]$, the point at infinity of Γ_L. Since $\wp$ takes all values, it follows that F is surjective; indeed, if (x, y) lies on the curve, we first find z such that $\wp(\pm z) = x$, and then $y = \pm\wp'(z) = \wp'(\pm z)$. Let us show that this map is bijective. Since only 0 goes to the point at infinity, it is enough to check bijectivity on $\mathbf{C}/L \setminus (0)$. If we exclude the points $\omega_i/2$, $\wp$ is $2--1$ on the remaining set, and $\wp'$ distinguishes between the pairs of points where $\wp$ takes the same value; but the three points $\omega_i/2$ are distinct and the $\wp$–values there are also distinct and different from those of $\wp$ at the other points. This proves the bijectivity.

This is already enough to finish the proof. An analytic map between complex manifolds of dimension 1 which is bijective, is an analytic isomorphism, i.e., the inverse map is also analytic (see problem # 1). But in this case we can explicitly verify this also. First consider a point $P_0 = (\wp(z_0), \wp'(z_0))$ of the affine cubic. If $z_0 \notin (1/2)L$, $\wp'(z_0) \neq 0$ and so z is an analytic function of $X = \wp$ in a neighborhood of P_0. If $z_0 \in (1/2)L \setminus L$, then $z_0 = \omega_i/2$ for some $i = 1, 2, 3$, and then $\wp'\prime(z_0) \neq 0$, so that z is an analytic function of $Y = \wp'(z)$ in a neighborhood of P_0. Finally, let $z_0 = 0$. The image point is the point $P_0 = (0, 1, 0)$ at infinity. The function X/Y is well defined and analytic on the curve around P_0 and coincides with $\wp/\wp'$ on the torus. The derivative

$$\Big(\frac{\wp}{\wp'}\Big)' = \frac{(\wp')^2 - \wp\wp''}{(\wp')^2} = 1 - \frac{\wp\wp''}{(\wp')^2}$$

is nonzero at $z = 0$; in fact, as $z \to 0$, this expression tends to $-1/2$. Hence z is an analytic function of X/Y around P_0. This finishes the proof.

4. The inversion problem. The theorem proved just now raises the following question: *is it true that every nonsingular cubic curve in* $\mathbf{CP}^2$ *is isomorphic to some torus* $\mathbf{C}/L$*?* This is the *inversion problem* in the theory of elliptic functions. The answer is affirmative and for that reason the nonsingular plane cubic curves are called *elliptic curves.* There are many different ways to prove this, and each method illustrates some profound aspect of the theory of elliptic functions. The simplest is the *method of modular functions* and we shall take it up first.

The functions g_2 and g_3 are defined on the space of all lattices in $\mathbf{C}$ and change as follows under the map $L \mapsto \lambda L$ ($\lambda \neq 0$):

$$g_2(\lambda L) = \lambda^{-4} g_2(L), \qquad g_3(\lambda L) = \lambda^{-6} g_3(L) \tag{7}$$

From this it is easy to construct functions on lattices that are invariant under these maps (we refer to this action $L \longmapsto \lambda L$ of $\mathbf{C}^\times = \mathbf{C} \setminus (0)$ on the space of lattices as complex multiplication). The discriminant $\Delta(L)$ defined by (6) is one such. Classically the following function plays a big role:

$$J(L) := \frac{g_2(L)^3}{\Delta(L)} = \frac{g_2(L)^3}{g_2(L)^3 - 27 g_3(L)^2} \tag{8}$$

Clearly J is invariant under $L \mapsto \lambda L$. Recall now the fact that any lattice is equivalent to a lattice $L_\tau (= \mathbf{Z}\tau \oplus \mathbf{Z}1)$ under complex multiplication, τ being in the upper half plane $\mathcal{H}$. So functions on the space of lattices that are invariant under complex multiplication of lattices may be viewed as functions on $\mathcal{H}$ that are invariant under the action of the modular group Γ. Here the modular group Γ is the group of all matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad a, b, c, d \in \mathbf{Z}, ad - bc = 1$$

and the action is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tau \longmapsto \frac{a\tau + b}{c\tau + d}$$

We put

$$j(\tau) = J(L_\tau) \tag{9}$$

The functions

$$G_k(\tau) = \sum_{(0,0)\neq(m,n)\in\mathbf{Z}^2} (m\tau + n)^{-2k}$$

are holomorphic in τ (we shall see this later) and so j is holomorphic. Since J is invariant under complex multiplication, we have

$$jBigl(\frac{a\tau + b}{c\tau + d}\Big) = j(\tau) \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix}) \in \Gamma)$$

If $L = L_i$, the lattice with $\mathbf{Z}$–basis $\{i, 1\}$, then $iL_i = L_i$ and so

$$g_3(L_i) = g_3(iL_i) = i^{-6} g_3(L_i) = -g_3(L_i)$$

hence

$$g_3(L_i) = 0 \tag{8}$$

A similar argument can be used to show that

$$g_2(L_\rho) = 0, \qquad \Big(\rho = -\frac{1}{2} + i\frac{\sqrt{3}}{2}\Big)$$

In fact, ρL_ρ is the lattice spanned by ρ and $\rho^2 = -1 - \rho$ and so is L_ρ. Hence

$$g_2(L_\rho) = g_2(\rho L_\rho) = \rho^{-4} g_2(L_\rho) = -\rho^{-1} g_2(L_\rho)$$

as $\rho^3 = 1$. Since $-\rho^{-1} \neq 1$, we are done. So

$$g_2(L_\rho) = 0$$

Thus

$$j(i) = 1, \qquad j(\rho) = 0 \tag{10}$$

The basic theorem on the j–function is the following.

Theorem 3. *The function j takes all values in* $\mathbf{C}$

We shall prove this in the next chapter. For now we shall complete the proof of the inversion theorem on the basis of this result.

Proof of inversion assuming Theorem 3. Let a_2, a_3 be complex numbers such that $\Delta = a_2^3 - 27a_3^2 \neq 0$. We must show that there is a lattice

L such that $g_2(L) = a_2, g_3(L) = a_3$. Clearly both a_2 and a_3 cannot be zero as $\Delta \neq 0$. If $a_3 = 0$, we know that $g_3(L_i) = 0$ and so, as $g_2(L_i) \neq 0$ we can choose $\lambda \neq 0$ such that $g_2(\lambda L_i) = \lambda^{-4} g_2(L_i) = a_2$. If $a_2 = 0$ we can likewise choose $\lambda \neq 0$ such that $g_3(\lambda L_\rho) = \lambda^{-6} g_3(L_\rho) = a_3$. We may therefore assume that both a_2 and a_3 are different from 0. This is equivalent to saying that $\Delta \neq 0, \neq 1$. By Theorem 3 there is a $\tau \in \mathcal{H}$ such that $j(\tau) = \Delta$. Write $A_2 = g_2(L_\tau), A_3 = g_3(L_\tau)$. Thus

$$\frac{A_2^3}{A_2^3 - 27A_3^2} = \frac{a_2^3}{a_2^3 - 27a_3^2} = k \neq 0, 1 \tag{11}$$

It is enough to prove that there is a $\lambda \neq 0$ such that

$$a_2 = \lambda^{-4} A_2, \qquad a_3 = \lambda^{-6} A_3 \tag{12}$$

for then we would have $a_2 = g_2(L), a_3 = g_3(L)$ for $L = \lambda L_\tau$. From (11) we get

$$\frac{A_2^3}{A_3^2} = \frac{a_2^3}{a_3^2}$$

Choose μ such that $a_2 = \mu^{-4} A_2$; this is possible, and together with μ, $\pm\mu, \pm i\mu$ are also valid choices. Then we get $a_3^2 = \mu^{-12} A_3^2$ so that $a_3 = \pm\mu^{-6} A_3$. If we have the plus sign here we can take $\lambda = \mu$ and obtain (12); if we have the minus sign here, we take $\lambda = i\mu$ to get (12). This finishes the argument.

The argument for proving Theorem 3 is a deeper one and takes us through the beginnings of the theory of modular functions and modular forms. We shall do this in the next chapter. However there is one more point. This argument proves only that every cubic curve in the special form (2) arises from a torus. There still remains the question whether the form (2) exhausts all nonsingular cubic curves. The answer again is in the affirmative and one can show that we can choose a suitable linear coordinate system in $\mathbf{CP}^2$ so that a given nonsingular cubic has the form (2). The form (2) is called the *Weierstrass normal form of the nonsingular cubic.* This also will be proved later.

It follows from Theorem 3 that the cubic curve (2) is parametrized by the map

$$z \longmapsto (\wp_L(z), \wp'_L(z))$$

for some lattice L. It can be shown that there is no *rational* parametrization of (2). For an algebraic proof see the book of Prasolov and Solovyev.

6. PLANE CUBIC CURVES AS COMPLEX TORI :

PROOF BY MODULAR FUNCTIONS

1. Modular group. We write $SL(2,\mathbf{Z})$ for the group of 2×2 matrices with integer entries and determinant 1. It acts on the upper half plane $\mathcal{H}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tau \longmapsto \frac{a\tau+b}{c\tau+d}$$

Since -1 acts trivially the action is actually of the group

$$\Gamma := SL(2,\mathbf{Z})/(\pm 1)$$

Γ is called the *modular group.* We shall often describe elements of γ as matrices but with the understnding that the choice of the matrix is ambiguous up to a sign. Let

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Clearly

$$T\tau = \tau + 1, \qquad S\tau = -\frac{1}{\tau}$$

Moreover we find

$$S^2 = 1, \qquad (ST)^3 = (TS)^3 = 1 \qquad (\text{ in } \Gamma)$$

$\mathcal{H}$ and the action of the modular group are important because $\mathcal{H}$ describes the space $\mathcal{L}$ of lattices in $\mathbf{C}$ up to complex multiplication. If L_τ is the lattice with $\mathbf{Z}$–basis $(\tau, 1)$ and we denote complex multiplication by $\sim$, then any $L \in \mathcal{L}$ is $\sim$ to some L_τ, while $L_\tau \sim L_{\tau'}$ if and only if there is a $\gamma \in \Gamma$ such that $\tau' = \gamma\tau$. We have

$$\Im\Big(\frac{a\tau+b}{c\tau+d}\Big) = \frac{1}{|c\tau+d|^2}\Im(\tau)$$

Furthermore. for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}))$, $\{a\tau + b, c\tau + d\}$ is also a $\mathbf{Z}$–basis for L_τ and so

$$L_{\tau'} = (c\tau + d)^{-1} L_\tau \qquad \tau' = \frac{a\tau + b}{c\tau + d}$$

Finally, let

$$\rho = \frac{-1 + i\sqrt{3}}{2}$$

Theorem 1. *Let*

$$D = \left\{ \tau \in \mathcal{H} \;\middle|\; |\Re(\tau)| \leq 1/2, |\tau| \geq 1 \right\}$$

Then we have:

(i) Every element of $\mathcal{H}$ can be moved to an element of D by Γ.

(ii) If $\tau, \tau' \in D$ and $\gamma \in \Gamma \setminus (1)$ is such that $\tau' = \gamma\tau$, then τ, τ', are both in the boundary of D and we have

$$\Re(\tau) = \pm 1/2, \quad \tau' = \tau \pm 1 = T^{\pm 1}\tau$$
$$|\tau| = 1, \tau' = -1/\tau = S\tau$$

(iii) The stabilizer at ρ (resp. $-\overline{\rho}$) is the cyclic group of order 3 generated by ST (resp. TS), the stbilizer at i is the 2–element group $\{1, S\}$, and all other stabilizers are trivial.

(iv) Γ is generated by S and T.

Proof. (See fig 1) Given τ, we have $\Im(\tau') \geq \Im(\tau)$ if and only if $|c\tau+d|^2 \leq 1$ where $\tau' = (a\tau + b)/(c\tau + d)$. Now $x\tau + y = 0$ for real x, y if and only if $x = y = 0$. Hence

$$|x\tau + y|^2 = x^2|\tau|^2 + 2\Re(\tau) + y^2$$

is a positive definite quadratic form in the real variables x and y and so $|m\tau+n|^2 \leq 1$ only for *finitely* many pairs (m, n) of integers. Thus there are points τ' in the orbit (under Γ) of τ such that $\Im(\tau')$ takes the maximum possible value for points on the orbit. Changing τ' to $\tau' + k$ where k is an integer does not change the imaginary part and so we may assume that the point τ' with maximum imaginary part satisfies $|\Re(\tau')| \leq 1/2$. We

now claim that $\tau' \in D$, i.e., $|\tau'| \geq 1$. If $|\tau'| < 1$ and $\tau_1 = S\tau' = -1\tau'$, then $\Im(\tau_1) = |\tau'|^{-2}\Im(\tau') > \Im(\tau')$ which is impossible because τ_1 is still in the orbit of τ and its imaginary part exceeds that of τ' which by construction had maximum imaginary part. This proves (i).

Suppose $\tau, \tau, \in D$ and $\tau' = \gamma\tau$ where $\gamma \in \Gamma$. Write $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. By interchanging τ and τ' we may assume that $\Im(\tau') \geq \Im(\tau)$. Then $c\tau + d|^2 \leq 1$. But if $\tau = u + iv$ where u, v are real, then $v \geq (\sqrt{3}/2)$ as $\tau \in D$, so that

$$1 \geq |c\tau + d|^2 = (cu + v)^2 + c^2v^2 \geq c^2v^2 \geq (3/4)c^2$$

Hence $c^2 \leq (4/3)$, showing that $c \in \{0, \pm 1\}$.

(a) $c = 0$. Then $ad = 1$ so that by changing the sign of the matrix representing γ we may assume that $a = d = 1$. Then $\gamma = T^k$, $\tau' = \tau + k$ where k is a non zero integer (as $\gamma \neq 1$), and as both τ and τ' have real parts in $[-1/2, +1/2]$ this means that we are in the first set of alternatives in (ii).

(b) $c = \pm 1$. The case $c = -1$ can be reduced to $c = 1$ by changing the sign of the matrix representing γ and so we may suppose that $c = 1$. Note that in this case $\gamma \neq 1$. Now, with $\tau = u + iv$ as before, $c\tau + d = \tau + d$ so that

$$1 \geq |\tau + d|^2 = |\tau|^2 + d^2 + 2du$$

But $d^2 + 2du \geq d^2 - |d|$ as $|u| \leq 1/2$ while $d^2 - |d| \geq 0$. Hence

$$1 \leq |\tau|^2 \leq |\tau|^2 + d^2 - |d| \leq |\tau|^2 + d^2 + 2du = |\tau + d|^2 \leq 1$$

which implies the set of relations

$$|\tau| = 1, |\tau + d| = 1, d^2 - |d| = 0 \Longleftrightarrow d \in \{0, \pm 1\}, |d| = -2du \qquad (*)$$

Since $ad - b = 1$ and $|\tau + d| = 1$, we have

$$\tau' = \frac{a\tau + ad - 1}{\tau + d} = a - \frac{1}{\tau + d} = a - (\overline{\tau} + d) = a - d - \overline{\tau}$$

So $\tau' = u' + iv'$ where $u' = a - d - u, v' = v$, and $|u| \leq 1/2, |u'| \leq 1/2$. Hence $a - d = 0, \pm 1$. If $a - d = 0$, then $\tau' = S\tau$; if $a - d = 1$, then

$u = u' = 1/2$ and so $\tau = \tau' = -\overline{\rho}$, while for $a - d = -1$, we have $u = u' = -1/2, \tau = \tau' = \rho$. The proof of (ii) is finished.

For (iii) the argument is a slight refinement of the preceding. Let $\tau' = \gamma\tau = \tau, \gamma \neq 1$ in the above. The treatment of the case $c = 0$ shows already that this possibility cannot arise. So we may suppose that $c = 1$. Then $|\tau| = 1$, $d \in \{0, \pm 1\}$ and the relation $|d| = -2du$ (cf. $(*)$) implies that for $|d| = 1$ the only possibilities are

$$d = 1, \tau = \rho;\ d = -1, \tau = -\overline{\rho}$$

If $a - d = 0$, then $\tau' = S\tau = \tau$ so that $\tau = i$; then $d = 0$ and so $\gamma = S$. If $a - d = 1$ we have $\tau = \tau' = -\overline{\rho}$ and d cannot be 1; so we have either $d = 0$ and $\gamma = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = TS$, or $d = -1$ and $\gamma = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} = (TS)^2$. If $a - d = -1$ we have $\tau = \tau' = \rho$ and d cannot be -1; so we have either $d = 0$ and $\gamma = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} = (ST)^2$, or $d = 1$ and $\gamma = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} = ST$.

We now prove (iv). Suppose $\gamma \in \Gamma$. Let Γ' be the subgroup of Γ generated by S and T. We wish to show that $\gamma' \in \Gamma'$. Let τ_0 be an *interior* point of D. The proof of (i) actually works with Γ' in place of Γ and so we can find $\gamma' \in \Gamma'$ such that $\gamma'\gamma\tau_0 \in D$. This is impossible by (ii) unless $\gamma'\gamma = 1$, i.e., $\gamma \in \Gamma'$. This finishes the proof.

Remark. If $\gamma \in \Gamma$, the domain $\gamma(D)$ has the same properties as D and D and $\gamma(D)$ meet only at their boundaries. Moreover, γ is the only element of Γ that moves D to a domain meeting $\gamma(D)$ at some interior point. So the various domains $\gamma(D)$ give a *paving* of $\mathcal{H}$ by fundamental domains (see figure 1).

