3. Historical remarks

The main difficulty in writing about the history of mathematics is that so much has to be left out. The mathematics we are studying has a richness which can never be conveyed in one article. For instance, our discussion of Gauss' proofs of Theorem 1.1 in no way does justice to the complexity of his mathematical thought; several important ideas were simplified or omitted altogether. This is not entirely satisfactory, yet to rectify such gaps is beyond the scope of this paper. As a compromise, we will explore the three following topics in more detail:

A. The history of the lemniscate,

B. Gauss' work on inverting lemniscatic integrals, and

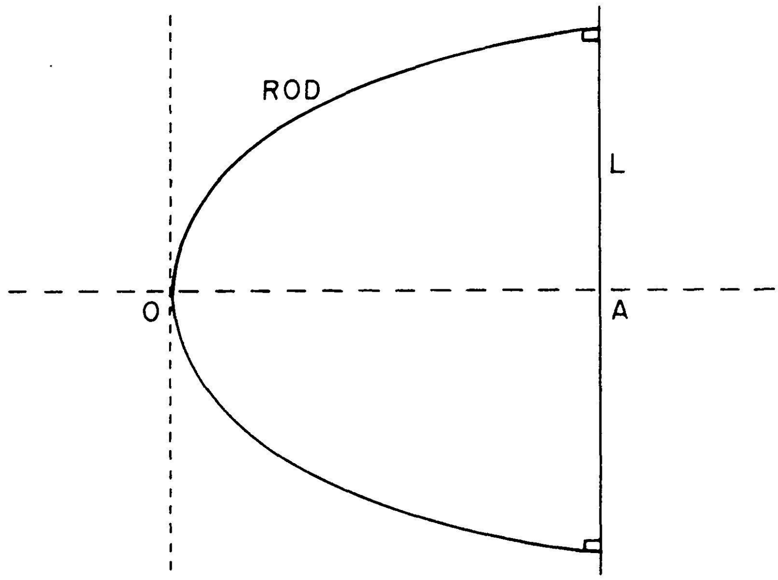
C. The chronology of Gauss' work on the agM and theta functions.

A. The lemniscate was discovered by Jacob Bernoulli in 1694. He gives the equation in the form

$$xx + yy = a\sqrt{xx - yy}$$

(in § 1 we assumed that $a = 1$), and he explains that the curve has "the form of a figure 8 on its side, as of a band folded into a knot, or of a lemniscus, or of a knot of a French ribbon" (see [2, p. 609]). "Lemniscus" is a Latin word (taken from the Greek) meaning a pendant ribbon fastened to a victor's garland.

More interesting is that the integral $\int_0^1 (1-z^4)^{-1/2}dz$, which gives one-quarter of the arc length of the lemniscate, had been discovered three years earlier in 1691! This was when Bernoulli worked out the equation of the so-called *elastic curve*. The situation is as follows: a thin elastic rod is bent until the two ends are perpendicular to a given line L.



After introducing cartesian coordinates as indicated and letting a denote 0A, Bernoulli was able to show that the upper half of the curve is given by the equation

$$y = \int_0^x \frac{z^2 dz}{\sqrt{a^4 - z^4}}, \tag{3.1}$$

where $0 \leqslant x \leqslant a$ (see [2, pp. 567-600]).

It is convenient to assume that $a = 1$. But as soon as this is done, we no longer know how long the rod is. In fact, (3.1) implies that the arc length from the origin to a point (x, y) on the rescaled elastic curve is $\int_0^x (1-z^4)^{-1/2} dz$. Thus the length of the whole rod is $2\int_0^1 (1-z^4)^{-1/2} dz$, which is precisely Gauss' ϖ!

How did Bernoulli get from here to the lemniscate? He was well aware of the transcendental nature of the elastic curve, and so he used a standard seventeenth century trick to make things more manageable: he sought "an algebraic curve... whose rectification should agree with the rectification of the elastic curve" (this quote is from Euler [9, XXI, p. 276]).

Jacob actually had a very concrete reason to be interested in arc length: in 1694, just after his long paper on the elastic curve was published, he solved a problem of Leibniz concerning the "isochrona paracentrica" (see [2, pp. 601-607]). This called for a curve along which a falling weight recedes from or approaches a given point equally in equal times. Since Bernoulli's solution involved the arc length of the elastic curve, it was natural for him to seek an algebraic curve with the same arc length. Very shortly thereafter, he found the equation of the lemniscate (see [2, pp. 608-612]). So we really can say that the arc length of the lemniscate was known well before the curve itself.

But this is not the full story. In 1694 Jacob's younger brother Johann independently discovered the lemniscate! Jacob's paper on the isochrona paracentrica starts with the differential equation

$$(xdx + ydy)\sqrt{y} = (xdy - ydx)\sqrt{a},$$

which had been derived earlier by Johann, who, as Jacob rather bluntly points out, hadn't been able to solve it. Johann saw this comment for the first time when it appeared in June 1694 in Acta Eruditorum. He took up the challenge and quickly produced a paper on the isochrona paracentrica which gave the equation of the lemniscate and its relation to the elastic curve. This appeared in Acta Eruditorum in October 1694 (see [3, pp. 119-

122]), but unfortunately for Johann, Jacob's article on the lemniscate appeared in the September issue of the same journal. There followed a bitter priority dispute. Up to now relations between the brothers had been variable, sometimes good, sometimes bad, with always a strong undercurrent of competition between them. After this incident, amicable relations were never restored. (For details of this controversy, as well as a fuller discussion of Jacob's mathematical work, see [18].)

