Elliptic galore

Master Course, Spring term 2025

Fernando Chamizo https://matematicas.uam.es/~fernando.chamizo/

Contents. Elliptic integrals. Elliptic functions. Elliptic curves. Klein's j-invariant.

4.1 Elliptic integrals, the starting point

Nowadays machines are able to get closed formulas for any reasonable integral admitting them, but few decades ago learning a big bunch of recipes for the integration in elementary terms was a nontrivial part of the basic education of scientists.

In many applications they appear integrals of the form:

$$I=\int R\big(t,\sqrt{P(t)}\big)\,dt$$

with R a rational function and P a polynomial.

If $\deg P=1$ a change of variables $u^2=P(t)$ transforms I into a rational integral. If $\deg P=2$ after some reductions essentially the only case to worry about is $P(t)=1-t^2$ and the change $t=\sin u$ followed by $v=\tan\frac{u}{2}$ reduces again I to a rational integral. Nowadays we know that this is the end of the story, when $\deg P>2$ it is impossible to write I in terms of *elementary functions* (roughly speaking, algebraic combinations of the functions appearing in standard scientific calculators) except in somewhat trivial situations [5].

Euler showed that the cases $\deg P=3$ and $\deg P=4$ come in the same pack, one can transform one into the other with a clever change of variables. These cases appear quite often in mathematics and physics. One of them was the length of an arc of ellipse and the integrals of the form I are called *elliptic integrals*, although most of them has nothing to do with ellipses. Sometimes the name is restricted to the case $R(x,y)=1/y$ and $P(t)=(1-t^2)(1-k^2t^2)$. As a matter of fact, the parameter k was called *modulus* and this is the origin of the name *modular forms* for functions that today mostly have not any relation to k. Euler also proved funny addition formulas, for instance,

$$F(t)=\int_0^x\frac{dt}{\sqrt{1-t^4}}\qquad\text{satisfies}\qquad F(x)+F(y)=F\Big(\frac{x\sqrt{1-y^4}+y\sqrt{1-x^4}}{1+x^2y^2}\Big)$$

and similar formulas for other elliptic integrals.

The big idea in the theory is due to Gauss (who did not publish it when he got it) and to Abel independently years later [10, §11.6]. Later Jacobi developed it greatly. To introduce

this idea, consider the elementary integral

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

Let us pretend that a mathematician with limited integration skills is not able to compute it, but with a complicated argument (as Euler did with elliptic integrals) arrives to the addition formula:

$$G(x) + G(y) = G\big(x\sqrt{1-y^2} + y\sqrt{1-x^2}\big).$$

This seems rather mysterious and the extension of it to integrals with $1-t^2$ replaced by an arbitrary quadratic polynomial would involve messy considerations about ranges and signs of square roots. Note for instance that G only makes real sense for $x \in [-1,1]$ and if we admit complex numbers to extend it to $x \in \mathbb{R}$ or even to $\mathbb{C}$, we have to take a choice about signs and paths. We know that, actually, $G(x) = \arcsin x$, which is an ugly function only well defined in an interval, but its inverse can be extended to $\sin x$, a nice entire function. Writing $x = \sin\alpha$ and $y = \sin\beta$ we see that the mysterious addition formula for G becomes

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha,$$

which is elementary, it admits a geometric proof. It makes sense even for $\alpha, \beta \in \mathbb{C}$. Part of the complications of the pretended theory for G and allied functions comes from the fact that a multivalued function is obtained when inverting a periodic function.

Coming back to elliptic integrals, the big idea was that inverting them one gets nice meromorphic functions with nice properties. The most noticeable one is that they have two periods. For instance, Gauss showed that the inverse of F has a real period p, which is related to the Gamma function by $p = \Gamma(1/4)^2/\sqrt{2\pi}$, and a purely imaginary period ip. Jacobi defined a kind of doubly periodic analogs of trigonometric functions. An amazing fact is that his research produced number theoretical outcomes. Namely, he proved using functions appearing in the theory of elliptic integrals that the number of representations as a sum of 4 squares $r_4(m) = \#\big\{\vec{n} \in \mathbb{Z}^4 \,:\, m = \|\vec{n}\|^2\big\}$ is given by

$$r_4(m) = 8\sum_{d|m} d \qquad \text{if } m \in \mathbb{Z}^+ \text{ is odd}$$

and a similar formula for m even. Nowadays it is more natural to obtain this result using modular forms. Although the simplicity of the formula, arguably simple elementary proofs are not known (read [7] for another opinion).

There are brief accounts of the history of the elliptic integrals in chapter 11 of [10] and in chapter 2 of [6].

4.2 Elliptic functions

The functions obtained when inverting elliptic integrals have two periods and it motivated to define an *elliptic function* as a meromorphic function f satisfying

$$f(z) = f(z+\omega_1) = f(z+\omega_2) \qquad \text{for some } \omega_1, \omega_2 \in \mathbb{C} \text{ with } \mathbb{C} = \omega_1\mathbb{R} + \omega_2\mathbb{R}.$$

The condition on ω_1 and ω_2 is a fancy way of saying that the periods are essentially distinct, they do not point in the same direction, they determine a parallelogram $\mathcal{P}$. Changing names we can always assume $\Im(\omega_1/\omega_2) > 0$. The open parallelogram $\mathcal{P}$ or any of its translates is said to be a *fundamental parallelogram*. It plays the same role as the fundamental domain and the condition $\Im(\omega_1/\omega_2) > 0$ specifies an orientation for $\mathcal{P}$.