2. Functions on the space of lattices. Modular forms and functions. We are interested in studying functions on the space of lattices. To come down to functions on $\mathcal{H}$ which parametrizes the lattices up to complex multiplication it is convenient to restrict ourselves to functions which behave in a simple manner under complex multiplication. A function $F : \mathcal{L} \longrightarrow \mathbf{C}$ is said to be of *weight* $2k$ if

$$F(\lambda L) = \lambda^{-2k} F(L)$$

If

$$G_k(L) = \sum_{0 \neq \omega \in L} \omega^{-2k} \qquad (k \geq 2)$$

then G_k is of weight $2k$; this is the motivation behind the definition. In general, if a space has an action by $\mathbf{C}^\times$, it is always important to study functions on that space which are homogeneous of some degree. If F is of weight $2k$ we get a function f on $\mathcal{H}$ if we set

$$f(\tau) = F(L_\tau)$$

Proposiiton 2. *The correspondence $f \longleftrightarrow F$ is a bijection between the set of functions on $\mathcal{L}$ of weight $2k$ and the set of functions f on $\mathcal{H}$ such that*

$$f(\tau) = (c\tau + d)^{-2k} f\Big(\frac{a\tau + b}{c\tau + d}\Big), \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}) \tag{1}$$

Moreover, f satisfies (1) if and only if

$$f(\tau + 1) = f(\tau), \qquad f(-1/\tau) = f(\tau) \tag{2}$$

Proof. If $\tau' = (a\tau + b)/(c\tau + d)$, then $L_{\tau'} = (c\tau + d)^{-1} L_\tau$. So if F is of weight $2k$, f satisfies (1). Conversely, let f satisfy (1). To define F, let L be a lattice. We can find a $\mathbf{Z}$–basis $\{\omega_1, \omega_2\}$ for L such that $\tau = \omega_1/\omega_2 \in \mathcal{H}$. We put $F(L) = \omega^{-2k} f(\tau)$. If we use another such basis $\{\omega_1', \omega_2'\}$, then $\omega_1' = a\omega_1 + b\omega_2, \omega_2' = c\omega_1 + d\omega_2$ where $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z})$; then, writing $\tau' = \omega_1'/\omega_2'$, we have, $\tau' = (a\tau + b)/(c\tau + d)$, and so,

$$\omega_2'^{-2k} f(\tau') = \omega_2'^{-2k} (c\tau + d)^{2k} f(\tau) = \omega_2^{-2k} f(\tau) \tag{3}$$

For the last statement it is a question of proving that if f satisfies (2), then it satisfies (1). Write

$$\theta(\gamma, \tau) = c\tau + d \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then it is an easy check that θ is a *cocycle,* i.e.,

$$\theta(\gamma_1\gamma_2, \tau) = \theta(\gamma_1, \gamma_2\tau)\theta(\gamma_2, \tau) \tag{4}$$

The equation (4) shows that if f satisfies (1) for γ_i $(i = 1, 2)$, then it satisfies (1) for $\gamma_1\gamma_2$. This finishes the proof because (2) is the statement that f satisfies (1) for T and S and T and S generate Γ.

Consider now a function f on $\mathcal{H}$ satisfying (2). If we set

$$q = e^{2i\pi\tau} \tag{5}$$

then the invariance $f(\tau + 1) = f(\tau)$ implies that f depends only on q. Moreover the map $\tau \longmapsto q$ takes $\mathcal{H}$ onto the punctured unit disk

$$D^{\times} = \{q \mid 0 < |q| < 1\}$$

So there is a function $f^{\sim}$ on $D^{\times}$ such that

$$f(\tau) = f^{\sim}(q) \tag{6}$$

We say that f is *meromorphic (resp. holomorphic) at infinity* if $f^{\sim}$ is meromorphic (resp. holomorphic) at $q = 0$. Notice that if f is meromorphic at infinity, then for some $r > 0$, $f^{\sim}$ has no singularity in $0 < |q| < r$; thus f has no poles in $\Im(\tau) > b$ for some $b > 0$. We say that f is a *modular form* if f is holomorphic everywhere, including infinity. In other words, f is a *modular form of weight* $2k$ if the following are satisfied:

(i) f is holomorphic on $\mathcal{H}$

(ii) f satisfies (1) or equivalently (2)

(iii) f is holomorphic at infinity.

The last condition is equivalent to saying that $f^{\sim}$ is bounded at $q = 0$ by Riemann's theorem on removable singularities. This is also equivalent to saying that f is bounded on $D_b = \{\tau \mid \Im(\tau) > b\}$ for some $b > 0$.

Proposition 3. *Let (by abuse of notation)*

$$G_k(\tau) = \sum_{(0,0)\neq(m,n)\in\mathbf{Z}^2} \frac{1}{(m\tau + n)^{2k}} \qquad (k \text{ an integer } \geq 2) \tag{7}$$

Then G_k is a modular form of weight $2k$. Moreover

$$G_k(\infty) = 2\zeta(2k) \tag{8}$$

where ζ is the Riemann zeta function.

Proof. We shall first show that the series for G_k converges uniformly in any region of the form

$$|u| \leq A,\ v \geq v_0 > 0 \qquad (\tau = u + iv, u, v \in \mathbf{R})$$

(In particular it will be uniformly convergent on D.) For this, note that

$$|m\tau + n|^2 = m^2(u^2 + v^2) + 2mnu + n^2 = Q(m, n)$$

where Q is the quadratic form corresponding to the symmetric matrix

$$\begin{pmatrix} u^2 + v^2 & u \\ u & 1 \end{pmatrix}$$

If $\alpha \geq \beta > 0$ are the eigenvalues of this matrix, we have

$$Q(m, n) \geq \beta(m^2 + n^2)$$

so that

$$\frac{1}{(m\tau + n)^{2k}} = \frac{1}{Q(m, n)^{k/2}} \leq \frac{1}{\beta^{k/2}} \frac{1}{(m^2 + n^2)^{k/2}}$$

and so it is enough to show that there is a constant $b > 0$ such that

$$\beta \geq b \qquad (|u| \leq A,\ v \geq v_0)$$

But

$$\alpha + \beta = 1 + u^2 + v^2, \quad \alpha\beta = v^2$$

Since $\alpha \leq \alpha + \beta \leq 1 + A^2 + v^2$ we get

$$\beta \geq \frac{v^2}{1 + A^2 + v^2} \geq \frac{v_0^2}{1 + A^2 + v_0^2} = b > 0$$

because the function $t \longmapsto t/(1 + A^2 + t)$ is increasing in $t > 0$. At this stage we know that G_k is holomorphic on $\mathcal{H}$. Notice that it would have been enough to prove uniform convergence on compact subsets of $\mathcal{H}$ to conclude that G_k is holomorphic on $\mathcal{H}$; however we need the sharper result to examine how G_k behaves when $\tau \to \infty$.

Since $G_k(\tau) = G_k(L_\tau)$ where the G_k on the right is the function on the space of lattices $L \longmapsto \sum_{0 \neq \omega \in L} \omega^{-2k}$, it is immediate that G_k satisfies (1). Hence to conclude that G_k is a modular form it is enough to verify that it is holomorphic at infinity. This would follow from (8) and so it is enough to prove (8). Since the series for G_k converges uniformly on D we have

$$\lim_{\tau \in D, |\tau| \to \infty} G_k(\tau) = \sum \lim_{\tau \in D, |\tau| \to \infty} \frac{1}{(m\tau + n)^{2k}} = \sum_{0 \neq n \in \mathbf{Z}} \frac{1}{n^{2k}} = 2\zeta(2k)$$

This finishes the proof.

Let f be a modular form of weight $2k$. We define the order function of f in the usual way; for any $p \in \mathcal{H}$, $m_p(f)$ is the order of the zero of f at p. We also define $m_\infty(f)$ as the order at $q = 0$ of $f^\sim$. Now, for any holomorphic function g on $\mathcal{H}$ and any $\gamma \in \Gamma$, the order of g at γp is the same as the order of $g \circ \gamma$ at p (check this). But the function $c\tau + d$ never vanishes on $\mathcal{H}$ if $(c, d) \neq (0, 0)$, so that by (1), the order of f at p is the same as the order of $f \circ \gamma$ at p which is, by the remark made just now, the order of f at $\gamma^{-1}p$. Hence

$$m_p(f) = m_{\gamma p}(f) \qquad (p \in \mathcal{H}) \tag{9}$$

Thus the order stays constant over each orbit. The following is the fundamental result we need.

Theorem 3. *For any nonzero modular form of weight $2k$ we have*

$$m_\infty(f) + \frac{1}{3}m_\rho(f) + \frac{1}{2}m_i(f) + \sum\nolimits^*_{p \in \Gamma\backslash\mathcal{H}} m_p(f) = \frac{k}{6} \tag{10}$$

where $\sum^*$ means that the sum is over all orbits of Γ in $\mathcal{H}$ other than the orbits of ρ and i.

Proof. The proof is very similar to the proof of the corresponding result for elliptic functions where we integrated f'/f over a fundamental parallelogram. By slightly moving that parallelogram we ensured that there were no zeros or poles on the boundary. Here the fundamental domain D is fixed and cannot be moved infinitesimally, and so we have to allow for singularities on the boundary of D. We take a contour $\mathcal{C}$ schematically shown in figure 2. We use the usual semicircular indentations at zeros of f on the vertical sides of D other than $\rho, -\overline{\rho}$, partially circular indentations around singularities at $\rho, -\overline{\rho}, i$ and at other points on the circular part of D. We shall pass to the limit when the indentations shrink to their centers. The top horizontal line is taken sufficiently high so that there are no singularities on that line and none beyond that line except at ∞. First of all we have, by Cauchy's theorem of residues,

$$\frac{1}{2i\pi}\oint_{\mathcal{C}} \frac{f'}{f} d\tau = \sum_{p \in \text{interior of } D} m_p(f)$$

Let us now compute this integral directly. The integral on a segment of the vertical line $\Re(\tau) = -1/2$ cancels with the integral on the corresponding vertical segment on the line $\Re(\tau) = 1/2$. If p is a zero of f on the vertical line $\Re(\tau) = -1/2$ different from ρ, the integral on the semicircular indentations around p and $p+1$ add up to $-m_p(f)$, the negative sign being taken because the orientation is the reverse of the standard one. The integral on a partially circular indentation around a singularity p on the boundary $|\tau| = 1$ (but different from $\rho, -\overline{\rho}$) needs a little more care. We write $f(\tau) = (\tau - p)^m g(\tau)$ where $m = m + p(f)$ and g is holomorphic around p with $g(p) \neq 0$. Then

$$\frac{f'}{f} = \frac{m}{\tau - p} + \frac{g'}{g}$$

Since the second term is holomorphic at p, the integral on the indentation tends to 0 in the limit. The integral of the first term is $-m\theta/2\pi$ where θ is the change of the argument on the boundary which tends to π in the limit. So the integral of f'/f tends to $-m_p(f)/2$. If $p = i$ this is the only contribution. If p is different from i, there are two contributions, from p and Sp, so that the contribution is $-m_p(f)$. The same argument applies at ρ and $\overline{\rho}$, except now the change of the argument along the indentation is $-2\pi/6$ in the limit. So the contribution is $-m_\rho(f)/6$ twice, i.e., $-m_\rho(f)/3$.

There remain the integral along the horizontal line segment and the integral over a circular segment to the left of i where there are no zeros of f, and the integral over the corresponding segment to the right of i. For the horizontal segment we use the transformation to $q = e^{2i\pi\tau}$. We have

$$f^\sim(q) = f(\tau)$$

and so

$$\frac{f^{\sim\prime}(q)}{f^\sim(q)} dq = \frac{f'(\tau)}{f(\tau)} d\tau$$

Since the segment from $\Re(\tau) = 1/2$ to $\Re(\tau) = -1/2$ goes over to the circle $|q| =$ constant with the reverse of the standard orientation, the integral is $-m_\infty(f)$. So we are finally left with the integral over a circular arc, say from B' to C, both B' and C being to the left of i and there are no zeros of f between them. Let C', D be the images of B', C respectively under the transformation S. It is enough to calculate

$$\frac{1}{2i\pi}\Big(\oint_{B'}^{C} + \oint_{C'}^{D}\Big)\frac{f'(\tau)}{f(\tau)} d\tau$$

Now, writing $w = -1/\tau$,

$$\frac{f'(w)}{f(w)}dw = \frac{f'(-1/\tau)}{f(-1/\tau)}\tau^{-2}d\tau$$

On the other hand,

$$f(-1/\tau) = \tau^{2k}f(\tau)$$

so that, differentiating with respect to τ,

$$f'(-1/\tau)\tau^{-2} = 2k\tau^{2k-1}f(\tau) + \tau^{2k}f'(\tau)$$

Hence

$$\frac{f'(-1/\tau)}{f(-1/\tau)} = \tau^2\Big(\frac{2k}{\tau} + \frac{f'(\tau)}{f(\tau)}\Big)$$

Hence

$$\frac{f'(w)}{f(w)}dw = \Big(\frac{2k}{\tau} + \frac{f'(\tau)}{f(\tau)}\Big)d\tau$$

This gives

$$\frac{1}{2i\pi}\oint_{C'}^{D}\frac{f'(w)}{f(w)}dw = -\frac{1}{2i\pi}\oint_{B'}^{C}\Big(\frac{2k}{\tau} + \frac{f'(\tau)}{f(\tau)}\Big)d\tau$$

Hence

$$\frac{1}{2i\pi}\Big(\oint_{B'}^{C} + \oint_{C'}^{D}\Big)\frac{f'(\tau)}{f(\tau)}d\tau = -2k\frac{1}{2i\pi}\oint_{B'}^{C}\frac{d\tau}{\tau}$$

which equals

$$\frac{k}{\pi} \times \text{ the angle from } C \text{ to } B'$$

There are possibly several such arcs from ρ to i and so the total contribution from these integrals is

$$\frac{k}{\pi}\frac{\pi}{6} = \frac{k}{6}$$

Hence we finally get

$$-\frac{1}{3}m_\rho(f) - \frac{1}{2}m_i(f) - \sum_{p\in]\rho,i[} m_p(f) - m_\infty(f) + \frac{k}{6} = \sum_{p\in\text{interior of } D} m_p(f)$$

This is just (10).

Let $\mathcal{M}_k$ be the space of modular forms of weight $2k$. It is a complex vector space.

Corollary 4. *$\mathcal{M}_k = 0$ for $k < 0, k = 1$ and $\mathcal{M}_k = \mathbf{C}$ for $k = 0$.*

Proof. For $k < 0$ this is immediate from (10) as all the orders are ≥ 0. If $k = 1$ and some order is > 0 the left side of (10) will be $> 1/6$ which is not possible; if all the orders are 0 the left side is 0 which is also impossible. For $k = 0$, $\mathcal{M}_k \supset \mathbf{C}$. If f is a nonzero element of $\mathcal{M}_0$, then all the orders are 0 so that f is nowhere 0. Then $f - f(i)$ is in $\mathcal{M}_0$ and if it is not identically zero, must have no zero, a contradiction since it vanishes at i.

Theorem 5. *We have the following.*

(i) The function $\Delta = g_2^3 - 27g_3^2$ is an element of $\mathcal{M}_6$. It has a simple zero at ∞ and has no zero on $\mathcal{H}$.

(ii) The function $j = g_2^3/(g_2^3 - 27g_3^2) = g_2^3/\Delta$ is invariant under the modular group and induces a bijection of $\Gamma\backslash\mathcal{H}$ onto $\mathbf{C}$ and has a simple pole at ∞.

Proof. We have

$$g_2(\infty) = 60G_2(\infty) = 120\zeta(4) = 120\frac{\pi^4}{90} = \frac{4}{3}\pi^4$$
$$g_3(\infty) = 140G_2(\infty) = 280\zeta(6) = 280\frac{\pi^6}{21 \times 45} = \frac{8}{27}\pi^6$$

From this it is immediate that

$$\Delta(\infty) = 0 \tag{11}$$

Now $\Delta \in \mathcal{M}_6$ and so we can use (10) with $f = \Delta, k = 6$. Since $m_\infty(\Delta) \geq 1$ it follows that $m_\infty(\Delta) = 1$ and all other orders are 0. This proves (i).

To prove (ii) let $\lambda \in \mathbf{C}$ and let $f = g_2^3 - \lambda\Delta$. Since $f(\infty) = g_2^3(\infty) \neq 0$, $m_\infty(f) = 0$ and so

$$1 = \frac{n_1}{3} + \frac{n_2}{2} + n^*$$

where n_1, n_2, n^* are integers ≥ 0 being respectively

$$m_\rho(f), m_i(f), \sum\nolimits^*_{p \in \Gamma\backslash\mathcal{H}} m_p(f)$$

The only possibilities for (n_1, n_2, n^*) are

$$(3,0,0),\ (0,2,0)\ (0,0,1)$$

which proves that f has exactly one zero in $\mathcal{H}$. This shows (as Δ is never zero on $\mathcal{H}$) that j takes the value λ exactly at one point of $\Gamma\backslash\mathcal{H}$.

Theorem 6. *Let $\mathbf{C}_3$ be the nonsingular cubic curve whose affine equation is in the Weierstrassian normal form*

$$Y^2 = 4X^3 - a_2X - a_3$$

Then there is a lattice L in $\mathbf{C}$ such that $g_2(L) = a_2, g_3(L) = a_3$. In particular, the map

$$z \longmapsto (\wp(z), \wp'(z))$$

ia a parametrization of $\mathbf{C}_3$, and the projective map

$$z \longmapsto [(\wp(z), \wp'(z), 1)]$$

is a complex analytic isomorphism of $\mathbf{C}/L$ with $\mathbf{C}_3$.

Proof. We have already seen how this follows from theorem 5.

Remark. This theorem can be informally described as asserting that the nonsingular plane cubic curves may be parametrized by elliptic functions. For this reason they are called *elliptic curves.* This is in striking contrast to conics which can be parametrized by rational functions. It can be proved that the elliptic curves *do not admit* any parametrization by rational functions.

7. GEOMETRY OF NONSINGULAR

PLANE CUBIC CURVES

1. Algebraic curves and compact Riemann surfaces. The isomorphism between complex tori and nonsingular plane cubic curves is remarkable because it says that two very different types of objects are exactly the same. The complex tori have a transcendental origin. They arise as the quotients $\mathbf{C}/L$ the complex plane $\mathbf{C}$ by lattices L which are discrete subgroups of full rank of $\mathbf{C}$. Thus their construction involves topology and function theory. They are, together with the extended complex plane, the simplest compact complex manifolds of dimension 1, i.e., compact Riemann surfaces. Historically compact Riemann surfaces of genus > 1 arose by uniformizing algebraic functions so that from the very beginning the transcendental and algebraic theories were deeply intertwined. From the algebraic point of view the goal was a complete understanding of the geometry of all algebraic curves. The curves can be singular, but it is possible to "desingularize"the curves and obtain compact Riemann surfaces. The transcendental theory was the creation of Riemann. It was he who first showed that there is no distinction between compact Riemann surfaces and irreducible algebraic curves. Riemann also discovered the basic idea that the transcendental theory should deal not only with meromorphic functions but also meromorphic differentials.