We need to mention one more thing before going on. Near the end of Jacob's paper on the lemniscate, he points out that the y-value $\int_0^x z^2(a^4-z^4)^{-1/2}dz$ of the elastic curve can be expressed as the difference of an arc of the ellipse with semiaxes $a\sqrt{2}$ and a, and an arc of the lemniscate (see [2, pp. 611-612]). This observation is an easy consequence of the equation

$$\int_0^x \frac{a^2dz}{(a^4-z^4)^{1/2}} + \int_0^x \frac{z^2dz}{(a^4-z^4)^{1/2}} = \int_0^x \left(\frac{a^2+z^2}{a^2-z^2}\right)^{1/2} dz\,. \tag{3.2}$$

What is especially intriguing is that the ratio $\sqrt{2}:1$, so important in Gauss' observation of May 30, 1799, was present at the very birth of the lemniscate.

Throughout the eighteenth century the elastic curve and the lemniscate appeared in many papers. A lot of work was done on the integrals $\int_0^1 (1-z^4)^{-1/2}dz$ and $\int_0^1 z^2(1-z^4)^{-1/2}dz$. For example, Stirling, in a work written in 1730, gives the approximations

$$\int_0^1 \frac{dz}{\sqrt{1-z^4}} = 1.31102877714605987$$

$$\int_0^1 \frac{z^2dz}{\sqrt{1-z^4}} = .59907011736779611$$

(see [31, pp. 57-58]). Note that the second number doubled is 1.19814023473559222, which agrees with $M(\sqrt{2}, 1)$ to sixteen decimal places. Stirling also comments that these two numbers add up to one half the circumference of an ellipse with $\sqrt{2}$ and 1 as axes, a special case of Bernoulli's observation (3.2).

Another notable work on the elastic curve was Euler's paper "De miris proprietatibus curvae elasticae sub equatione $y = \int \frac{xxdx}{\sqrt{1-x^4}}$ contentae" which appeared posthumously in 1786. In this paper Euler gives approximations to the above integrals (not as good as Stirling's) and, more importantly, proves the amazing result that

$$\int_0^1 \frac{dz}{\sqrt{1-z^2}} \cdot \int_0^1 \frac{z^2 dz}{\sqrt{1-z^4}} = \frac{\pi}{4} \tag{3.3}$$

(see [9, XXI, pp. 91-118]). Combining this with Theorem 1.1 we see that

$$M(\sqrt{2}, 1) = 2\int_0^1 \frac{z^2 dz}{\sqrt{1-z^4}},$$

so that the coincidence noted above has a sound basis in fact.

We have quoted these two papers on the elastic curve because, as we will see shortly, Gauss is known to have read them. Note that each paper has something to contribute to the equality $M(\sqrt{2}, 1) = \pi/\varpi$: from Stirling, we get the ratio $\sqrt{2}$: 1, and from Euler we get the idea of using an equation like (3.3).

Unlike the elastic curve, the story of the lemniscate in the eighteenth century is well known, primarily because of the key role it played in the development of the theory of elliptic integrals. Since this material is thoroughly covered elsewhere (see, for example, [1, Ch. 1-3], [8, pp. 470-496], [19, § 1-§ 4] and [21, § 19.4]), we will mention only a few highlights. One early worker was C. G. Fagnano who, following some ideas of Johann Bernoulli, studied the ways in which arcs of ellipses and hyperbolas can be related. One result, known as Fagnano's Theorem, states that the sum of two appropriately chosen arcs of an ellipse can be computed algebraically in terms of the coordinates of the points involved. He also worked on the lemniscate, starting with the problem of halving that portion of the arc length of the lemniscate which lies in one quadrant. Subsequently he found methods for dividing this arc length into n equal pieces, where $n = 2^m$, $3 \cdot 2^m$ or $5 \cdot 2^m$. These researches of Fagnano's were published in the period 1714-1720 in an obscure Venetian journal and were not widely known. In 1750 he had his work republished, and he sent a copy to the Berlin Academy. It was given to Euler for review on December 23, 1751. Less than five weeks later, on January 27, 1752, Euler read a paper giving new derivations for Fagnano's results on elliptic and hyperbolic arcs as well as significantly new results on lemniscatic arcs. By 1753 he had a general addition theorem for lemniscatic integrals, and by 1758 he had the addition theorem for elliptic integrals (see [9, XX, pp. VII-VIII]). This material was finally published in 1761, and for the first time there was a genuine theory of elliptic integrals. For the next twenty years Euler and Lagrange made significant contributions, paving the way for Legendre to cast the field in its classical form which we glimpsed at the end of § 1. Legendre published his definitive treatise on elliptic integrals in two volumes in 1825 and 1826. The irony is that in 1828 he had to publish a third volume describing the groundbreaking papers of Abel and Jacobi which rendered obsolete much of his own work (see [23]).