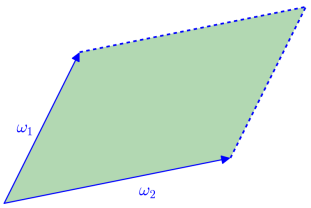
An elliptic function f as before is also trivially invariant by any element of the *period lattice*

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

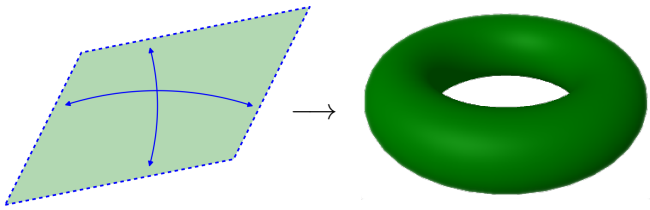
We also introduce the notation $\Lambda^* = \Lambda - \{0\}$. Note that Λ does not determine ω_1 and ω_2. For instance, $1\mathbb{Z} + i\mathbb{Z} = 1\mathbb{Z} + (1+i)\mathbb{Z}$.

Instead of considering a fixed elliptic function and obtaining its period lattice, in the sequel we fix the period lattice Λ and consider the linear space of elliptic functions, defined as meromorphic functions such that $f(z) = f(z+\omega)$ for every $\omega \in \Lambda$. In this way some technical issues are avoided (for instance, what the periods of a constant functions are) and the fundamental parallelogram is associated to Λ without reference to the elliptic function.

An elliptic function f is completely determined by its values in a fundamental parallelogram $\mathcal{P}$ because f is Λ-invariant. In connection with this, it can be said that an elliptic function is a meromorphic function $\mathbb{C}/\Lambda \longrightarrow \mathbb{C}$ because the natural Riemann surface structure of $\mathbb{C}/\Lambda$ allows to talk about meromorphic functions there. The topology of $\mathbb{C}/\Lambda$ is the quotient topology of $\overline{\mathcal{P}}$ when we use the translations corresponding to the periods to identify the boundaries. The result is homeomorphic to a torus.



A fundamental paralelogram



Geometric view of $\mathbb{C}/\Lambda$ as a torus

If g is a meromorphic function, formally $\sum_{\omega\in\Lambda} g(z+\omega)$ is elliptic, but the possible divergence constitutes a serious issue. Choosing g as a rational function we avoid essential singularities at ∞ (recall Casorati-Weierstrass theorem), but me still need the total degree (the difference of the degrees of denominator and numerator) to be negative enough. With some basic skills in analysis it can be proved that total degree -2 is the limiting case to get absolute convergence: $\sum_{\omega\in\Lambda} |z+\omega|^{-p}$ with $p \in \mathbb{R}$ converges if and only if $p > 2$. It turns out that it is important to save the case of degree -2 and Weierstrass subtracted infinities (like modern particle physicists do) to force the convergence. He introduced the *Weierstrass $\wp$ function*

$$\wp(z) = z^{-2} + \sum_{\omega\in\Lambda^*} \big((z+\omega)^{-2} - \omega^{-2}\big).$$

It is formally elliptic and, again, with some basic analytic ability it is proved that it is uniformly convergent on compact sets not including the points of Λ, where $\wp$ has double poles. The

underlying idea is that when ω is large, $(z+\omega)^{-2}-\omega^{-2}$ goes like $-2z\omega^{-3}$ which guarantees the convergence.

The theory of elliptic functions is plenty of surprising identities essentially because Liouville's theorem implies that there are not (nonconstant) elliptic function without poles. In fact, Liouville stated his theorem for elliptic functions (in a memoir in 1847). The sketch of the proof of the following result can give an idea about how to use it.

Lemma 4.1. *If f is an even elliptic function, then it is a rational function of $\wp$.*

Recall that we consider Λ fixed, with such Λ we define $\wp$ and we assume that it is a set of periods of f i.e., $f(z)=f(z+\omega)$ for $\omega\in\Lambda$.

Proof (sketchy). Multiply f by factors of the form $\wp(z)-\wp(p)$, which vanish at $z=p$, to cancel all the poles p not in Λ. Subtracting a polynomial in $\wp$ the principal part at $z=0$ is also canceled. The result has no poles and must be constant by Liouville's theorem. □

The parity of the function is not essential if we complicate a little the statement.

Proposition 4.2. *If f is an elliptic function, then there exist $R_1, R_2\in\mathbb{C}(z)$, rational functions, such that $f=R_1(\wp)+\wp' R_2(\wp)$.*

Proof. Write

$$f=f_1+\wp' f_2 \qquad \text{with} \quad f_1(z)=\frac{f(z)+f(-z)}{2}, \quad f_2(z)=\frac{f(z)-f(-z)}{2\wp'(z)}.$$

Clearly, f_1 and f_2 are both elliptic and even (note that $\wp'$ is odd) and the result follows from Lemma 4.1. □

The functions $\wp$ and $\wp'$ are not algebraically independent, because $(\wp')^2$ is even and by Lemma 4.1 it must be a rational function function of $\wp$. In fact, revising the proof it must be a polynomial in $\wp$. Let us work out the numerical details.