Thus the whole theory has two faces to it. It is like a book, where the odd numbered pages are written in the transcendental language while the even numbered pages are in the algebraic language. Some questions and their solutions are natural in the transcendental picture while other concepts and ideas are more natural in the algebraic picture. But in the end both tell the same story. This bilingual aspect is what gives the entire theory its great beauty and depth.

2. Addition theorem for the Weierstrass functions. Addition on a cubic curve. If L is a lattice in $\mathbf{C}$ the manifold $\mathbf{C}/L$ has a group structure. It is therefore an interesting question to ask if we can compute the values of the elliptic functions at $z_1 + z_2$ if we know them at z_1 and

z_2. If this could be done, we can use the map

$$z + L \longmapsto [(\wp(z), \wp'(z), 1)]$$

to transfer the additive structure on $\mathbf{C}/L$ to the cubic curve. It will turn out that this structure can be given a completely geometric interpretation that makes sense on any nonsingular cubic curve.

The idea behind the problem of determining the coordinates of $z_1 + z_2$ is very simple. If $z_3 = -(z_1 + z_2)$, then $z_1 + z_2 + z_3 = 0$. But from the properties of elliptic functions we know that if (a_i) and (b_i) are the zeros and poles of an elliptic function with period lattice L, then

$$\sum_i a_i - \sum_i b_i = 0 \text{ in } \mathbf{C}/L$$

In particular,

$$z_1 + z_2 + z_3 = 0 \text{ in } \mathbf{C}/L$$

if the z_i are the zeros of an elliptic function with a triple pole at 0 and no other singularity. In other words, the group structure on $\mathbf{C}/L$, or, what comes to the same thing, the determination of the triples (z_1, z_2, z_3) such that $z_1 + z_2 + z_3 = 0$, is given by the assertion that they are the zeros of an elliptic function of order 3 with a triple pole at 0 and no other singularity.

Let $z_1, z_2, z_1 \pm z_2$ be all nonzero. Let $f = \wp' - (a\wp + b)$ where a and b are constants. Then f has a triple pole at 0 and no other singularity. So its zeros add up to 0. We choose the constants a and b such that $f(z_1) = f(z_2) = 0$. This gives

$$a = \frac{\wp'(z_1) - \wp'(z_2)}{\wp(z_1) - \wp(z_2)}, \qquad b = \frac{\wp'(z_2)\wp(z_1) - \wp'(z_1)\wp(z_2)}{\wp(z_1) - \wp(z_2)}$$

Notice that $\wp(z_1) \neq \wp(z_2)$ because, otherwise, that common value will be taken at the 4 distinct points $\pm z_1, \pm z_2$. Then z_3 is the third zero of f. To determine $\wp(z_3)$ we consider

$$g(X) = 4X^3 - g_2X - g_3 - (aX + b)^2$$

Since the z_i satisfy $\wp'(z_i)^2 = (a\wp(z_i) + b)^2$, it follows that

$$g(\wp(z_i)) = 0 \qquad (i = 1, 2, 3)$$

So $\wp(z_i)(i=1,2,3)$ are the three roots of g. Therefore

$$\wp(z_1)+\wp(z_2)+\wp(z_3)=\frac{a^2}{4}$$

Substituting for a we therefore obtain, since $\wp(z_1+z_2)=\wp(-z_3)=\wp(z_3)$,

$$\wp(z_1+z_2)=-\wp(z_1)-\wp(z_2)+\frac{1}{4}\Big(\frac{\wp'(z_1)-\wp'(z_2)}{\wp(z_1)-\wp(z_2)}\Big)^2 \qquad (1)$$

which is the *addition theorem for the $\wp$–function.* This has been derived under the assumption that $z_1, z_2, z_1 \pm z_2$ are all nonzero. If we now let z_1 approach z_2 we get

$$\wp(2z)=-2\wp(z)+\frac{1}{4}\Big(\frac{\wp''(z)}{\wp'(z)}\Big)^2 \qquad (2) \qquad (2z \notin L)$$

Let us go over to the cubic curve corresponding to $\mathbf{C}/L$ and denote by P_i the points on the curve corresponding to z_i. The discussion shows that the line $Y=aX+b$ contains the points $(\wp(z_i),\wp'(z_i),1)(i=1,2)$ and so is the line P_1P_2. Moreover, the point $(\wp(z_3),\wp'(z_3),1)$ is the third point of this line on the cubic curve. Thus this third point is P_3. Since

$$(\wp(z_1+z_2),\wp'(z_1+z_2),1)=(\wp(-z_3),\wp'(-z_3),1)=(\wp(z_3),-\wp'(z_3),1)$$

it is clear that the point P_1+P_2 on the curve corresponding to z_1+z_2 is obtained by reflecting P_3 in the X–axis (see figure 1).

This construction of the addition of points is valid for any nonsingular projective cubic curve, and leads to a group structure for the points on the curve. The group structure has the following elegant description: three points $P_1, P_2.P_3$ satisfy $P_1+P_2+P_3=0$ if and only if they are collinear. See figure 1 where P_0 is the identity element.

3. Reduction of nonsingular cubic to Weierstrassian normal form. The main result is that if a nonsingular cubic is given in $\mathbf{CP}^2$ we can choose linear coordinates so that it appears in the form

$$Y^2W=4X^3-a_2XW^2-a_3W^3$$

where

$$a_2^3-27a_3^2\neq 0$$

The proof depends on the fact that a nonsingular cubic has at least one flex or inflection point.

The definition of a flex is simple. Let C be an algebraic curve in $\mathbf{CP}^2$ and let $P = (p_i)$ be a nonsingular point of C. Let the curve C be given by an equation

$$F(X_1, X_2, X_3) = 0$$

where F is a homogeneous polynomial in X, X_2, X_3. Let

$$F_i = \partial F/\partial X_i, \qquad F_{ij} = \partial^2 F/\partial X_i \partial X_j$$

Since P is nonsingular, at least one of the numbers $F_i(P)$ is nonzero and so the equation

$$\sum_i X_i \partial F/\partial X_i(P) = 0$$

defines a line. We shall see presently that this is the tangent line to C at P. If $A = (a_i)$ is a point, the points of the line joining P and A are of the form $P + tA$ (except for A) and to find which of these are on the curve we have to find the values t such that

$$F(P + tA) = 0$$

By expanding F as a Taylor series we get

$$F(P) + t\sum_i a_i F_i(P) + \frac{t^2}{2!}\sum_{ij} a_i a_j F_{ij}(P) + \ldots = 0$$

Since $F(P) = 0$ as P is on C, this equation, which is a polynomial equation in t, has $t = 0$ as a root. The line PA is said to be *tangent to C at P* if $t = 0$ is a *double root.* The condition for this is that

$$\sum_i a_i F_i(P) = 0$$

which was introduced above. The point P is called a *flex* or an *inflection point* if (1) is satisfied and $t = 0$ is a root of multiplicity at least 3. The condition for this is that

$$\sum_i a_i F_i(P) = 0 \Longrightarrow \sum_{ij} a_i a_j F_{ij}(P) = 0 \qquad (*)$$

Proposition 1. *Let H be the 3×3 matrix of the second partial derivatives of F, namely, $H = (F_{ij})$. Then P is a flex if and only if* $\det H(P) = 0$.

Proof. Let ℓ be the linear function $\sum_i x_i F_i(P)$ and q the quadratic function $\sum_{ij} x_i x_j F_{ij}(P)$. In suitable linear coordinates (z_i) we can take ℓ as $z_1 = 0$; then it is clear that if $(*)$ holds there is no term in the expression for q that does not contain z_1, i.e., z_1 divides q. It is then easy to verify that the determinant of the matrix of q is 0. This must be true in the original coordinate system because vanishing of the determinant of H is an invariant condition, as may be easily seen. Conversely, suppose that $\det H(P) = 0$. Then we can find a nonzero vector (a_i) such that

$$\sum_j F_{ij}(P)a_j = 0 \qquad (i = 1, 2, 3) \tag{$\dagger$}$$

We now use Euler's theorem which states that if f is a homogeneous polynomial in n variables z_i of degree N, then

$$\sum_i z_i f_i = Nf$$

(This is proved by differentiating the relation $f(tX) = t^N f(X)$ with respect to t and taking $t = 1$.) We note first that $P \neq A$. In fact, if $P = A$, then, applying Euler's theorem to F_i we get

$$F_i(P) = \sum_j F_{ij}(P)p_j = 0 \qquad (i = 1, 2, 3)$$

which is impossible. Further, multiplying $(\dagger)$ by p_i and adding we get, using Euler's Theorem again,

$$\sum_j a_j F_j(P) = 0$$

Thus the line joining P and the point A with coordinates (a_i) is tangent to C. But then

$$\sum_{ij} a_i a_j F_{ij}(P) = \sum_i a_i (\sum_j a_j F_{ij}(P)) = 0$$

showing that P is a flex.

Remark. Let the affine curve have the equation $Y = f(X)$ where f is a polynomial of degree ≥ 3, and let P be $(0,0)$; let the tangent line to the curve at P be $Y = 0$. Then the above criterion for P to be a flex reduces to $f''(0) = 0$, which is the classical elementary criterion.

The flexes of the curve C are thus the common points of the curves with equations

$$F = 0, \qquad \det H = 0$$

It is a general fact that any two curves have a common point in $\mathbf{CP}^2$ so that the existence of flexes is an immediate corollary. The existence of common points follows from the following classical result (which we shall not prove).

Proposition 2. *Let k be a field and $a_0, b_0 \in k$ be two nonzero elements. Then the polynomials*

$$f(X) = a_0X^m + a_1X^{m-1} + \ldots + a_m, \qquad g(X) = b_0X^n + b_1X^{n-1} + \ldots + b_n$$

have a common nonconstant factor if and only if

$$R(a_0, a_1, \ldots, a_m, b_0, b_1, \ldots, b_n) \neq 0$$

where R is a polynomial with integer coefficients.

Returning to our two curves in $\mathbf{CP}^2$ let us choose coordinates so that the point $(0,0,1)$ is on neither of the curves. If the curves are of degrees m and n, their equations can be written as

$$a_0X_3^m + \ldots + a_m = 0, \qquad b_0X_3^n + \ldots + b_n = 0$$

where a_i (resp. b_j) is a homogeneous polynomial in X_1, X_2 of degree $m-i$ (resp. $n-j$). In particular a_0 and b_0 are *constants* while the other a_i, b_j are of positive degrees; and, as $(0,0,1)$ is on neither of the curves, a_0, b_0 are nonzero. Computing the polynomial R and taking a point (x_1, x_2) at which R vanishes, we see that there is some x_3 such that (x_1, x_2, x_3) is a common point to the two curves. Note that $(x_1, x_2) \neq (0,0)$ as otherwise $(0,0,1)$ will be on both curves. This is only a sketch of a proof and one should refer to a book on algebra for a more detailed discussion.

We shall now prove that any nonsingular cubic curve C in $\mathbf{CP}^2$ can be put in the Weierstrass form in a suitable coordinate system. We know

that C has a flex P. We take P to be $(0,1,0)$ and the tangent to C at P as the line $W=0$. Since P is on C the equation to C contains no Y^3 term. Let $F=0$ be the equation. Since the tangent at P is $X_3=0$, we must have

$$F_1(P)=F_2(P)=0, F_3(P)\neq 0$$

The last condition means that the coefficient of Y^2W in F is nonzero, while the first condition means that the coefficient of XY^2 is 0 and that of X^3 is not zero. The second condition is automatic since there is no Y^3 term. We now write the condition that the line $W=0$ meets the curve in 3 points that are coincident. The condition for this is that the coefficient of X^3 is nonzero and that the coefficient of X^2Y is zero. So the curve has the equation

$$Y^2W-b_5XYW-b_4YW^2=b_0X^3+b_1X^2W+b_2XW^2+b_3W^3 \qquad (b_0\neq 0)$$

which can be written as

$$Y'^2W=b_0X^3+b_1'X^2W+b_2'XW^2+b_3'W^3$$

where

$$Y'=Y-(1/2)b_5X-(1/2)b_4W$$

If we replace X by $X'=pX+q$ we can, since $b_0\neq 0$, choose p and q so that the coefficient of X'^3 is 4 and the coefficient of X'^2W is 0. The resulting equation is in the Weierstrass form. We have thus proved the following.

Theorem 3. *Given a nonsingular cubic curve in* $\mathbf{CP}^2$ *we can choose coordinates such that the curve has the equation*

$$Y^2W-4X^3-a_2XW^2-a_3W^3$$

Then

$$\Delta=a_2^3-27a_3^2\neq 0$$

Moreover the point at infinity $(0,1,0)$ *is a flex, and the tangent to the curve at this point is* $W=0$*, the line at infinity.*

4. Flexes on nonsingular cubic curves. Let C be a nonsingular cubic curve in $\mathbf{CP}^2$. Our aim is to prove the following theorem.

Theorem 4. *C has exactly 9 flexes, all distinct. The line joining any two of them contains a unique additional flex. The flexes lie, 3 by 3, on exactly 12 lines.*

Proof. We take C in the Wierstrass normal form with one flex at $(0,1,0)$ and equation

$$Y^2 - (4X^3 - a_2X - a_3) = 0 \qquad (\Delta = a_2^3 - 27a_3^2 \neq 0) \tag{1}$$

in the affine XY–plane. The line at infinity is $W = 0$ and it contains only one flex, namely $P = (0,1,0)$.

We shall show that among the lines through P distinct from the line at infinity there are exactly 4 which contain flexes of C, and on each of these there are exactly 2 flexes other than P. Clearly we can work over the XY–plane to prove this. The lines through P other than the line at infinity ($W = 0$) have equations $X = c$ where c is a constant. If

$$F = Y^2W - (4X^3 - a_2XW^2 - a_3W^3) = 0 \tag{2}$$

is the homogeneous equation of the curve C, the flexes are the points of intersection of C with the curve $\det H = 0$ where H is the matrix

$$\begin{pmatrix} -24X & 0 & 2a_2W \\ 0 & 2W & 2Y \\ 2a_2W & 2Y & 2a_2X + 6a_3W \end{pmatrix}$$

Hence the flexes are the points common to C and the curve with equation

$$-12X(a_2XW + 3a_3W^2 - Y^2)a_2^2W^3 = 0$$

Putting $W = 1$ and taking the line $X = c$ there are flexes on this line in the affine plane only for those values of c such that the equations

$$Y^2 - (4c^3 - a_2c - a_3) = 0, \qquad -12a_2c^2 - 36a_3c + 12cY^2 - a_2^2 = 0$$

have a common solution Y. Substituting for Y^2 from the first of these equations eliminates Y and we get the equation to be satisfied by c, namely

$$f(c) = 4c^4 - 2a_2c^2 - 4a_3c - \frac{a_2^2}{12} = 0 \tag{3}$$

We must prove two things: first, this equation has 4 distinct roots, and second, for each root c, the expression $4c^3 - a_2c - a_3$ does not vanish so that the equation $Y^2 = 4c^3 - a_2c - a_3$ has 2 distinct solutions. But

$$f'(c) = 4(4c^3 - a_2c - a_3)$$

so that our equations become

$$Y^2 = f'(c), \qquad f(c) = 0$$

For a simple root c of $f(c) = 0$ one has automatically $f'(c) \neq 0$, so that the equation $Y^2 = f'(c)$ has 2 distinct solutions. Thus it is only a question of showing that f and f' have no common roots. This is equivalent to showing that the g.c.d of

$$f = 4X^4 - 2a_2X^2 - 4a_3X - \frac{a_2^2}{12} \text{ and } g = 4X^3 - a_2X - a_3$$

in the ring $\mathbf{C}[X]$ is a nonzero constant. The euclidean algorithm leads to the following formulae:

$$\begin{aligned}
f &= Xg + f_1, \qquad f_1 = -a_2X^2 - 3a_3X - \frac{a_2^2}{12} \\
a_2g &= -4Xf_1 + f_2, \qquad f_2 = -12a_3X^2 - \frac{4}{3}a_2^2X - a_2a_3 \\
12a_3f_1 &= a_2f_2 + f_3, \qquad f_3 = \frac{4}{3}\Delta X \qquad (\Delta = a_2^3 - 27a_3^2) \\
f_2 &= -\Delta^{-1}(9a_3X + a_2^2)f_3 - a_2a_3
\end{aligned}$$

If $a_2a_3 \neq 0$ then this shows that the g.c.d of f and g is a nonzero constant and we are done. Suppose that $a_2 = 0$. Then as $\Delta \neq 0$ we must have $a_3 \neq 0$. The algorithm now reduces to

$$\begin{aligned}
f &= Xg + f_1, \qquad f_1 = -3a_3X \\
g &= \Big(-\frac{4}{3a_3}X^2\Big)f_1 - a_3
\end{aligned}$$

so that the g.c.d is again a nonzero constant. If $a_3 = 0$ then $a_2 \neq 0$ and the algorithm becomes

$$\begin{aligned}
f &= Xg + f_1, \qquad f_1 = -a_2X^2 - \frac{a_2^2}{12} \\
a_2g &= -4Xf_1 + f_2, \qquad f_2 = -\frac{4}{3}a_2^2X \\
f_1 &= \Big(\frac{3}{4a_2}X\Big)f_2 - \frac{a_2^2}{12}
\end{aligned}$$

and again the g.c.d is a nonzero constant. We have thus proved our claim.

This already shows that apart from P there are 8 distinct flexes and so there are 9 altogether. These lie on 4 lines through P, each line containing 3 of the flexes if we include P also. But the choice of P was completely arbitrary. So, through *any* flex Q, there are exactly 4 lines that contain the flexes and there are exactly 3 flexes on each of these four lines including Q. There are 36 pairs of points among the 9 and on the line containing such a pair there is a third flex. Hence each of the 36 lines is counted 3 times so that there are 12 lines in all. This proves the remaining statements of the theorem.

Let k be a field which we assume to be of characteristic 0 for simplicity, and let $X(k) = \mathbf{P}^2(k)$ be the projective plane over k. We consider a configurations Γ of lines and points in $X(k)$. We denote it by

$$[p, \pi; \ell, \lambda]$$

if it consists of p points and π lines such that on every line of the configuration Γ there are ℓ points of Γ and through every point of Γ there are λ lines. The flexes of a nonsingular cubic curve in the complex projective plane $X(\mathbf{C})$ is a configuration

$$[9, 12; 3, 4]$$

Since $X(k)$ can be viewed as the set of lines (=one dimensional linear subspaces) of k^3, the projective group $PGL(3, k) = GL(3, k)/k^\times$) acts on $X(k)$ and hence on the set of configurations of a given type in an obvious manner. The following theorem places the above results on the flexes in a better perspective.