An important problem not mentioned so far is that of computing tables of elliptic integrals. Such tables were needed primarily because of the many applications of elliptic integrals to mechanics. Legendre devoted the entire second volume of his treatise to this problem. Earlier Euler had computed these integrals using power series similar to (1.8) (see also [9, XX, pp. 21-55]), but these series often converged very slowly. The real breakthrough came in Lagrange's 1785 paper "Sur une nouvelle méthode de calcul intégral" (see [22, pp. 253-312]). Among other things, Lagrange is concerned with integrals of the form

$$\int \frac{M\,dy}{\sqrt{(1+p^2y^2)\,(1+q^2y^2)}}, \tag{3.4}$$

where M is a rational function of y^2 and $p \geqslant q > 0$. He defines sequences $p, p', p'', \ldots, q, q', q'', \ldots$ as follows:

$$\begin{aligned} p' &= p + (p^2-q^2)^{1/2}, & q' &= p - (p^2-q^2)^{1/2}, \\ p'' &= p' + (p'^2-q'^2)^{1/2}, & q'' &= p' - (p'^2-q'^2)^{1/2}, \\ &\vdots & &\vdots \end{aligned} \tag{3.5}$$

and then, using the substitution

$$y' = \frac{y((1+p^2y^2)\,(1+q^2y^2))^{1/2}}{1 + q^2y^2} \tag{3.6}$$

he shows that

$$((1+p^2y^2)\,(1+q^2y^2))^{-1/2}dy = ((1+p'^2y'^2)\,(1+q'^2y'^2))^{-1/2}dy'. \tag{3.7}$$

Two methods of approximation are now given. The first starts by observing that the sequence $p, p', p'', \ldots$ approaches $+\infty$ while $q, q', q'', \ldots$ approaches 0. Thus by iterating the substitution (3.6) in the integral of (3.4), one can eventually assume that $q = 0$, which gives an easily computable integral. The second method consists of doing the first backwards: from (3.5) one easily obtains

$$p = (p'+q')/2\,, \qquad q = (p'q')^{1/2}\,.$$

Lagrange then observes that continuing this process leads to sequences $p', p, {}'p, {}''p, \ldots, q', q, {}'q, {}''q, \ldots$ which converge to a common limit (see [22, p. 271]). Hence iterating (3.6) allows one to eventually assume $p = q$, again giving an easily computable integral.

So here we are in 1785, staring at the definition of the arithmetic-geometric mean, six years before Gauss' earliest work on the subject. By setting $py = \tan\phi$, one obtains

$$((1+p^2y^2)\,(1+q^2y^2))^{-1/2}dy = (p^2\cos^2\phi+q^2\sin^2\phi)^{-1/2}d\phi\,,$$

so that (3.6) and (3.7) are precisely (1.5) and (1.6) from the proof of Theorem 1.1. Thus Lagrange not only could have defined the agM, he could have also proved Theorem 1.1 effortlessly. Unfortunately, none of this happened; Lagrange never realized the power of what he had discovered.

One question emerges from all of this: did Gauss ever see Lagrange's article? The library of the Collegium Carolinum in Brunswick had some of Lagrange's works (see [4, p. 9]) and the library at Gottingen had an extensive collection (see [12, X.2, p. 22]). On the other hand, Gauss, in the research announcement of his 1818 article containing the proof of Theorem 1.1, claims that his work is independent of that of Lagrange and Legendre (see [12, III, p. 360]). A fuller discussion of these matters is in [12, X.2, pp. 12-22]. Assuming that Gauss did discover the agM independently, we have the amusing situation of Gauss, who anticipated so much in Abel, Jacobi and others, himself anticipated by Lagrange.

The elastic curve and the lemniscate were equally well known in the eighteenth century. As we will soon see, Gauss at first associated the integral $\int (1-z^4)^{-1/2}dz$ with the elastic curve, only later to drop it in favor of the lemniscate. Subsequent mathematicians have followed his example. Today, the elastic curve has been largely forgotten, and the lemniscate has suffered the worse fate of being relegated to the polar coordinates section of calculus books. There it sits next to the formula for arc length in polar coordinates, which can never be applied to the lemniscate since such texts know nothing of elliptic integrals.

B. Our goal in describing Gauss' work on the lemniscate is to learn more of the background to his observation of May 30, 1799. We will see that the lemniscatic functions played a key role in Gauss' development of the arithmetic-geometric mean.

Gauss began innocently enough in September 1796, using methods of Euler to find the formal power series expansion of the inverse function of first $\int (1-x^3)^{-1/2}dx$, and then more generally $\int (1-x^n)^{-1/2}dx$ (see [12, X.1, p. 502]). Things became more serious on January 8, 1797. The 51st entry in his mathematical diary, bearing this date, states that "I have begun to investigate the elastic curve depending on $\int (1-x^4)^{-1/2}dx$." Notes written at the same time show that Gauss was reading the works of Euler and Stirling on the elastic curve, as discussed earlier. Significantly, Gauss later struck out the word "elastic" and replaced it with "lemniscatic" (see [12, X.1, pp. 147 and 510]).

Gauss was strongly motivated by the analogy to the circular functions. For example, notice the similarity between $\varpi/2 = \int_0^1 (1-z^4)^{-1/2}dz$ and $\pi/2 = \int_0^1 (1-z^2)^{-1/2}dz$. (This similarity is reinforced by the fact that many eighteenth century texts used ϖ to denote π — see [12, X.2, p. 33].) Gauss then defined the lemniscatic functions as follows:

$$\text{sinlemn}\left(\int_0^x (1-z^4)^{-1/2}dz\right) = x$$

$$\text{coslemn}\left(\varpi/2 - \int_0^x (1-z^4)^{-1/2}dz\right) = x$$

(see [12, III, p. 404]). Gauss often used the abbreviations sl ϕ and cl ϕ for sinlemn ϕ and coslemn ϕ respectively, a practice we will adopt. From Euler's addition theorem one easily obtains

$$\text{sl}^2\phi + \text{cl}^2\phi + \text{sl}^2\phi\,\text{cl}^2\phi = 1 \tag{3.8}$$

$$\text{sl}(\phi+\phi') = \frac{\text{sl}\,\phi\,\text{cl}\,\phi' + \text{sl}\,\phi'\,\text{cl}\,\phi}{1 - \text{sl}\,\phi\,\text{sl}\,\phi'\,\text{cl}\,\phi\,\text{cl}\,\phi'} \tag{3.9}$$

(see [12, X.1, p. 147]).

