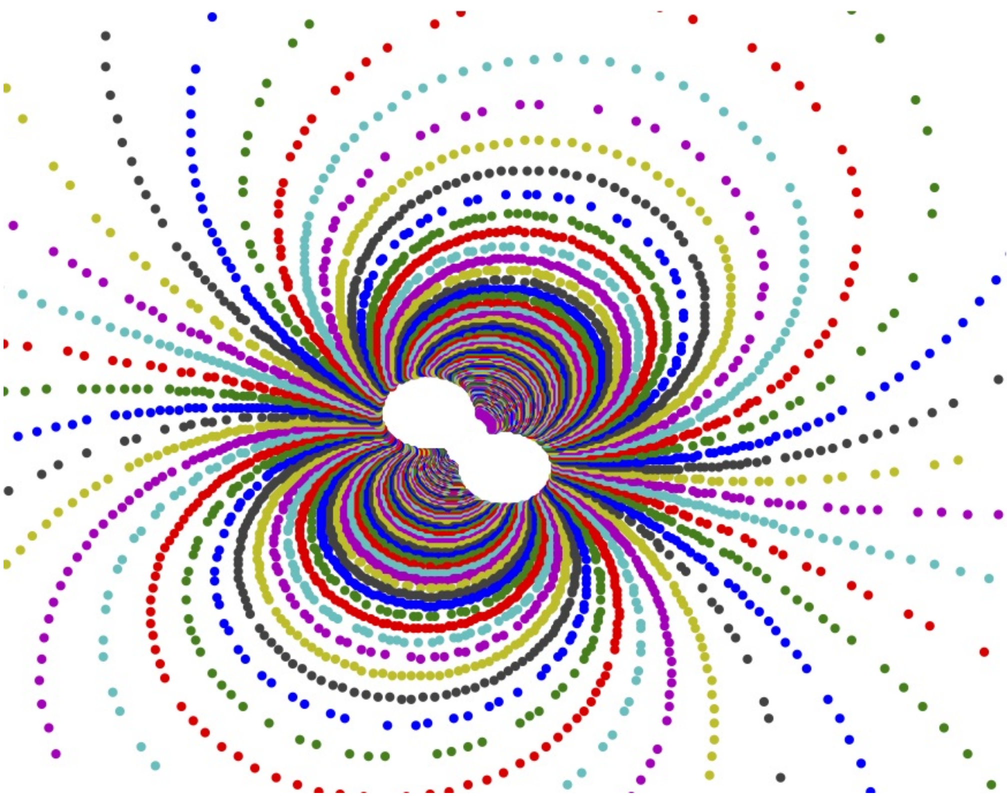
The Arithmetic-Geometric Mean of Gauss

Tomack Gilmore
080106489

Supervised by Professor S. R. Bullett
Queen Mary, University of London



Abstract

At the end of the eighteenth century Gauss observed in his diary "We have established that the arithmetic-geometric mean between 1 and $\sqrt{2}$ is $\pi/\tilde{\omega}$ to the eleventh decimal place; the demonstration of this fact will surely open an entirely new field of analysis." This project aims to examine what lead Gauss to make such a remark and explore some of the fascinating consequences of this relationship.

Acknowledgements

There are many people who have contributed to this project in a number of ways that I would like to acknowledge. First and foremost I would like to thank my supervisor, Professor Bullett, whose guidance and insights have proved invaluable to me over the past year.

I would also like to thank Professor Vivaldi and Professor Müller who have both been a source of great inspiration and advice.

Finally I would like to express my sincere gratitude to Dr Touchette and Dr Bandtlow for their continued support and encouragement throughout my career at Queen Mary.

Contents

1 Introduction

The arithmetic-geometric mean is the common limit of two sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$. Each a_n term is the arithmetic mean of a_{n-1} and b_{n-1} and each b_n term is the geometric mean of a_{n-1} and b_{n-1}. We will explore the properties of this mean for real numbers and also consider its behaviour on the complex plane. This leads to a surprising theorem, the proof of which incorporates many interesting concepts from complex analysis and the theory of modular forms.

We begin by defining the arithmetic-geometric mean for real numbers and exploring a few of its important properties. We will also see our first interesting theorem that hints at the link between the arithmetic-geometric mean and elliptic integrals. Much of this section takes its cue from Cox [1] and the Borwein brothers [4].

We continue with a discussion similar to Siegel's [6] of how elliptic integrals are derived using basic geometry. We then explore the properties of elliptic integrals of the first kind and realise how they are connected to the arithmetic-geometric mean. This is the same connection that Gauss noted over two hundred years ago. We then offer an alternative proof of Theorem 2.6 that borrows from Almkvist and Berndt [5].

The next section examines how we can use our knowledge of the arithmetic-geometric mean and elliptic integrals to calculate π. To do this we exploit the relationship between elliptic integrals of the first and second kind. Almkvist and Berndt's excellent paper [5] provided much of the inspiration for this section.

In the final section we consider the arithmetic-geometric mean on the complex plane. The results in this section are deep and the method used to prove them incorporates many profound and beautiful ideas from the theory of modular forms. We rely heavily on Cox [1] to fill in the gaps in this last section where we will finally see just how Gauss's observation really did open up an entirely new field of analysis.

2 The Real Arithmetic-Geometric Mean

In this section we will define the arithmetic-geometric mean for real numbers and explore some of its properties.

Consider first the arithmetic-geometric mean inequality:

Lemma 2.1. Given a, $b \in \mathbb{R}$ such that $a \geqslant b > 0$ it follows that

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

Proof. To prove this we utilise the finite form of Jensen's inequality. Consider the function $\ln(x)$ which is strictly convex. We have that

$$\ln\left(\frac{a+b}{2}\right) > \frac{\ln(a)}{2} + \frac{\ln(b)}{2} = \ln(\sqrt{ab}).$$

As the natural logarithm is strictly increasing it follows that

$$\frac{a+b}{2} > \sqrt{ab},$$

with equality when $a = b = 1$. $\square$

Now let a and b be real numbers such that $a \geqslant b > 0$. Consider the sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ where

$$\begin{aligned} a_0 &= a, & b_0 &= b \\ a_n &= \frac{a_{n-1}+b_{n-1}}{2}, & b_n &= \sqrt{a_{n-1}b_{n-1}}, \quad n = 1, 2, ..., \end{aligned} \tag{1}$$

where b_n is always taken to be positive. By Lemma 2.1 can immediately see that $a_n \geqslant b_n$ for all $n \geqslant 0$, however there is much more that we can observe.

Corollary 2.2. Let $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ be defined as in (1). Then

1. $a \geqslant a_1 \geqslant \cdots \geqslant a_{n-1} \geqslant a_n \geqslant b_n \geqslant b_{n-1} \geqslant \cdots \geqslant b_1 \geqslant b$;
2. $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$.

Proof. By Lemma 2.1 we have that $a_{n-1} \geqslant b_{n-1}$ for all $n \geqslant 1$. So it follows that $2a_{n-1} \geqslant a_{n-1} + b_{n-1}$ and hence

$$a_{n-1} \geqslant \frac{a_{n-1}+b_{n-1}}{2}. \tag{2}$$

So $a_{n-1} \geqslant a_n$ for all $n \geqslant 1$. Similarly we also have that $a_{n-1}b_{n-1} \geqslant b_{n-1}^2$ whence

$$\sqrt{a_{n-1}b_{n-1}} \geqslant b_{n-1}, \tag{3}$$

so $b_n \geqslant b_{n-1}$ for all $n \geqslant 1$. Putting (2) and (3) together proves the first assertion.

To prove the second part we show inductively that $a_n - b_n \leqslant \dfrac{a-b}{2^n}$. Consider the equation

$$a_n - b_{n-1} = \frac{a-b}{2^n}.$$

This clearly holds for $n = 1$. Now as $b_n \geqslant b_{n-1}$ it follows that

$$a_n - b_n \leqslant a_n - b_{n-1} = \frac{a_{n-1} - b_{n-1}}{2} \leqslant \frac{a_{n-1} - b_{n-2}}{2} = \frac{1}{2}\left(\frac{a-b}{2^{n-1}}\right)$$

by the induction hypothesis. So we have

$$0 \leqslant a_n - b_n \leqslant \frac{a-b}{2^n}$$

for all $n \geqslant 0$. Clearly by part 1 the limits of a_n and b_n exist. Moreover the term on the right tends to zero as n tends to infinity so

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

□

We can now define the arithmetic-geometric mean for real numbers.

Definition 2.3. Let $a, b \in \mathbb{R}$ such that $a \geqslant b > 0$, where the sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are defined as above. Then the *arithmetic-geometric mean* of a and b is defined to be

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

Example 1. One particularly important example that appears in [1, §1] is the arithmetic-geometric mean of $\sqrt{2}$ and 1,

$$M(\sqrt{2}, 1) = 1.1981402347355922074...$$

with accuracy to nineteen decimal places. This was computed by Gauss who produced the following table in "De origine proprietatibusque generablis numerorum mediorum arithmetico-geometricorum" [2, vol. III, pp. 361-371].

n	a_n	b_n
0	1.41421356237390504880 2	1.00000000000000000000
1	1.20710678118654752440 1	1.18920711500272106671 7
2	1.19815694809463429555 9	1.19812352149312012260 7
3	1.19814023479387720908 3	1.19814023467730720579 8
4	1.19814023473559220744 1	1.19814023473559220743 9

We will see just how important this value is when we come to explore properties of elliptic integrals.

It appears from the example above that this algorithm converges very quickly. In fact we can quantify the rate of convergence using the following definition.

Definition 2.4. Suppose a sequence $\{\alpha_n\}_{n=0}^{\infty}$ converges to some A and suppose there exist constants $C > 0$ and $m \geqslant 1$ such that

$$|\alpha_{n+1} - A| \leqslant C|\alpha_n - A|^m,$$

where $n \geqslant 1$. Then we say that the convergence of $\{\alpha_n\}_{n=0}^{\infty}$ is of the *mth order*.

Setting $c_n = \sqrt{a_n^2 - b_n^2}$ for $n \geqslant 0$ where a_n and b_n are defined as in (1) we have that

$$c_{n+1} = \frac{a_n - b_n}{2} = \frac{(a_n + b_n)(a_n - b_n)}{2(a_n + b_n)} = \frac{c_n^2}{4a_{n+1}} \leqslant \frac{c_n^2}{4M(a,b)}.$$

We know that $M(a, b) \geqslant a_n$ is constant so it follows that the sequence $\{c_n\}_{n=0}^{\infty}$ tends to zero quadratically, or the convergence is of the second order. This is the main reason why the arithmetic-geometric mean provides such an efficient method for calculating constants such as π.

There are two other properties of the arithmetic-geometric mean that are also noteworthy. They are:

$$\begin{aligned} &M(a,b) = M(a_1,b_1) = M(a_2,b_2) = ... \\ &M(\lambda a, \lambda b) = \lambda M(a,b). \end{aligned} \tag{4}$$

These both follow easily from the definition. The first property above shows that

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

so letting $a = 1$ and using the fact that the arithmetic-geometric mean is homogeneous we obtain

$$M(1,b) = \frac{1+b}{2} M\left(1, \frac{2\sqrt{b}}{1+b}\right).$$

The arithmetic-geometric mean in the form of a single variable iteration is known as the *Legendre form* [4, §1]. This is interesting as it shows that the arithmetic-geometric mean of 1 and b is the arithmetic-geometric mean of 1 and m, where m is the ratio of the geometric and arithmetic mean of 1 and b. We now use the fact that the arithmetic-geometric mean is homogeneous in the following corollary.

Corollary 2.5. Let $x = \frac{1}{a}\sqrt{a^2 - b^2}$. Then

$$M(1+x, 1-x) = \frac{1}{a} M(a,b).$$

Proof. Letting $a_0 = 1 + x$ and $b_0 = 1 - x$ we have that $a_1 = 1$ and $b_1 = \frac{b}{a}$ according to (1). The result follows by extracting $\frac{1}{a}$ on the right hand side and applying the property $M(a_0, b_0) = M(a_1, b_1)$. □

An important thing to note in the above corollary is the fact that x is the eccentricity of an ellipse with semi-major axis a and semi-minor axis b, a fact we shall return to later. For now our first taste of the real depth of the arithmetic-geometric mean comes in the form of the following theorem.