Theorem 2. *In order that $X(k)$ contain a configuration of type $[9, 12; 3, 4]$ it is necessary and sufficient that k contains the cube roots of 1. In this case all such configurations are mutually conjugate under $PGL(3, k)$. Moreover we can choose coordinates such that the 9 points have the following coordinates:*

$$\begin{pmatrix} (0,1,-1) & (0,1,-\omega) & (0,1,-\omega^2) \\ (1,0,-1) & (-\omega,0,1) & (-\omega^2,0,1) \\ (1,-1,0) & (1,-\omega,0) & (1,-\omega^2,0) \end{pmatrix} \tag{4}$$

Proof. Let Γ be a configuration of type $[9, 12; 3, 4]$ in $X(k)$. Let ξ_1 be a line of Γ and let $P1, Q_1, R_1$ be the three points of Γ on ξ_1. Through

each of these points there are 3 lines other than ξ_1, and so there are 2 additonal lines, ξ_2, ξ_3. Notice that we have accounted for the 12 lines of Γ. Let P_j, Q_j, R_j be the points of Γ on ξ_j $(j = 2, 3)$. The line P_1P_2 has an additional point of Γ on it; this has to be one of P_3, Q_3, R_3, and we can choose the notation that it is P_3. Similarly we assume that Q_3 is the point where the lines Q_1Q_2 and ξ_3 meet, and R_3 is the point where the lines R_1R_2 and ξ_3 meet.

We can take coordinates such that ξ_i has the equation $X_i = 0$ $(i = 1, 2)$ and the line $P_1P_2P_3$ to have the equation $X_3 = 0$. Moreover we can choose the line $Q_1Q_2Q_3$ to have the equation $X_1 + X_2 + X_3 = 0$ and $R_1R_2R_2$ to have the equation $aX_1 + bX_2 + X_3 = 0$. Let the line ξ_3 have the equation $\alpha X_1 + \beta X_2 + \gamma X_3 = 0$.

We now compute the coordinates of the 9 points. We arrange the points as the matrix

$$\begin{pmatrix} P_1 & Q_1 & R_1 \\ P_2 & Q_2 & R_2 \\ P_3 & Q_3 & R_3 \end{pmatrix}$$

Their coordinates are then represented by the matrix (the correspondence is the obvious one)

$$\begin{pmatrix} (0,1,0) & (0,1,-1) & (0,1,-b) \\ (1,0,0) & (1,0,-1) & (1,0,-a) \\ (\beta,-\alpha,0) & (\beta-\gamma,\gamma-\alpha,\alpha-\beta) & (b\gamma-\beta,\alpha-a\gamma,a\beta-b\alpha) \end{pmatrix} \quad (5)$$

In addition to the 6 lines described above there are 6 other lines of the configuration, and these are

$$P_1Q_2R_3, P_1R_2Q_3, Q_1P_2R_3, Q_1R_2P_3, R_1P_2Q_3R_1Q_2P_3$$

We now write down the condition that each of these triplets of points are collinear. We get the following 6 conditions:

$$\begin{aligned} b\gamma - \beta + a\beta - b\alpha &= 0 && (a) \\ \alpha - \beta + a(\beta - \gamma) &= 0 && (b) \\ a\beta - b\alpha + \alpha - a\gamma &= 0 && (c) \\ -a\beta + \alpha &= 0 && (d) \\ \alpha - \beta + b(\gamma - \alpha) &= 0 && (e) \\ -\beta + b\alpha &= 0 && (f) \end{aligned}$$

Obviously (e)-(d)=(a). Using (d) and (f) we get, in place of (c) and (e) the equations

$$a\gamma = 2\alpha - \beta, \qquad b\gamma = 2\beta - \alpha \tag{6}$$

and then (b) reduces to the first of these two equations. We thus get the equations

$$\alpha = a\beta, \qquad \beta = b\alpha, \qquad a\gamma = 2\alpha - \beta, \qquad b\gamma = 2\beta - \alpha \tag{7}$$

We now solve for α and β from the last two of the above equations to get

$$\alpha = \frac{\gamma(2a+b)}{3} \qquad \beta = \frac{\gamma(2b+a)}{3} \tag{8}$$

If $\gamma = 0$ then the last two of the equations (7) imply that $\alpha = \beta = 0$ which is impossible as $\alpha X_1 + \beta X_2 + \gamma X_3 = 0$ is the equation to the line ξ_3. Hence $\gamma \neq 0$. The first pair of equations (7) then lead to the relations

$$a(a+2b) = 2a + b, \qquad b(2a+b) = a + 2b$$

So

$$ab(a+2b) = a + 2b$$

which implies that either $ab = 1$ or $a + 2b = 2a + b = 0$, i.e.,

$$ab = 1 \text{ or } a = b = 0$$

But $a = b = 0$ is impossible as $aX_1 + bX_2 + X_3 = 0$ is the equation to the line $R_1R_2R_3$. Hence

$$ab = 1$$

so that the equation $a(a+2b) = 2a + b$ leads to

$$a^3 - 2a^2 + 2a - 1 = (a-1)(a^2 - a + 1) = 0$$

The other equation $b(2a+b) = a + 2b$ leads to the same equation for a. But if $a = 1$ then $b = 1$ which is impossible as the lines $Q_1Q_2Q_3$ and $R_1R_2R_3$ would become identical. Hence

$$a^2 - a + 1 = 0$$

But then $-a$ is a cube root of 1. Thus the field k must contain a cube root of 1 if there is a configuration of type $[9, 12; 3, 4]$ in $X(k)$.

For the sufficiency we take $a = -\omega$ where ω is a nontrivial cube root of 1; then by the above discussion to note that once a is given, we find $b = 1/a$, and then α, β are determined in terms of a, b, and γ by (8). Since (α, β, γ) can be replaced by any nonzero multiple of it we can take $\gamma = 1$. The coordinates of the points of the configuration, given by the matrix (4), are now as in the matrix

$$\begin{pmatrix} (0,1,0) & (0,1,-1) & (0,1,\omega^2) \\ (1,0,0) & (1,0,-1) & (1,0,\omega) \\ (1-\omega^2,-1+\omega,0) & (-2-\omega^2,2+\omega,-\omega+\omega^2) & (1,1,-1) \end{pmatrix} \tag{9}$$

Since there are no arbitrary constants in these expressions it is clear that every configuration is conjugate to this particular one.

The lines $P_iQ_iR_i (i = 1, 2, 3)$ are now given by the equations

$$X_1 = 0, \qquad X_2 = 0, \qquad , (1-\omega)X_1 + (1-\omega^2)X_2 + 3X_3 = 0$$

while the lines $P_1P_2P_3, Q_1Q_2Q_3, R_1R_2R_3$ have the equations

$$X_3 = 0, \qquad X_1 + X_2 + X_3 = 0, \qquad \omega X_1 + \omega^2 X_2 - X_3 = 0$$

The other six lines are given by

$$\begin{aligned} P_1Q_2R_3 &: X_1 + X_3 = 0, & P_1R_2Q_3 &: \omega X_1 - X_3 = 0 \\ Q_1P_2R_3 &: X_2 + X_3 = 0, & Q_1R_2P_3 &: \omega X_1 - X_2 - X_3 = 0 \\ R_1P_2Q_3 &: \omega^2 X_2 - X_3 = 0, & R_1Q_2P_3 &: X_1 - \omega^2 X_2 + X_3 = 0 \end{aligned}$$

It is possible to simplify the appearance of this scheme by applying a suitable element of $PGL(3, k)$. We shall choose the new coordinates Y_i so that the lines $P_iQ_iR_i (I = 1, 2, 3)$ become $Y_i = 0$ and the line $P_1P_2P_3$ becomes $Y_1 + Y_2 + Y_3 = 0$. Then a simple calculation shows that

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1-\omega & 0 & 0 \\ 0 & 1-\omega^2 & 0 \\ -(1-\omega) & -(1-\omega^2) & -3 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

Applying this transformation the new matrix of coordinates then becomes

$$\begin{pmatrix} (0,1,-1) & (0,1,-\omega) & (0,1,-\omega^2) \\ (1,0,-1) & (-\omega,0,1) & (-\omega^2,0,1) \\ (1,-1,0) & (1,-\omega,0) & (1,-\omega^2,0) \end{pmatrix} \tag{10}$$

The lines $P_iQ_iR_i (i = 1, 2, 3)$ are now given by the equations

$$Y_1 = 0, \qquad Y_2 = 0, \qquad , Y_3 = 0 \tag{11}$$

while the lines $P_1P_2P_3, Q_1Q_2Q_3, R_1R_2R_3$ have the equations

$$Y_1 + Y_2 + Y_3 = 0, \qquad Y_1 + \omega^2 Y_2 + \omega Y_3 = 0, \qquad Y_1 + \omega Y_2 + \omega^2 Y_3 = 0 \tag{12}$$

The other six lines are given by

$$\begin{aligned} P_1Q_2R_3 &: Y_1 + \omega(Y_2 + Y_3) = 0, \quad P_1R_2Q_3 : Y_1 + \omega^2(Y_2 + Y_3) = 0 \\ Q_1P_2R_3 &: Y_1 + \omega Y_2 + Y_3 = 0, \quad Q_1R_2P_3 : \omega(Y_1 + Y_2) + Y_3 = 0 \\ R_1P_2Q_3 &: Y_1 + \omega^2 Y_2 + Y_3 = 0, \quad R_1Q_2P_3 : Y_1 + Y_2 + \omega Y_3 = 0 \end{aligned} \tag{13}$$

It is interesting to compute the equation to a cubic that contains the 9 flexes given by the coordinate scheme (10). A simple calculation shows that these form an 1–parameter family with equations

$$Y_1^3 + Y_2^3 + Y_2^3 - \lambda Y_1 Y_2 Y_3 = 0$$

The parameter λ can be given a geometric interpretation.

Remark. The condition that k should contain the cube roots of 1 means that there is no configuration of type $[9, 12; 3, 4]$ in the *real projective plane.* This is why we cannot draw the configuration on paper or blackboard!

8. ELLIPTIC INTEGRALS

1. Fagnano, Euler, and Gauss. The lemniscatic functions of Gauss. The first appearance of elliptic integrals goes back to the work of the Italian count Fagnano in the years 1714– 1718 (indeed he was the first to coin the term elliptic integral). His work contained a remarkable formula for doubling the arc of the *lemniscate* (see below for the definition and elementary properties of the lemniscate). Fagnano worked with the lemniscatic arc length, namely the integral (see below)

$$\int_0^x \frac{dr}{\sqrt{(1-r^4)}}$$

and his result was that

$$2\int_0^x \frac{dr}{\sqrt{(1-r^4)}} = \int_0^w \frac{dr}{\sqrt{(1-r^4)}}$$

where

$$w = \frac{2u\sqrt{(1-u^4)}}{(1+u^4)}$$

Around 1750 he published a book containing his discoveries and sent it to the Berlin Academy for review where it came to the attention of Euler. Euler's imagination was fired by Fagnano's work and he obtained a series of results, including a remarkable generalization of Fagnano's result on the doubling of the lemniscatic arc length. Euler considered integrals of the form

$$\Phi(x) = \int_0^x \frac{du}{\sqrt{P(u)}}$$

where P is a polynomial of degree ≤ 4 and obtained for them a remarkable formula showing that

$$\Phi(x) + \Phi(y) = \Phi(w)$$

where $w = w(x, y)$ is *an algebraic function of* x *and* y. For instance, if $P(u) = 1 - u^2$, then $\Phi(x) = \arcsin(x)$, so that

$$\Phi(x) + \Phi(y) = \Phi(w) \qquad (\mathbf{A})$$

where

$$w = x\sqrt{1-y^2} + y\sqrt{1-x^2}$$

More generally, let

$$P(u) = 1 + mu^2 + nu^4 \qquad (n \neq 0)$$

Then Euler proved that (**A**) is true for

$$w = \frac{x\sqrt{1+my^2+ny^4} + y\sqrt{1+mx^2+nx^4}}{1-nx^2y^2}$$

For $m = 0, n = -1, y = x$ this reduces to the Fagnao formula for doubling the lemniscatic arc. The formula (**A**) is known as *Euler's addition theorem.*

The analogy with the circular arcs suggest that it is the function *inverse* to the arc length function that should be expected to have nice properties. This is in fact so as we shall see later. In any case the next step in this story was taken by Gauss. Gauss pursued the investigation of the lemniscate at the analytical level in contrast to the formal level of the treatment of Euler. Gauss defined the lemniscatic sine function as the inverse function to the arc length function of the lemniscate. He then porceeded to take the remarkable step of extending the definition of this function to the *complex plane* and made the momentous discovery that it had —it two periods, namely 2ω and $\omega + i\omega$ where 2ω is the total length of the lemniscate. This is the original elliptic function. To motivate Gauss's work and also to give more details to the discussion above let us begin with the trigomometric and exponential functions.

Consider the circle in the plane of diameter 1 which is standing on the x–axis touching it at the origin. Let A be the point $(0, 1)$. If P is a variable point on the semicircle to the right of A such that the arc OP subtends an angle $2t$ at the center, the arc OP has length t while the chord OP has length $\sin t$ (figure 1). The angle between the x–axis and OP is t. The equation of the circle in polar coordinates is

$$r = \sin\theta$$

Since

$$dx^2 + dy^2 = r^2 d\theta^2 + dr^2, \qquad \frac{dr}{d\theta} = \sqrt{1-r^2}$$

the arc length is

$$\int \sqrt{1+(d\theta/dr)^2}dr$$

we get

$$\int_0^{\sin t} \frac{dr}{\sqrt{1-r^2}}dr = t$$

More precisely we consider the function

$$\int_0^u \frac{dr}{\sqrt{1-r^2}}dr = f(u)$$

and note that it increases from $-\pi/2$ to $\pi/2$ as u increases from -1 to 1. The inverse function is $\sin t$ which increases from -1 to 1 as t increases from $-\pi/2$ to $\pi/2$. The function is extended to all of $\mathbf{R}$ by $\sin(t) = \sin(\pi - t)$ and by requiring that it be odd and periodic of period 2π. This is the ancient way of treating the sin; the other trigonometric functions are obtained by other processes, algebraic and infinitesimal.

For the hyperbolic functions an analogous method applies although one uses areas rather than arclengths. Take a hyperbola with equation

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

and take a variable point P with coordinates $(\cosh t, \sinh t)$. If $A = (1, 0)$, then a simple calculation shows that (figure 1)

$$2 \times \text{area of region } OAP = \int_0^{\sinh t} \frac{dy}{\sqrt{1+y^2}} = t$$

The function *sinh* is thus the function inverse to the function

$$f(u) = \int_0^u \frac{dy}{\sqrt{1+y^2}}$$

i.e.,

$$\int_0^{\sinh t} \frac{du}{\sqrt{1+u^2}} = t$$

Since the integral over the infinite range diverges, f increases from $-\infty$ to ∞ as u increases from $-\infty$ to ∞. So $\sinh t$ increases from $-\infty$ to ∞

as t increases from ∞ to ∞. As usual the other hyperbolic functions arte obtained by algebraic and infinitesimal processes.

Clearly there is no *real* period for $\sinh t$. However a formal change of variables from u to $iu (i^2 = -1)$ shows that

$$\int_0^{i \sinh t} \frac{du}{\sqrt{1-u^2}} = it$$

and suggests that $\sinh t$ has an *imaginary period* $2\pi i$.

Sometime before 1799 Gauss started to think along these lines for the functions that can be defined by the rectification of the *lemniscate.* The lemniscate is the locus of a point which moves so that the product of its distances from two fixed points remains constant. If the points are taken as $(\pm 1/\sqrt{2}, 0)$ and we take the constant to be $1/2$, the origin is on the locus and the locus has the equation

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

The graph looks like a figure 8 lying on its side and looks like a ribbon, hence its name, derived from the Latin expression *lemniscatus* meaning ribboned. If we rotate the lemniscate by 45° we get the equation in polar coordinates (see figure 2)

$$r^2 = \sin 2\theta$$

In this case

$$1 + r^2 (d\theta/dr)^2 = \frac{1}{1-r^4}$$

so that the arc length is

$$\int_0^y \frac{dr}{\sqrt{1-r^4}}$$

Let the total length of the lemniscate in the first quadrant be ω. In analogy with the case of the sin Gauss defined the *lemniscate sin* as the function defined by inverting the arc length function, and wrote it as *sin lemn*; we shall denote it by sl so that

$$\int_0^{\mathrm{sl}\ t} \frac{dr}{\sqrt{(1-r^4)}} = t \qquad (0 \le t \le \omega/2)$$

Exactly as in the case of the sin we extend the function to $[0, 2\omega]$ by

$$sl\ (t) = sl\ (\omega - t)\ (\omega/2 \le t \le \omega), \qquad sl\ (t) = -sl\ (t - \omega)\ (\omega \le t \le 2\omega)$$

The function sl is then extended to all of $\mathbf{R}$ by making it periodic of period 2ω. The value of ω is given by

$$\frac{\omega}{2} = \int_0^1 \frac{dr}{\sqrt{1-r^4}}$$

The lemniscatic cosine function cl can be defined by analogy with cos by

$$cl\ t = sl\ (\frac{\omega}{2} - t), \qquad \int_{cl\ t}^1 \frac{dr}{\sqrt{1-r^4}} = \frac{\omega}{2} - t$$

Sometime around 1799 Gauss discovered that the function sl t can be extended to the *complex plane* and that it had *two* periods, one real, namely 2ω, and the other $\omega + i\omega$. The functions sl, cl are the original elliptic functions. Gauss went far in their development, and obtained results about double periodicity, division points, representation of the functions as ratios of entire power seires (which later would be independently discovered by Jacobi as theta functions), and so on.