Other formulas can now be derived in analogy with the trigonometric functions (see [25, pp. 155-156] for a nice treatment), but Gauss went much, much farther. A series of four diary entries made in March 1797 reveal the amazing discoveries that he made in the first three months of 1797. We will need to describe these results in some detail.

Gauss started with Fagnano's problem of dividing the lemniscate into n equal parts. Since this involved an equation of degree n^2, Gauss realized that most of the roots were complex (see [12, X.1, p. 515]). This led him to define $\mathrm{sl}\,\phi$ and $\mathrm{cl}\,\phi$ for complex numbers ϕ. The first step is to show that

$$\mathrm{sl}(iy) = i\,\mathrm{sl}\,y\,, \qquad \mathrm{cl}(iy) = 1/\mathrm{cl}(y)\,,$$

(the first follows from the change of variable $z = iz'$ in $\displaystyle\int (1-z^4)^{-1/2}dz$, and the second follows from (3.8)). Then (3.9) implies that

$$\mathrm{sl}(x+iy) = \frac{\mathrm{sl}\,x + i\,\mathrm{sl}\,y\,\mathrm{cl}\,x\,\mathrm{cl}\,y}{\mathrm{cl}\,y - i\,\mathrm{sl}\,x\,\mathrm{sl}\,y\,\mathrm{cl}\,x}$$

(see [12, X.1, p. 154]).

It follows easily that $\mathrm{sl}\,\phi$ is doubly periodic, with periods 2ϖ and $2i\varpi$. The zeros and poles of $\mathrm{sl}\,\phi$ are also easy to determine; they are given by $\phi = (m+in)\varpi$ and $\phi = \big((2m-1)+i(2n-1)\big)(\varpi/2)$, $m, n \in \mathbf{Z}$, respectively. Then Gauss shows that $\mathrm{sl}\,\phi$ can be written as

$$\mathrm{sl}\,\phi = \frac{M(\phi)}{N(\phi)}$$

where $M(\phi)$ and $N(\phi)$ are entire functions which are doubly indexed infinite products whose factors correspond to the zeros and poles respectively (see [12, X.1, pp. 153-155]). In expanding these products, Gauss writes down the first examples of Eisenstein series (see [12, X.1, pp. 515-516]). He also obtains many identities involving $M(\phi)$ and $N(\phi)$, such as

$$N(2\phi) = M(\phi)^4 + N(\phi)^4 \tag{3.10}$$

(see [12, X.1, p. 157]). Finally, Gauss notices that the numbers $N(\varpi)$ and $e^{\pi/2}$ agree to four decimal places (see [12, X.1, p. 158]). He comments that a proof of their equality would be "a most important advancement of analysis" (see 12, X.1, p. 517]).

Besides being powerful mathematics, what we have here is almost a rehearsal for what Gauss did with the arithmetic-geometric mean: the observation that two numbers are equal, the importance to analysis of proving this, and the passage from real to complex numbers in order to get at the real depth of the subject. Notice also that identities such as (3.10) are an important warm-up to the theta function identities needed in § 2.

Two other discoveries made at this time require comment. First, only a year after constructing the regular 17-gon by ruler and compass, Gauss found a ruler and compass construction for dividing the lemniscate into five equal pieces (see [12, X.1, p. 517]). This is the basis for the remarks concerning $\int (1-x^4)^{-1/2}dx$ made in Disquisitiones Arithmeticae (see [11, § 335]). Second, Gauss discovered the complex multiplication of elliptic functions when he gave formulas for $\mathrm{sl}(1+i)\phi$, $N(1+i)\phi$, etc. (see [12, III, pp. 407 and 411]). These discoveries are linked: complex multiplication on the elliptic curve associated to the lemniscate enabled Abel to determine all n for which the lemniscate can be divided into n pieces by ruler and compass. (The answer is the same as for the circle! See [28] for an excellent modern account of Abel's theorem.)

After this burst of progress, Gauss left the lemniscatic functions to work on other things. He returned to the subject over a year later, in July 1798, and soon discovered that there was a better way to write $\mathrm{sl}\,\phi$ as a quotient of entire functions. The key was to introduce the new variable $s = \sin\left(\frac{\pi}{\varpi}\phi\right)$. Since $\mathrm{sl}\,\phi$ has period 2ϖ, it can certainly be written as a function of s. By expressing the zeros and poles of $\mathrm{sl}\,\phi$ in terms of s, Gauss was able to prove that

$$\mathrm{sl}\,\phi = \frac{P(\phi)}{Q(\phi)},$$

where

$$P(\phi) = \frac{\varpi}{\pi} s\left(1 + \frac{4s^2}{(e^{\pi}-e^{-\pi})^2}\right)\left(1 + \frac{4s^2}{(e^{2\pi}-e^{-2\pi})^2}\right)\cdots$$

$$Q(\phi) = \left(1 - \frac{4s^2}{(e^{\pi/2}+e^{-\pi/2})^2}\right)\left(1 - \frac{4s^2}{(e^{3\pi/2}+e^{-3\pi/2})^2}\right)\cdots$$

(see [12, III, pp. 415-416]). Relating these to the earlier functions $M(\phi)$ and $N(\phi)$, Gauss obtains (letting $\phi = \psi\varpi$)

$$M(\psi\varpi) = e^{\pi\psi^2/2}\, P(\psi\varpi),$$

$$N(\psi\varpi) = e^{\pi\psi^2/2}\, Q(\psi\varpi),$$

(see [12, III, p. 416]). Notice that $N(\varpi) = e^{\pi/2}$ is an immediate consequence of the second formula.