Proposition 4.3. *We have the identity*

$$(\wp')^2=4\wp^3-g_2\wp-g_3 \qquad \textit{where} \quad g_2=60\sum_{\omega\in\Lambda^*}\omega^{-4} \quad \textit{and} \quad g_3=140\sum_{\omega\in\Lambda^*}\omega^{-6}.$$

Note that this can be rephrased saying that $\wp$ inverts the elliptic integral $\int \frac{dt}{\sqrt{4t^3-g_2t-g_3}}$ because the change $t=\wp(u)$ reduces it to $\int du$.

For the proof, we need the Laurent expansion of $\wp$ at 0.

Lemma 4.4. *The Laurent expansion of $\wp$ at the origin is*

$$\wp(z)=z^{-2}+\sum_{m=1}^{\infty}(2m+1)G_{2m+2}z^{2m} \qquad \textit{where} \quad G_k=\sum_{\omega\in\Lambda^*}\omega^{-k}$$

and it converges absolutely for $0\neq|z|<\inf\{|\omega| : \omega\in\Lambda^*\}$.

Proof. The result follows from the identity

$$(z+\omega)^{-2}-\omega^{-2}=\omega^{-2}\big((1+z/\omega)^{-2}-1\big)$$

substituting the Taylor expansion $(1+x)^{-2}-1=\sum_{m=1}^{\infty}(m+1)(-x)^m$ with convergence radius 1. Note that $G_m=0$ for m odd because $\omega^{-m}+(-\omega)^{-m}=0$. □

Proof of Proposition 4.3. By Lemma 4.4,

$$\wp(z)=z^{-2}+3G_4z^2+5G_6z^4+\dots$$

where the dots indicate higher order terms. Then

$$\big(\wp'(z)\big)^2=4z^{-6}-24G_4z^{-2}-80G_6+\dots \quad\text{and}\quad \big(\wp(z)\big)^3=z^{-6}+9G_4z^{-2}+15G_6+\dots$$

This implies that $\big(\wp'(z)\big)^2-4\big(\wp(z)\big)^3+g_2\wp+g_3$ is regular at $z=0$ and takes the value 0 there. By the double periodicity, it is also regular at any $z\in\Lambda$ and hence in $\mathbb{C}$. By Liouville's theorem, it is identically 0. □

Each equation $\wp(z)=s_0$ has two solutions in $\mathbb{C}/\Lambda$ essentially because $\wp$ has a double pole. For further reference, let us state a result in this direction.

Lemma 4.5. *For any $s_0\in\mathbb{C}$ the set of zeros of $\wp(z)-s_0$ is $\pm z_0+\Lambda$ for some $z_0\in\mathbb{C}$. The zeros are double if and only if $s_0\in\big\{\wp(\lambda_j)\big\}_{j=1}^3$ with $\lambda_1=\omega_1/2$, $\lambda_2=\omega_2/2$, $\lambda_3=(\omega_1+\omega_2)/2$.*

Proof. Let $\mathcal{P}$ be a fundamental parallelogram $0\in\mathcal{P}$ such that its boundary ∂P does not contain zeros or poles of $\wp(z)-s_0$. By the argument principle,

$$\frac{1}{2\pi i}\int_{\partial P}\frac{\wp'(z)}{\wp(z)-s_0}\,dz=Z-P=Z-2$$

where Z and P over the number of zeros and poles of $\wp(z)-s_0$ in $\mathcal{P}$ counting multiplicities (the double pole at $z=0$ gives $P=2$). The double periodicity implies that the integral vanishes and, consequently, $Z=2$. If z_0 is one of the zeros, $(-z_0+\Lambda)\cap\mathcal{P}$ is also a zero because the function is even. They coincide if and only if $2z_0\in\Lambda$, as $\wp(\Lambda)=\infty$, this is equivalent to $z_0\in\bigcup_{j=1}^3(\lambda_j+\Lambda)$. At these points there are double zeros because $\wp'(\lambda_j)=0$ is obtained differentiating $\wp(z)=\wp(2\lambda_j-z)$ at $z=\lambda_j$, which follows from the parity of $\wp$. □

4.3 Elliptic curves are actually toric

An *elliptic curve* is a nonsingular cubic (projective) curve. It can be proved [8] that after a birrational change of variables it can be written in the *Weierstrass form*

$$y^2=x^3+ax+b. \tag{1}$$

In principle we have $\mathbb{C}$ in mind, but this reduction can be done in any field K of definition excluding the cases $\operatorname{char}(K)=2,3$ (this means fields with $1+1=0$ or $1+1+1=0$) in which

some problems may appear [8, III.2]. For an elliptic curve E the usual notation is to write $E(K)$ to mean the points over K or to emphasize that the field of definition of E is K.

The nonsigularity implies that the right side of (1) has three distinct (complex) zeros. Computing the *discriminant* [11], the nonsingularity is equivalent to

$$4a^3 + 27b^2 \neq 0. \tag{2}$$

As a projective curve, (1) contain a point at infinity that we call O.