2. Elliptic integrals of Jacobi and Legendre. Gauss did not publish any of his significant results during his lifetime. After Gauss, Abel, Jacobi, and Legendre were the classical masters of the theory of elliptic integrals. Legendre in particular spent a huge part of his scientific life working on the theory of elliptic functions and integrals and wrote a great treatise on the subject. In Legendre's notation the elliptic integrals are the indefinite itegrals of the form

$$\int \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}}$$

where k is known as the *modulus.* In most of the formulae k is real and lies between 0 and 1, but the general theory requires the treatment of these functions for complex k such that $k^2 \neq 0, 1$. In fact, the lemniscatic integral of Gauss is obtained for the value $k^2 = -1$. Actually these are know as the *elliptic integrals of the first kind;* the general elliptic integral is an expression of the form

$$\int R(u, y)dx$$

where R is a rational function of u and y and y is the quadratic irrationality

$$y = \sqrt{P(u)}$$

where P is a polynomial of degree ≤ 4. The case treated by Jacobi and Legendre is when

$$P(u) = \sqrt{(1-u^2)(1-k^2u^2)}$$

Legendre classified the elliptic integrals into three kinds, showed how their evaluations can be reduced to the evaluation of three basic types. The definite integrals

$$K(k) = \int_0^1 \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}}, \qquad E(k) = \int_0^1 \sqrt{\frac{1-k^2u^2}{1-u^2}}du$$

are the *complete ellitptic integrals of the first and second kind.* Legendre's monumental work contains many remarkable properties of these.

The reason for calling these integrals and the functions obtained from them elliptic is that they arise in the problem of the *rectification of an ellipse.* Let us take the equation of the ellipse in the form

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

where

$$a > b > 0, \qquad b^2 = a^2(1-e^2) \qquad (e = \text{ eccentricity of the ellipse })$$

Let $B = (0, b)$ and $A = (a, 0)$ be the points of the ellipse in the first quadrant (see figure 3). Then the length of the arc of the ellipse in the first quadrant from B to the point $S = (s, t)$ is gien by the integral

$$\int_0^s \sqrt{1+y'^2}dx \qquad (y' = dy/dx)$$

Now

$$y = b\sqrt{1-(x^2/a^2)}$$

and so the integral becomes, after some simplification,

$$a\int_0^{s/a} \sqrt{\frac{1-e^2x^2}{1-x^2}}dx$$

which is an elliptic integral according to our definition. In particular, the total length of the ellipse is

$$L = 4a \times E(e)$$

where

$$E(k) = \int_0^1 \sqrt{\frac{1-k^2x^2}{1-x^2}}dx = \int_0^1 \frac{(1-k^2x^2)}{\sqrt{(1-k^2x^2)(1-x^2)}}dx$$

The problem of determining the motion of a *simple pendulum* also leads to elliptic integrals. Let θ be the deviation from the vertical of the pendulum (see figure 3). Then the equation of motion for θ is

$$\ddot{\theta} + \frac{g}{\ell}\sin\theta = 0$$

where ℓ is the length of the pendulum. For small oscillations, one usually replaces $\sin\theta$ by θ to obtain the equation of the *linear pendulum:*

$$\ddot{\theta} + \frac{g}{\ell}\theta = 0$$

The solution for the linear pendulum is

$$\theta = A\cos\theta\sqrt{\frac{g}{\ell}} + B\sin\theta\sqrt{\frac{g}{\ell}}$$

where A and B are determined by the initial conditions. Clearly, the solution is periodic with period

$$T_{\text{lin}} = 2\pi\sqrt{\frac{\ell}{g}}$$

and it is *independent of the initial conditions.* The integration of the non-linear pendulum equation leads to elliptic integrals in a very well known manner. Multiplying the equation of motion by $\dot{\theta}$ one gets

$$\frac{d}{dt}\left(\dot{\theta}^2 - 2\frac{g}{\ell}\cos\theta\right) = 0$$

If the vertical displacement at time $t = 0$ is θ_0, then, as $\dot{\theta}(0) = 0$, we get

$$\dot{\theta} = \pm\sqrt{\frac{2g}{\ell}}\,(\cos\theta - \cos\theta_0) \qquad (|\theta| \le \theta_0)$$

The condition on θ arises fromthe fact that $\cos\theta - \cos\theta_0$ has to be ≥ 0 during motion. The sign of $\dot{\theta}$ changes every time $\dot{\theta}$ becomes 0. The usual way to integrate this is to write it as

$$\pm\frac{\dot{\theta}}{\sqrt{\cos\theta - \cos\theta_0}} = \sqrt{\frac{2g}{\ell}}dt$$

to get t as a function of θ:

$$t = \sqrt{\frac{\ell}{2g}}\int_{\theta}^{\theta_0}\frac{d\theta}{\sqrt{\cos\theta - \cos\theta_0}} \qquad (0 \leq \theta \leq \theta_0)$$

We now make the trigonometric substitution $\sin(\psi/2) = u\sin(\theta_0)/2$ to obtain

$$t = \sqrt{\frac{\ell}{g}}\int_{u}^{1}\frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}} \qquad (k = \sin(\theta_0/2))$$

which is an elliptic integral. The period T is then given by

$$T = 4 \times \sqrt{\frac{\ell}{g}}\int_{0}^{1}\frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}} \qquad (k = \sin(\theta_0/2))$$

Thus

$$T = T_{\text{lin}}(2/\pi)K(k) \qquad (k = \sin(\theta_0/2))$$

where

$$K(k) = \frac{\pi}{2}\left(1 + (1/4)k^2 + \ldots\right)$$

The interesting thing is that the period *depends on the intial conditions.* To a first approximation,

$$T = T_{\text{lin}}\left(1 + (1/4)k^2 + O(k^4)\right)$$

The period of the nonlinear pendulum is thus *slightly larger than* the period of the linearized version.

3. The Jacobian elliptic functions. The integrand

$$\frac{1}{\sqrt{(1-u^2)(1-k^2u^2)}}$$

appearing in the definition of elliptic integral is a special case of a class of functions that occur in the theory of conformal mappings, namely the *Schwarz–Christoffel functions.* These are the functions which give the conformal mapping of the disk or the upper half plane onto the interior of a closed polygon. The Jacobian function is obtained when the closed polygon is a rectangle. The inverse map, from the rectangle to the upper half plane, can be extended meromorphically to the whole plane by repeated reflection in the sides of the rectangle, and leads to a doubly periodic meromorphic function whose periods, when the sides of the rectangle are parallel to the coordinate axes, are $4K$ and $2iK'$ where $2K, K'$ are the sides of the original rectangle. This is the Jacobian elliptic function

$$\mathrm{sn}(z,k)$$

nd the entire theory of elliptic functions can be erected on it. This was what Jacobi did in the 1820's. Weierstrass's lectures which came later in the 1890's, presented the theory through the $\wp$–function, which is the modern way of introducing the student to the theory of elliptic functions and curves.

It must be pointed out that the concept of analytic continuation, which is fundamental to the problem of extending these functions beyond their original domains of definition, was not available before Riemann and Weierstrass. Abel and Jacobi circumvented it by clever arguments. For the Jacobian functions let

$$\mathrm{cn}(z,k) = \sqrt{1-\mathrm{sn}(z,k)^2}$$

Then one can prove by a formal argument that

$$\mathrm{sn}(it,k) = i\frac{\mathrm{sn}(t,k')}{\mathrm{cn}(t,k')} \qquad (k' = \sqrt{1-k^2})$$

So one can extend the definition of sn to the imaginary axis; one can then extend it to all of $\mathbf{C}$ using Euler's addition theorem to define $\mathrm{sn}(s+it)$. It was only after Riemann and Weierstrass that the definition of the Jacobian elliptic functions on all of $\mathbf{C}$ was given a secure foundation.

An analytic function mapping a domain (open connected set) U into a domain U' of $\mathbf{C}$ is called *conformal* if $f'(z) \neq 0$ for all $z \in U$. This is the same as saying that f is locally one–one. If f is globally one–one (*univalent* in older terminology) it is a *conformal equivalence* or a *conformal*

isomorphism of U with $f(U')$ (recall that any analytic function is an *open mapping* on its domain so that $f(U')$ is open and connected). For any domain U we write Aut (U) for the group of all conformal isomorphisms of U onto itself. The group Aut (U) is known for the classical domains and is described as follows. Here $S^2 = \mathbf{C} \cup \{\infty\}$ is the extended complex plane, D is the unit disk, and $\mathcal{H}$ is the upper half plane.

(i) Aut (S^2)=$SL(2, \mathbf{C})/\{\pm 1\}$; the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ acts by $z \longmapsto \frac{az+b}{cz+d}$. Here and in what follows we refer to the group of matrices as the automorphism group when we actually mean its quotient by $\{\pm 1\}$.

(ii) Aut $(\mathbf{C})$=the subgroup of $SL(2, \mathbf{C})$ of matrices of the form $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ and so consists of the affine maps $z \longmapsto az + b$.

(iii) Aut (D)=the subgroup $SU(1, 1)$ of $SL(2, \mathbf{C})$, namely, the group of all transformations of the form $z \longmapsto \frac{\alpha z+\beta}{\bar{\beta} z+\bar{\alpha}}$, with $\alpha\bar{\alpha} - \beta\bar{\beta} = 1$.

(iv) Aut $(\mathcal{H})$=the subgroup $SL(2, \mathbf{R})$ of $SL(2, \mathbf{C})$, namely the matrices with *real* entries.

The basic theorem is the Riemann mapping theorem for simply connected domains of the complex plane. Let us recall this theorem.

Theorem (Riemann mapping theorem). *Let U be a simply connected domain in $\mathbf{C}$, with $U \neq \mathbf{C}$. Then U is conformally isomorphic to the disk D. More precisely, given $z_0 \in U$, there is a unique conformal isomorphism F of U with D,*

$$F \simeq D$$

such that

(a) $F(z_0) = 0$

(b) $F'(z_0) > 0$

Moreover $D \not\simeq \mathbf{C}$.

For any simply connected domain in $\mathbf{C}$ which is not $\mathbf{C}$, the functions from D to U that define conformal isomorphisms are called *mapping functions.*

In constructing or identifying mapping functions it is convenient to use the following principles. These are proved in the theory of conformal mappings. We recall that a *Jordan domain* is a bounded domain whose boundary is a simple closed curve (a Jordan curve).

(i) If U, U' are Jordan domains and $f(U \simeq U')$ is a conformal isomorphism, then f extends to a *homeomorphism* of cl (U) with cl (U'), cl denoting closure.

(ii) Let f be as above. If σ is a line or circular arc of ∂U, then f extends analytically across σ by reflection.

(iii) If U, U' are Jordan regions and f is a continuous map cl $(U) \longrightarrow$ cl (U') which is conformal on U,and also on ∂U except at a finite number or points, and if f is a homeomorphism of ∂U with $\partial U'$, then f is a conformal isomorphism of U with U'.

As a simple example, consider the mapping F of the strip

$$S_\ell = \{z \mid |\Re(z)| < \ell\}$$

onto the disk D. We may assume that $F(0) = 0, F'(0) > 0$. Then F extends anlytically across the bounding lines $x = \pm\ell$ by reflection. The extension has the property that two values of z that are reflections of one another on the lines $x = \pm\ell$ go over to values $w, 1/\bar{w}$ (which are symmetric with respect to the unit circle. Thus for the extension we have

$$F(\pm 2\ell) = \infty, \qquad F(\pm 2\ell - \bar{z}) = \frac{1}{\overline{F(z)}}$$

The second reflection is in the line $x = 3\ell$ and is given by $z \longmapsto 6\ell - \bar{z}$ so that the product of these two reflections is $z \longmapsto 6\ell - \overline{(2\ell - \bar{z})} = z + 4\ell$. Hence we get

$$F(4\ell + z) = F(z)$$

So F has period 4ℓ, and is meromorphic with zeros at points of $4\ell\mathbf{Z}$ and poles at $2\ell + 4\ell\mathbf{Z}$. Moreover F stays bounded as z goes to infinity on the strip. It is then not difficult to show that that

$$F(z) = \tan\frac{\pi z}{4\ell}$$

Thus the periodicity comes in from the relection principle.

Let P be a closed polygon with vertices $z_1, z_2, \ldots z_n$ and interior angles $\alpha_j\pi$ at z_j where $0 < \alpha_j < 2$; some of the α_j are allowed to be > 1 so that the polygon need not be convex. The exterior angle at z_j is $\beta_j\pi$ where $\beta_j = 1 - \alpha_j$ so that $-1 < \beta_j < 1$. We have

$$\beta_1 + \beta - 2 + \ldots + \beta_n = 2$$

For a convex polygon

$$0 < \beta_j < 1 \qquad (j = 1, 2, \ldots, n)$$

The classical result on the mapping functions from the disk to the closed polygon P with exterior angles $\beta_j \pi$ is as follows (see Ahlfors's book).

Theorem. *The conformal isomorphisms $f(D \simeq P)$ are of the form*

$$f(w) = C \int^w (w - w_1)^{-\beta_1} \ldots (w - w_n)^{-\beta_n} dw + C'$$

where $|w_j| = 1, f(w_j) = z_j$ and C, C' are nonzero complex constants.

Remark. The lower limit of integration is left open; we can take it to be any point in the interior of the disk or even on the boundary but different from the w_j. The constants C, C' depend on the location of P in the complex plane. As they vary the polygon gets moved by a rigid motion, translation, or a dilation, and so remains a closed polygon similar to itself. Thus the map

$$w \longmapsto \int^w (w - w_1)^{-\beta_1} \ldots (w - w_n)^{-\beta_n} dw$$

is a conformal isomorphism of D with a closed polygon with interior angles $\beta_j \pi$. From the main results on the boundary behaviour of conformal mappings that we reviewed above, the determination of the image polygon is done by looking at the image of f of the boundary of the disk.

When D is replaced by $\mathcal{H}$, the upper half plane, it is easy to get the corresponding result by the Cayley transform applied to the above result. There are two formulations depending on whether one of the $w_j = 1$ or not. The transformation is

$$z = i\frac{1 + w}{1 - w}, \qquad w = \frac{z - i}{z + i}$$

and the conformal isomorphism $F(\mathcal{H} \simeq P)$ is of the form

$$F(z) = \int^z (u - \xi_1)^{-\beta_1} \ldots (u - \xi_n)^{-\beta_n} du$$

where

$$\xi_j \in \mathbf{R}, -\infty < \xi_1 < \ldots < \xi_n < \infty$$

The alternative form is when one of the ξ_j is thrown to the point at infinity. Then the mapping takes the form

$$G(z) = \int^z (u-\xi_1)^{-\beta_1} \ldots (u-\xi_{n-1})^{-\beta_{n-1}} du$$

where

$$\xi_j \in \mathbf{R}, -\infty < \xi_1 < \ldots < \xi_{n-1} < \infty$$

Here we must remember that $\beta_n = 2 - (\beta_1 + \ldots + \beta_{n-1})$.

The case when $n = 4$ and $\alpha_j = 1/2$ is important for us; then P is a rectangle. Then the integrand is $\sqrt{P}$ where P is either a quartic or a cubic polynomial with distinct real roots. In the quartic case we can map the ξ_j to $-1/k, -1, 1, 1/k$ where $k > 1$ by a fractional real linear transformation by choosing k so that we have

$$[\xi_1, \xi_2, \xi_3, \xi_4] = [-1/k, -1, 1, 1/k]$$

where [, , ,] denotes the cross ratio. Then the mapping function is

$$f(z) = \int_0^z \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}} \qquad (k > 1)$$

which is the *Legendre normal form.* In the cubic case the mapping function is

$$f(z) = \int_0^z \frac{du}{\sqrt{(1-u)(1-k^2u^2)}} \qquad (k > 1)$$

which is often called the *Riemann normal form.* The inverse maps of these functions are conformal isomorphisms of rectangles with $\mathcal{H}$. We formulate the result here for the Legendre normal form.

Theorem. *Suppose P is a rectangle with vertices $p, p+\omega_1, p+\omega_1+i\omega_1, p+i\omega_1$ (in natural order and orientation). Then a conformal isomorphism $g(P \simeq \mathcal{H})$ extends to $\mathbf{C}$ as a mermomorphic doubly periodic function (also written g) with periods $2\omega_1, 2i\omega_1$. In any fundamental parallelogram the function g has 2 simple zeros and 2 simple poles. In particular, if*

$$f(z) = \int_0^z \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}} \qquad (z \in \mathcal{H})$$

then f is a conformal isomorphism of $\mathcal{H}$ with the interior of the recatngle R with vertices $-K, K, K+iK', -K+iK'$ where

$$K = K(k) = \int_0^1 \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}}$$
$$K' = K'(k) = \int_1^{1/k} \frac{du}{\sqrt{(u^2-1)(1-k^2u^2)}}$$

The inverse function g extends to $\mathbf{C}$ as doubly periodic function meromorphic function g with periods $4K, 2iK'$ whose zeros are simple and precisely at points of $L, L+2K$ and whose poles are simple and precisely at the points of $L+iK', L+iK'+2K$.