Many other things were going on at this time. The appearance of π/ϖ sparked Gauss' interest in this ratio. He found several ways of expressing ϖ/π, for example

$$\text{(3.11)} \qquad \frac{\varpi}{\pi} = \frac{\sqrt{2}}{2}\left(1 + \left(\frac{1}{2}\right)^2\frac{1}{2} + \left(\frac{3}{8}\right)^2\frac{1}{4} + \left(\frac{5}{16}\right)^2\frac{1}{8} + \dots\right),$$

and he computed ϖ/π to fifteen decimal places (see [12, X.1, p. 169]). He also returned to some of his earlier notes and, where the approximation $2\int_0^1 z^2(1-z^4)^{-1/2}dz \approx 1.198$ appears, he added that this is π/ϖ (see [12, X.1, pp. 146 and 150]). Thus in July 1798 Gauss was intimately familiar with the right-hand side of the equation $M(\sqrt{2}, 1) = \pi/\varpi$. Another problem he studied was the Fourier expansion of $\operatorname{sl}\phi$. Here, he first found the numerical value of the coefficients, i.e.

$$\operatorname{sl}\psi\varpi = .95500599 \sin\psi\pi - .04304950 \sin 3\,\psi\pi + \dots,$$

and then he found a formula for the coefficients, obtaining

$$\operatorname{sl}\psi\varpi = \frac{4\pi}{\varpi(e^{\pi/2}+e^{-\pi/2})}\sin\psi\pi - \frac{4\pi}{\varpi(e^{3\pi/2}+e^{-3\pi/2})}\sin 3\,\psi\pi + \dots$$

see [12, X.1, p. 168 and III, p. 417]).

The next breakthrough came in October 1798 when Gauss computed the Fourier expansions of $P(\phi)$ and $Q(\phi)$. As above, he first computed the coefficients numerically and then tried to find a general formula for them. Since he suspected that numbers like $e^{-\pi}$, $e^{-\pi/2}$, etc., would be involved, he computed several of these numbers (see [12, III, pp. 426-432]). The final formulas he found were

$$P(\psi\varpi) = 2^{3/4}(\pi/\varpi)^{1/2}\left(e^{-\pi/4}\sin\psi\pi - e^{-9\pi/4}\sin 3\,\psi\pi + e^{-25\pi/4}\sin 5\psi\pi - \dots\right)$$

$$\text{(3.12)} \qquad Q(\psi\varpi) = 2^{-1/4}(\pi/\varpi)^{1/2}\left(1 + 2e^{-\pi}\cos 2\psi\pi + 2e^{-4\pi}\cos 4\psi\pi + 2e^{-9\pi}\cos 6\psi\pi + \dots\right)$$

(see [12, X.1, pp. 536-537]). A very brief sketch of how Gauss proved these formulas may be found in [12, X.2, pp. 38-39].

These formulas are remarkable for several reasons. First, recall the theta functions Θ_1 and Θ_3:

$$
\begin{aligned}
\Theta_1(z, q) &= 2q^{1/4} \sin z - 2q^{9/4} \sin 3z + 2q^{25/4} \sin 5z - \dots \\
\Theta_3(z, q) &= 1 + 2q \cos 2z + 2q^4 \cos 4z + 2q^9 \cos 6z + \dots
\end{aligned}
\tag{3.13}
$$

(see [36, p. 464]). Up to the constant factor $2^{-1/4}(\pi/\varpi)^{1/2}$, we see that $P(\psi\varpi)$ and $Q(\psi\varpi)$ are precisely $\Theta_1(\psi\pi, e^{-\pi})$ and $\Theta_3(\psi\pi, e^{-\pi})$ respectively. Even though this is just a special case, one can easily discern the general form of the theta functions from (3.12). (For more on the relation between theta functions and sl ϕ, see [36, pp. 524-525]).

Several interesting formulas can be derived from (3.12) by making specific choice for ψ. For example, if we set $\psi = 1$, we obtain

$$\sqrt{\varpi/\pi} = 2^{-1/4}(1+2e^{-\pi}+2e^{-4\pi}+2e^{-9\pi}+\dots).$$

Also, if we set $\psi = 1/2$ and use the nontrivial fact that $P(\varpi/2) = Q(\varpi/2) = 2^{-1/4}$ (this is a consequence of the formula $Q(2\phi) = P(\phi)^4 + Q(\phi)^4$ – see (3.10)), we obtain

$$
\begin{aligned}
\sqrt{\varpi/\pi} &= 2(e^{-\pi/4}+e^{-9\pi/4}+e^{-25\pi/4}+\dots) \\
\sqrt{\varpi/\pi} &= 1 - 2e^{-\pi} + 2e^{-4\pi} - 2e^{-9\pi} + \dots .
\end{aligned}
\tag{3.14}
$$

Gauss wrote down these last two formulas in October 1798 (see [12, III, p. 418]). We, on the other hand, derived the first and third formulas as (2.21) in § 2, only after a very long development. Thus Gauss had some strong signposts to guide his development of modular functions.

