Modular Functions and Picard's Theorem

Richard Chapling*

v1 1 March 2016

We have all met the following result of supreme coolness in complex anaylsis:

Theorem 1 (Liouville's theorem). *A bounded entire function is constant.*

There are various very simple proofs of this, but perhaps the easiest one to understand conceptually is Nelson's proof, which works for any function with the mean-value property, that the value at a point is the average value of the function over a ball centred at this point. Imagine enclosing two points in such balls. Then if we take the radii equal and extremely large, the vast majority of the points inside one ball are inside the other, and we can make the proportion as close to 1 as we like. Hence the integrals can be made as close as we like, and the values at the two original points must be equal.[1]

But this isn't a deep result at all! Can we do better? The answer is very much yes, as was discovered by Picard in the late nineteenth century:[2]

Theorem 2 (Picard's little theorem). *Let f be an entire function whose image omits two points. Then f is constant.*

The usual two thoughts will cross your mind if you've not seen this before:

1. WHAT?!

2. How on Earth could you prove that?

This handout is dedicated to answering the second of these.[3] In particular, we shall prove this the classical way: by conjuring up a function that has the properties we would like.

1 Initial Progress

At the moment, the problem looks totally intractable. Let's think about how we can simplify it into something we have a hope of solving.

Firstly, we can simplify to taking 0 and 1 as the missing points. This is clear: if $f(z)$ misses a and b, $g(z) = \frac{f(z)-a}{b-a}$ misses 0 and 1.

How do we show that a function is constant? Well, we have Liouville's theorem; can we use that? The usual way to apply Liouville's theorem, which you probably remember from early complex analysis, is to compose $f : \mathbb{C} \to A$ with a function that maps A to a bounded set.

* Trinity College, Cambridge

[1] This is not much shorter than Nelson's original exposition, in the paper *A Proof of Liouville's' Theorem*, Proceedings of the American Mathematical Society, **12**, (1961), p. 995.

[2] Picard, É, "Sur une propriété des fonctions entières" *C.R. Acad. Sci. Paris* 88 (1879) 1024–1027.

[3] Resolving the first is left as an exercise to the reader.

A prototypical bounded set is the unit ball, which is not always easy to work with, but the unit ball is conformally equivalent to the upper half-plane. (Consider the Möbius transformation $h(z) = (z+i)/(z-i)$, for example.)

If we take a function $\mathbb{C} \to \mathbb{H}$, it will necessarily be multivalued (otherwise it would be analytic and bounded, which we know is no good!). Hence it is more natural to take the inverse to start with, and produce a function on the upper half-plane, that misses 0 and 1. Such a function must have a natural boundary to analytic continuation on the real axis, and so have infinitely many singularities on the real axis.

Therefore, we want an analytic function $\mathbb{H} \to \mathbb{C} \setminus \{0, 1\}$. What are functions that naturally live on the upper half-plane? Modular functions!

2 Summary of Results About the Weierstrass Elliptic Function

This section is included for completeness. Those already satisfied with their elliptic function limit points can skip to the next section.

2.1 Lattices

A *Lattice* is a discrete additive subgroup of $\mathbb{C}$ with two generators, i.e.

$$\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \{m\omega_1 + n\omega_2 : n, m \in \mathbb{Z}\}. \tag{1}$$

In particular, the quantity $\tau = \omega_2/\omega_1$, which we call the *modulus* of the lattice, is not real. By choosing the order of ω_1 and ω_2 appropriately, we can take $\tau \in \mathbb{H}$, the upper half-plane $\{\tau : \Im(\tau) > 0\}$.

One can show that two lattices are equivalent up to rotation and scaling if and only if their is a Möbius transformation $T(z) = (az+b)/(cz+d)$ with $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$, such that $T(\tau) = \tau'$. (T is called a *modular transformation*, for obvious reasons. The group of such transformations is called the *modular group*, denoted Γ.)

2.2 Elliptic functions

It is easy to show the following results about elliptic functions using contour integration:

Theorem 3. *An analytic elliptic function is constant.*

Hence we take meromorphic doubly-periodic functions as our elliptic functions. The total order of the poles of an elliptic function is called its *order*.

Lemma 4. *Let f be elliptic. Then the sum of the residues of f in a fundamental domain is zero.*

Lemma 5. *Let f be elliptic of order N. Then f takes on every complex value exactly N times in a fundamental domain.*

2.3 The Weierstrass elliptic function

Definition 6 (Weierstrass pe-function). The Weierstrass elliptic function associated with a lattice Λ is

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda}{}' \frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}, \tag{2}$$

where the prime on the sum shall denote that the sum excludes terms with zero denominator.