If we apply the change $(x, y) \mapsto (4x, 4y)$ to (1) at take into account (2), we obtain the alternative equation

$$y^2 = 4x^3 - Ax - B \qquad \text{with} \quad A^3 - 27B^2 \neq 0. \tag{3}$$

This form recalls to Proposition 4.3 and suggests that $(\wp, \wp')$ parametrizes elliptic curves. This is so and in fact the double periodicity of $\wp$ and $\wp'$ allows to show that any elliptic curve over $\mathbb{C}$ is actually a tours. This sounds good because topologists say that a torus has *genus* one and algebraic geometers say that an elliptic curve is a curve of genus one.

Our goal is:

Theorem 4.6. *For each elliptic curve E over $\mathbb{C}$ of the form (3) there exists a period lattice Λ such that $\phi : \mathbb{C}/\Lambda \longrightarrow E$ given by $\phi(z) = \big(\wp(z), \wp'(z)\big)$ is bijective.*

This establishes a conformal equivalence between the Riemann surface $\mathbb{C}/\Lambda$ and the compact Riemann surface corresponding to the curve. We understand that 0 goes to the point at infinity O.

A natural question is if actually the polynomial in Proposition 4.3 is always valid to have an elliptic curve.

Lemma 4.7. *The discriminant of $4x^3 - g_2x - g_3$ does not vanish. In other words, for any period lattice Λ the condition $g_2^3 - 27g_3^2 \neq 0$ is always fulfilled.*

Proof. By Lemma 4.5, we know that $\wp'$ vanishes at λ_1, λ_2 and λ_3. Hence, recalling Proposition 4.3,

$$4\wp^3 - g_2\wp - g_3 = 4\big(\wp - \wp(\lambda_1)\big)\big(\wp - \wp(\lambda_2)\big)\big(\wp - \wp(\lambda_3)\big).$$

Then $\{\wp(\lambda_j)\}_{j=1}^3$ are the zeros of $4x^3 - g_2x - g_3$ and they are distinct by Lemma 4.5. □

A second question is if it is possible to reach all valid A and B satisfying the condition in (3). We state the affirmative answer here but postpone the proof because it is related to the first modular function we are going to introduce in the next subsection.

Lemma 4.8. *Given $A, B \in \mathbb{C}$ with $A^3 - 27B^2 \neq 0$, there exists Λ such that $A = g_2$ and $B = g_3$.*

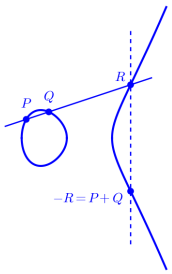
Proof of Theorem 4.6. The existence of Λ is guaranteed by Lemma 4.8 and Proposition 4.3. The map is injective because $\phi^{-1}(O) = 0$ and if $\wp(z_1) = \wp(z_2)$ for $z_1, z_2 \neq 0$ distinct then $z_1 = -z_2$ in $\mathbb{C}/\Lambda$ by Lemma 4.5 with $z_1 \neq \lambda_1, \lambda_2, \lambda_3$ (in these cases $z_1 = z_2$) and $\wp'(z_1) = \wp'(z_2) = -\wp(z_1)$ cannot hold.

The surjectivity follows similar lines. Given $(x_0, y_0) \in E - \{O\}$, the equation $\wp(z) = x_0$ has solutions $\pm z_0$ in $\mathbb{C}/\Lambda$ by Lemma 4.5 and $y_0 = \wp(z_0)$ or $y_0 = -\wp(z_0) = \wp(-z_0)$ by Proposition 4.3. Hence, for some choice of the sign, $\phi(\pm z_0) = (x_0, y_0)$. □

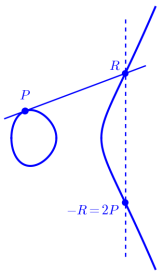
We have proved that an elliptic curve E over $\mathbb{C}$ is a torus. On the other hand, the flat torus $\mathbb{C}/\Lambda$ is trivially an abelian group with the addition modulo Λ. Then the curve (3) inherits the abelian group structure. Using $\phi(0) = O$ and $\phi(-z) = \big(\wp(z), -\wp'(z)\big)$ it is deduced that O is the identity element and $(x_0, -y_0) \in E$ is the inverse of $(x_0, y_0) \in E$. This is also preserved in the original form (1) because the parametrization would be $(4\wp, 4\wp')$. It turns out that the group law on the elliptic curve has a simple geometric meaning.

Proposition 4.9. *Let E be an elliptic curve of the form (1). It becomes an abelian group taking the point O as the identity element, applying the symmetry $(x, y) \mapsto (x, -y)$ to get the inverse element and defining $P + Q$ for $P, Q \in E - \{O\}$ as the inverse of the third intersection point of the secant line connecting P and Q with the curve. If $P = Q$ this line is considered to be tangent to the curve.*

Taking the geometric interpretation as the definition of the group law would lead to problems to prove associativity. Via Theorem 4.6 it follows from the associativity in $(\mathbb{C}, +)$.



Adding $P + Q$



Adding $P + P$

$$P + Q + R = O$$

$$(P_1 + P_2) + P_3 = P_1 + (P_2 + P_3)$$
$$P_1 + P_2 = P_2 + P_1$$
$$-P = \text{symmetric of } P$$
$$O + P = P$$

After this geometric interpretation, it is clear that the group law works finely in any subfield of $\mathbb{C}$, in particular in $E(\mathbb{Q})$ when the elliptic curve is defined over $\mathbb{Q}$. Noting that the actual formulas are rational functions involving integer numbers, the group law also extends to finite fields $E(\mathbb{F}_q)$. In general, the third intersection of a line and a cubic can be computed in any field.