Theorem 2.6. If $a \geqslant b > 0$ then

$$M(a,b) \cdot \int_0^{\frac{\pi}{2}} (a^2 \cos^2\phi + b^2 \sin^2\phi)^{-1/2} d\phi = \frac{\pi}{2}.$$

There are many different proofs of this theorem, most of which can be found in [1, §1], [4, §1] and [5, §2] . The proof we present here appears slightly differently in [1, §1] and makes use of the following lemma.

Lemma 2.7. Consider ϕ and ϕ' such that

$$\sin\phi = \frac{2a\sin\phi'}{a+b+(a-b)\sin^2\phi'}.$$

Then it follows that

$$\cos\phi = \frac{2\cos\phi'(a_1^2\cos^2\phi' + b_1^2\sin^2\phi')^{1/2}}{a+b+(a-b)\sin^2\phi'} \tag{5}$$

and

$$(a^2\cos^2\phi + b^2\sin^2\phi)^{1/2} = a\frac{a+b-(a-b)\sin^2\phi'}{a+b+(a-b)\sin^2\phi'}. \tag{6}$$

Proof. We first prove (5).

$$\begin{aligned}
\sin^2\phi &= \frac{((a+b+(a-b))\sin\phi')^2}{(a+b+(a-b)\sin^2\phi')^2} \\
&= 1 - \frac{\cos^2\phi'((a+b)^2-(a-b)^2\sin^2\phi')}{(a+b+(a-b)\sin^2\phi')^2} \\
\Rightarrow \quad \cos^2\phi &= \frac{4\cos^2\phi'(((a+b)/2)^2\cos^2\phi' + ab\sin^2\phi')}{(a+b+(a-b)\sin^2\phi')^2} \\
\Rightarrow \quad \cos\phi &= \frac{2\cos\phi'(a_1^2\cos^2\phi' + b_1^2\sin^2\phi')^{1/2}}{a+b+(a-b)\sin^2\phi'}.
\end{aligned}$$

The intermediary steps have been omitted to save paper and ink as they are straightforward manipulations that can be gleaned from the above. To prove the second identity we simply plug in our values for $\cos\phi$ and $\sin\phi$:

$$
\begin{aligned}
(a^2\cos\phi + b^2\sin\phi)^{1/2} &= a\left(1 + \frac{(b^2-a^2)4\sin^2\phi'}{(a+b+(a-b)\sin^2\phi')^2}\right)^{1/2} \\
&= a\frac{a+b-(a-b)\sin^2\phi'}{a+b+(a-b)\sin^2\phi'},
\end{aligned}
$$

where again we have foregone the gory details as they are simple manipulations. □

The above expression of $\sin\phi$ as a function of $\sin\phi'$ is often referred to as *Gauss's transformation*. Cox [1, §1] references Jacobi [3, vol. I, p.152] for a more explicit proof of the above identity, though it is possible to derive this independently, so we now turn to our proof of Theorem 2.6.

Proof of Theorem 2.6. Proving Theorem 2.6 is equivalent to showing that

$$
I(a,b) = \frac{\pi}{2\mu}, \tag{7}
$$

where $M(a,b) = \mu$ and

$$
I(a,b) = \int_0^{\pi/2} (a^2\cos^2\phi + b^2\sin^2\phi)^{-1/2} d\phi. \tag{8}
$$

The entire proof hinges on showing that

$$
I(a,b) = I(a_1,b_1) \tag{9}
$$

where a, a_1, b and b_1 are defined according to (1).

Consider the identity in Lemma 2.7,

$$
\sin\phi = \frac{2a\sin\phi'}{a+b+(a-b)\sin^2\phi'},
$$

and note that $0 \leqslant \phi' \leqslant \pi/2$ corresponds to $0 \leqslant \phi \leqslant \pi/2$. Differentiating with respect to ϕ' gives

$$
\cos\phi\, d\phi = \frac{2a\cos\phi'(a+b-(a-b)\sin^2\phi')}{(a+b+(a-b)\sin^2\phi')^2}\, d\phi'.
$$

Replacing $\cos\phi$ with our expression (5) gives

$$
(a_1^2\cos^2\phi' + b_1^2\sin^2\phi')^{1/2}\, d\phi = a\frac{a+b-(a-b)\sin^2\phi'}{a+b+(a-b)\sin^2\phi'}\, d\phi',
$$

Finally substituting in (6) we see that

$$
(a_1^2\cos^2\phi' + b_1^2\sin^2\phi')^{1/2}\, d\phi = (a^2\sin^2\phi + b^2\sin^2\phi)^{1/2}\, d\phi',
$$

or equivalently

$$(a^2 \sin^2 \phi + b^2 \sin^2 \phi)^{-1/2}\, d\phi = (a_1^2 \cos^2 \phi' + b_1^2 \sin^2 \phi')^{-1/2}\, d\phi'. \tag{10}$$

This means that we can iterate (9) so that

$$I(a, b) = I(a_1, b_1) = I(a_2, b_2) = \ldots \tag{11}$$

As the function $(a_n^2 \cos^2 \phi + b_n^2 \sin^2 \phi)^{-1/2}$ converges uniformly to μ^{-1} we have

$$I(a, b) = \lim_{n\to\infty} I(a_n, b_n) = \lim_{n\to\infty} \int_0^{\pi/2} (a_n^2 \cos^2 \phi + b_n^2 \sin^2 \phi)^{-1/2}\, d\phi = \frac{1}{\mu} \int_0^{\pi/2} 1\, d\phi = \frac{\pi}{2\mu}.$$

□

We shall return to Theorem 2.6 shortly. What is most remarkable about the above proof is the fact that in (11) we can replace a and b with their arithmetic and geometric means respectively. This is sometimes referred to as *Landen's transformation* though it was also independently discovered by Gauss years later. We shall now move on to consider how the arithmetic-geometric mean is related to certain elliptic integrals. The clue to this relation lies in the integral from the above theorem.

3 Elliptic Integrals

To give a flavour of the history of elliptic integrals we will consider a particular polar curve known as the *lemniscate*, or the *lemniscate of Bernoulli*. This is a special type of Cassini oval discovered by Bernoulli towards the end of the seventeenth century. The general properties of the lemniscate were discovered by G. Fagnano in 1750 and investigations into the arc length of this curve by Gauss and Euler led to later work on elliptic functions.

Let us fix two points α_1 and α_2 in a cartesian coordinate system such that these points lie at $(\pm a, 0)$ whence the distance between them is $2a$. Then the lemniscate is the locus of a point α in the plane such that the product of the distances from α to α_1 and α_2 has constant value c^2. Suppose α has coordinates (x, y), α_1 is at $(a, 0)$ and α_2 is at $(-a, 0)$ for positive a. Letting r_1 be the distance from α to α_1 and r_2 be the distance from α to α_2 we have

$$\begin{aligned} r_1^2 &= (x+a)^2 + y^2 = r^2 + a^2 + 2ax \\ r_2^2 &= (x-a)^2 + y^2 = r^2 + a^2 - 2ax \end{aligned}$$

where

$$r^2 = x^2 + y^2. \tag{12}$$

Multiplying the above two equations together we obtain

$$r^4 + 2a^2r^2 + a^4 - 4a^2x^2 = (r_1r_2)^2 = c^4. \tag{13}$$

So we see that every value of c generates a class of lemniscates. We consider those such that $c = a$, that is those that take the shape of a horizontal figure 8 and pass through the origin. In fact it is customary to let $2a^2 = 1$ so that the distance between α_1 and α_2 is $\sqrt{2}$. Under this construction (13) becomes

$$2x^2 = r^2 + r^4. \tag{14}$$

Rearranging (12) and substituting this into (14) we have

$$2y^2 = r^2 - r^4. \tag{15}$$

It is straightforward to derive a parametric representation of the lemniscate using (14) and (15), however we are concerned with finding its arc length s. As the lemniscate is symmetric about the axes of the cartesian plane we need only be concerned with the first quadrant where r varies independently over the interval [0,1]. Differentiating (14) and (15) with respect to r we obtain

$$\frac{dx}{dr} = \frac{r + 2r^3}{2x} \quad \text{and} \quad \frac{dy}{dr} = \frac{r - 2r^3}{2y}.$$

Substituting these into the equation for finding the arc length, $s(r)$, of a curve parametrically we have

$$\left(\frac{ds}{dr}\right)^2 = \left(\frac{r + 2r^3}{2x}\right)^2 + \left(\frac{r - 2r^3}{2y}\right)^2,$$

or equivalently

$$(2xy)^2 \left(\frac{ds}{dr}\right)^2 = y^2(r+2r^3)^2 + x^2(r-2r^3)^2.$$

Bearing in mind that $(2xy)^2 = (r^2+r^4)(r^2-r^4) = r^4(1-r^4)$ and substituting in our expressions (14) and (15) we get

$$\begin{aligned} r^4(1-r^4)\left(\frac{ds}{dr}\right)^2 &= \frac{r^2-r^4}{2}(r^2+4r^4+4r^6) + \frac{r^2+r^4}{2}(r^2-4r^4+4r^6) \\ \Rightarrow \qquad (1-r^4)\left(\frac{ds}{dr}\right)^2 &= 1. \end{aligned}$$

So rearranging, solving for $s(r)$ and multiplying by 4 we see that the total arc length of the lemniscate of Bernoulli is

$$s(r) = 4\int_0^1 \frac{1}{\sqrt{1-r^4}}\, dr. \tag{16}$$

This particular integral is referred to by Siegel in [6, vol. I, p. 3] as the *lemniscatic integral* and it is a special case of a more general class of integrals.

In essence an elliptic integral is of the form

$$\int_0^u R(x,y)\, dx,$$

where R is a rational function of x and y and y^2 is a quartic polynomial in x. The above integral is complete when $u = 1$. Integrals of this nature are divided into three different kinds, the first of which we will now define. For a rigorous account of the theory behind elliptic integrals consult Whittaker and Watson [7, §22.7].

Definition 3.1. The *complete elliptic integral of the first kind* is defined to be

$$K(x) = \int_0^1 \frac{1}{\sqrt{(1-t^2)(1-x^2t^2)}}\, dt = \int_0^{\pi/2} \frac{1}{\sqrt{1-x^2\sin^2\phi}}\, d\phi,$$

where $|x| < 1$ is known as the *modulus*.

Integrals such as this arise in the computation of the arc length of a lemniscate or the calculation of the period of the pendulum. Often these integrals involve nonelementary or transcendental functions, and hence can only be numerically approximated. A robust explanation of how to calculate the period of a pendulum can be found in [8], it is mentioned here as an example of where such integrals may occur.

Example 2. Say T is the period of a pendulum with amplitude α and length h. Then we have

$$T = 4\sqrt{\frac{h}{g}} K\left(\sin\left(\frac{\alpha}{2}\right)\right),$$

where g is the gravitational constant. Notice here that if we let $\alpha \to 0$ then $K(\sin(\alpha/2)) \to \pi/2$ and we are left with a simple harmonic motion.