These results, all dating from 1798, were recorded in Gauss' mathematical diary as the 91st and 92nd entries (in July) and the 95th entry (in October). The statement of the 92nd entry is especially relevant: "I have obtained most elegant results concerning the lemniscate, which surpasses all expectation—indeed, by methods which open an entirely new field to us" (see [12, X.1, p. 535]). There is a real sense of excitement here; instead of the earlier "advancement of analysis" of the 63rd entry, we have the much stronger phrase "entirely new field." Gauss knew that he had found something of importance. This feeling of excitement is confirmed by the

95th entry: "A new field of analysis is open before us, that is, the investigation of functions, etc." (see [12, X.1, p. 536]). It's as if Gauss were so enraptured he didn't even bother to finish the sentence.

More importantly, this "new field of analysis" is clearly the same "entirely new field of analysis" which we first saw in § 1 in the 98th entry. Rather than being an isolated phenomenon, it was the culmination of years of work. Imagine Gauss' excitement on May 30, 1799: this new field which he had seen grow up around the lemniscate and reveal such riches, all of a sudden expands yet again to encompass the arithmetic-geometric mean, a subject he had known since age 14. All of the powerful analytic tools he had developed for the lemniscatic functions were now ready to be applied to the agM.

C. In studying Gauss' work on the agM, it makes sense to start by asking where the observation $M(\sqrt{2}, 1) = \pi/\varpi$ came from. Using what we have learned so far, part of this question can now be answered: Gauss was very familiar with π/ϖ, and from reading Stirling he had probably seen the ratio $\sqrt{2} : 1$ associated with the lemniscate. (In fact, this ratio appears in most known methods for constructing the lemniscate—see [24, pp. 111-117].) We have also seen, in the equation $N(\varpi) = e^{\pi/2}$, that Gauss often used numerical calculations to help him discover theorems. But while these facts are enlightening, they still leave out one key ingredient, the idea of taking the agM of $\sqrt{2}$ and 1. Where did this come from? The answer is that every great mathematical discovery is kindled by some intuitive spark, and in our case, the spark came on May 30, 1799 when Gauss decided to compute $M(\sqrt{2}, 1)$.

We are still missing one piece of our picture of Gauss at this time: how much did he know about the agM? Unfortunately, this is a very difficult question to answer. Only a few scattered fragments dealing with the agM can be dated before May 30, 1799 (see [12, X.1, pp. 172-173 and 260]). As for the date 1791 of his discovery of the agM, it comes from a letter he wrote in 1816 (see [12, X.1, p. 247]), and Gauss is known to have been sometimes wrong in his recollections of dates. The only other knowledge we have about the agM in this period is an oral tradition which holds that Gauss knew the relation between theta functions and the agM in 1794 (see [12, III, 493]). We will soon see that this claim is not as outrageous as one might suspect.

It is not our intention to give a complete account of Gauss' work on the agM. This material is well covered in other places (see [10], [12, X.2, pp. 62-114], [13], [14] and [25]—the middle three references are especially complete), and furthermore it is impossible to give the full story of what happened. To explain this last statement, consider the following formulas:

$$B + (1/4)B^3 + (9/64)B^5 + \dots = (2z^{1/2}+2z^{9/2}+\dots)^2 = r^2 ,$$

(3.15)

$$\frac{a}{M(a, b)} = 1 + (1/4)B^2 + (9/64)B^4 + \dots ,$$

where $B = (1-(b/a)^2)^{1/2}$. These come from the first surviving notes on the agM that Gauss wrote after May 30, 1799 (see [12, X.1, pp. 177-178]). If we set $a = 1$ and $b = k' = \sqrt{1-k^2}$, then $B = k$, and we obtain

$$\frac{1}{M(1, k')} = 1 + (1/4)k^2 + (9/64)k^4 + \dots$$

(3.16)

$$\frac{k}{M(1, k')} = (2z^{1/2}+2z^{9/2}+\dots)^2 = r^2 .$$

The first formula is (1.8), and the second, with $z = e^{\pi i \tau/2}$, follows easily from what we learned in § 2 about theta functions and the agM. Yet the formulas (3.15) appear with neither proofs nor any hint of where they came from. The discussion at the end of § 1 sheds some light on the bottom formula of (3.15), but there is nothing to prepare us for the top one.

It is true that Gauss had a long-standing interest in theta functions, going back to when he first encountered Euler's wonderful formula

$$\sum_{n=-\infty}^{\infty} (-1)^n x^{(3n^2+n)/2} = \prod_{n=1}^{\infty} (1-x^n) .$$

The right-hand side appears in a fragment dating from 1796 (see [12, X.1, p. 142]), and the 7th entry of his mathematical diary, also dated 1796, gives a continued fraction expansion for

$$1 - 2 + 8 - 64 + \dots .$$

Then the 58th entry, dated February 1797, generalizes this to give a continued fraction expansion for

$$1 - a + a^3 - a^6 + a^{10} - \dots$$

(see [12, X.1, pp. 490 and 513]). The connection between these series and lemniscatic functions came in October 1798 with formulas such as (3.14).