A more advanced way of bypassing the problems with the associativity and to consider general fields is to use Riemann-Roch theorem to prove that $P \mapsto (P) - (O)$ establishes a bijection between $E(K)$ and $\mathrm{Pic}^0(E)$, the group of divisor of degree zero under equivalence.

Proof of Proposition 4.9. After the previous comments about O and the inverse, we have to prove that given $P_1, P_2, P_3 \in E - \{O\}$ collinear points and under the tangency condition if two of them coincide, it holds $z_1 + z_2 + z_3 = 0$ in $\mathbb{C}/\Lambda$ where $z_j = \phi^{-1}(P_j)$ with ϕ as in Theorem 4.6.

Say that the P_j lie on the line $Ax + By + C = 0$. Then z_1, z_2 and z_3 are zeros of $F(z) = A\wp(z) + B\wp'(z) + C$ and they are still three counting multiplicities if two of the points coincide by the tangency condition. Let $\mathcal{P}$ be a fundamental parallelogram $\mathcal{P}$ without zeros and poles of F on the boundary. With a slight abuse of notation we identify the z_j with complex numbers in $\mathcal{P}$ with the same name. For $g = F'/F$ we have

$$\frac{1}{2\pi i}\int_{\partial\mathcal{P}} z g(z)\,dz = z_1 + z_2 + z_3 \qquad \text{and} \qquad z g(z) - (z+\omega)g(z+\omega) = -\omega g(z) \quad \text{for } \omega \in \Lambda.$$

The integral follows from residue theorem (the poles of $zg(z)$ are at the z_j) and the identity is a trivial consequence of $g(z+\omega) = g(z)$.

The boundary of $\mathcal{P}$ is composed by four oriented segments c_1, c_2, c_3 and c_1. The identity partially cancels the integrals in parallel sides, resulting

$$z_1 + z_2 + z_3 = \frac{1}{2\pi i}\int_{\partial\mathcal{P}} zg(z)\,dz = -\frac{\omega_2}{2\pi i}\int_{c_1} g(z)\,dz - \frac{\omega_1}{2\pi i}\int_{c_2} g(z)\,dz.$$

Note that the last integrals divided by $2\pi i$ are the *winding number* [1] of the closed curves $F\circ c_1$ and $F\circ c_2$ (they are closed by the periodicity). Then $z_1+z_2+z_3 \in \Lambda$, equivalently, this sum is zero in $\mathbb{C}/\Lambda$. □

4.4 At last something modular

The elements of $\mathrm{SL}_2(\mathbb{Z})$ are linear isomorphism of $\mathbb{Z}^2$. In the same way, they preserve period lattices acting on the generators. Namely, if $\Lambda = \omega_1\mathbb{Z} + \omega_2\mathbb{Z}$ and $\Lambda' = \omega_1'\mathbb{Z} + \omega_2'\mathbb{Z}$ then

$$\begin{pmatrix}\omega_1'\\ \omega_2'\end{pmatrix} = \gamma\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix} \qquad \text{with} \quad \gamma\in\mathrm{SL}_2(\mathbb{Z}) \quad \Rightarrow \quad \Lambda = \Lambda'.$$

Under the normalization condition $\Im(\omega_1/\omega_2)$, $\Im(\omega_1'/\omega_2') > 0$, the converse is also true because it rules out the possibility of $\det(\gamma) = -1$.

The dilations change the lattice but they do not essentially change the associate elliptic curve. This is due to the lack of uniqueness of the equations (1) and (3). For instance, the change $(x,y)\mapsto(-x,iy)$ transform the equation $y^2 = x^3+1$ into $y^2 = x^3-1$ and in principle there is no reason to prefer one of these equations. If $y^2 = 4x^3 - g_2x - g_3$ is the equation corresponding to a lattice, the corresponding to $\Lambda' = \lambda\Lambda$ is $y^2 = 4x^3 - \lambda^{-4}g_2x - \lambda^{-6}g_3$ (recall the definitions of g_2 and g_3) and the latter comes from the former after a change of variables $(x,y)\mapsto(\lambda^2x,\lambda^3y)$. This suggests to consider something invariant, not depending on the particular form of the equation. Any function of g_3^2/g_2^3 does the job. Taking the inverse of $1 - 27g_3^2/g_2^3$ we have a infinite free function thanks to Lemma 4.7. A normalization factor $1728 = 12^3$ is introduced by arithmetic reasons to define

$$J(\omega_1,\omega_2) = \frac{1728g_2^3}{g_2^3 - 27g_3^2}.$$

By previous comments this function is invariant by $(\omega_1,\omega_2)\mapsto(a\omega_1+b\omega_2, c\omega_1+d\omega_2)$ where a, b, c, d are the entries of any matrix of the modular group, and it is also invariant under $(\omega_1,\omega_2)\mapsto(\lambda\omega_1,\lambda\omega_2)$. In particular, it can be considered as a function of $z = \omega_1/\omega_2 \in \mathbb{H}$. The formulas for g_2 and g_3 show that they are holomorphic on ω_1 when ω_2 is fixed. Then we have proved

Lemma 4.10. *The so called* j function *or* Klein's j-invariant *defined as* $j(z) = J(z,1)$ *is an holomorphic function* $j:\mathbb{H}\longrightarrow\mathbb{C}$ *satisfying*

$$j(\gamma z) = j(z) \qquad \textit{for any} \quad \gamma\in\mathrm{SL}_2(\mathbb{Z}).$$

In a somewhat more explicit form, we have

$$j(z) = \frac{34560\big(\sum{}'(mz+n)^{-4}\big)^3}{20\big(\sum{}'(mz+n)^{-4}\big)^3 - 49\big(\sum{}'(mz+n)^{-6}\big)^2} \tag{4}$$

where the prime indicates summation over $(m,n) \in \mathbb{Z}^2 - \{(0,0)\}$.