We will now derive the power series for $K(x)$. Assuming $|x| < 1$ it follows that $K(x)$ is analytic so we can expand $(1 - x^2 \sin^2 \phi)^{-1/2}$ to obtain

$$(1 - x^2 \sin^2 \phi)^{-1/2} = 1 + \frac{x^2 \sin^2 \phi}{2} + \frac{3x^4 \sin^4 \phi}{2^2 2!} + \frac{15x^6 \sin^6 \phi}{2^3 3!} + ... = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_k}{k!} x^{2k} (\sin \phi)^{2k},$$

where $(\alpha)_k = \alpha \cdot (\alpha + 1) \cdots (\alpha + k - 1)$, and $(\alpha)_0 = 1$. Hence we can rewrite the integral in Definition 3.1 as

$$\sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_k}{k!} x^{2k} \int_0^{\pi/2} (\sin \phi)^{2k} \, d\phi. \tag{17}$$

In order to deal with the integral in (17) we split the integrand up as follows

$$\int_0^{\pi/2} (\sin \phi)^{2k} \, d\phi = \int_0^{\pi/2} (\sin \phi)^{2k-1} \sin \phi \, d\phi$$

and integrate by parts so that

$$\begin{aligned} \int_0^{\pi/2} (\sin \phi)^{2k} \, d\phi &= \left[-\cos \phi (\sin \phi)^{2k-1}\right]_0^{\pi/2} + (2k-1) \int_0^{\pi/2} (\sin \phi)^{2(k-1)} \cos^2 \phi \, d\phi \\ &= (2k-1) \int_0^{\pi/2} (\sin \phi)^{2(k-1)} (1 - \sin^2 \phi) \, d\phi \\ &= (2k-1) \int_0^{\pi/2} (\sin \phi)^{2(k-1)} \, d\phi - (2k-1) \int_0^{\pi/2} (\sin \phi)^{2k} \, d\phi. \end{aligned} \tag{18}$$

Letting $u(2k) = \int_0^{\pi/2} (\sin \phi)^{2k} \, d\phi$ we see that (18) satisfies the recurrence relation

$$u(2k) = \frac{2k-1}{2k} u(2(k-1)) \tag{19}$$

Iterating (19) we see that

$$u(2k) = \frac{2k-1}{2k} \cdot \frac{2k-3}{2(k-1)} \cdots \frac{1}{2} \cdot 1 \cdot u(0) = \left(\frac{1}{2}\right)_k \frac{1}{k!} \frac{\pi}{2}. \tag{20}$$

Combining (17) and (20) we see that the power series expansion for $K(x)$ is

$$K(x) = \frac{\pi}{2} \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_k^2}{(k!)^2} x^{2k}. \tag{21}$$

Elliptic integrals of this kind are intrinsically linked to the arithmetic-geometric mean. Before we examine this relationship formally we first consider the integral derived for the length of the lemniscate of Bernoulli above. This is an elliptic integral of the first kind and at first glance we might assume that we could use the above power series for to approximate this length. However if we set $r = \cos\phi$ then the integral varies over $[\pi/2, 0]$ and we have

$$\int_0^1 \frac{1}{\sqrt{1-r^4}}\, dr = \int_{\pi/2}^0 \frac{-\sin\phi}{\sqrt{1-\cos^4\phi}}\, d\phi = \int_0^{\pi/2} \frac{1}{\sqrt{1+\cos^2\phi}} d\phi.$$

Finally we use the fact that $\cos^2\phi + \sin^2\phi = 1$ to show

$$\int_0^1 \frac{1}{\sqrt{1-r^4}}\, dr = \int_0^{\pi/2} (2\cos^2\phi + \sin^2\phi)^{-1/2}\, d\phi.$$

This is precisely our integral $I(a,b)$ from Theorem 2.6 where $a = \sqrt{2}$ and $b = 1$, which means that we can calculate the length of the lemniscate using the arithmetic-geometric mean. Note from the above that a is the distance between the foci of the lemniscate and also that b is the point at which the lemniscate intersects the x-axis.

This particular integral was the source of much mathematical research throughout the eighteenth century and Gauss even had a separate notation for it, namely

$$\tilde{\omega} = 2\int_0^1 \frac{1}{\sqrt{1-z^4}}\, dz,$$

so that the relationship between the arc length of the lemniscate and the arithmetic-geometric mean can be expressed as

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

Gauss's observation that these two numbers are essentially the same was the discovery that he claimed would "open up an entirely field of analysis" and it has inspired some particularly beautiful mathematics.

We shall now formally cement the relationship between the arithmetic-geometric mean and elliptic integrals of the first kind. What follows is an alternative proof of Theorem 2.6, motivated by results found in [5].

Second proof of Theorem 2.6. Consider again the integral

$$I(a,b) = \int_0^{\pi/2} \frac{1}{\sqrt{a^2\cos^2\phi + b^2\sin^2\phi}}\, d\phi = \frac{1}{a}\int_0^{\pi/2} \frac{1}{\sqrt{1 - \left(\frac{a^2-b^2}{a^2}\right)\sin^2\phi}}\, d\phi.$$

Now setting $x = \frac{1}{a}\sqrt{a^2 - b^2}$ we have that

$$I(a,b) = \frac{1}{a}\int_0^{\pi/2} \frac{1}{\sqrt{1 - x^2\sin^2\phi}}\, d\phi = \frac{1}{a}K(x).$$

Replacing $M(a,b)$ with $aM(1+x, 1-x)$ from Corollary 2.5 we see that proving Theorem 2.6 is equivalent to showing that

$$aM(1+x, 1-x) \cdot \frac{1}{a}K(x) = \frac{\pi}{2}. \tag{22}$$

By rearranging and replacing $K(x)$ with its power series expansion (21) our proof of Theorem 2.6 reduces to showing that

$$\frac{1}{M(1+x, 1-x)} = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_k^2}{(k!)^2} x^{2k}. \tag{23}$$

Letting $a = (1+t^2)$ and $b = (1-t^2)$, it follows from Corollary 2.5 that

$$M(1+x, 1-x) = \frac{1}{1+t^2} M(1+t^2, 1-t^2), \tag{24}$$

where $x = 2t/(1+t^2)$.

Now we can assume that the function $1/M(1+x, 1-x)$ has a power series expansion of the following form

$$\frac{1}{M(1+x, 1-x)} = \sum_{k=0}^{\infty} C_k x^{2k}, \tag{25}$$

so combining (24) and (25) we have

$$(1+t^2)\sum_{k=0}^{\infty} C_k t^{4k} = \sum_{k=0}^{\infty} C_k \left(\frac{2t}{(1+t^2)}\right)^{2k}.$$

After expanding $(1+t^2)^{-2k-1}$ and equating like powers of t it becomes apparent that

$$\frac{1}{M(1+x, 1-x)} = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_k^2}{(k!)^2} x^{2k} = \frac{2}{\pi}K(x)$$

and we are done. $\square$

There are a few things to point out about the above proof. Firstly the assumption in (25) is completely valid: it is clear that $M(1+x, 1-x)$ is an even function of x and so the x^{2k} is justified; also given the convergence of $M(a,b)$ and observing by (4) that

$M(a,b) = M(a_1, b_1) = ...$, it is possible to find an $M(a_n, b_n)$ such that $|x| < 1$. Secondly if we let $x' = \sqrt{1-x^2}$, known as the *complementary modulus*, then using the fact that $M(a,b) = M(a_1, b_1)$ we can rewrite (22) as

$$\frac{1}{M(1,x')} = \frac{2}{\pi} \int_0^{\pi/2} \frac{1}{\sqrt{1-x^2\sin^2\phi}}\, d\phi.$$

Gauss interpreted this equation as showing that the average value of the integrand on the interval $[0, \pi/2]$ is the reciprocal of the arithmetic-geometric mean of the reciprocals of the minimum and maximum values of the function [2, vol. III, p.371].

The above proof of Theorem 2.6 highlights the important relationship between the arithmetic-geometric mean and elliptic integrals of the first kind. Given that these integrals are non-elementary and generally tricky to calculate, Theorem 2.6 provides an incredibly fast and efficient algorithm for calculating $K(x)$. As we will see in the next section this is particularly useful for calculating elliptic integrals of the second kind and consequently provides us with an effective method for calculating π.

4 Calculating π

In this section we will examine how the arithmetic-geometric mean is used to efficiently calculate π. As motivation we first consider earlier techniques that were employed to pin down this elusive constant at the end of the sixteenth century. We then turn to more modern methods that incorporate many of the ideas we have seen so far.

We begin by considering the function $(\sin\theta)\theta^{-1}$ which appears in many areas of mathematics, particularly calculus. The following results and an interesting discussion of their applications can be found in [9]. By employing the double angle formula for the sine function we obtain

$$\frac{\sin\theta}{\theta} = \frac{2\sin(\theta/2)\cos(\theta/2)}{\theta} = \cos(\theta/2)\frac{\sin(\theta/2)}{\theta/2},$$

and after successive applications we see that

$$\frac{\sin\theta}{\theta} = \cos(\theta/2)\cos(\theta/4)\cdots\cos(\theta/2^n)\frac{\sin(\theta/2^n)}{(\theta/2^n)}. \tag{26}$$

Now $\phi = \theta/2^n$ tends to 0 as n tends to infinity, so by L'Hôpital we have

$$\lim_{\phi\to 0}\frac{\sin\phi}{\phi} = 1,$$

whence (26) becomes Euler's formula

$$\frac{\sin\theta}{\theta} = \prod_{n=1}^{\infty}\cos(\theta/2^n),$$

where the left hand side is defined to be 1 when $\theta = 0$. In 1593 Francois Viéte, preceding Euler, produced the following beautiful expression for π

$$\frac{2}{\pi} = \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{2+\sqrt{2}}}{2}\cdot\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2}\cdots,$$

which is precisely Euler's formula for $\theta = \pi/2$. It may be written more succinctly as

$$\frac{2}{\pi} = \lim_{n\to\infty} 2^{-n}\prod_{i=1}^{n} a_i, \tag{27}$$

where $a_0 = 0$ and $a_i = \sqrt{2 + a_{i-1}}$. Although aesthetically pleasing this formula is inefficient for the purposes of calculating π to a high degree of accuracy. To see this note that by letting $b_i = 2 - a_i$ for $i \geqslant 0$ we can rearrange (27) in the following way

$$\frac{\pi}{2} = \lim_{n\to\infty} 2^{n-2} b_{n-1}\prod_{i=1}^{n} a_i.$$

As the b_n term converges very rapidly to 0 it is clear that this method is inaccurate.

In order to develop a more sophisticated algorithm for the calculation of π we must first introduce the following theorems and definitions. We shall continue with a_n and b_n being defined as in (1). Also recall how we defined c_n to be $\sqrt{a_n^2 - b_n^2}$ and $I(a,b)$ in Section 2.

Theorem 4.1. Given $a > b > 0$ define

$$J(a,b) = \int_0^{\frac{\pi}{2}} (a^2 \cos^2\theta + b^2 \sin^2\theta)^{1/2} d\theta.$$

Then it follows that

$$J(a,b) = \left(a^2 - \frac{1}{2} \sum_{n=0}^{\infty} 2^n c_n^2 \right) I(a,b).$$

We forego the proof of this theorem here but a lengthy and cumbersome argument may be found in [10]. An alternative and much more elegant proof is alluded to by Salamin in [12]. As with the integral $I(a,b)$ the above result is derived by exploiting the relationship between $J(a_n, b_n)$ and $J(a_{n+1}, b_{n+1})$ that can be found in [11], i.e.,

$$J(a_n, b_n) = 2J(a_{n+1}, b_{n+1}) - a_n b_n I(a,b).$$

Definition 4.2. Given x such that $|x| < 1$ we define the *complete elliptic integral of the second kind* to be

$$E(x) = \int_0^{\frac{\pi}{2}} (1 - x^2 \sin^2\psi)^{1/2}\, d\psi.$$

Elliptic integrals of the second kind describe the circumference of an ellipse. Liouville proved in 1834 that elliptic integrals of the first and second kind are nonelementary, meaning that they are integrals that have an antiderivative that cannot be expressed in terms of elementary functions. They are often evaluated using Taylor series, however if a given function is not infinitely differentiable such integrals are evaluated numerically.

Elliptic integrals of the second kind are not analogous to elliptic integrals of the first kind inasmuch as they cannot be directly expressed as a function of the arithmetic-geometric mean. Note, however, that if we let $x = \frac{1}{a}\sqrt{a^2 - b^2}$ then

$$E(x) = \int_0^{\frac{\pi}{2}} \left(1 - \left(\frac{a^2 - b^2}{a^2} \right) \sin^2\theta \right)^{1/2} d\theta = \int_0^{\frac{\pi}{2}} \frac{(a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}}{a} d\theta = \frac{1}{a} J(a,b).$$

We have one more theorem to consider before we can address the problem of calculating π. It is known as *Legendre's relation* and is incredibly useful as it describes the relationship between elliptic integrals of the first and second kind. Different proofs of this theorem can be found in [7] and [4] which both employ different techniques. The more direct proof that is reproduced below is due to Almkvist and Berndt [5].

Theorem 4.3. Let x be real such that $0 < x < 1$. Then

$$K(x)E(x') + K(x')E(x) - K(x)K(x') = \frac{\pi}{2},$$

where $K(x)$ is defined according to Definition 3.1 and $x' = \sqrt{1 - x^2}$.