$\wp$ is an even elliptic function, and has order 2. We can easily show that it has Laurent series

$$\wp(z) = \frac{1}{z^2} + \sum_{k=1}^{\infty} (2n+1) z^{2k} G_{2k+2}, \tag{3}$$

where G_{2k} denotes the Eisenstein series

$$G_{2k} = \sum_{\omega \in \Lambda}{}' \frac{1}{\omega^{2k}}. \tag{4}$$

These turn out to be prototypical *modular forms*, but we don't have space to discuss such things. Writing $g_2 = 60G_4, g_3 = 140G_6$, we can show by cancelling the poles that $\wp$ satisfies the following differential equation:

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3. \tag{5}$$

By the above results, $\wp'$ is an odd elliptic function of order 3. Hence it has 3 roots in a fundamental domain, and it is easy to check that they are at

$$e_i = \wp(\omega_i/2), \tag{6}$$

where $\omega_3 = -\omega_1 - \omega_2$. Moreover, e_1, e_2, e_3 are distinct, because otherwise $\wp(z) - e_i$ would have too many zeros in a fundamental domain.

3 The Modular Lambda-Function

3.1 Definition, modular properties and fundamental domain

With the elliptic functions out of the way, we make the following

Definition 7. Let the periods of a lattice be $(1, \tau)$. The *modular λ-function* is

$$\lambda(\tau) = \frac{e_3 - e_2}{e_1 - e_2}, \tag{7}$$

where e_i are given in the previous section.

Because thte e_i are unequal, we immediately have

Lemma 8. *$\lambda : \mathbb{H} \to \mathbb{C}$ misses 0 and 1.*

Ah-ha! Right, now we just need to prove that $\lambda : \mathbb{H} \to \mathbb{C}$ is surjective ....

One thing that we need to know about λ is that it is not an invariant of the lattice: permuting the ω_i gives six different values of λ, namely

$$\lambda, 1-\lambda, \frac{1}{\lambda}, 1-\frac{1}{\lambda}, \frac{1}{1-\lambda}, \frac{\lambda}{1-\lambda}, \tag{8}$$

which are of course related by the Möbius transformations that permute the set $\{0, 1, \infty\}$.

On the other hand, it is easy to check that any Möbius transformation congruent to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ modulo 2 preserves the half-period lattice, and thus the ordering of the e_i. Such matrices form a proper subgroup $\Gamma(2)$ of the modular group; one can verify that it is generated by

$$\tau \mapsto \tau + 2, \qquad \tau \mapsto \frac{\tau}{1+2\tau}. \tag{9}$$

This is in fact the total group under whose action on the upper half-plane λ is preserved.

Let us look for a fundamental domain for this group (i.e., a set which contains exactly one member of each orbit of $\Gamma(2)$). It is clear that it must lie between two lines a distance 2 apart, since λ has real period 2; we may as well take these as $\Re(\tau) = \pm 1$. On the other hand, the other generator preserves the circle $|\tau - 1/2| = 1/2$, so we find that the fundamental region is bounded by the lines $\Im(\tau) = \pm 1$ and the circles $|\tau \pm 1/2| = 1/2$.

3.2 Fourier series and large-$\Im(\tau)$ behaviour

To prove that λ is surjective, we'll need to know how λ behaves as $\Im(\tau) \to \infty$; the simplest way to find this is to calculate the first few terms of its Fourier series. Recall the trigonometrical identity

$$\pi^2 \operatorname{cosec}^2 \pi z = \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2}; \tag{10}$$

this means that we can calculate $\wp$ as a sum of cosecants:

$$\begin{aligned}
\wp(z) &= \frac{1}{z^2} + \sum_{m,n}{}' \frac{1}{((z-n\tau)-m)^2} - \frac{1}{(-n\tau - m)^2} &(11)\\
&= \pi^2 \operatorname{cosec}^2 \pi z - 2\sum_{n=0}^{\infty} \frac{1}{n^2} + \sum_{n}{}' \left(\pi^2 \operatorname{cosec}^2 \pi(z-n\tau) - \pi^2 \operatorname{cosec}^2 n\tau\right) &(12)\\
&= \pi^2 \left(\operatorname{cosec}^2 \pi z - \frac{1}{3}\right) + \pi^2 \sum_{n}{}' \left(\operatorname{cosec}^2 \pi(z-n\tau) - \operatorname{cosec}^2 n\tau\right). &(13)
\end{aligned}$$

A simpler formula exists for $\wp(z) - \wp(w)$:

$$\begin{aligned}
\wp(z) - \wp(w) &= \sum_{m,n} \frac{1}{(z-m-n\tau)^2} - \frac{1}{(w-m-n\tau)^2} &(14)\\
&= \pi^2 \sum_{n} \operatorname{cosec}^2 \pi(z-n\tau) - \operatorname{cosec}^2 \pi(w-n\tau), &(15)
\end{aligned}$$

where we don't have to exclude terms, so the calculation is much simpler.