This function is employed below to prove Lemma 4.8 and its interest goes largely beyond. It is the function giving the conformal equivalence between $X(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^*$ and the Riemann sphere, it allows to give a short proof [2, §2.9] of *Picard's little Theorem* (a deep result in complex analysis: if f is entire, the cardinality of $\mathbb{C} - f(\mathbb{C})$ is at most 1) and it also has arithmetic significance when studying factorization in quadratic fields. In 1992 it was discovered an astonishing relation between j and the *monster group* (the largest sporadic simple group, with order greater than $8\cdot 10^{53}$) that was a Fields medal worth and has generated the so called *moonshine theory*.

The key to prove Lemma 4.8 is the surjectivity of j.

Proposition 4.11. *The map $j : \mathbb{H} \longrightarrow \mathbb{C}$ is surjective.*

Proof. If $j(\mathbb{H}) \neq \mathbb{C}$ choose w_0 on the boundary of the open set $j(\mathbb{H})$ (nonconstant holomorphic functions are open). There exists a sequence $\{z_n\}_{n=1}^{\infty}$ such that $j(z_n) \to w_0$. By Lemma 4.10 we may assume $z_n \in \overline{\mathcal{F}}$. If this sequence remains bounded, by Bolzano-Weierstrass theorem we get $z_{n_k} \to z_0$ with $j(z_0) = w_0$, which contradicts $w_0 \notin j(\mathbb{H})$. Then there exists z_{n_k} unbounded in $\overline{\mathcal{F}}$ such that $j(z_{n_k}) \to w_0$. The numerator in (4) goes to $34560\big(\sum_{n\in\mathbb{Z}-\{0\}} n^{-4}\big)^3 > 0$ and the denominator to

$$20\Big(\sum_{n\in\mathbb{Z}-\{0\}} n^{-4}\Big)^3 - 49\Big(\sum_{n\in\mathbb{Z}-\{0\}} n^{-6}\Big)^2 = 160\zeta(4)^3 - 196\zeta(6)^2 = 0$$

where have been used the special values [12] $\zeta(4) = \pi^4/90$ and $\zeta(6) = \pi^6/945$ of the *Riemann zeta function* $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$. This proves $j(z_{n_k}) \to \infty$ leading again to a contradiction. □

Proof of Lemma 4.8. Assume $A \neq 0$. By Proposition 4.11 there exists z_0 such that

$$j(z_0) = \frac{1728A^3}{A^3 - 27B^2} = \frac{1728}{1 - 27B^2/A^3}.$$

Let $g_2(\Lambda)$ and $g_3(\Lambda)$ be the values of g_2 and g_3 corresponding to the lattice $\Lambda = z_0\mathbb{Z} + \mathbb{Z}$, then $B^2/A^3 = g_3^2(\Lambda)/g_2^3(\Lambda)$. For $\lambda \in \mathbb{C} - \{0\}$, $g_2(\lambda\Lambda) = \lambda^{-4}g_2(\Lambda)$. Choosing as λ any of the fourth roots of $g_2(\Lambda)/A$ we have $g_2(\lambda\Lambda) = A$ that implies $B^2 = g_3^2(\lambda\Lambda)$ and selecting the fourth root we can force $B = g_3(\lambda\Lambda)$.

The case $A = 0$ is much simpler because necessarily $g_2(\Lambda) = 0$ and it is enough to adjust λ to get $B = g_3(\lambda\Lambda)$. □

Of course, j is not injective because $j(\gamma z) = j(z)$, but it becomes injective when we identify the points in the same orbit.

Theorem 4.12. *Klein's j-invariant defines a bijection $j : \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H} \longrightarrow \mathbb{C}$.*

Proof. The map is well defined by Lemma 4.10 and it is surjective by Proposition 4.11. It remains to show that $j(z_1) = j(z_2)$ implies that there exists $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $z_2 = \gamma z_1$.