Proof. To avoid confusing notation we shall let $c = x^2$ and $c' = 1 - c$. Also we shall denote $E(c)$ and $K(c)$ as E and K respectively. Similarly we shall let K' and E' denote $K(c')$ and $E(c')$ respectively.

We begin by differentiating $(E - K)$ with respect to c,

$$\begin{aligned} \frac{d}{dc}(E - K) &= -\frac{d}{dc}\int_0^{\pi/2} \frac{c \sin^2 \theta}{(1 - c \sin^2 \theta)^{1/2}} d\theta \\ &= \frac{E}{2c} - \frac{1}{2c}\int_0^{\pi/2} \frac{d\theta}{(1 - c\sin^2\theta)^{3/2}}. \end{aligned}$$

Now we have

$$\frac{d}{d\theta}\left(\frac{\sin\theta\cos\theta}{(1 - c\sin^2\theta)^{1/2}}\right) = \frac{1}{c}(1 - c\sin^2\theta)^{1/2} - \frac{c'}{c}(1 - c\sin^2\theta)^{-3/2},$$

so it follows that

$$\begin{aligned} \frac{d}{dc}(E - K) &= \frac{E}{2c} - \frac{E}{2cc'} + \frac{1}{2c'}\int_0^{\pi/2} \frac{d}{d\theta}\left(\frac{\sin\theta\cos\theta}{(1 - c\sin^2\theta)^{1/2}}\right) d\theta \\ &= \frac{E}{2c}\left(1 - \frac{1}{c'}\right) \\ &= -\frac{E}{2c'}. \end{aligned} \tag{28}$$

Since $c' = 1 - c$ we have

$$\frac{d}{dc}(E' - K') = \frac{E'}{2c}. \tag{29}$$

Finally it can be shown that

$$\frac{dE}{dc} = \frac{E - K}{2c} \text{ and } \frac{dE'}{dc} = -\frac{E' - K'}{2c'}. \tag{30}$$

Notice that the left hand side of Legendre's relation can be written as

$$L = EE' - (E - K)(E' - K').$$

Differentiating L with respect to c and applying (28)-(30) we find

$$\frac{dL}{dc} = \frac{(E - K)E'}{2c} - \frac{E(E' - K')}{2c'} + \frac{E(E' - K')}{2c'} - \frac{(E - K)E'}{2c} = 0,$$

from which we can deduce L is constant. We determine the value of L by letting c approach zero.

Note that as c tends to zero we have

$$E - K = -c \int_0^{\pi/2} \frac{\sin^2 \theta}{(1 - c \sin^2 \theta)^{1/2}} d\theta = O(c)$$

and

$$K' = \int_0^{\pi/2} (1 - c' \sin^2 \theta)^{-1/2} d\theta \leqslant \int_0^{\pi/2} (1 - c')^{-1/2} d\theta = O(c^{-1/2}).$$

Therefore it follows that

$$\begin{aligned} \lim_{c \to 0} L &= \lim_{c \to 0} \left\{ (E - K)K' + E'K \right\} \\ &= \lim_{c \to 0} \left\{ O(c^{1/2}) + 1 \cdot \frac{\pi}{2} \right\} = \frac{\pi}{2} \end{aligned}$$

which concludes the proof. □

We are now in a position to develop a sophisticated algorithm for calculating π. As motivation we first consider an example of Legendre's relation.

Example 3. Let $x = \frac{1}{\sqrt{2}}$. Then we have that $x' = \sqrt{1 - x^2} = \frac{1}{\sqrt{2}}$ which reduces Legendre's relation to the following

$$K\left(\frac{1}{\sqrt{2}}\right) \left[2E\left(\frac{1}{\sqrt{2}}\right) - K\left(\frac{1}{\sqrt{2}}\right) \right] = \frac{\pi}{2}.$$

The fact that our choice of x can reduce Legendre's relation in such a way is an indication of just how useful these theorems are. We now combine Theorem 4.1 and Theorem 4.3 to produce the following result which provides an incredibly efficient method for calculating π, courtesy of Almkvist and Berndt [5].

Theorem 4.4. Let c_n and $M(a, b)$ be defined as in Section 2. Then

$$\pi = \frac{4M^2(1, 1/\sqrt{2})}{1 - \displaystyle\sum_{n=1}^{\infty} 2^{n+1} c_n^2}.$$

Proof. We begin by recalling the relationship between elliptic integrals:

$$I(a, b) = \frac{1}{a} K(x) \quad \text{and} \quad E(x) = \frac{1}{a} J(a, b), \tag{31}$$

where $x = \frac{1}{a}\sqrt{a^2 - b^2}$.

Now if we let $a = 1$ and $b = 1/\sqrt{2}$ then $x = x' = 1/\sqrt{2}$, so employing Theorem 4.1 and (31) it follows that

$$E\left(\frac{1}{\sqrt{2}}\right) = \left(1 - \frac{1}{2}\sum_{n=0}^{\infty} 2^n c_n^2\right) K\left(\frac{1}{\sqrt{2}}\right). \tag{32}$$

Now consider the result obtained in Example 3. Inserting (32) into this expression we obtain

$$K\left(\frac{1}{\sqrt{2}}\right)\left[2\left[\left(1 - \frac{1}{2}\sum_{n=0}^{\infty} 2^n c_n^2\right) K\left(\frac{1}{\sqrt{2}}\right)\right] - K\left(\frac{1}{\sqrt{2}}\right)\right] = \frac{\pi}{2}. \tag{33}$$

Finally by Theorem 2.6 we have

$$K\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{2M(1, 1/\sqrt{2})}$$

so after substituting this into (33), noting that $c_0^2 = 1/2$ and solving for π we obtain the required result. □

It is this theorem that provides the basis for almost all modern computing algorithms that calculate π. The reason it is so efficient is the incredibly fast rate at which the arithmetic-geometric mean converges. Borwein and Borwein use a variation of Theorem 4.4 in [4], offering an in-depth discussion of the efficiency of the algorithm . It is also the vital component in the algorithm produced by Salamin in [12]. For a discussion of how a similar method can be used to calculate various elementary functions see Brent [13].

Although this section has provided an important insight into how the arithmetic-geometric mean can be applied to modern day computing problems we have only really seen the tip of the iceberg. We shall now begin to explore how the arithmetic-geometric mean behaves on the complex plane.

5 The Complex Arithmetic-Geometric Mean

We will now examine what happens to the arithmetic-geometric mean function if we relax the condition that a and b are real and positive. This is leads us to modifying our definition for the arithmetic-geometric mean and subsequently considering the convergence of sequences produced by such an algorithm for complex a and b. We will see that in the complex case the arithmetic-geometric mean is a multi-valued function. Finally we consider a fascinating theorem from which it is possible to generate the various values for the arithmetic-geometric mean for any given a and b. This incorporates much of the theory behind theta functions and modular forms. The following results have been taken almost exclusively from Cox's paper [1].

For now consider our algorithm from Section 2 again

$$\begin{aligned} a_0 &= a, & b_0 &= b \\ a_n &= \frac{a_{n-1}+b_{n-1}}{2}, & b_n &= \sqrt{a_{n-1}b_{n-1}}, \quad n = 1, 2, ... \end{aligned} \tag{34}$$

If we let a and b be complex it is no longer clear which value we take for b_n when n is greater than or equal to 1. In fact for given a and b it is obvious that there are uncountably many sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ and we do not know whether any of these sequences converge to the same limit.

Example 4. We can deduce a common limit for some particular a and b. By letting $a = 0$ or $b = 0$ or $a = \pm b$ we see that the sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ both converge to either 0 or a.

The above example is trivial and not particularly interesting and so for the rest of the section we will only consider complex a and b such that $a \neq 0$, $b \neq 0$ and $a \neq \pm b$.

We would like to define the arithmetic-geometric mean for such a and b and to do this we must first establish a method to distinguish between the possible values for each b_n. We make the following definition.

Definition 5.1. Let $a, b \in \mathbb{C} \setminus \{0\}$ such that $a \neq \pm b$. Then a square root b_1 of ab is called the *right choice* if the following two conditions hold:

1. $|a_1 - b_1| \leqslant |a_1 + b_1|$;
2. If $|a_1 + b_1| = |a_1 - b_1|$ we have $\Im(b_1/a_1) > 0$;

where a_1 and b_1 are defined according to (34) and $\Im(a)$ denotes the imaginary part of a.

Notice how if we let a and b both be real then by the above definition we would always take b_1 to be the positive square root of ab, thus agreeing with our algorithm in Section 2. Note also that we can swap a and b and the right choice will remain unchanged. To see that this definition is logically sound suppose that $\Im(b_1/a_1) = 0$. Then it follows that

b_1/a_1 is real, say r, and we have

$$|a_1 - b_1| = |a_1||1 - r| \neq |a_1||1 + r| = |a_1 + b_1|$$

since $r \neq 0$. We now examine more closely some of the properties of the right choice of b_1.

Lemma 5.2. Let $\angle(a, b) \in [0, \pi]$ denote the unoriented angle between the two complex numbers a and b and suppose that b_1 is the right choice according to Definition 5.1. Then it follows that:

1. $|a_1 - b_1| \leqslant \dfrac{1}{2}|a - b|$
2. $\angle(a_1, b_1) \leqslant \dfrac{1}{2}\angle(a, b)$.

Proof. To prove 1 note how

$$\begin{aligned} |a_1 - b_1||a_1 + b_1| &= |(a + b)/2 - \sqrt{ab}||(a + b)/2 + \sqrt{ab}| \\ &= \frac{1}{4}|(a + b)^2 - 4ab| \\ &= \frac{1}{4}|a - b|^2. \end{aligned}$$

We know that b_1 is the right choice for $\sqrt{ab}$ so it follows that $|a_1 - b_1| \leqslant |a_1 + b_1|$. Multiplying both sides of this inequality by $|a_1 - b_1|$ and taking square roots yields the result.

In order to prove 2 we consider the numbers a, b, a_1 and b_1 geometrically. Letting θ_1 denote the angle between a_1 and b_1, $\angle(a_1, b_1)$, then by the cosine rule we have:

$$|a_1 \pm b_1|^2 = |a_1|^2 + |b_1|^2 \pm 2|a_1||b_1| \cos\theta_1,$$

so as b_1 is the right choice we have $|a_1 - b_1| \leqslant |a_1 + b_1|$, so it is immediately clear that $\theta_1 \leqslant \pi/2$, whence

$$\angle(a_1, b_1) = \theta_1 \leqslant \pi - \theta_1 = \angle(a_1, -b_1).$$

Now it can be shown that one of $\pm b_1$, say b_1', will bisect θ (the angle between a and b). It is obvious that a_1 is the vector from the origin to the midpoint of the parallelogram with sides a and b , so it immediately follows that the angle between a_1 and b_1', $\angle(a_1, b_1')$, is less than or equal to $\theta/2$. Putting these inequalities together and bearing in mind that $b_1' = \pm b_1$ we see that

$$\angle(a_1, b_1) = \theta_1 \leqslant \angle(a_1, b_1') \leqslant \frac{\theta}{2} = (1/2)\angle(a, b)$$

□

With this new definition in hand we could be forgiven for defining the arithmetic-geometric mean for complex numbers by consistently making the right choice for b_n. However the following table shows that it is possible to produce interesting results despite not always making the right choice for b_n:

n	a_n	b_n
0	6.0000000000	5.0000000000
1	5.5000000000	-5.4772255751
2	4.9930339887	$5.4886009750i$
3	$2.4965169944 + 2.7443004875i$	$3.7016733526 + 3.7016733526i$
4	$3.0990951735 + 3.2229869201i$	$3.0417307149 + 3.1889401524i$
5	$3.0704129442 + 3.2059635362i$	$3.0702919901 + 3.2059308051i$
6	$3.0703524671 + 3.2059471706i$	$3.0703524667 + 3.2059471708i$

Although we have not made the right choice for b_1, the sequences $\{a_n\}_{n=0}^{6}$ and $\{b_n\}_{n=0}^{6}$ clearly converge incredibly quickly to the same number correct to 8 decimal places. It appears that the arithmetic-geometric mean will still converge even if we do not make the right choice for b_n all the time. Indeed, given a pair of sequences such that for some N, b_n is the right choice for all $n > N$ then by Lemma 5.2 we know that the sequences $\{a_n\}_{n=N+1}^{\infty}$ and $\{b_n\}_{n=N+1}^{\infty}$ will converge to a common limit, hence $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ will also converge. We therefore make the following definition.