This seems to have piqued his interest in the subject, for at this time he also set himself the problem of expressing

$$(3.17) \qquad 1 + x + x^3 + x^6 + x^{10} + \dots$$

as an infinite product (see [12, X.1, p. 538]). Note also that the first formula of (3.14) gives r with $z = e^{-\pi/2}$.

Given these examples, we can conjecture where (3.15) came from. Gauss could easily have defined p, q and r in general and then derived identities (2.8)-(2.9) (recall the many identities obtained in 1798 for $P(\phi)$ and $Q(\phi)$—see (3.10) and [12, III, p. 410]). Then (3.15) would result from noticing that these identities formally satisfy the agM algorithm, which is the basic content of Lemma 2.3. This conjecture is consistent with the way Gauss initially treated z as a purely formal variable (the interpretation $z = e^{-\pi i\tau/2}$ was only to come later—see [12, X.1, pp. 262-263 and X.2, pp. 65-66]).

The lack of evidence makes it impossible to verify this or any other reasonable conjecture. But one thing is now clear: in Gauss' observation of May 30, 1799, we have not two but three distinct streams of his thought coming together. Soon after (or simultaneous with) observing that $M(\sqrt{2}, 1) = \pi/\varpi$, Gauss knew that there were inimate connections between lemniscatic functions, the agM, and theta functions. The richness of the mathematics we have seen is in large part due to the many-sided nature of this confluence.

There remain two items of unfinished business. From § 1, we want to determine more precisely when Gauss first proved Theorem 1.1. And recall from § 2 that on June 3, 1800, Gauss discovered the "mutual connection" among the infinitely many values of $M(a, b)$. We want to see if he really knew the bulk of § 2 by this date. To answer these questions, we will briefly examine the main notebook Gauss kept between November 1799 and July 1800 (the notebook is "Scheda Ac" and appears as pp. 184-206 in [12, X.1]).

The starting date of this notebook coincides with the 100th entry of Gauss' mathematical diary, which reads "We have uncovered many new things about arithmetic-geometric means" (see [12, X.1, p. 544]). After several pages dealing with geometry, one all of a sudden finds the formula (3.11) for ϖ/π. Since Gauss knew (3.15) at this time, we get an immediate proof of $M(\sqrt{2}, 1) = \pi/\varpi$. Gauss must have had this in mind, for otherwise why would he so carefully recopy a formula proved in July 1798? Yet one could also ask why such a step is necessary: isn't Theorem 1.1 an immediate consequence of (3.15)? Amazingly enough, it appears that Gauss wasn't yet aware of this connection (see [12, X.1, p. 262]). Part of the problem is that he had been distracted by the power series, closely related to (3.15), which gives the arc length of the ellipse (see [12, X.1, p. 177]). This distraction was actually a bonus, for an asymptotic formula of Euler's for the arc length of the ellipse led Gauss to write

$$M(x, 1) = \frac{(\pi/2)(x - \alpha x^{-1} - \beta x^{-3} - \ldots)}{\log(1/z)} \tag{3.18}$$

where $x = k^{-1}$, and z and k are as in (3.16) (see [12, X.1, pp. 186 and 268-270]). He was then able to show that the power series on top was $(k\, M(1, k'))^{-1}$, which implies that

$$z = \exp\left(-\frac{\pi}{2} \cdot \frac{M(1, k')}{M(1, k)} \right)$$

(see [12, X.1, pp. 187 and 190]). Letting $z = e^{\pi i \tau/2}$, we obtain formulas similar to (2.20). More importantly, we see that Gauss is now in a position to uniformize the agM; z is no longer a purely formal variable.

In the process of studying (3.18), Gauss also saw the relation between the agM and complete elliptic integrals of the first kind. The formula

$$\frac{1}{M(1, k')} = \frac{2}{\pi} \int_0^1 ((1 - x^2)(1 - k^2 x^2))^{-1/2} dx$$

follows easily from [12, X.1, p. 187], and this is trivially equivalent to (1.7). Furthermore, we know that this page was written on December 14, 1799 since on this date Gauss wrote in his mathematical diary that the agM was the quotient of two transcendental functions (see (3.18)), one of which was itself an integral quantity (see the 101st entry, [12, X.1, 544]). Thus Theorem 1.1 was proved on December 14, 1799, nine days earlier than our previous estimate.

Having proved this theorem, Gauss immediately notes one of its corollaries, that the "constant term" of the expression $(1 + \mu \cos^2\phi)^{-1/2}$ is $M(\sqrt{1+\mu}, 1)^{-1}$ (see [12, X.1, p. 188]). By "constant term" Gauss means the coefficient A in the Fourier expansion

$$(1 + \mu \cos^2\phi)^{-1/2} = A + A' \cos\phi + A'' \cos 2\phi + \ldots .$$

Since A is the integral $\frac{2}{\pi} \int_0^{\pi/2} (1 + \mu \cos^2\phi)^{-1/2} d\phi$, the desired result follows from Theorem 1.1. This interpretation is important because these coefficients are useful in studying secular perturbations in astronomy (see [12, X.1, pp. 237-242]). It was in this connection that Gauss published his 1818 paper [12, III, pp. 331-355] from which we got our proof of Theorem 1.1.