Let $\Lambda = z_0\mathbb{Z}+\mathbb{Z}$ and $\Lambda = \lambda z_2\mathbb{Z}+\lambda\mathbb{Z}$. Proceeding as in the proof of Lemma 4.8 we can adjust the value of λ in such a way that $g_2(\Lambda) = g_2(\Lambda')$ and $g_3(\Lambda) = g_3(\Lambda')$. Let $\wp$ the Weierstrass function corresponding to Λ. Differentiating in Lemma 4.3, canceling $\wp'$ and substituting the Laurent expansion of Lemma 4.4, it is deduced for $m > 2$ (Exercise 7)

$$m(2m-1)a_m = 6a_m + 3\sum_{k=1}^{m-2} a_k a_{m-1-k} \qquad \text{with} \quad a_m = (2m+1)G_{2m+2}$$

This proves that $\wp$ is determined by $a_1 = 3G_4 = \frac{1}{20}g_2(\Lambda)$ and $a_2 = 5G_6 = \frac{1}{28}g_3(\Lambda)$. In particular, $g_2(\Lambda)$ and $g_3(\Lambda)$ determine Λ, the set of poles of $\wp$, and we conclude $\Lambda = \Lambda'$ that implies

$$\begin{pmatrix}\lambda z_2\\ \lambda\end{pmatrix} = \gamma\begin{pmatrix}z_1\\ 1\end{pmatrix} \qquad \text{for some} \quad \gamma \in \mathrm{SL}_2(\mathbb{Z}).$$

Dehomogenizing (dividing both coordinates) $z_2 = \gamma z_1$. □

Most of the authors would agree saying that Lemma 4.10 implies that j is a *weakly modular function*, but the actual definition of what a *modular function* actually is, varies a lot. Here we follow [9] and [2] defining a modular function as a meromorphic function $f : \mathbb{H} \longrightarrow \mathbb{C}$ such that $f(\gamma z) = f(\gamma)$ which is also meromorphic at infinity.

Now we have to clarify the meaning of the last concept. A 1-periodic function $f : \mathbb{H} \longrightarrow \mathbb{C}$ is said to be *meromorphic at infinity* if it admits a Fourier expansion of the form

$$f(z) = \sum_{n=-N}^{\infty} a_n\, e^{2\pi i n z} \qquad \text{converging in some region} \quad \Im(z) > c. \tag{5}$$

If $a_{-N} \neq 0$ it is said to have a *pole of order* N at infinity if $N \in \mathbb{Z}^+$ and a *zero of order* $-N$ at infinity if $N \in \mathbb{Z}^-$. This weird definition becomes clear if we note that $z \mapsto q(z) = e^{2\pi z}$ maps any width 1 vertical strip onto the punctured unit disk with $q(i\infty) = 0$. It defines a chart at infinity for the Riemann surface $\langle T\rangle\backslash\mathbb{H}^*$ in which f becomes $\sum_{n=-N}^{\infty} a_n q^n$ and the concepts of zeros and poles are as usual.

We have said that j is weakly modular because we do not know if it is meromorphic at infinity. If fact, it is. We postpone the proof of this result to the next chapter.

Theorem 4.13. *The function j is a modular function with a pole of order 1 at infinity.*

We shall see that the Fourier coefficients in (5) are integers (it is a reason a introduce the constant 1728 in the definition) being the first terms in the Fourier expansion

$$j(z) = e^{-2\pi i z} + 744 + 196884\, e^{2\pi i z} + 21493760\, e^{4\pi i z} + \ldots$$

In connection with this, let us briefly mention one of the amazing deep properties of j. Recall that we considered the equivalence of positive definite binary primitive quadratic forms. It can be proved (but it is quite difficult [4, §12]) that if for a discriminant there is only a class then for the $z_Q \in \mathbb{H}$ corresponding to these quadratic forms $j(z)$ is the cube of an integer!

The largest discriminant with only a class is -163. Namely, every quadratic form of discriminant -163 is equivalent to

$$Q = x^2 + xy + 41 \qquad \text{that has} \quad z_Q = \frac{-1 + i\sqrt{163}}{2}.$$

The Fourier coefficients of j grow quickly but if $\Im(z)$ is large the imaginary exponential wins. Taking the two first terms, one deduces that $e^{-2\pi i z_Q} + 744$ must be very close to and integer, actually, to a cube. The numerical calculations are quite convincing revealing $(-640320)^3$ as that cube:

$$e^{\pi\sqrt{163}} = 6253741264076874 3.99999999999925007\cdots \approx 640320^3 + 744$$

and, extracting cubic roots,

$$\sqrt[3]{e^{\pi\sqrt{163}} - 744} = 640319,999999999999999999999939031\ldots$$

For a quick self-contained view of these topics, see [3].

After this detour, assuming Theorem 4.13, we are going to show that j generates in some sense all the modular functions. We start with a result whose proof is essentially a repetition of that of Proposition 4.11.

Proposition 4.14. *The only bounded modular functions are constants.*

Proof. Let $f : \mathbb{H} \longrightarrow \mathbb{C}$ be a nonconstant bounded modular form. Then we can take $N = 0$ in (5) and $\lim f(z) = a_0$ when $\Im(z) \to \infty$. The boundedness implies that f is holomorphic and $f(\mathbb{H})$ is a bounded open set. Let w_0 be a point on its boundary different from a_0. There exists a sequence $\{z_n\}_{n=1}^{\infty}$ that we can assume to be in $\overline{\mathcal{F}}$ such that $f(z_n) \to w_0$. This sequence is bounded because $w_0 \neq a_0$ and Bolzano-Weierstrass theorem allows to get a subsequence $z_{n_k} \to z_0$ with $f(z_0) = w_0$, which contradicts $w_0 \notin f(\mathbb{H})$. □

Theorem 4.15. *Any modular function f is of the form $f = R(j)$ with R a rational function. In fact, if f is holomorphic, R is a polynomial.*

Proof. If f has a pole of order N at infinity then $g(z) = f(z) - \sum_{n=-N}^{0} a_n \big(j(z)\big)^n$, with the notation as in (5), tends to 0 when $\Im(z) \to \infty$ and Proposition 4.14 proves $g = 0$ giving the last part of the result (g bounded in $\mathcal{F}$ implies g bounded in $\mathbb{H}$).