Definition 5.3. Let a, b be complex as before. We call a pair of sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ generated by (34) *good* if b_{n+1} is the right choice for $\sqrt{a_n b_n}$ for all but finitely many $n \geqslant 0$.

Despite our new definitions and the convergence that is apparent in the above table we cannot be sure that every pair of sequences generated for complex a and b converge to a common limit. We shall now consider good sequences and "bad" sequences (i.e. sequences that are not good) in order to bring us closer to defining the arithmetic-geometric mean of two complex numbers.

First we will suppose that $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are both good according to Definition 5.3. This means that for sufficiently large N we may neglect the first N terms of the sequence $\{b_n\}_{n=0}^{\infty}$ so that b_n is the right choice for all $n \geqslant 0$. We can assume by the second part of Lemma 5.2 that θ, the angle between a and b, is less than π. By letting $\theta_n = \angle(a_n, b_n)$ and applying Lemma 5.2 recursively we see that

$$|a_n - b_n| \leqslant \frac{|a-b|}{2^n} \quad \text{and} \quad \theta_n \leqslant \frac{\theta}{2^n} \tag{35}$$

for all $n \geqslant 1$. We now use the fact that $a_n - a_{n+1} = (a_n - b_n)/2$ to rearrange (35) so that

$$|a_n - a_{n+1}| \leqslant \frac{|a-b\,|}{2^{n+1}}.$$

If $m > n$ we have

$$|a_n - a_m| \leqslant \sum_{k=1}^{m-1} |a_k - a_{k+1}| \leqslant \left(\sum_{k=n}^{m-1} 2^{-(k+1)} \right) |a - b| < \frac{|a-b|}{2^n}.$$

We see that $\{a_n\}_{n=0}^{\infty}$ is a Cauchy sequence and is therefore convergent. By (35) this immediately implies

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

We now know that any two good sequences generated by (34) will converge to a common limit, but there is more to tell about what this limit may be. If we now let $m_n = \min\{|a_n|, |b_n|\}$ it should be obvious that $|b_{n+1}| \geqslant m_n$ and in order to relate $|a_{n+1}|$ and m_n we use the cosine rule again to obtain

$$\begin{aligned} (2|a_{n+1}|)^2 &= |a_n|^2 + |b_n|^2 + 2|a_n||b_n| \cos\theta_n \\ &\geqslant 2m_n^2(1 + \cos\theta_n) = 4m_n^2 \cos^2(\theta_n/2). \end{aligned}$$

Given $\theta < \pi$ and as b_n will always be the right choice by construction, it follows from (35) that $0 \leqslant \theta_n < \pi$ and hence $m_{n+1} \geqslant m_n \cos(\theta_n/2)$. By repeated application of (35) we see that

$$m_n \geqslant \left(\prod_{k=1}^{n} \cos(\theta/2^k) \right) m_0.$$

The product on the right is precisely Euler's formula from Section 4 so we conclude that for all $n \geqslant 1$

$$m_n \geqslant \left(\frac{\sin\theta}{\theta} \right) m_0.$$

Now as $0 \leqslant \theta < \pi$ and given that a and b are distinct and non-zero we arrive at the startling fact that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n \neq 0.$$

This is a remarkable result. We now know that so long as b_n is not the right choice for finitely many n then the two sequences will not only converge to a common limit, but that limit will be non-zero.

The last thing we must establish is whether the sequences generated by (34) that are not good converge to a common limit and what that limit may be. So suppose $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are both not good sequences and let $M_n = \max\{|a_n|, |b_n|\}$. Clearly for all $n \geqslant 0$ we have $M_{n+1} \leqslant M_n$.

Now suppose that for some n, b_{n+1} is not the right choice for $(a_n b_n)^{1/2}$. This means that there is some b_{n+1} which is the right choice, say b'_{n+1}. It follows that $b_{n+1} = -b'_{n+1}$ and hence the a_{n+2} term in the sequence that is not good can be written as

$$a_{n+2} = \frac{1}{2}(a_{n+1} + b_{n+1}) = \frac{1}{2}(a_{n+1} - b'_{n+1}).$$

Now given that b'_{n+1} by construction is the right choice for $(a_n b_n)^{1/2}$ we can apply Lemma 5.2 so

$$|a_{n+2}| = \frac{1}{2}|a_{n+1} - b_{n+1}| \leqslant \frac{1}{4}|a_n - b_n| \leqslant \frac{1}{2}M_n.$$

But it is also true that $|b_{n+2}| \leqslant M_n$ so by adding $|b_{n+2}|$ to each side and observing that $2M_{n+3} \leqslant |a_{n+2}| + |b_{n+2}|$ we conclude that

$$M_{n+3} \leqslant \frac{3}{4}M_n. \tag{36}$$

Given that $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are not good sequences (36) must occur infinitely often, so it must be the case that

$$\lim_{n\to\infty} M_n = 0.$$

So we know that every pair of sequences generated by (34) that are not good for some distinct non-zero complex numbers a and b will converge to a common limit of 0, and every pair of good sequences will converge to some non-zero limit. We have therefore shown the following:

Theorem 5.4. Given two non-zero complex numbers a and b such that $a \neq \pm b$, then any pair of sequences generated by (34) will converge to a common limit. Furthermore this limit is non-zero if and only if $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are good sequences.

This is fascinating because it implies that of all the possible sequences we could generate for some complex a and b, countably many will converge to a non-zero limit. We can now define the arithmetic-geometric mean for complex numbers.

Definition 5.5. Let a, b be as before. We say that the non-zero complex number μ is the arithmetic-geometric mean, $M(a, b)$, of a and b if there are good sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ generated by (34) such that

$$\mu = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

So $M(a, b)$ is a multi-valued function of a and b and it has a countable number of values. We will distinguish the common limit of $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ where b_n is the right choice for every n as the *simplest value* of $M(a, b)$. This distinction plays an important role in the next theorem. Clearly if $\Im(a) = \Im(b) = 0$ and a and b are positive then the simplest value of $M(a, b)$ is simply the arithmetic-geometric mean defined in Section 2.

We now arrive at the final result of this project. It is a beautiful theorem that provides us with a method for finding all possible values of $M(a, b)$.

Theorem 5.6. Let a and b be such that $|a| \geqslant |b|$. Let μ and λ denote the simplest values of $M(a, b)$ and $M(a + b, a - b)$ respectively. Then all possible values μ' of $M(a, b)$ are given by the formula

$$\frac{1}{\mu'} = \frac{d}{\mu} + \frac{ic}{\lambda},$$

where d and c are arbitrary relatively prime integers satisfying $d \equiv 1 \mod 4$ and $c \equiv 0 \mod 4$.

This is theorem remarkable because it provides us with a formula that will produce all the possible values for $M(a,b)$. It is a particularly deep result and the proof uses many ideas from many different areas of mathematics. Below we sketch the proof that can be found in [1] which is similar to that of Geppert [15], though an alternative may also be found by David [14].

We begin by considering properties of the Jacobi theta functions defined on the upper half plane. Letting $\mathfrak{H} = \{\tau \in \mathbb{C} : \Im(\tau) > 0\}$ we have

$$\begin{aligned}
\Theta_3(\tau,0) &= 1 + 2\sum_{n=1}^{\infty} \mathrm{q}^{n^2} &&= p(\tau),\\
\Theta_4(\tau,0) &= 1 + 2\sum_{n=1}^{\infty} (-1)^n \mathrm{q}^{n^2} &&= q(\tau),\\
\Theta_2(\tau,0) &= 2\sum_{n=1}^{\infty} \mathrm{q}^{\frac{1}{4}(2n-1)^2} &&= r(\tau),
\end{aligned}$$

where q$= e^{\pi i \tau}$. Gauss is responsible for the notation on the right, whereas the more common notation on the left will be found in [7] and [4]. Note that as $|\mathrm{q}| < 1$ for all $\tau \in \mathfrak{H}$ these theta functions are holomorphic functions of τ.

The Jacobi theta functions can also be expressed as infinite products which show that they are non-vanishing on $\mathfrak{H}$:

$$\begin{aligned}
p(\tau) &= \prod_{n=1}^{\infty} (1-\mathrm{q}^{2n})(1+\mathrm{q}^{2n-1})^2,\\
q(\tau) &= \prod_{n=1}^{\infty} (1-\mathrm{q}^{2n})(1-\mathrm{q}^{2n-1})^2,\\
r(\tau) &= 2\mathrm{q}^{1/4} \prod_{n=1}^{\infty} (1-\mathrm{q}^{2n})(1+\mathrm{q}^{2n})^2.
\end{aligned}$$

There are a huge amount of formulas associated with these functions and the proofs of these can be found in Whittaker and Watson [7]. In particular we will use the following transformations where it shall be assumed that $\Re(-i\tau)^{1/2} > 0$, where $\Re(z)$ denotes the real part of z:

$$\begin{array}{rclrcl}
p(\tau+1) &=& q(\tau), & p(-1/\tau) &=& (-i\tau)^{1/2} p(\tau),\\
q(\tau+1) &=& p(\tau), & q(-1/\tau) &=& (-i\tau)^{1/2} r(\tau),\\
r(\tau+1) &=& e^{\pi i/4} r(\tau), & r(-1/\tau) &=& (-i\tau)^{1/2} q(\tau).
\end{array} \tag{37}$$

In order to motivate our discussion of theta functions in relation to the arithmetic-geometric mean consider the following:

$$\begin{aligned} p(\tau)^2 + q(\tau)^2 &= 2p(2\tau)^2, \\ p(\tau)^2 - q(\tau)^2 &= 2r(2\tau)^2, \\ p(\tau)q(\tau) &= q(2\tau)^2. \end{aligned} \tag{38}$$

Proof of these identities may be found in [16]. Consider the first and last equations above. Clearly $p(2\tau)^2$ is the arithmetic mean of $p(\tau)^2$ and $q(\tau)^2$ and $q(2\tau)$ is the geometric mean of $p(\tau)$ and $q(\tau)$. Cox formalises this relation in the following lemma by introducing the function $k'(\tau) = q(\tau)^2/p(\tau)^2$.

Lemma 5.7. Suppose there is $\tau \in \mathfrak{H}$ such that $k'(\tau) = b/a$ for some complex a and b. Let $\mu = a/p(\tau)^2$ and for $n \geqslant 0$ let $a_n = \mu p(2^n\tau)^2$ and $b_n = \mu q(2^n\tau)^2$. Then

1. $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are good sequences satisfying (34);
2. $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \mu$.

Proof. Clearly $a_0 = a$ and by definition of $k'(\tau)$ it is obvious that $b_0 = b$. Given our above observation that $p(2\tau)^2$ is the arithmetic mean of $p(\tau)^2$ and $q(\tau)^2$ and $q(2\tau)$ is the geometric mean of $p(\tau)$ and $q(\tau)$ it follows easily that the other conditions of (34) are also satisfied. Note also that $e^{\pi i 2^n \tau}$ tends to 0 as n tends to infinity, so we have $\lim_{n\to\infty} p(2^n\tau)^2 = \lim_{n\to\infty} q(2^n\tau)^2 = 1$, from which 2 follows directly. Finally observe that we have $\mu \neq 0$, so by Theorem 5.4 both sequences are necessarily good. $\square$

We may conclude then that every solution of $k'(\tau) = b/a$ will provide us with a value of $M(a, b)$. In order to begin studying all possible solutions of $k'(\tau) = b/a$ we now consider the region R_1 contained in $\mathfrak{H}$ defined to be

$$R_1 = \{\tau \in \mathfrak{H} : |\Re(\tau)| \leq 1 \text{ and } |\Re(1/\tau)| \leq 1\}.$$

There is a well-known lemma related regarding this region and the function $k'(\tau)$ that may be found in [7, pp. 481-484]. We state it here without proof:

Lemma 5.8. The function $k'(\tau)^2$ assumes every value in $\mathbb{C} \setminus \{0, 1\}$ exactly once in the region $R_1' = R_1 \setminus \{\partial R_1 \cap \{\tau \in \mathfrak{H} : \Re(\tau) < 0\}\}$, where ∂R_1 denotes the boundary of R_1.