Proof. Regard g as a map into the disk. Then g extends by reflection in the sides repeatedly to all of $\mathbf{C}$, the extension also being written as g. The extension is doubly periodic, the rectangle obtained by reflection once in each pair of adjacent sides being a fundamental parallelogram. A rotation leads us to the situation where the larger rectangle has sides parallel to the coordinate axes and has vertices $0, 2a, 2a+2ib, 2ib$. The larger rectangle contains two simple zeros and two simple poles of the extension. The epriods are $2a, 2ib$ and so before the rotation they are $2\omega_1, 2i\omega_1$. In the special case we know *a priori* that the image of the map is a rectangle, and so it is a question of describing the map on the real axis. Let the branch of the square root of the integrand be the one which is 1 at the origin. This is also the branch which is real and positive on the imaginary axis in $\mathcal{H}$. We wish to determine how this branch determines uniquely the branch on $\mathbf{R} \setminus \{\pm 1, \pm 1/k\}$. A simple calculation* leads to the following;

* To see this, note that on $\mathcal{H}$ we have a unique branch of $\sqrt{\zeta}$ such that it is 1 at $\zeta = 1$; it is $re^{i\theta} \longmapsto r^{1/2}e^{i\theta/2}$ $(0 < \theta < \pi)$. This branch induces a unique branch on $\mathbf{R} \setminus (0)$ which is $[\sqrt{u}]_+$ for $u > 0$ and $i[\sqrt{u}]_+$ for $u < 0$. If we change $\mathcal{H}$ to the lower half plane in the above, the induced branch on $\mathbf{R} \setminus (0)$ remains the same for $u > 0$ and becomes $-i[\sqrt{u}]_+$ for $u < 0$. Hence, the branch induced on $\mathbf{R} \setminus (-1)$ by the unique branch of $\sqrt{1+z}$ which is 1 at $z = 0$ is $[\sqrt{1+u}]_+$ for $u > -1$ and $i[\sqrt{-(1+u)}]_+$ for $u < -1$. For $\sqrt{1-z}$ the branch induced on $\mathbf{R} \setminus (1)$ is $[\sqrt{-i(u-1)}]_+$ for $u > 1$ and $[\sqrt{(1-u)}]_+$ for $u < 1$.

we write $[\sqrt{a}]_+$ for the positive square root of a if $a > 0$.

$$\frac{1}{\sqrt{(1-u^2)(1-k^2u^2)}} = \begin{cases} \frac{1}{[\sqrt{(1-u^2)(1-k^2u^2)}]_+} & \text{if } -1 < u < 1 \\ \frac{i}{[\sqrt{(u^2-1)(1-k^2u^2)}]_+} & \text{if } 1 < u < 1/k \\ \frac{-1}{[\sqrt{(u^2-1)(k^2u^2-1)}]_+} & \text{if } 1/k < u \text{ or } u < -1/k \\ \frac{-i}{[\sqrt{(u^2-1)(1-k^2u^2)}]_+} & \text{if } -1/k < u < -1 \end{cases}$$

As u increases from 0 to 1, $f(u)$ increases from 0 to K; when u increases from 1 to $1/k$, $f(u)$ goes *up* from K to $K + iK'$; when u increases from $1/k$ to ∞, $f(u)$ turns *left* and goes from $K + iK'$ to

$$K + iK' - \int_{1/k}^{\infty} \frac{du}{\sqrt{(u^2-1)(k^2u^2-1)}} = iK'$$

in view of the relation

$$\int_{1/k}^{\infty} \frac{du}{\sqrt{(u^2-1)(k^2u^2-1)}} = K$$

which can be seen by the substitution $u = 1/kv$. Thus as u increases from 0 to ∞, $f(u)$ traverses the right half of the rectangle R. Similarly as u goes from 0 to $-\infty$, $f(u)$ traverses the left half of the rectangle R. Thus if u goes from $-\infty$ to ∞, $f(u)$ traverses the boundary of R in the usual orientation. Hence f is a conformal isomorphism of $\mathcal{H}$ onto the interior of R. This finishes the proof.

The elliptic function with periods $4K, 2iK'$ defined above is the Jacobi function

$$\text{sn } (z : k)$$

Between the periods K and K' there is a relation:

$$K'(k) = K(k') \qquad (k' = \sqrt{1-k^2})$$

To prove this we make the substitution

$$u = \frac{1}{\sqrt{1-k'^2t^2}}$$

in the integral for $K'(k)$ to get $K(k')$. The trigonometric substitution

$$u = \sin\phi$$

gives

$$K(k) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}}$$

The periods K, K' are *hypergeometric functions.* This is a deep relationship between periods and solutions of certain types of differential equations (the *regular singular ones*) that goes back to Legendre and was pushed much farther by Fuchs. We expand $(1-k^2u^2)^{-1/2}$ by the binomial theorem and use Euler's formula

$$\int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \qquad \Gamma(1/2) = \sqrt{\pi}$$

where Γ is the classical gamma function of Euler, and obtain

$$\begin{aligned} K(k) &= \sum_n \binom{1/2}{n} \int_0^1 u^{2n}(1-u^2)^{-1/2} du k^{2n} \\ &= (1/2) \sum_n \binom{1/2}{n} \int_0^1 t^{n-(1/2)}(1-t)^{-1/2} dt k^{2n} \\ &= (\pi/2) \sum_n \Big(\frac{(1/2)(3/2)\dots(2n-1)/2}{n!} \Big)^2 k^{2n} \end{aligned}$$

We now recall the hypergeometric series

$$F(a,b,c;z) = \sum_n \frac{a_{(n)} b_{(n)}}{c_{(n)} n!} z^n$$

where

$$u_{(n)} = u(u+1)\dots(u+n-1)$$

So

$$K(k) = \frac{\pi}{2} F(1/2, 1/2, 1; k^2)$$

Letting $k \to 0$ we have

$$K(0+) = \frac{\pi}{2}$$

which can also be obtained directly from the integral.

It is classical that the hypergeometric series $F(1/2, 1/2, 1; z)$ is the solution regular at $z = 0$ of the differential equation

$$z(1-z)\frac{d^2F}{dz^2} + (1-2z)\frac{dF}{dz} - \frac{1}{4} = 0$$

This equation is invariant under $z \longmapsto 1-z$ and so has the solution $F(1-z)$. We have

$$K(k) = \frac{\pi}{2} F(k^2), K'(k) = K(k')$$

so that K and K' are solutions to

$$k(1-k^2)\frac{d^2K}{dk^2} + (1-3k^2)\frac{dK}{dk} - kK = 0$$

One can find expressions for the solutions by the Frobenius method. But this is a case (as already noticed by Euler!) where the indicial equation has roots differing by an integer and the second solution has a logarithmic term.

One can show that

$$\lim_{k\downarrow 0} K(k) = \frac{\pi}{2}, \qquad \lim_{k\uparrow 1} K(k) = \infty$$

Hence also

$$\lim_{k\downarrow} K'(k) = \infty, \qquad \lim_{k\uparrow 1} K'(k) = \frac{\pi}{2}$$

Hence

$$\lim_{k\downarrow 0} \frac{K(k)}{K'(k)} = 0, \qquad \lim_{k\uparrow 1} \frac{K(k)}{K'(k)} = \infty$$

This shows that the ratio

$$\frac{K(k)}{K'(k)}$$

takes *all* values betwqeen 0 and ∞ as k varies between 0 and 1. The parameter τ for the elliptic function $\mathrm{sn}(z,k)$ is

$$\tau = i\frac{K}{2K'}$$

and so varies over the entire imaginary axis in $\mathcal{H}$ as k varies from 0 to 1. Notice that the period lattices are all *rectangular.* For nonrectangular period lattices we must consider *complex* k. These however require working over the Riemann surface of the function

$$\sqrt{(1-u^2)(1-k^2u^2)}$$

from the beginning, and hteir treatment became clear only after Riemann's epoch-making work on algebraic functions.

It can be shown that

$$K(k) = 2\log 2 + \log k + o(1) \qquad (k \uparrow 1)$$

The appearance of the logarithmic divergence is not surprising in view of the remarks we made on the second solution of the differential equation having a logarithmic term. The above relation can ndeed be obtained by that technique.

Finally mention must be made of the famous relation of Legendre

$$KE' + EK' - KK' = \frac{\pi}{2}$$

where, for any function G on $(0,1)$, the function G' is defined by $G'(k) = G(k')$ where, as usual, $k' = \sqrt{1-k^2}$.

RECTIFICATION OF CIRCLE

$$r = \sin t \qquad \int_0^{\sin t} \frac{du}{\sqrt{1-u^2}} = t$$

$$2 \times \text{ area } OAP = \int_0^{\sinh t} \frac{du}{\sqrt{1+u^2}} = t$$

FIGURE 1

RECTIFICATION OF LEMNISCATE

$$r^2 = \sin 2\theta$$

$$\int_0^{slt} \frac{dr}{\sqrt{1-r^4}} = t$$

$$\int_0^1 \frac{dr}{\sqrt{1-r^4}} = \frac{\omega}{2} \qquad \text{(half length of lemniscate in the first quadrant)}$$

FIGURE 2

RECTIFICATION OF ELLIPSE

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

$$\begin{aligned}\text{Length of arc } BS &= \int_0^s \sqrt{1 + (dy/dx)^2}\,dx \\ &= a\int_0^{s/a} \sqrt{\frac{1 - e^2u^2}{1 - u^2}}\,du\end{aligned}$$

$$\text{Total length of ellipse} = 4a \times \int_0^1 \sqrt{\frac{1 - e^2u^2}{1 - u^2}}\,du = 4aE(e)$$

FIGURE 3

PERIOD OF THE SIMPLE PENDULUM

$$0 < \theta_0 < \pi$$

$$T_{\text{lin}} = 2\pi\sqrt{\frac{\ell}{g}}$$

$$T = T_{\text{lin}}\cdot(2/\pi)K(k) = T_{\text{lin}}F(1/2, 1/2, 1; k^2) = T_{\text{lin}}\left(1 - (1/4)k^2 + \ldots\right)$$

FIGURE 4

9. THETA FUNCTIONS

1. Introduction and definition of theta functions. Theta functions were first introduced by Jacobi. They are entire functions defined by power series with a parameter $\tau \in \mathcal{H}$ and play a decisive role in the theory of elliptic functions because the elliptic functions can be expressed as a quotient of theta functions. Jacobi discovered extraordinary identities involving theta functions and showed that the entire theory of ellipti functions can be developed from their point of view. He solved fundamental problems of number theory with their help, namely the number of representations of an integer as a sum of k squares, especially $k = 2, 4, 8$.

It turned out that theta functions were equally critical in the theory of algebraic functions of higher genus. This was due to Riemann and Weierstrass who developed the theory of theta functions of several variables. Exactly as in the case of a single variable, ratios of theta functions give rise to multiply periodic functions of several variables, and indeed this is one of the methods for developing a generalization of the theory of elliptic functions to several variables. Theta functions of several variables are attached to lattices in $\mathbf{C}^g$ and one of the fundamental questions became the following: how to recognize, through the theta functions, whether the lattice is the lattice of periods of al algebraic function of genus g? The theta functions associated to the period lattice are called *Riemann theta functions.* It was only in recent years that this question, often called the *Shottky problem*, was solved. One of the solutions is due to Shiota, who solved it in the form conjectured by Novikov: a theta function is a Riemann theta function if and only if it satisfies the KdV equation.

There are many theta functions and we shall confine ourselves to the simplest aspects of the theory, mainly to give a sample of what is possible. In what follows all sums are over the set of all integers, unless explicitly indicated. We define

$$\theta(z, \tau) = \theta(z) = \sum_{n \in \mathbf{Z}} (-1)^n e^{2\pi i n z + \pi i n(n+1)\tau} \qquad (z \in \mathbf{C}, \tau \in \mathcal{H})$$

If we write

$$z = x + iy, \quad \tau = u + iv \qquad (x, y, u, v \in \mathbf{R})$$

then, in the region $|y| \leq Y, v \geq V$ we have the estimate

$$|(-1)^n e^{2\pi inz+\pi in(n+1)\tau}| \leq e^{-[n(n+1)V-2\pi nY]}$$

For $n \geq n_0$ we have $(n+1)V - 2\pi Y > 1$ and so the RHS of the estimate above is majorized by e^{-n} for such n. So the series defining theta converges absolutely and uniformly in the region specified. Hence θ is a holomorphic function on $\mathbf{C} \times \mathcal{H}$.

The critical property of theta is that as a function of z it has period 1 and a *quasi period* τ. First of all since $n(n+1)$ is even it is clear that changing z to $z+1$ does not change the terms of the series and so it follows that theta has period 1. On the other hand,

$$\begin{aligned}\theta(z+\tau,\tau) &= \sum (-1)^n e^{2\pi inz+\pi i(n^2+3n)\tau} \\ &= -\sum_m (-1)^m e^{2\pi i(m-1)z+\pi i(m^2+m)\tau-2\pi\tau} \\ &= -e^{2\pi(z+\tau)} \sum_m e^{2\pi imz+\pi i(m^2+m)\tau}\end{aligned}$$

Thus we have

$$\theta(z+\tau,\tau) = -e^{2\pi(z+\tau)}\theta(z,\tau)$$

In other words, changing z to $z+\tau$ (with θ fixed), changes θ to a multiple of itself by the exponential of a linear function. This is usually described by saying that θ has τ as a *quasi period.* Of course theta cannot have period τ also as then it would be doubly periodic and so would be a constant. If we take $z = x$ to be real, then

$$\theta(x,\tau) = \sum (-1)^n e^{\pi n(n+1)\tau} e^{2\pi inx}$$

showing that $\theta(x,\tau)$ has Fourier coefficients

$$(-1)^n e^{\pi n(n+1)\tau)}$$

(which are rapidly decreasing in n) so that it cannot be a constant. We also have, by a similar manipulation of series

$$\begin{aligned}\theta(-z,\tau) &= \sum (-1)^n e^{-2\pi inz+\pi n(n+1)\tau} \\ &= \sum (-1)^n e^{2\pi inz+\pi n(n-1)\tau} \\ &= -e^{2\pi iz} \sum_m e^{2\pi imz+\pi m(m+1)\tau}\end{aligned}$$

so that

$$\theta(-z,\tau) = -e^{2\pi iz}\theta(z,\tau)$$

In particular

$$\theta(0,\tau) = 0$$

Let P be a fundamental parallelogram for the lattice L_τ in the complex plane. We can translate P slightly so that no zero of θ lies on the boundary of P.

Proposition 1. *We have*

$$\frac{1}{2\pi i}\oint_P \frac{\theta'}{\theta}dz = 1$$

In particular, inside P there is exactly one zero for θ, and it is a simple one.

Proof. The proof is essentially the same as that of the Liouville theorems in the theory of doubly periodic functions. Let the vertices of P be $p, p+1, p+1+\tau, p+\tau$. Since θ and hence θ', has period 1, the integrals over the line segments from $p+\tau$ to p and from $p+1$ to $\mathbf{P}+\tau+\mathbf{1}$ cancel. The other two integrations do not quite cancel but almost. In fact,

$$\frac{1}{2\pi i}\oint_{p+\tau+1}^{p+\tau} \frac{\theta'(z)}{\theta(z)}dz = \frac{1}{2\pi i}\oint_{p+1}^{p} \frac{\theta'(w+\tau)}{\theta(w+\tau)}dw$$

But

$$\frac{\theta'(w+\tau)}{\theta(w+\tau)} = -2\pi i + \frac{\theta'(w)}{\theta(w)}$$

Hence,

$$\frac{1}{2\pi i}\oint_P \frac{\theta'}{\theta}dz = \frac{1}{2\pi i}\oint_{p+1}^{p} (-2\pi i)dw = 1$$

Corollary 2. *The zeros of θ are all simple and are precisely at the points of L_τ.*

Proof. This is because the zero set of θ is invariant under translations by 1 and τ and in a fundamental parallelogram there is just one simple zero.

2. Elliptic functions as the ratio of products of thetas. Abel's theorem. We now observe that the ratio of two functions, both entire with period 1 and *same* quasiperiod τ is a meromorphic function with periods $1, \tau$. So, if a, b are complex numbers,

$$\frac{\theta(z-a+\tau)}{\theta(z-b+\tau)} = e^{2\pi i(a-b)\tau}\frac{\theta(z-a)}{\theta(z-b)}$$

Hence, if $a_j, b_j (1 \leq j \leq k)$ are arbitrary complex numbers (equality among the a_j and b_j separately is allowed but no a_j is equal to any b_r) and we put

$$f(z) = \frac{\theta(z-a_1)\dots\theta(z-a_k)}{\theta(z-b_1)\dots\theta(z-b_k)}$$

then f has period 1 and

$$f(z+\tau) = e^{2\pi i(\sum_j a_j - \sum_j b_j)\tau} f(z)$$

Suppose now $\alpha_j, \beta_j (1 \leq j \leq k)$ are points on the torus $\mathbf{C}/L_\tau$ such that no α_j is equal to any β_r, although among themselves the α's and β's may be equal, with

$$\sum_j \alpha_j - \sum_j \beta_j = 0$$

Clearly we can choose a_j, b_j in $\mathbf{C}$ above α_j, β_j respectively such that

$$\sum_j a_j - \sum_j b_j = 0$$

Then, for the corresponding function f constructed above we have,

$$f(z+1) = f(z), \qquad f(z+\tau) = f(z)$$

Hence f is a meromorphic function with period lattice L_τ. From the construction it is clear that f has zeros exactly at the a_j and poles exactly at the b_j mod L_τ. We have thus proved *Abel's theorem:*

Theorem 3 (Abel). *If α_j, β_j are elements of $\mathbf{C}/L_\tau$ as above satisfying $\sum_j \alpha_j - \sum_j \beta_j = 0$, there is an elliptic function with period lattice L_τ which has zeros exactly at the α_j and poles exactly at the β_j.*

Remark. We already know (Liouville's theorems) that the condition

$$\sum_j \alpha_j - \sum_j \beta_j = 0$$

is necessary. So Abel's theorem is the sufficiency of this condition.

3. Sums of squares. Theta as a modular form. We now take up two other aspects of thetas, both due to Jacobi, which show the depth at which the thetas interact with arithmetic. First we introduce a variant of the theta function defined above. Let

$$\theta_0(z,\tau) = \sum_{n\in\mathbf{Z}} e^{2\pi i n z + \pi n^2 \tau}$$

The convergence of this series is similar to the earlier one and so θ_0 is holomorphic on $\mathbf{C}\times\mathcal{H}$. Again

$$\theta_0(z+1) = \theta_0(z)$$

while

$$\theta_0(z+\tau) = e^{-2\pi i z - \pi i \tau}\theta_0(z)$$

so that 1 is a period and τ is a quasi period for this theta. We are now interested in the value of θ_0 when $z = 0$. The null value, $\theta_0(0)$, is a function of θ. In fact

$$\theta_0(0)(\tau) = \sum_{n\in\mathbf{Z}} e^{\pi i n^2 \tau} \qquad (\tau\in\mathcal{H})$$

We now write

$$x = q^{1/2} = e^{\pi i \tau}$$

Then $\theta_0(0)$ is a function of x and we have

$$\theta_0(0)(\tau) = \sum_n x^{n^2}$$

We shall presently show that $\theta_0(0)$ is a *modular form* in the sense we have defined earlier, but with a small modification, namely that the properties of invariance are not with respect to the modular group $PSL(2,\mathbf{Z})$ but

with a *congruence subgroup* of it 9see below). The above power series representation of $\theta_0(0)$ shows that for any integer $k \geq 1$,

$$\theta_0(0)(x)^k = \sum_{m \geq 1} r_k(m) x^m$$

where $r_k(m)$ for any integer $m \geq 1$ is the number of $(n_1, n_2, \ldots, n_k)$ such that the n_j are in $\mathbf{Z}$ and

$$n_1^2 + n_2^2 + \ldots + n_k^2 = m$$

i.e., *the number of representations of m as a sum of squares of k integers.* It must be remembered that $r_k(m)$ counts as different representations where the signs are changed and the numbers permuted.