Now, we use this to find the large-$\Im(\tau)$ behaviour of λ. We have $e_1 = \wp(1/2)$, $e_2 = \wp(\tau/2)$, $e_3 = \wp(1/2 + \tau/2)$, so

$$\begin{aligned}
e_3 - e_2 &= \pi^2 \sum_{n} \sec^2 \pi(n-1/2)\tau - \operatorname{cosec}^2 \pi(n-1/2)\tau &(16)\\
e_1 - e_2 &= \pi^2 \sum_{n} \sec^2 \pi n\tau - \operatorname{cosec}^2 \pi(n-1/2)\tau &(17)
\end{aligned}$$

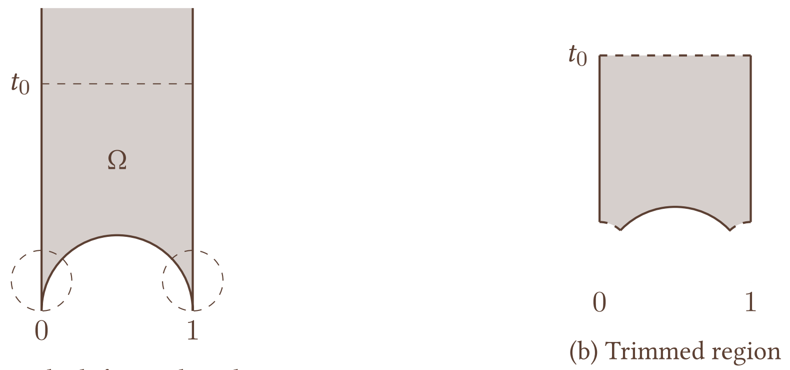
It is now just a matter of writing these in terms of $e^{in\pi\tau}$ and expanding the sums. It turns out that

$$\lambda(\tau) \sim 16 e^{i\pi\tau} \tag{18}$$

as $\Im(\tau) \to \infty$; hence we discover that $\lambda(\tau) \to 0$ as well.

If τ is pure imaginary, it is easy to see that λ is real, since the terms in the sums above just turn into real hyperbolic functions. Since

$$\lambda(\tau+1) = 1 - \frac{1}{\lambda(\tau) - 1}, \tag{19}$$



(a) Contour on which λ is real, with interior Ω

(b) Trimmed region

Figure 1: Regions considered in the proof of the surjectivity of λ

we also find that λ is real on all lines of integer real part. Similarly, $\tau \mapsto 1 - 1/\tau$ maps the imaginary axis to the circle $|\tau - 1/2| = 1/2$, so λ is real on the boundary of the region Ω contained between these, pictured in Figure 1a. We can also verify using these transformations and the result about large imaginary parts to derive $\lambda(\tau) \to 1$ as $\tau \to 0$ and $\lambda(\tau) \to \infty$ as $\tau \to 1$.

We trim the region Ω along the dashed lines (Figure 1b): the circles are the images of the line $\Im(\tau) = t_0$ under the group action.

Considering the action of the group elements on λ, we find that for large t_0, the three segments we added all map to approximate circles in the upper half-plane: $e^{i\pi\tau}$, $1 - e^{i\pi\tau}$, and $e^{-i\pi\tau} - 1$. Even better, it is apparent by shifting t_0 that the interior of the contour must be the part contained in the upper half-plane (Figure 2), and as we send $t_0 \to \infty$, the contour encloses more and more of the upper half-plane without limit. Moreover, the way we have considered the rest of the contour, along with some simple real analysis, shows us that it only encircles a point at most once. It follows that any $w \in \mathbb{H}$ is the image of precisely one $\tau \in \Omega$, and a bit more real analysis (in the form of the Intermediate Value Theorem) convinces us that each real value (excluding 0 and 1) is taken once.

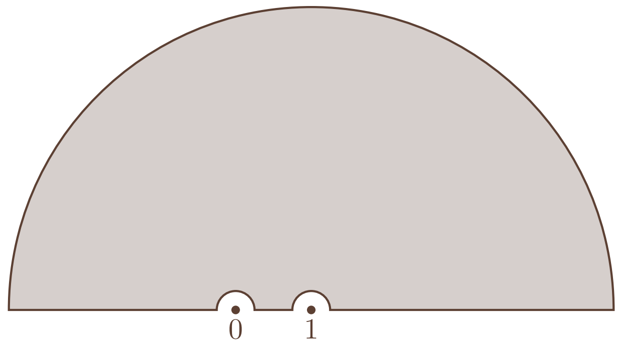


Figure 2: Image of the contour in Figure 1b and its interior under λ.