See Figure 1 for a graphical representation of R_1'. Keeping in mind the restrictions we placed on a and b in the statement of Theorem 5.6 it is clear that $(b/a)^2 \in \mathbb{C} \setminus \{0, 1\}$, so according to Lemma 5.8 we will always be able to solve $k'(\tau)^2$ or equivalently $k'(\tau) = \pm b/a$. The next stage of the proof relies on showing that

$$k'\left(\frac{\tau}{2\tau + 1}\right) = -k'(\tau). \tag{39}$$

To prove (39) we move into the realm of modular forms. We will consider subgroups of the special linear group of degree 2 over the integers that act on $\mathfrak{H}$ by linear fractional

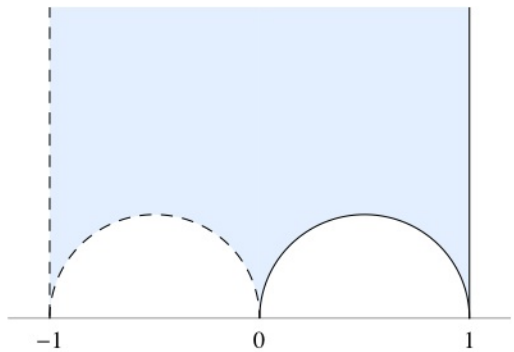


Figure 1: The region R_1' is shaded.

transformations. In other words if $\tau \in \mathfrak{H}$ and $\phi \in SL(2,\mathbb{Z})$ then we have

$$\phi\tau = \begin{pmatrix} a & b \\ c & d \end{pmatrix}(\tau) = \frac{a\tau + b}{c\tau + d}.$$

The first subgroup of $SL(2,\mathbb{Z})$ that we will consider is the principal congruence subgroup of level 2:

$$\Phi(2) = \{\phi \in SL(2,\mathbb{Z}) : \phi \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod 2\}.$$

It is worth observing that $-1 \in \Phi(2)$ and that the normal subgroup $\Phi(2)/\{\pm 1\}$ acts on $\mathfrak{H}$. The following properties of $\Phi(2)$ are important in bringing us closer to showing the invariance of $k'(\tau)$. We state them here without proof though the argument can be found in [1]:

1. $\Phi(2)/\{\pm 1\}$ acts freely on $\mathfrak{H}$.
2. $\Phi(2)$ is generated by -1, $U = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $V = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$.
3. For any $\tau \in \mathfrak{H}$ there exists $\phi \in \Phi(2)$ such that $\phi\tau \in R_1$.

There are a couple of things to note about the above properties. The third property is an important step in realising the solutions of $k'(\tau)$. The fact that we can map any point in $\mathfrak{H}$ into the region R_1 brings us one step closer to proving the invariance of $k'(\tau)$. The other thing to note is that within the proof of the second property it emerges that $\phi \in \Phi(2)$ is in the subgroup generated by U and V which is particularly useful as we come to consider how $p(\tau)$ and $q(\tau)$ transform under elements of $\Phi(2)$. It is also important to highlight the fact that U and V commute modulo 4.

Lemma 5.9. Let $\phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Phi(2)$ and suppose $a \equiv d \equiv 1 \mod 4$. Then we have

1. $p(\phi\tau)^2 = (c\tau + d)p(\tau)^2$,
2. $q(\phi\tau)^2 = i^c(c\tau + d)q(\tau)^2$.

Proof. It is straightforward to see that by (37)

$$p(U\tau)^2 = p(\tau)^2 \quad \text{and} \quad q(U\tau)^2 = q(\tau)^2. \tag{40}$$

We use the fact that $V = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} U^{-1} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ to show that again by (37)

$$p(V\tau)^2 = (2\tau + 1)p(\tau)^2 \quad \text{and} \quad q(V\tau)^2 = -(2\tau + 1)q(\tau)^2. \tag{41}$$

Now given that ϕ is in the subgroup generated by U and V and as we have just shown 1 and 2 above hold for U and V we may proceed inductively on the length of ϕ as a word in U and V.

To prove 1 note that we have that $U\phi = \begin{pmatrix} u_1 & u_2 \\ c & d \end{pmatrix}$ and $V\phi = \begin{pmatrix} v_1 & v_2 \\ 2a + c & 2b + d \end{pmatrix}$ for any $\phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where u_1, u_2, v_1, and v_2 are arbitrary. Now suppose that $p(\phi\tau)^2 = (c\tau + d)p(\tau)^2$. Then by (40) and (41) we have

$$p(U\phi\tau)^2 = p(\phi\tau)^2 = (c\tau + d)p(\tau)^2,$$

and

$$\begin{aligned} p(V\phi\tau)^2 &= (2\phi\tau + 1)p(\phi\tau)^2 = (2\phi\tau + 1)(c\tau + d)p(\tau)^2 \\ &= ((2a + c)\tau + (2b + d))p(\tau)^2, \end{aligned}$$

showing that 1 holds for all $U\phi$ and $V\phi$.

To prove 2 note that if $\phi = U^{i_1}V^{j_1}...U^{i_n}V^{j_n}$ then by repeated application of (40) and (41) it follows that

$$q(\phi\tau)^2 = (-1)^{\sum j_n}(c\tau + d)q(\tau)^2.$$

As U and V commute modulo 4 we have

$$\phi \equiv \begin{pmatrix} 1 & 2\sum i_n \\ 2\sum j_n & 1 \end{pmatrix} \mod 4.$$

Hence 2 follows from the fact that $c \equiv 2\sum j_n \mod 4$. □

From this lemma it follows immediately that with V as above

$$k'(V\tau) = \frac{q(V\tau)^2}{p(V\tau)^2} = \frac{i^2(2\tau + 1)q(\tau)^2}{(2\tau + 1)p(\tau)^2} = -k'(\tau),$$

which proves (39). Another important thing to note about Lemma 5.9 is that it is fairly straightforward to modify the proof to show that if $\phi \in \Phi(2)$ we have $p(\phi\tau)^4 = (c\tau +$

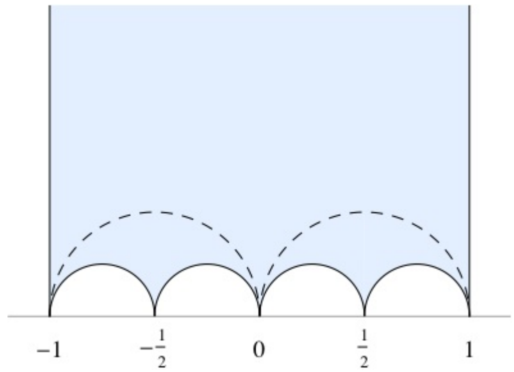


Figure 2: Clearly the shaded region, R, properly contains R_1 (the region above the dotted lines).

$d)^2 p(\tau)^4$ and $q(\phi\tau)^4 = (c\tau + d)^2 q(\tau)^4$. This implies immediately that $k'(\tau)^2$ is invariant under $\Phi(2)$.

So what have we done so far? We have shown through Lemma 5.7 that we can express the arithmetic-geometric mean of two complex numbers using the Jacobi theta functions. We have introduced the function $k'(\tau) = q(\tau)^2/p(\tau)^2$ and shown that every solution of $k'(\tau) = b/a$ will provide us with a value for $\mu = a/p(\tau)^2$ which is a value of $M(a,b)$. We then restricted ourselves to the region R_1', on which we know that $k'(\tau)^2$ assumes every value in $\mathbb{C} \setminus \{0,1\}$ exactly once. The conditions on a and b in Theorem 5.6 ensure that $k'(\tau) \notin \{0,1\}$, so in order to study solutions of $k'(\tau)$ we need only be concerned with this strip of $\mathfrak{H}$ defined by R_1'. Although $k'(\tau)^2$ assumes every value in $\mathbb{C} \setminus \{0,1\}$ exactly once on R_1 the same cannot be said for $k'(\tau)$. However by considering the transformations of the theta functions by way of the group $\Phi(2)$ acting on $\mathfrak{H}$ we are able to deduce that for any solution of $k'(\tau)$ that does not lie in $\mathfrak{H}$, say $-b/a$, we have a corresponding value in $\mathfrak{H}$ such that $k'\left(\frac{\tau}{2\tau+1}\right) = b/a$.

The upshot of this is that we now have to show three things to prove Theorem 5.6. We need to find every solution τ of $k'(\tau)$, we need to study how these values for τ relate to $\mu = a/p(\tau)^2$ and we need to show that every value of $M(a,b)$ arises in this way.

We will now introduce two subgroups of $\Phi(2)$:

$$\begin{aligned} \Phi(2)_0 &= \{\phi \in \Phi(2) : a \equiv d \equiv 1 \mod 4\}, \\ \Phi_2(4) &= \{\phi \in \Phi(2)_0 : c \equiv 0 \mod 4\}. \end{aligned}$$

Directly employing Lemma 5.9 we have

$$\begin{aligned} p(\phi\tau)^2 &= (c\tau + d)p(\tau)^2, \qquad \phi \in \Phi(2)_0 \\ q(\phi\tau)^2 &= (c\tau + d)q(\tau)^2, \qquad \phi \in \Phi_2(4). \end{aligned} \tag{42}$$

Given that $\Phi_2(4) < \Phi(2)_0 < \Phi(2)$ it follows immediately from (42) that $k'(\tau)$ is invariant

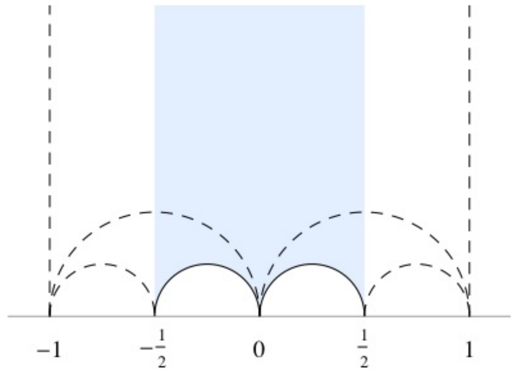


Figure 3: The shaded region is R_2. The dashed lines indicate R and R_1

under $\Phi_2(4)$ and $\Phi(2)_0$. This is an important point that will be useful later.

For now we will turn our attention to the quotients of $\mathfrak{H}$ by $\Phi(2)$ and $\Phi_2(4)$. Recall the region R_1 we defined earlier and consider the following larger region (see Figure 2):

$$R = \{\tau \in \mathfrak{H} : |\Re(\tau)| \leqslant 1, |\tau \pm 1/4| \geqslant, |\tau \pm 3/4| \geqslant 1/4\}.$$

In a similar fashion to how we defined R_1' before we now set R' to be

$$R' = R \setminus \{\partial R \cap \{\tau : \Re(\tau) < 0\}\}.$$

Having defined these regions we come to the following important lemma.

Lemma 5.10. The fundamental domains for $\Phi(2)$ and $\Phi_2(4)$ are R_1' and R' respectively. Moreover the functions $k'(\tau)^2$ and $k'(\tau)$ induce the following conformal maps

$$\begin{aligned} k_*'(\tau)^2 \quad &: \mathfrak{H}/\Phi(2) \to \mathbb{C} \setminus \{0, 1\} \\ k_*'(\tau) \quad &: \mathfrak{H}/\Phi_2(4) \to \mathbb{C} \setminus \{0, \pm 1\}, \end{aligned}$$

where $k_*'(\tau)$ denotes the ratio of the theta functions of the complex conjugate of q.

Proof. We will prove that R_1' is a fundamental domain for $\Phi(2)$ and the first conformal mapping in the above lemma. Proof of the other half can be found again in Cox's paper [1].

Suppose we have $\tau \in \mathfrak{H}$. Then by the previously mentioned properties we know that $\phi\tau \in R_1$ for some $\phi \in \Phi(2)$. Now consider the action of U and V on ∂R_1. Clearly U maps the line at -1 to 1 and V maps the left semi-circle to the right one, so $\phi\tau \in R_1'$. Now suppose we have $\gamma \in \Phi(2)$ such that $\gamma\tau \in R_1'$. Then we have $k'(\gamma\tau)^2 = k'(\tau)^2 = k'(\phi\tau)^2$ so that by Lemma 5.8, $\gamma\tau = \phi\tau$. Given that $k'(\tau)^2$ is invariant under $\Phi(2)$ we conclude that R_1' is a fundamental domain for $\Phi(2)$.