Theorem 4. *For any positive integer m let $d_r(m)$ $(r = 1, 3)$ be the number of divisors u of m such that $u \equiv r(4)$. Then*

$$r_2(m) = 4 \times (d_1(m) - d_3(m)$$

Further,

$$r_4(m) = 8 \times \text{ sum of divisors of } m \text{ that are not divisible by } 4$$

Note that for any positive integer m we count 1 and m as divisors of m. Suppose first that $m = p$ is an odd prime. Then if $p \equiv 1(4)$, we have $d_1(p) = 2, d_3(p) = 0$ so that $r_2(p) = 8$. But if $p = a^2 + b^2$, then neither of a, b is 0 and $\pm a \neq \pm b$. Hence each solution $p = a^2 + b^2$ where a, b are positive integers contributes 8 to $r_2(p)$. Thus we see that there is a unique solution $p = a^2 + b^2$ where a, b are positive integers and $a < b$. If on the other hand $p \equiv 3(4)$, then $d_1(p) = 1, d_3(p) = 1$ so that $r_2(p) = 0$. Thus in this case p cannot be represented as a sum of two squares. This is a classical result of Fermat. Suppose next that m is any positive integer., Then there is at least one divisor of m not divisible by 4, namely 1. hence

$$r_4(m) > 0 \text{ for all } m$$

Thus every positive integer can be expressed as a sum of 4 squares. This is a famous theorem that Lagrange had proved before Jacobi.

There is no question of proving the theorem of Jacobi here except to mention that it depends on the theory of the theta functions and their relationship to the $\wp$–functions.

We mentioned that $\theta_0(0)$ is a modular form. We shall now discuss this aspect of theta functions. From the definition of θ_0 it is immediate that

$$\theta_0(0)(\tau) = \theta_0(0)(\tau + 2) \tag{1}$$

To get the modular property we must investigate what happens when we make the transformation

$$S : \tau \longmapsto -\frac{1}{\tau}$$

We have the famous formula

$$\theta_0(0)(-1/\tau) = (-i\tau)^{1/2}\theta_0(0)(\tau) \qquad (\tau \in \mathcal{H}) \tag{2}$$

where $(-i\tau)^{1/2}$ is the branch of the square root that takes the value 1 at $\tau = i$. This is usually proved as an application of the *Poisson summation formula* and we shall give this proof.

Let $\mathcal{S}$ be the space of rapidly decreasing functions on $\mathbf{R}$, the so called *Schwartz space* of the real line. This is the space of smooth functions which, together with each of their derivatives, go to 0 at infinity faster than $(1 + x^2)^{-N}$ for any $N \geq 0$. For $f \in \mathcal{S}$ its Fourier transform $f^\sim$ is defined by

$$f^\sim(k) = \int e^{-2\pi ixk} f(x)dx \qquad (k \in \mathbf{R})$$

(all integrations are over $\mathbf{R}$). Then the basic result of Fourier analysis is that $f^\sim$ is also in $\mathcal{S}$ and f can be recovered from $f^\sim$ by

$$f(x) = (f^\sim)^\sim(-x)$$

i.e.,

$$f(x) = \int e^{2\pi ikx} f^\sim(k)dk \qquad (x \in \mathbf{R})$$

The *Poisson summation formula* is the statement that for any $f \in \mathcal{S}$ we have

$$\sum_{n\in\mathbf{Z}} f(n) = \sum_{n\in\mathbf{Z}} f^\sim(n)$$

We now take

$$f(x) = e^{\pi i x^2 \tau} \quad (x \in \mathbf{R}, \tau \in \mathcal{H})$$

It is a standard calculation that

$$f^{\sim}(k) = (-i\tau) - 1/2 e^{\pi i k^2)(-1\tau)} \qquad (k \in \mathbf{R})$$

To prove this we first take $\tau = iy$ where $y > 0$ and compute $f^{\sim}$ to be given by the above formula; $-i\tau$ is equal to y and we take the positive square root. The full formula is obtained by observing that $f^{\sim}$ is analytic on $\mathcal{H}$. The Poisson summation formula applied to this f then gives

$$\sum_n e^{\pi i n^2 \tau} = (-i\tau)^{-1/2} \sum_n e^{(\pi i n^2)(-1/\tau)}$$

which gives (2). We can also write (2) as

$$\theta_0(0)(-1/\tau) = e^{-\pi i/4} \tau^{1/2} \theta_0(0)(\tau) \tag{3}$$

where $\tau^{1/2}$ is the branch of the square root that takes the value $e^{\pi i/4}$ at $\tau = i$.

The equations (1) and (2) imply that $\theta_0(0)$ is a *modular form* for the subgroup of the modular group generated by T^2 and S. The subgroups of the modular group of the greatest interest are the *congruence subgroups;* the congruence subgroup of the modular group of level n is the subgroup Γ_n of $SL(2, \mathbf{Z})$ of all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{n}$$

It can be shown that the subgroup of the modular group generated by T^2 and S contains Γ_4.

The subject has an unbelievable number of remarkable and deep lying identities. We mention the product formula for the discriminant of the elliptic curve, namely

$$(2\pi)^{-12} \Delta(\tau) = q \prod_{n=1}^{\infty} (1 - q^n)^{24} \qquad (q = e^{2\pi i \tau}) \tag{4}$$

where, as usual, $\Delta = g_2^3/g_2^3 - 27g_3^2$. The infinite series

$$(2\pi)^{-12}\Delta(\tau) = \sum_{n=1}^{\infty} \tau(n)q^n \tag{5}$$

was singled out by Ramanujan in a famous paper as the generating function of the numbers $\tau(n)$ and so $n \longmapsto \tau(n)$ is called the *Ramanujan function'* and the $\tau(n)$ the *Ramanujan numbers.* Ramanujan conjectured remarkable properties of them, among them the result that they are *multiplicative,* i.e.,

$$\tau(mn) = \tau(m)\tau(n) \qquad ((m,n) = 1) \tag{6}$$

where $(m,n) = 1$ means that the integers m, n are mutually prime. This was an entirely novel concept of multiplicativity for arithmetically defined functions because up to that time only sequences which were *unrestrictedly multiplicative* were known, namely those for which $f(mn) = f(m)f(n)$ for *all* integer pairs m, n. The multiplicativity means that the knowledge of τ for powers of primes is enough to determine it completely, and Ramanujan conjectured the form of the generating function of the $\tau(n)$ in the form

$$\sum_{n=1}^{\infty} \frac{\tau(n)}{n^s} = \prod_p \frac{1}{(\tau(p)p^{-s} + \tau^{11-2s})} \tag{7}$$

where the product is over all primes. Finally he conjectured that

$$|\tau(p)| < 2p^{11/2} \tag{8}$$

for all primes p, the statement being equivalent to the statement that the roots of the quadratic polynomial

$$g_p(T) = 1 - \tau(p)T + p^{11}T^2 \tag{9}$$

are complex and have absolute value $p^{11/2}$. The multiplicativity of τ and the formula for the generating function were proved by Mordell a little time after Ramanujan made the conjecture. Mordell's work was the inspiration for Hecke who realized that Mordell's ideas could be vastly generalized and applied to *all* modular forms, thus inaugurating his celebrated theory of *Hecke operators.* The location of the roots of $g+p(T)$ was another story altogether. It resisted proof for a long time until Deligne proved first that

it will follow from the so called Weil conjectures in algebraic geometry, and then a few years later proved the Weil conjectures themselves. No *elementary* proof of (8) is known. The result (8) itself has been vastly generalized, in conjectural form, by various people, and is an important aspect of the modern program for modular forms pioneered by Langlands, the so called *Langlands program.*

One final historical note. The infinite product

$$\prod_{n=1}^{\infty}(1-q^n)$$

already was encountered by Euler who treated this and similar products in a famous paper *Partitio Numerorum.* He established the absolutely remarkable identity

$$\prod_{n=1}^{\infty}(1-q^n)=\sum_{n\in\mathbf{Z}}(-1)nq^{n(3n+1)/2}=\sum_{n\in\mathbf{Z}}(-1)^nq^{n(3n-1)/2} \qquad (10)$$

The numbers $n(3n-1)/2$ are known as the *pentagonal numbers.* If one builds regular pentagons starting at the origin in an expanding manner where the sides of the n^{th} pentagon have n lattice points, the number of lattice points that one gets upto the n^{th} pentagon is $n(3n-1)/2$, hence the name. The relation (120 is known as the *pentagonal number theorem of Euler.* It is one of the more dramatic achievements of Euler.

10. RIEMANN SURFACES OF ALGEBRAIC FUNCTIONS

1. Concept of an algebraic function. At various places earlier we have discussed Riemann surfaces and the examples of $\mathbf{C} \cup (\infty)$ and $\mathbf{C}/L$ (L a lattice in $\mathbf{C}$) as compact Riemann surfaces. We have also mentioned that the proper foundation for treating the Jacobian elliptic functions for *complex moduli* k is to view

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

as a holomorphic differential on the Riemann surface of the function

$$y = \sqrt{(1-z^2)(1-k^2z^2)} \tag{1}$$

In this section we shall take up such matters in a little more detail but still briefly.

By an *algebraic function of the complex variable* z we mean a function y satisfying a relation

$$y^k + a_1(z)y^{k-1} + \ldots + a_k(z) = 0 \tag{2}$$

where the a_j are rational functions of z. The function (1) is an example of an algebraic function. Let us write $a_j = f_j/f_0$ where the f_j are polynomials (we may always assume that they have no common factor). Then (2) is the same as

$$F(y,z) = f_0(z)y^k + f_1(z)y^{k-1} + \ldots + f_k(z) = 0 \tag{3}$$

A *branch* of (2) on a domain U is an analytic function y on U that satisfies (3) on U. However branches may not exist unless the domain U is chosen with care. For instance, no branch of $\sqrt{z}$ exists on any domain containing 0 while on any simply connected domain not containing 0 there are two branches differing only by a sign factor. The basic question of the theory of algebraic functions is to devise a method of treating them as single–valued functions. It was Riemann's great idea that this can be done if we

abandon the requirement that analytic functions be defined on domains in the complex plane but accept that they could be defined on what we now call Riemann surfaces. He realized that the topology of these Riemann surfaces influences the function theory in a decisive manner. We have already seen the difference between function theory on the extended complex plane and on the complex tori.

The construction of a Riemann surface for an algebraic function (3) on which y becomes single–valued becomes especially transparent if we follow the path taken by Weierstrass, namely, the method of *function elements.* In modern terminology, given a point $a \in \mathbf{C}$, a *germ* of an analytic function at a is an equivalence class of analytic functions defined around a, two such functions being called equivalent if they coincide in a little disk with a as center. Weierstrass called the germs *function elements.* The germs at a given point form an algebra and one can introduce the totality $\mathbf{G}$ of all germs at all possible points. If g_a is a germ at a, we may view g_a as defining an analytic function (also written as g_a) on a disk D_a around a; this function defines at each point b of the disk a germ g_b at b. The set of all germs g_b is thus a subset of $\mathbf{G}$ containing g_a. We denote this by $D(g_a)$ and view it as a neighborhood of g_a in the space $\mathbf{G}$. The space $\mathbf{G}$ thus becomes a topological space. More than that, the map

$$Z : g_b \longmapsto b \tag{4}$$

is bijective from $D(g_a)$ onto the disk D_a, and we give to $D(g_a)$ the complex structure that makes this map an analytic isomorphism. In this manner $\mathbf{G}$ becomes *a complex manifold.* Moreover, Z is a complex analytic map from $\mathbf{G}$ to $\mathbf{C}$.

But unlike usual complex manifolds, $\mathbf{G}$ is *not* connected. Nevertheless the concept of paths in $\mathbf{G}$ has an intrinsic significance. A path in $\mathbf{G}$ lies above, through Z, a path γ in $\mathbf{C}$. One can then view the germs g, h located at the ends of the path as germs that are connected to each other by *analytic continuation* along the path γ in the complex plane. Indeed, this is essentially the definition of analytic continuation along paths in the complex plane. A *connected component* of $\mathbf{G}$ is thus the totality of all germs that can be obtained by analytic continuation from a given germ.

2. The Riemann surface of an algebraic function. Let us now start with an algebraic function defined by the equation (1). It makes perfectly good sense to speak about germs that satisfy (1). The basic question is

now to look at the complex manifold of all such germs and determine completely its structure and the manner it lies above $\mathbf{C}$. Notice that on this manifold y becomes a single–valued function:

$$y : g_a \longmapsto g_a(a) \tag{5}$$

The key observation is that if a is a generic point of the complex plane, there are k distinct branches satisfying (3).

Proposition 1. *Let Δ be the set of all points $t \in \mathbf{C}$ such that $f_0(t) \neq 0$ and the polynomials $F(y,t), F_y(y,t)$ do not have a common root, F_y being the partial derivative of F with respect to w, so that the equation $f_0(t)X^k + f_1(t)X^{k-1} + \ldots + f_k(t) = 0$ has k distinct roots $u_1, u_2, \ldots, u_k$. Then there are k distinct branches $y_j (1 \leq j \leq k)$ of (3) such that $y_j(t) = u_j (1 \leq j \leq k)$. Moreover, if h is a continuous function that is defined around t and satisfies $F(h(z), z) = 0$ around t, then $h = y_j$ for some j around t.*

Proof. Fix $j = 1, 2, \ldots, k$ and write $u = u_j$. Since $F_y(u,t) \neq 0$ we can find $\rho > 0, \delta > 0$ such that $F_y(y,z) \neq 0$ for $|z-t| \leq \rho, |y-u| \leq \delta$. By choosing ρ smaller we may assume that $F(y,t)$ has no zero in $0 < |y-u| \leq \delta$, in particular on $|y-u| = \delta$. A simple compactness argument shows that we can choose a smaller ρ so that $F(y,z) \neq 0$ for $|y-u| = \delta, |z-t| \leq \rho$. The integral

$$n(z) = \frac{1}{2\pi i} \oint_{|y-u|=\delta} \frac{F_y(y,z)}{F(y,z)} dy$$

is the number of roots of $F(y,z) = 0$ inside the circle $|y-u| = \delta$ which is 1 for $z = t$. On the other hand, it follows from a general argument that it is continuous. So $n(z) = 1$ for $|z-t| \leq \rho$. So there is a unique root of $F(y,z) = 0$ inside this circle for $|z-t| \leq \rho$. Let us denote it by $y(z)$. By the residue calculus we have

$$y(z) = \frac{1}{2\pi i} \oint_{|y-u|=\delta} y \frac{F_y(y,z)}{F(y,z)} dy$$

and from this formula it is clear that $y(z)$ is analytic in z. y thus defines a branch of (3). We denote y as y_j to remind us that its value at $z = t$ was $u = u_j$. If h is as in the theorem, then $h(t) = u_j$ for some j, and so, as $h(t)$ is different from the other y_r, we see that for z sufficiently near t, $h(z)$ has to coincide with $y_j(z)$, for reasons of continuity.

Proposition 2. *The complement of Δ is finite.*

Proof. This is a consequence of the theory of resultants or the use of the euclidean algorithm in the ring of polynomials in y, z. We omit the argument.

The points outside Δ, including ∞, are called *critical points.* Then at each point of Δ there are k branches located at that point. Let R' be the set of all branches at the noncritical points. Then R' is a complex manifold and the map

$$Z : R' \longrightarrow \Delta$$

is a *covering map* such that each fiber has exactly k elements. Z is called a *k–sheeted covering map.* The second key point, which we shall not prove here, is the following.

Proposition 3. *R' is connected if F is irreducible. In particular, in this case, analytic continuation of any branch is possible along any path in Δ.*

One would like to call R' the Riemann surface of attached to the algebraic function y satisfying (3). But one can do more. For each critical point t including ∞ we can add points $p_1(t), p_2(t), \ldots p_r(t)$ to R' such that $R = R' \cup_t \{p_1(t), \ldots, p_r(t)\}$ such that R is a Riemann surface which is compact. R is called the *Riemann surface of the algebraic function defined by (3).*

The addition of the p_j above the critical point t is seen to be possible as follows. Although R' is connected, if we take a small disk D around t (not containing any critical point other than t), the open set $Z^{-1}(D^\times)$ is not connected in general, $D^\times$ being the punctured disk $D \setminus (t)$. Let

$$Z^{-1}(D^\times) = K_1 \cup K_2 \cup \ldots \cup K_r \tag{6}$$

be its decomposition into connected components. One can give an invariant description of the branches that lie in the various K_j. Fix a point $z_0 \in D^\times$ and let η be a branch of (3) at z_0. If we analytically continue this branch along a closed path beginning and ending at z_0, the end result will be another branch η_1. The branch η_1 depends only on the homotopy class of the path in $D^\times$. So the fundamental group of $D^\times$ with base point z_0, which is isomorphic to $\mathbf{Z}$, acts on the set B_0 of branches at z_0. This action will split B_0 into a finite set of disjoint orbits $\Sigma_j (1 \leq j \leq r)$:

$$B_0 = \Sigma_1 \cup \Sigma_2 \cup \ldots \cup \Sigma_r \tag{7}$$

Then we can take K_j to be the set of all branches at points of $D^\times$ that have an analytic continuation into some element of Σ_j. Let

$$n_j = \text{number of elements of } \Sigma_j \tag{8}$$

It is then easy to show that

$$Z : K_j \longrightarrow D^\times$$

is a covering map also, mapping K_j onto $D^\times$ and having n_j sheets. Clearly

$$n_1 + n_2 + \ldots + n_r = k$$

Now comes the decisive observation, that the punctured disk has only one m–sheeted covering map for any positive integer m, namely the map that locally looks like

$$z \longmapsto z^m$$

So we can find a disk D_j and a complex analytic isomorphism

$$t_j : D_j^\times \simeq K_j$$

such that the composition

$$t_j \circ Z$$

becomes the map

$$\zeta \longmapsto \zeta^{n_j}$$

of $D_j^\times$ onto $D^\times$. We glue D_j to R' so that $D_j^\times$ gets glued on K_j via the identification t_j. The construction of R is complete.

The points $p_j(t)$ above the various critical points are called *ramification points.* The integer n_j is called the *ramification index* at the point which corresponds to the center of D_j (that represents the additional point corresponding to Σ_j). The Riemann surface R defined above has two maps Z, y both into $\mathbf{C} \cup (\infty)$. y is the algebraic function satisfying (3) while Z gives the location of the branch of y.