Lastly, since λ is analytic on $\mathbb{H}$ and real on the line $\Re(\tau) = 0$, we must have $\lambda(-a + bi) = \overline{\lambda(a + bi)}$, because these functions are both analytic and their difference vanishes on the line. (This is an application of the Schwarz Reflection Principle.) Hence λ maps the corresponding region to the left of Ω (call it Ω' onto the lower half-plane. We conclude that

Theorem 9. *λ is a bijection* $(\bar{\Omega} \cup \Omega') \cap \mathbb{H} \to \mathbb{C} \setminus \{0, 1\}$.

Therefore we can use the inverse of this function to prove Picard's Little Theorem!

There is one sticking point, which is to ensure the analyticity of $\lambda^{-1} \circ g$. The simplest thing to say here is that the Monodromy Theorem saves us by allowing us to just define this by analytic continuation starting from a specific point in $\mathbb{C}$. Bluntly, because $\lambda'(\tau) \neq 0$ for $\tau \in \mathbb{H}$, λ is an infinitely-branched covering map of $\hat{\mathbb{C}} \setminus \{0, 1, \infty\}$ by $\mathbb{H}$, so there is a well-defined analytic way to walk around on the Riemann surface of its inverse (viz., $\mathbb{H}$) using g: a simpler but similar phenomenon occurs with the complex exponential, which you can think of as an infinitely-branched covering of $\hat{\mathbb{C}} \setminus \{0, \infty\}$ by $\mathbb{C}$.

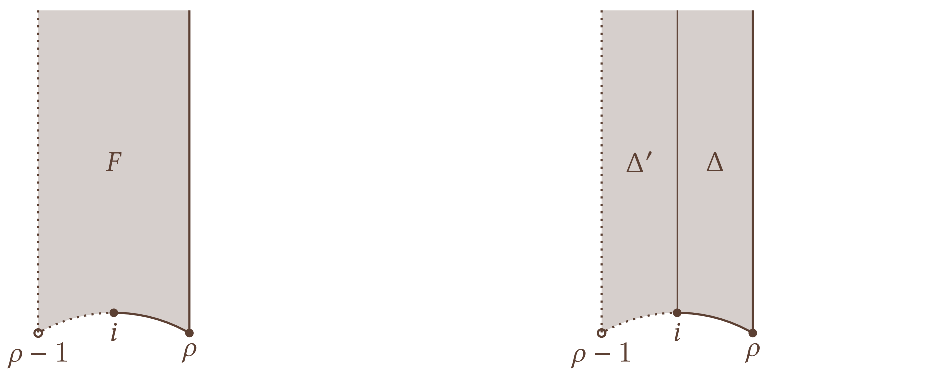
Exercise There is another possible proof using a similar method, but a different function. Define the *j-invariant* of a lattice,

$$j(\tau) = \frac{g_2^3}{g_2^3 - 27 g_3^2}. \tag{20}$$

1. Show that j is fully modular, in that $j(T.\tau) = j(\tau)$ for any $T \in \Gamma$.
2. Show that $j(\tau) \to \infty$ as $\Im(\tau) \to \infty$. (You'll need to calculate the asymptotics of the series g_2 and g_3. You should find that $j(\tau) \sim e^{-i\pi\tau}$ for large $\Im(\tau)$.)
3. Show that $j(i) = 1$ and $j(\rho) = 0$, where $\rho = e^{i\pi/3}$ (Hint: a quick way to do this is to consider the symmetries of the corresponding Weierstrass elliptic functions under rotations: consistency in the differential equation allows you to determine that g_2 or g_3 is zero.)
4. Show that a fundamental domain for j (and hence the modular group) is

   $$F := (\{|\tau| \geqslant 1\} \cap \{0 \leqslant \Re(\tau) \leqslant 1/2\}) \cup (\{|(|\tau) > 1\} \cap -1/2 < \Re(\tau) < 0\}) =: \Delta \cup \Delta', \tag{21}$$

   as shown in Figure 3a.
5. Consider the right-hand half of the fundamental domain, $\Delta = \{\Re(\tau) \geqslant 0\} \cap F$. Show that j is real on its boundary.
6. Show that the interior of the domain Δ is mapped to the upper half-plane.
7. Proceed from here to show that j is a bijection from $F \setminus \{i, \rho\}$ to $\mathbb{C} \setminus \{0, 1\}$. (Notice that therefore, two lattices with moduli τ and τ' are equivalent if and only if $j(\tau) = j(\tau')$.)



(a) Fundamental domain for the modular group Γ. (b) The regions Δ and Δ'; the thin line is in Δ.

Figure 3: Significant domains for j. Dotted edges are excluded from the shaded region.