By Lemma 5.8 we have a bijection $k_*'(\tau)^2 : \mathfrak{H}/\Phi(2) \to \mathbb{C} \setminus \{0, 1\}$. We also have that $\mathfrak{H}/\Phi(2)$ is a complex manifold as $\Phi(2)/\{\pm 1\}$ acts freely on $\mathfrak{H}$. As $k_*'(\tau)^2$ is holomorphic it is therefore a conformal map. $\square$

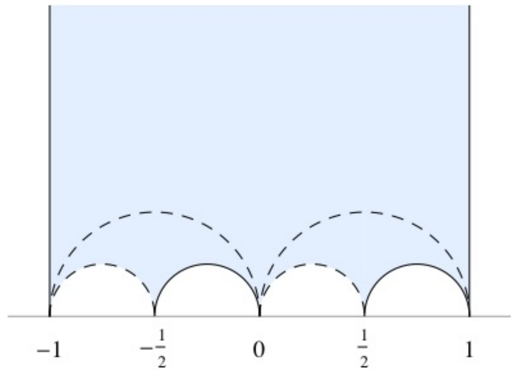


Figure 4: The shaded region is $\hat{R}$, the larger dashed semi-circles depict R_1 and the smaller dashed semi-circles is the boundary of $\hat{R}$.

We are gradually building up a repertoire of tools with which we can examine the solutions of $k'(\tau)$. We have two more lemmas to consider before we can begin to embark on the proof of Theorem 5.6 proper. Before stating them we must first define another region of the complex plane which we shall denote $R_2 = (1/2)R_1$, so that $R_2 \subseteq R_1$ (see Figure 3).

Lemma 5.11. Let R_1 and R_2 be defined as above. Then we have

$$\begin{aligned} k'(R_1) &= \{z \in \mathbb{C} \setminus \{0, \pm 1\} : \Re(z) \geqslant 0\}, \\ k'(R_2) &= \{z \in \mathbb{C} \setminus \{0, \pm 1\} : |z| \leqslant 1\}. \end{aligned}$$

We will forego the proof, though Cox provides the argument for this lemma. The fact that $k'(\tau)$ maps the region R_1 into the right half of the complex plane and the region R_2 into the unit circle about the origin plays a significant role in the final result of this project. We will now piece together what we have already seen and outline the proof of Theorem 5.6. First we must define another region of the complex plane closely related to the region R (see Figure 4):

$$\hat{R} = \{\tau \in R : |\tau - 1/4| > 1/4, |\tau + 3/4| > 1/4\}.$$

Now let a, $b \in \mathbb{C} \setminus \{0\}$ be such that $a \neq \pm b$ and suppose $\tau \in \mathfrak{H}$ satisfies $k'(\tau) = b/a$. We have established by Lemma 5.7 that $\mu = a/p(\tau)^2$ is a value of $M(a, b)$. We will now show the following.

Lemma 5.12. If $\tau \in \hat{R}$ then μ is the simplest value of $M(a, b)$.

Proof. Consider the sequences $a_n = \mu p(2^n\tau)^2$ and $b_n = \mu q(2^n\tau)^2$ where $n \in \mathbb{N}$. We have already shown by way of Lemma 5.7 that these are good sequences that converge to a common limit μ. We will now show that b_{n+1} is the right choice for every $n \geqslant 0$. It is

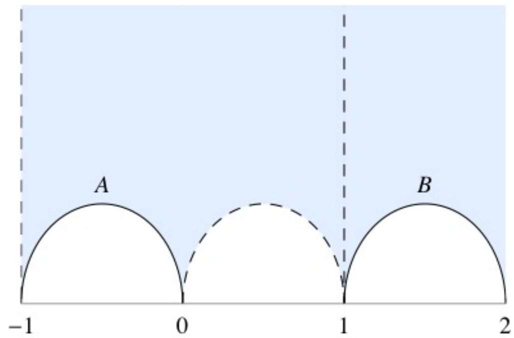


Figure 5: The dashed lines denote the region R_1.

straightforward to prove that $\Re\left(\frac{b_{n+1}}{a_{n+1}}\right) \geqslant 0$ if and only if $|a_{n+1}+b_{n+1}| \geqslant |a_{n+1}-b_{n+1}|$ and also that $\Re\left(\frac{b_{n+1}}{a_{n+1}}\right) = 0$ if and only $|a_{n+1}+b_{n+1}| = |a_{n+1}-b_{n+1}|$. Therefore proving that b_{n+1} is the right choice is equivalent to showing that $\Re\left(\frac{b_{n+1}}{a_{n+1}}\right) \geqslant 0$ for all $n \geqslant 0$ and that if $\Re\left(\frac{b_{n+1}}{a_{n+1}}\right) = 0$ then $\Im\left(\frac{b_{n+1}}{a_{n+1}}\right) > 0$ by the definition of the right choice for b_{n+1}. By our definition of a_{n+1} and b_{n+1} we have that

$$\frac{b_{n+1}}{a_{n+1}} = \frac{q(2^{n+1}\tau)^2}{p(2^{n+1})^2} = k'(2^{n+1}\tau).$$

If we let $\tau \in \hat{R}$ then it naturally follows that we must show for all $n \geqslant 0$, $\Re(k'(2^{n+1}\tau)) \geqslant 0$ and that if $\Re(k'(2^{n+1}\tau)) = 0$ then we must have $\Im(k'(2^{n+1}\tau)) > 0$.

Now let $\tilde{R_1}$ denote the region covered by translating R_1 by $2m$ to the left or right for all integers m, so that $k'(\tau)$ has period 2. It follows from Lemma 5.11 that the real part of $k'(\tau)$ is nonnegative on R_1, and hence is nonnegative on all of $\tilde{R_1}$. Also we have that the only points on R_1 that satisfy $\Re(k'(\tau)) = 0$ are those that lie on the boundary, ∂R_1. Considering the product expansions of $p(\tau)$ and $q(\tau)$ we see that $k'(\tau)$ is real when $\Re(\tau) = \pm 1$, whence we infer that $\Re(k'(\tau)) = 0$ may only occur on the boundary semi-circles of R_1. Due to the periodicity of $k'(\tau)$ it follows immediately that $\Re(k'(\tau)) > 0$ only on the interior of $\tilde{R_1}$.

It should be clear from Figure 4 that if $\tau \in \hat{R}$ then clearly for $n \geqslant 0$ we have $2^{n+1}\tau \in \tilde{R_1}$ and furthermore for all $n \geqslant 1$ we have that $2^{n+1}\tau$ lies within the interior of $\tilde{R_1}$. By the above argument it emerges that $\Re(2^{n+1}\tau) > 0$ for all $n \geq 0$, except when $n = 0$ and 2τ lies on the boundary of $\tilde{R_1}$. Thus we will be done once we have shown that when $\tau \in \hat{R}$ and $2\tau \in \partial\tilde{R_1}$ we have $\Im(k'(2\tau)) > 0$. This means that $k'(\tau)$ must lie on one of the semi-circles A or B in Figure 5. However we know that $k'(\tau)$ assumes the same values on B as it does on A by the periodicity of $k'(\tau)$ so all we need to show is that $\Im(k'(2\tau)) > 0$ for $2\tau \in A$.

Consider now the linear fractional transformation $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ which maps the line $\Re(\tau') = 1$ onto A. This enables us to write $2\tau = -1/\tau'$ with $\Re(\tau') = 1$. Then by (37) we see that

$$k'(2\tau) = k'(-1/\tau') = \frac{q(-1/\tau')^2}{p(-1/\tau')^2} = \frac{r(\tau')}{p(\tau')}.$$

The product expansions for $r(\tau')$ and $p(\tau')$ when $\Re(\tau') = 1$ show straight away that $\Im(r(\tau')/p(\tau')) > 0$ and we are done.

□

We now have everything we need to examine every solution of $k'(\tau)$, we can also now study how they relate to $a/p(\tau)^2$ and moreover we can now show that every value of $M(a, b)$ must arise in this way.

Proof of Theorem 5.6. Suppose that we have $a,\ b \in \mathbb{C} \setminus \{0\}$ such that $a \neq \pm b$ and $|a| \geqslant |b|$. Given that $|b/a| \leqslant 1$ we can apply Lemma 5.11, so there exists $\tau' \in R_2$ such that $k'(\tau') = b/a$. We must ensure that τ' also lies in $\hat{R}$, so that we can apply Lemma 2.9. Suppose $\tau' \notin \hat{R}$. Then τ_0 must lie on the semi-circle from 0 to 1/2 (see Figure 5). However given that $k'(\tau')$ is invariant under $\Phi_2(4)$ by (42), we have $k'(\phi\tau_0) = k'(\tau') = b/a$ for some $\phi \in \Phi_2(4)$. So all we need to do is find a ϕ that maps the semi-circle between 0 and 1/2 onto the semi-circle from $-1/2$ to 0. By letting $\phi = \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}$ we can replace any τ' by $\phi\tau' \in R_2 \cap \hat{R}$.

Now we know from Lemma 5.10 that $k'(\tau)$ generates a bijection between $\mathfrak{H}/\Phi_2(4)$ and $\mathbb{C} \setminus \{0\}$, so it follows that every possible solution τ of $k'(\tau) = b/a$ will be given by $\tau = \phi\tau'$ for some $\phi \in \Phi_2(4)$. We now have the following set of values for $\mu' = M(a, b)$:

$$\left\{ \frac{a}{p(\phi\tau')^2} : \phi \in \Phi_2(4) \right\}.$$

Now by letting $\mu = a/p(\tau')^2$ we see immediately from Lemma 5.12 that μ is the simplest value of $M(a, b)$ as $\tau' \in \hat{R}$. By (42) we have $p(\phi\tau')^2 = (c\tau' + d)p(\tau')^2$ with ϕ as before, which we now substitute together with μ into the reciprocal of the above set. Then we have the following:

$$T = \{(c\tau' + d)/\mu : \phi \in \Phi_2(4)\}.$$

Now consider the bottom row $(c,\ d)$ of some $\phi \in \Phi_2(4)$. It is straightforward to see that these are pairs (c, d) such that $c \equiv 0 \bmod 4$, $d \equiv 1 \bmod 4$ and that the greatest common divisor of c and d is 1. Now if we set $\lambda = i\mu/\tau'$ we can re-write T as

$$T = \left\{ \frac{d}{\mu} + \frac{ic}{\lambda} : \gcd(c, d) = 1,\ c \equiv 0 \mod 4,\ d \equiv 1 \mod 4 \right\}.$$

We will now show that λ is the simplest value of $M(a+b, a-b)$. In Lemma 5.7 we defined $a = \mu p(\tau')^2$ and $b = \mu q(\tau')^2$. So by the properties of theta functions mentioned earlier we have

$$a + b = \mu(p(\tau')^2 + q(\tau')^2) = 2\mu p(2\tau')^2 = 2\mu\left(\frac{i}{2\tau'}\right) p\left(\frac{-1}{2\tau'}\right)^2 = \lambda p\left(\frac{-1}{2\tau'}\right)^2$$

and also

$$a - b = \mu(p(\tau')^2 - q(\tau')^2) = 2\mu r(2\tau')^2 = 2\mu\left(\frac{i}{2\tau'}\right) q\left(\frac{-1}{2\tau'}\right)^2 = \lambda q\left(\frac{-1}{2\tau'}\right)^2.$$

This shows that λ is a value of $M(a+b, a-b)$. Given that $\tau' \in R_2$ it is clear that $2\tau' \in R_1$ by a similar argument used in Lemma 5.12. However under the transformation $U = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ sends lines in R_1 to lines and semi-circles to semi-circles, so R_1 is stable under U. Therefore we have that $U\tau' = -1/2\tau' \in R_1$ and given that $R_1 \subseteq \hat{R}$ we immediately see that λ is the simplest value of $M(a+b, a-b)$ by Lemma 5.12. We can now state that if $\mu' = M(a, b)$ for some a, b as before then

$$\mu' \in T = \left\{\frac{d}{\mu} + \frac{ic}{\lambda} : \gcd(c, d) = 1,\ c \equiv 0 \mod 4,\ d \equiv 1 \mod 4\right\},$$

where μ and λ are the simplest values of $M(a, b)$ and $M(a+b, a-b)$.