To add some additional insight into these considerations, we remark that if f is a local nonconstant complex analytic map defined around 0 taking 0 to 0, then there is an integer $m \geq 1$ such that in suitable local coordinates f becomes the map $z \longmapsto z^m$; m is the *ramification index* of the map f at the origin.

For any compact smooth manifold X of dimension 2 we have the finite dimensional vector space

$$H^1(X) = Z(X)/B(X)$$

where $Z(X)$ is the vector space of all real closed 1-forms on X and $B(X)$ is the subspace of all forms df where f is a real smooth function on X. If α, β are two closed forms, $\alpha \wedge \beta$ is a 2–form and so

$$B'(\alpha, \beta) = \int_X \alpha \wedge \beta = -B'(\beta, \alpha)$$

is well defined and vanishes whenever either α or β is exact. Hence B' induces a bilinear form

$$B : H^1(X) \times H^1(X) \longrightarrow \mathbf{R}$$

which is skew symmetric. It is a basic result that this form is *nondegenerate.* As only an even dimensional vector space can carry a nondegenerate skew symmetric bilinear form we conclude that $\dim H^1(X)$ must be even,

$$\dim H^1(X) = 2g$$

where g is called the *genus* of X. If X is a Riemann surface Riemann's work showed the decisive role played by the genus in the function theory on X. The genus is 1 if and only if X is a torus while the genus is 0 if and only if $X \simeq S^2$.

The Riemann surface R constructed above for an algebraic function (3) has a genus g given by the so called *Riemann–Hurwitz formula:*

$$2 - 2g = 2k - \sum(n - 1) \tag{8}$$

where the sum is over all the ramification points of the map $Z : R \longrightarrow \mathbf{C} \cup (\infty)$. In classical texts for various simple forms of (3) the genus g is computed explicitly by the so called glue and scissors method.

As an example consider the algebraic function defined by

$$y^2 = (z - e_1)(z - e_2) \dots (z - e_N)$$

where $e_1, e_2, \dots, e_N$ are N distinct complex numbers. We first claim that the Riemann surface R is ramified above the points e_j. Indeed there are 2

sheets and so if there is no ramification above e_j, then any branch returns to itself if continued analytically around e_j (inside a small disk around e_j). So the branch must be meromorphic at e_j and, as it is bounded near e_j, it must be holomorphic. But it is easy to see that there is no holomorphic branch around e_j. The same argument applies at ∞ if N is odd, while for N even there is no ramification above ∞ because there is a branch meromorphic at ∞ as is easily verified. Hence, by (8) the genus is

$$g = \left[\frac{N-1}{2}\right] \qquad ([\,\cdot\,] \text{ is the integral part }) \tag{9}$$

Thus if $N = 3, 4$ the genus is 1, while for higher N the genus is > 1. Riemann surfaces with genus 1 are complex tori and so are called elliptic. The surfaces for $N > 4$ are called *hyperelliptic.*

For the Jacobi case (1) we introduce the Riemann surface X corresponding to (1). On it the form

$$\frac{dZ}{y} \tag{10}$$

is a holomorphic differential and so its integral over paths make sense. The indefinite integral however depends on the path but is uniquely determined upto a lattice L in the complex plane, namely the lattice of periods of the form (10). One thus obtains a map of X on $\mathbf{C}/L$. The inversion theorem of Jacobi is that this map is an analytic isomorphism.

3. The algebraic point of view. If R is the Riemann surface of y constructed above, it turns out that the meromorphic functions on R are precisely the *rational functions* of y, Z. They form a field, which we denote by $\mathbf{C}(R)$. The rational functions of Z form a subfield, $\mathbf{C}(Z)$, and $\mathbf{C}(R)$ is a finite (hence algebraic) extension of $\mathbf{C}(Z)$. Conversely, if R is any Riemann surface, there is a nonconstant meromorphic function Z on R and another meromorphic function y on R, such that y is an algebraic function of Z and the field of meromorphic functions on R is precisely the field of rational functions of y, Z. On the other hand, if K is a finite algebraic extension of $\mathbf{C}(z)$, there is a Riemann surface R, unique upto analytic isomorphism, such that $K = \mathbf{C}(R)$; R can be recovered from K as the set of its *places.*. This last result suggests that the entire theory of Riemann surfaces can be developed from the algebraic point of view; this was historically done by Dedekind and Weber.

This last point can be illustrated by an example. We have seen that nonsingular plane cubic curves

$$Y^2W = 4X^3 - a_2XW^2 - a_3W^3 \qquad (a_2^3 - 27a_3^2 \neq 0) \tag{11}$$

are isomorphic to complex tori. The space defined by this equation is a submanifold of $\mathbf{CP}^2$ and can be identified with the Riemann surface of (11). If however we consider a plane *quartic curve,* such as (1) in projective form

$$Y^2W^2 = (W^2 - X^2)(W^2 - k^2X^2) \tag{12}$$

we see that it has singular points; indeed, the point at infinity $(0, 1, 0)$ is a singular point, although it is the only singular point. So the Riemann surface of (12) cannot be identified with this curve. Nevertheless the Riemann surface exists and its field of meromorphic functions is the same as the restriction to the curve of the rational functions p/q in X, Y such that q does not vanish identically on the curve. It is on this surface that the elliptic integral is defined. We should regard the Riemann surface as a *nonsingular model* of the curve; the nonsingular model is unique.

In the 19^{th} century people developed the theory of algebraic curves in great depth and obtained results of great beauty (recall for instance the geometry of plane cubics). The great figures were Clebsch, Noether (the father of Emmy Noether), Brill, etc. The work of Riemann, and the subsequent work of Dedekind–Weber showed conclusively that all three points of view lead to the same theory. However, the algebraic and geometric points of view have an additional aspect, namely that they can be developed over an *arbitrary field*, instead of the complex field over which the transcendental theory takes place. Nevertheless the formal structures have a great deal of resemblance. In particular, As Schmidt and then Artin did, one can develop the theory over *finite fields.* One may think of the theory of algebraic function fields over finite fields as a theory of Riemann surfaces with finite fields of constants.

Let us start with the field $\mathbf{C}(z)$ and then consider the tower of all finite algebraic extensions of it. From our remarks above it is clear that this tower is more or less the same as the tower of Riemann surfaces over $\mathbf{C} \cup (\infty)$. Now this structure has a great similarity to the structure of all finite algebraic extensions of $\mathbf{Q}$, the field of rational numbers. These are the *algebraic number fields.* Thus algebraic number theory and algebraic function theory are built along very similar lines, and the mathematicians of the 19^{th} century obtained great insights into both theories by exploring

this analogy. The analogy becomes particularly close when we compare the algebraic number fields with algebraic function fields over finite fields of constants. What came out of this analogy and how it has led to a revolution in modern mathematics is another story.

11. JUGENDTRAUM

1. Cyclotomic fields. Theorem of Kronecker–Weber. Cyclotomy means division of the circle. The points of division of the unit circle into n equal parts are the n^{th} roots of unity,

$$\omega_r = \omega_n^r \ (0 \leq r \leq n-1), \qquad \omega_n = e^{2\pi i/n} \tag{1}$$

These are the vertices of a regular n–gon inscribed in the unit circle. The algebraic and arithmetic importance of these numbers has probably been recognized from ancient times. The formulae

$$\cos\frac{\pi}{3} = \frac{\sqrt{3}}{2}, \qquad \sin\frac{\pi}{10} = \frac{\sqrt{5}-1}{4}$$

show that cyclotomy is closely related to quadratic irrationalities. When he was 19 years old, Gauss made a profound discovery, namely that the quantities

$$\cos\frac{2\pi}{n}$$

are not accessible by euclidean geometric constructions except for certain special values of n. For n a power of 2 this was already known to Archimedes, but Gauss showed, in a famous discussion, that this is possible for $n = 17$, and more generally, that this is possible if and only if n is of the form

$$2^k p_1 p_2 \dots p_r$$

where the p_i are distinct *Fermat primes,* i.e., primes of the form

$$2^{2^s} + 1$$

Since $17 = 2^4 + 1$ this result contains the theory of euclidean construction of regular 17–gons as a special case.

Let us write F_n for the extension of $\mathbf{Q}$ generated by ω_n where $\mathbf{Q}$ is as usual the field of rational numbers. F_n is called the n^{th} *cyclotomic field.* This is the splitting field of the polynomial

$$X^n - 1$$

and so is a Galois extension. Let

$$G_n = \text{Gal } (F_n/\mathbf{Q})$$

be the Galois group of F_n over $\mathbf{Q}$. Now ω_n is a *primitive* n^{th} root of unity, namely a root of unity such that its r^{th} power is not unity for any r with $1 \leq r < n$. It is clear that the primitive n^{th} roots of unity are those of the form

$$\omega_n^r \qquad ((r,n) = 1)$$

and so are $\phi(n)$ in number where $\phi(n)$ is Euler's number, i.e., the number of integers that are $< n$ and prime to n. If $\sigma \in G_n$, it is clear that $\sigma(\omega_n)$ is also a primitive root of unity and so there is an integer $r = r(\sigma)$ such that

$$\sigma(\omega_n) = \omega_n^{r(\sigma)} \qquad ((r(\sigma), n) = 1)$$

The integer $r(\sigma)$ is determined only mod n and so we have a map

$$r : G_n \longrightarrow \mathbf{Z}/n\mathbf{Z} := \mathbf{Z}_n$$

Moreover the values of r are residue classes prime to n so that we have indeed a map

$$r : G_n \longrightarrow P_n \qquad \sigma(\omega_n) = \omega_n^{r(\sigma)}$$

where P_n is the *multiplicative group* of residue classes mod n that are prime to n. This map is a homomorphism. Indeed, if σ, τ are two elements of G_n, we have

$$\sigma\tau(\omega_n) = \sigma(\omega_n^{r(\tau)}) = \omega_n^{r(\sigma)r(\tau)}$$

Since ω_n generates F_n it is immediate that this map is injective. It is a deep fact that this map is surjective also so that

$$G_n \simeq P_n$$

In particular

$$[F_n : \mathbf{Q}] = \phi(n)$$

The last fact can be described in another way. The polynomial $X^n - 1$ is not irreducible, it has $X - 1$ as a factor. The polynomial

$$\Phi_n(X) = \prod_{r:(r,n)=1} (X - \omega_n^r)$$

is called the n^{th} *cyclotomic polynomial.* Thus

$$\Phi_3 = X^2 + X + 1, \Phi_4 = X^2 + 1, \Phi_5 = X^4 + X^3 + X^2 + X + 1$$

and one can write down a formula for computing Φ_n for any n. The degree of this polynomial is $\phi(n)$ and one can show that

$$\Phi_n(X) \in \mathbf{Z}[X]$$

The fact that the degree of F_n over $\mathbf{Q}$ is $\phi(n)$ shows that $\Phi_n(X)$ is irreducible.

The fact that the Galois group of F_n is isomorphic to P_n implies in particular that it is abelian. The converse, which is very surprising, is also true. It was discovered by Kronecker but is known as the *Kronecker–Weber theorem.* It asserts that if K is a Galois extension of $\mathbf{Q}$ such that its Galois group is abelian, or as it is usual to say, K is an *abelian extension* of $\mathbf{Q}$, then there is an integer n such that

$$K \subset F_n$$

In other words, every abelian extension of $\mathbf{Q}$ is a subextension of a cyclotomic field extension.

2. Jugendtraum. In a letter to Dedekind written on March 15, 1880, Kronecker referred to the problem of describing explicitly all the abelian extensions of a quadratic imaginary field $\mathbf{Q}(\sqrt{-D})$ where D is a positive square–free integer (this means that D is a product of distinct primes). Clearly he thought of this as a vast generalization of the theory of abelian extensions of $\mathbf{Q}$ which were known to him to be cyclotomic. In the letter, which is in German, he referred to this as *meinen liebsten Jugendtraum,* my dearest dream of youth. In fact he writes

> *... I have overcome today the last of many difficulties which have been preventing me the conclusion of an investigation with which I have been occupied again in the past months. The subject is my dearest dream of youth, namely the proof that the abelian equations with square roots of rational numbers as coefficients are given by the transformation equations of elliptic functions with singular moduli, just as the abelian equations with integer coefficients are given by the cyclotomic equations...*

Kronecker's vision was remarkable and he obtained most of the fundamental results in this program. But there were some gaps in methods as well as results, and his work was completed by Weber after Kronecker had died. Ever since that time, this circle of problems has always been referred to as the Jugendtraum problems. For us, the most interesting aspect of this question is the role played by elliptic and modular functions. We shall not describe fully the results of Kronecker and Weber but confine ourselves to just one aspect. For this some preparation is necessary, especially in the arithmetic of number fields.

Everyone knows that the basis of ordinary arithmetic is the *unique factorization theorem* in $\mathbf{Q}$ which asserts that any positive integer is a product of primes and that this representation is unique except for the order of the prime factors. Although the concept of primes and integers can be extended to arbitrary number fields, and although for a few of them such as $\mathbf{Q}(i)$ the unique factorization is verifiable, people like Dirichlet and Gauss knew that this is not true in general. For instance, in the field $\mathbf{Q}(\sqrt{-5})$ one has the two distinct decompositions

$$6 = 2 \times 3, \qquad 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

and it is easy to verify that $2, 3, 1 \pm \sqrt{-5}$ are all prime numbers in the sense that they are divisible only by themselves and units (integers whose inverses are also integers, integers being numbers of the form $a + b\sqrt{-5}$). It was Kummer who overcame this hiatus and discovered the path toward a general arithmetical theory of number fields. His idea was to insist on unique factorization but change the concept of prime so as to secure unique factorization. This is not the place to go into Kummer's epoch–making discoveries but we confine ourselves to the statement that Kummer replaced primes by *prime ideals.* Let K be an algebraic number field. An integer in K is an element which satisfies a monic equation with integer coefficients, i.e., any element θ such that

$$\theta^k + a_1\theta^{k-1} + \ldots + a_k = 0$$

where $a_1, \ldots, a_k$ are ordinary integers. The integers form a ring whose quotient field is K. Let R be the ring of integers in K. For instance, the ring of integers in $\mathbf{Q}(\sqrt{-5})$ is $\mathbf{Z} + \mathbf{Z}\sqrt{-5}$ while the integer ring in $\mathbf{Q}(\sqrt{5})$ is $\mathbf{Z} + \mathbf{Z}\theta$ where $\theta = (1 + \sqrt{5})/2$. Kummer* realized that one should work

* Actually Kummer worked only with cyclotomic fields. The extension of the Kum-

with prime ideals in R instead of prime numbers. In the case of $\mathbf{Z}$ the two concepts coincide; the prime ideals in $\mathbf{Z}$ are precisely the principal ideals generated by prime numbers. For an arbitrary number field, one should consider ideals whether they are prime or not, and more generally, *fractional ideals,* namely R–modules in K of the form xJ where J is an ideal in R. The fractional ideals admit an operation of *multiplication,* namely,

$$J_1 J_2 = R - \text{module generated by } a_1 a_2 \text{ where } a_i \in J_i$$

The fundamental theorem of algebraic number theory, proved by Kummer for the cyclotomic fields and Dedekind and Kronecker extended to all number fields, is the following.

Theorem. *Under multiplication the set of fractional ideals form a group, and that this group is the free abelian group generated by the prime ideals. In particular if J is any ideal in R,*

$$J = P_1^{k_1} \dots P_r^{k_r}$$

where the k_j are integers ≥ 0 and the P_j are prime ideals, the representation being unique except for the order of the factors P_j.

Clearly the group property and its freeness is the abstract expression of the unique factorization property. How do the fields with unique factorization property in the old sense fit in here? Clearly when all ideals of R are principal. This is however not always true. Kummer proved (again for cyclotomic fields, and Dedekind and Kronecker for arbitrary number fields) that if G is the group of fractional ideals and G_0 is the subgroup of *principal* fractional ideals, then

$$C(K) := G/G_0$$

is a *finite group.* The elements of $C(K)$ are called *ideal classes* and $C(K)$ itself is called the *ideal class group* or simply, the *class group.* For instance

$$C(\mathbf{Q}(\sqrt{-5})) = 2$$

---

mer theory to all number fields was carried out by Dedekind and Kronecker in independent and different ways. We adopt the Dedekind point of view that the proper generalization of a prime number to a number field is a prime ideal.

The number of elements in $C(K)$ is known as the *class number* of K. The class number is a deep lying global invariant of a number field.

Let us return to the jugendtraum. Let

$$K = \mathbf{Q}(\sqrt{-D})$$

where $D > 0$ is a square free integer. The key observation that shows that the theory is somehow related to elliptic functions is that the integers of K form a *lattice* in $\mathbf{C}$. In fact, each fractional ideal in K is a lattice, and two fractional ideals are in the same class if and only if they are equivalent under complex multiplication. Now let J be the modular function

$$J = \frac{g_2^3}{g_2^3 - 27g_3^2}$$

Then as J gives the same value to all lattices that are equivalent under complex multiplication, it follows that the J–value of all the fractional ideals in a given ideal class $\mathfrak{k}$ are the same. We denote this by $J(\mathfrak{k})$. Then Kronecker's beautiful theorem, which is one of the many that he obtained is the following.

Theorem (Kronecker). *Let $\mathfrak{k}_1, \ldots \mathfrak{k}_d$ be the distinct ideal classes of $K = \mathbf{Q}(\sqrt{-D})$, d being the class number of K. Then the values*

$$J(\mathfrak{k}_1), \ldots, J(\mathfrak{k}_d)$$

are all algebraic numbers, and they generate an abelian extension of K. Moreover this is the maximal unramified abelian extension of K.

It is necessary to comment only on the last statement of the theorem. The point is, and we have mentioned this earlier, the tower of extensions of a given number field is very similar in structure to the tower of Riemann surfaces above a given Riemann surface. Clearly it is possible that for Riemann surfaces X, X' where there is a nonconstant map $X' \longrightarrow X$, this map may be everywhere unramified, i.e., it is conformal. Then it is a covering map. For instance, if L is a lattice in $\mathbf{C}$ and $L' = rL$ where r is a positive integer > 1, the natural map

$$\mathbf{C}/L' \longrightarrow \mathbf{C}/L$$

is a covering map and so is unramified everywhere. It is possible to develop the notion when a number field K' that contains a number field K is ramified at a prime ideal of K'. One says that K'/K is unramified if the extension is unramified at all primes. The field obtained by Kronecker in the above theorem is the so called *Hilbert classfield.*