If f is not holomorphic, apply the same argument replacing $f(z)$ by $f(z)\prod_k \big(j(z) - j(p_k)\big)$ where p_k are the poles of f in $\overline{\mathcal{F}}$ repeated according their multiplicities. □

Exercises

Exercise 1. Prove that for any $z \notin \Lambda$ the series defining $\wp$ converges.

Exercise 2. Prove that $\wp'' + \frac{1}{2}g_2 = 6\wp^2$.

Exercise 3. Using the previous exercise, prove the unbelievable identity $3G_4^2 = 7G_8$ with the notation as in Lemma 4.4.

Exercise 4. Consider an elliptic curve in the form (3). Prove that if $(x_1, y_1), (x_2, y_2) \in E - \{O\}$ with $x_1 \neq x_2$, the first coordinate x_3 of the third point of intersection of the secant line satisfies $4(x_1 + x_2 + x_3) = m^2$ where m is the slope of the line connecting (x_1, y_1) and (x_2, y_2). Deduce from it the following addition formula for $\wp$:

$$\wp(z+w) = -\wp(z) - \wp(w) + \frac{1}{4}\Big(\frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)}\Big)^2.$$

Exercise 5. Prove the identity

$$16\wp(2z) + 32\wp(z) = \frac{\big(12\wp(z)^2 - g_2\big)^2}{4\wp(z)^3 - g_2\wp(z) - g_3}$$

Exercise 6. Assuming $\omega_1/\omega_2, \omega_1'/\omega_2' \in \mathbb{H}$, give a detailed proof of $\omega_1\mathbb{Z} + \omega_2\mathbb{Z} = \omega_1'\mathbb{Z} + \omega_2'\mathbb{Z}$ if and only if there exists $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\left(\begin{smallmatrix}\omega_1'\\ \omega_2'\end{smallmatrix}\right) = \gamma \left(\begin{smallmatrix}\omega_1\\ \omega_2\end{smallmatrix}\right)$.

Exercise 7. Proceeding as suggested in the proof of Theorem 4.12 prove for $m > 2$

$$(2m^2 - m)a_m = 6a_m + 3\sum_{k=1}^{m-2} a_k a_{m-1-k} \qquad \text{with} \quad a_m = (2m+1)G_{2m+2}$$

and explain how to deduce that for each m there exists a polynomial $P_m \in \mathbb{Q}[x, y]$ such that $G_{2m+2} = P_m(g_2, g_3)$.

Exercise 8. Prove $j(i) = 1728$.

Exercise 9. Prove $j(e^{2\pi i/3}) = 0$.

Exercise 10. Write a sagemath code to approximate $j(z)$ for any $z \in \mathbb{H}$ following the scheme: Find $z' \in \overline{\mathcal{F}}$ in the same orbit as z and apply (4) with $z = z'$ and the sums restricted to $|m|, |n| \leq 20$. Check the approximation trying $j\big(\frac{17}{10} + \frac{i}{10}\big) = 1728$ and $j\big(\frac{23}{14} + \frac{i}{14}\sqrt{3}\big) = 0$ that follow from the previous exercises.

References

[1] L. V. Ahlfors. *Complex analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York, third edition, 1978. An introduction to the theory of analytic functions of one complex variable.

[2] T. M. Apostol. *Modular functions and Dirichlet series in number theory*, volume 41 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[3] F. Chamizo and D. Raboso. Formas modulares y números casi enteros (Spanish). *Gac. R. Soc. Mat. Esp.*, 13(3):539–555, 2010.

[4] D. A. Cox. *Primes of the form $x^2 + ny^2$*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1989. Fermat, class field theory and complex multiplication.

[5] G. H. Hardy. *The integration of functions of a single variable*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 2. Hafner Publishing Co., New York, 1971. Reprint of the second edition, 1916.

[6] Y. Hellegouarch. *Invitation to the mathematics of Fermat-Wiles*. Academic Press, Inc., San Diego, CA, 2002. Translated from the second (2001) French edition by L. Schneps.

[7] M. D. Hirschhorn. A simple proof of Jacobi's four-square theorem. *Proc. Amer. Math. Soc.*, 101(3):436–438, 1987.

[8] A. W. Knapp. *Elliptic curves*, volume 40 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 1992.

[9] G. Shimura. *Introduction to the arithmetic theory of automorphic functions*. Kanô Memorial Lectures, No. 1. Iwanami Shoten Publishers, Tokyo; Princeton University Press, Princeton, NJ, 1971. Publications of the Mathematical Society of Japan, No. 11.

[10] J. Stillwell. *Mathematics and its history*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1989.

[11] Wikipedia contributors. Discriminant — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Discriminant&oldid=1276987913`, 2025. [Online; accessed 24-February-2025].

[12] Wikipedia contributors. Particular values of the Riemann zeta function — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Particular_values_of_the_Riemann_zeta_function&oldid=1270810862`, 2025. [Online; accessed 25-February-2025].