We have thus found the set of all solutions of $k'(\tau) = b/a$ and seen how the set consists of values of the reciprocal of $M(a, b)$. Furthermore we have shown that the elements of this set can be written in terms of the simplest values λ and μ of $M(a+b, a-b)$ and $M(a, b)$ respectively. All that remains to be seen is that the reciprocal of every value of $M(a, b)$ belongs to this set.

Let $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ be good sequences that satisfy $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \mu'$ for some value μ' of $M(a, b)$. Therefore there is exists an m such that b_{n+1} is the right choice for $(a_n b_n)^{1/2}$ for all $n \geqslant m$, and hence μ' is the simplest value for $M(a_m, b_m)$. By Lemma 5.10 we can assume there exists $\tau \in R'$ such that $k'(\tau) = b_m/a_m$. Given that R' is contained in $\hat{R}$ and μ' is the simplest value for $M(a_m, b_m)$ it follows that $\tau \in \hat{R}$. We can now employ Lemma 5.12 to show that $\mu' = a_m/p(\tau)^2$ and also for $n \geqslant m$,

$$a_n = \mu' p(2^{n-m}\tau)^2 \quad \text{and} \quad b_n = \mu' q(2^{n-m}\tau). \tag{43}$$

We now consider a_{m-1} and b_{m-1}. By considering the following equation

$$x^2 - (a_{m-1} + b_{m-1})x + a_{m-1}b_{m-1} = 0$$

and using the fact that $a_{m-1} + b_{m-1} = 2a_m$ and $a_{m-1}b_{m-1} = b_m^2$. So by the quadratic formula we see that

$$\{a_{m-1}, b_{m-1}\} = \{a_m \pm \sqrt{(a_m^2 - b_m^2)}\}.$$

Now by applying the properties of the theta functions we have

$$(a_m^2 - b_m^2) = \mu'^2(p(\tau)^4 - q(\tau)^4) = \mu' r(\tau)^4,$$

and furthermore

$$a_m \pm \sqrt{(a_m^2 - b_m^2)} = \mu'(p(\tau)^2 \pm r(\tau)^2) = \left\{ \begin{array}{l} \mu' p(\tau/2)^2 \\ \mu' q(\tau/2)^2 \end{array} \right. .$$

So either we have $a_{m-1} = \mu' p(\tau/2)^2$ and $b_{m-1} = \mu' q(\tau/2)^2$ or vice versa.

Now let $\tau_0 = \tau/2$ and consider the case when $a_{m-1} = \mu' p(\tau/2)^2$ and $b_{m-1} = \mu' q(\tau/2)^2$. Then for $n \geqslant m-1$ it follows from (43) that

$$a_n = \mu' p(2^{n-m+1}\tau_0)^2 \quad \text{and} \quad b_n = \mu' q(2^{n-m+1}\tau_0)^2 \tag{44}$$

Now if $a_n = \mu' q(2^{n-m+1}\tau_0)^2$ and $b_n = \mu' p(2^{n-m+1}\tau_0)^2$, we let $\tau_0 = \tau/2+1$ and so by (37) we have $a_{m-1} = \mu' p(\tau_0)^2$ and also $b_{m-1} = \mu' q(\tau_0)^2$. We also see that $p(2^{n-m+1}\tau_0) = p(2^{n-m}\tau)$ and similarly $q(2^{n-m+1}\tau_0) = p(2^{n-m}\tau)$ for all $n \geqslant m$ so we conclude that (44) holds for our choice of τ_0 and $n \geqslant m-1$.

Applying this argument inductively we see that there is $\tau_m \in \mathfrak{H}$ such that

$$a_n = \mu' p(2^n \tau_m)^2 \quad \text{and} \quad b_n = \mu' q(2^n \tau_m)^2 \ .$$

This shows that we have $\mu' = a/p(\tau_m)^2$ and also $k'(\tau_m) = b/a$. We can conclude then that $(\mu')^{-1} = p(\tau_m)^2/a$ belongs to the set T defined above. Therefore the reciprocal of every single value of $M(a, b)$ belongs to this set T, and the proof is complete. $\square$

The formula given by Theorem 5.6 lends itself fairly easily to modern mathematical computing. Figure 6 (bottom) shows a plot of the arithmetic-geometric means for $a = 12 + 32i$ and $b = 2 - i$, where d takes values from [-403,397] and c takes values between [-400,400]. The elliptical nature of the plot can be easily explained if we analyse Theorem 5.6. The values for $1/\mu'$ lie a lattice on the complex plane (see top of Figure 6). Naturally there will be points missing that correspond to when c and d are not co-prime. The orientation of this lattice is dependent on our initial values of a and b and varying these values will contort the lattice. The values of c and d correspond directly to the vertical and horizontal axes of the lattice respectively. Now if we consider the map $f : \tau \to 1/\tau$ which sends $0 \to \infty$, $\infty \to 0$ and fixes 1 and -1 we can see that under such a map"lines are sent to circles". So when we plot the reciprocal of the values generated by Theorem 5.6 we are in fact sending the lines of the lattice to circles, however they look like ellipses in Figure 6 because of the difference in scale between the two axes. In Figure 6 the colour of each ellipse corresponds to a value of d and the number of points that lie on each ellipse is

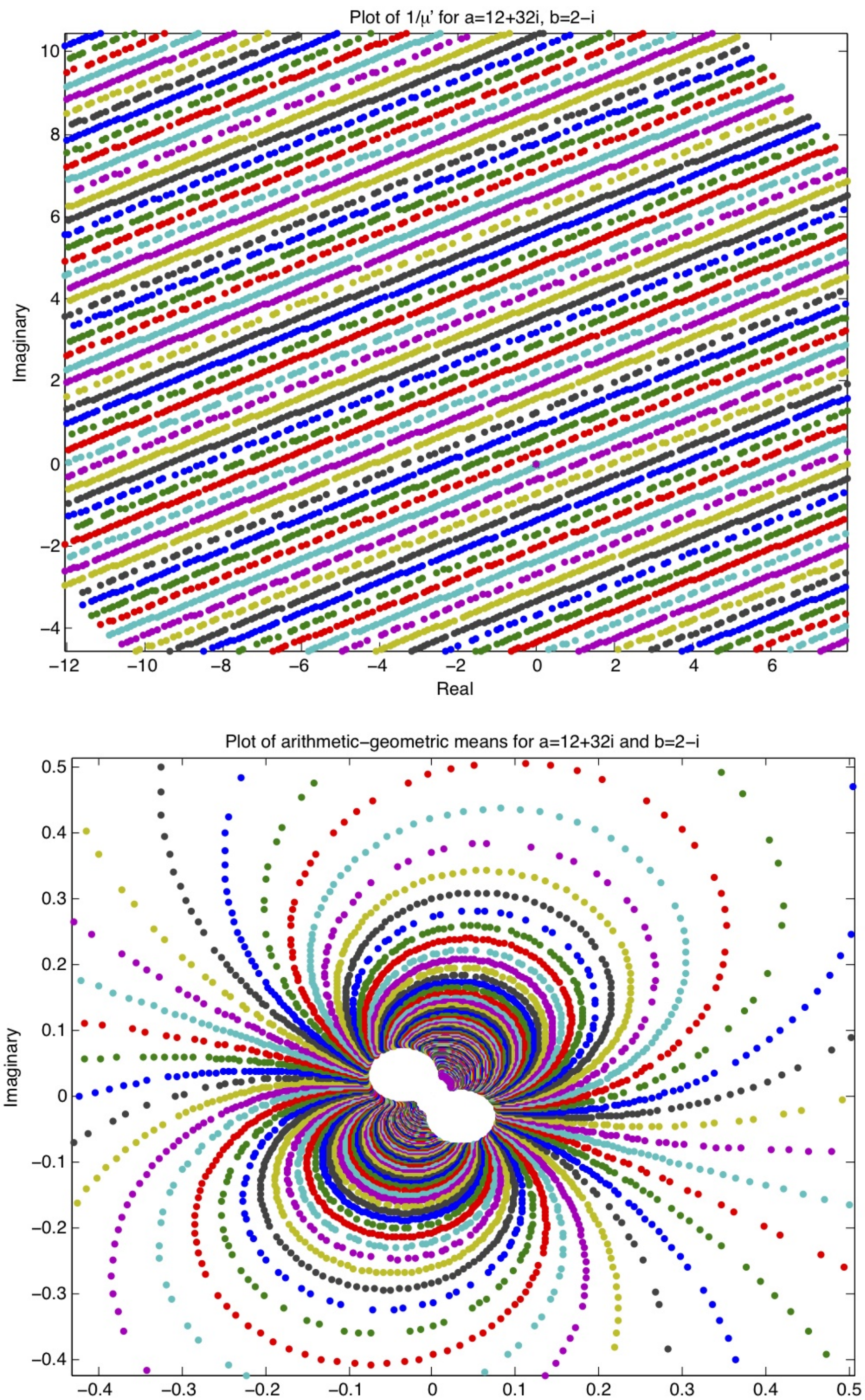


Figure 6: Plot of $1/\mu'$ (top) and μ' (bottom) for $a = 12 + 32i$ and $b = 2 - i$.

determined by the range of values which c may take. We can also see from the plot that as we increase the range of values for c, the ellipses will move closer and closer towards zero, so the limit point of these limit points is zero.

It is worth taking a minute to consider the amount of work it has taken to prove Theorem 5.6. It truly takes a mammoth amount of knowledge to produce such an interesting result. Although Gauss was aware of a great deal of the theorems and lemmas have been used to prove Theorem 5.6 Cox states how the closest he came to realising the result was noting that there is a "mutual connection" between the infinitely many values of $M(a, b)$. Despite this it really is Gauss that we have to thank for the arithmetic-geometric mean. Without his observation of the link between the arithmetic-geometric mean and the arc length of the lemniscate we cannot say that we would understand the nature of this fascinating limit as well as we do now.

References

[1] D. A. Cox, "The Arithmetic-Geometric Mean of Gauss", *L'Enseignement Mathematique*, Vol. 30, 1984, pp. 275-330.

[2] C. F. Gauss, *Werke*, Göttingen, 1866-1933.

[3] C. C. J. Jacobi, *Gesammelte Werke*, G. Reimer, Berlin, 1881.

[4] J. M. and P. B. Borwein, *Pi and the arithmetic-geometric mean*, John Wiley and Sons, 1987.

[5] G. Almkvist and B. Berndt,"Gauss, Landen, Ramanujan, the Arithmetic-Geometric Mean, Ellipses, π, and the *Ladies Diary*", *The American Mathematical Monthly*, Vol. 95, 1988, pp. 585-608.

[6] C. L. Siegel, *Topics in Complex Function Theory*, Wiley- INTERSCIENCE, 1969.

[7] E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge University Press, 1946.

[8] S. Bullett and J. Stark, *"Renormalising the Simple Pendulum"*, SIAM Review, Vol. 35, Dec. 1993, 631-640.

[9] W. B. Gearhart and H. S. Schultz, "The Function $\frac{\sin x}{x}$", *The College Mathematics Journal*, Volume 21, Number 2, March 1990, pp. 90-99.

[10] L. V. King, *On the Direct Numerical Calculation of Elliptic Functions and Integrals*, Cambridge University Press, Cambridge 1924.

[11] H. and B. S. Jeffreys, *Methods of Mathematical Physics (3rd edition)*, Cambridge University Press, Cambridge 1966.

[12] E. Salamin, "Computation of π Using Arithmetic-Geometric Mean", *Mathematics of Computation*, Vol. 30, No. 135, July 1976, pp. 565-570.

[13] R. P. Brent, "Fast Multiple-Precision Evaluation of Elementary Functions", *Journal of the Association for Computing Machinery*, Vol. 23, No. 2, April 1976, pp. 242-251.

[14] L. David, "Arithmetisch-geometrisches Mittel und Modulfunktion", *J. für die Reine u. Ange. Math.*, 159, pp.154-170 (1928).

[15] H. Geppert, "Zur theorie des arithmetisch-geometrischen Mittels", *Math. Annalen*, 90, pp. 162-180 (1928).

[16] J. Tannery and J. Molk, *Eléments de la Théorie des Fonctions Elliptiques*, Vol. 2, Gauthiers-Villars, Paris, 1893.