What Gauss did next is unexpected: he used the agM to generalize the lemniscate functions to arbitrary elliptic functions, which for him meant inverse functions of elliptic integrals of the form

$$\int (1+\mu^2 \sin^2\phi)^{-1/2} d\phi = \int ((1-x^2)(1+\mu^2 x^2))^{-1/2} dx .$$

Note that $\mu = 1$ corresponds to the lemniscate. To start, he first set $\mu = \tan\nu$,

$$\varpi = \frac{\pi \cos\nu}{M(1, \cos\nu)}, \qquad \varpi' = \frac{\pi\cos\nu}{M(1, \sin\nu)}$$

and finally

$$z = \exp\left(-\frac{\pi}{2}\cdot\frac{\varpi'}{\varpi}\right) = \exp\left(-\frac{\pi}{2}\cdot\frac{M(1,\cos\nu)}{M(1,\sin\nu)}\right). \tag{3.19}$$

Then he defined the elliptic function $S(\phi)$ by $S(\phi) = \dfrac{T(\phi)}{W(\phi)}$ where

$$\begin{aligned} T(\psi\varpi) &= 2\mu^{-1/2}\sqrt{M(1,\cos\nu)}\,(z^{1/2}\sin\psi\pi - z^{9/2}\sin 3\psi\pi + \ldots) \\ W(\psi\varpi) &= \sqrt{M(1,\cos\nu)}\,(1+2z^2\cos 2\psi\pi + 2z^8\cos 4\psi\pi + \ldots) \end{aligned} \tag{3.20}$$

(see [12, X.1, pp. 194-195 and 198]). In the pages that follow, we find the periods 2ϖ and $2i\varpi'$, the addition formula, and the differential equation connecting $S(\phi)$ to the above elliptic integral. Thus Gauss had a complete theory of elliptic functions.

In general, there are two basic approaches to this subject. One involves direct inversion of the elliptic integral and requires a detailed knowledge of the associated Riemann surface (see [17, Ch. VII]). The other more common approach defines elliptic functions as certain series ($\mathfrak{P}$-functions) or quotients of series (theta functions). The difficulty is proving that such functions invert all elliptic integrals. Classically, this uniformization problem is solved by studying a function such as $k(\tau)^2$ (see [36, § 20.6 and § 21.73]) or $j(\tau)$ (as in most modern texts—see [30, § 4.2]). Gauss uses the agM to solve this problem: (3.19) gives the desired uniformizing parameter! (In this connection, the reader should reconsider the from [12, VIII, p. 101] given near the end of § 2.)

For us, the most interesting aspect of what Gauss did concerns the functions T and W. Notice that (3.20) is a direct generalization of (3.12); in fact, in terms of (3.13), we have

$$T(\psi\varpi) = \mu^{-1/2}\sqrt{M(1, \cos\nu)}\;\Theta_1(\psi\pi, z^2),$$

$$W(\psi\varpi) = \sqrt{M(1, \cos\nu)}\;\Theta_3(\psi\pi, z^2).$$

Gauss also introduces $T(\varpi/2-\phi)$ and $W(\varpi/2-\phi)$, which are related to the theta functions Θ_2 and Θ_4 by similar formulas (see [12, X.1, pp. 196 and 275]). He then studies the squares of these functions and he obtains identities such as

$$2\Theta_3(0, z^4)\,\Theta_3(2\phi, z^4) = \Theta_3(\phi, z^2)^2 + \Theta_4(\phi, z^2)^2$$

(this, of course, is the modern formulation—see [12, X.1, pp. 196 (Eq. 14) and 275]). When ϕ =' 0, this reduces to the first formula

$$p(\tau)^2 + q(\tau)^2 = 2p(2\tau)^2$$

of (2.8). The other formulas of (2.8) appear similarly. Gauss also obtained product expansions for the theta functions (see [12, X.1, pp. 201-205]). In particular, one finds all the formulas of (2.6). These manipulations yielded the further result that

$$1 + z + z^3 + z^6\; z^{10} + \ldots - \prod_{n=1}^{\infty} (1-z)^{-1}(1-z^2),$$

solving the problem he had posed a year earlier in (3.17).

From Gauss' mathematical diary, we see that the bulk of this work was done in May 1800 (see entries 105, 106 and 108 in [12, X.1, pp. 546-549]). The last two weeks were especially intense as Gauss realized the special role played by the agM. The 108th entry, dated June 3, 1800, announces completion of a general theory of elliptic functions ("sinus lemniscatici universalissime accepti"). On the same day he recorded his discovery of the "mutual connection" among the values of the agM!

This is rather surprising. We've seen that Gauss knew the basic identities (2.6), (2.8) and (2.9), but the formulas (2.7), which tell us how theta functions behave under linear fractional transformations, are nowhere to be seen, nor do we find any hint of the fundamental domains used in § 2. Reading this notebook makes it clear that Gauss now knew the basic observation of

Lemma 2.3 that theta functions satisfy the agM algorithm, but there is no way to get from here to Theorem 2.2 without knowing (2.7). It is not until 1805 that this material appears in Gauss' notes (see [12, X.2, pp. 101-103]). Thus some authors, notably Markushevitch [25], have concluded that on June 3, 1800, Gauss had nothing approaching a proof of Theorem 2.2.

Schlesinger, the last editor of Gauss' collected works, feels otherwise. He thinks that Gauss knew (2.7) at this time, though knowledge of the fundamental domains may have not come until 1805 (see [12, X.2, p. 106]). Schlesinger often overestimates what Gauss knew about modular functions, but in this case I agree with him. As evidence, consider pp. 287-307 in [12, X.1]. These reproduce twelve consecutive pages from a notebook written in 1808 (see [12, X.1, p. 322]), and they contain the formulas (2.7), a clear statement of the basic observation of Lemma 2.3, the infinite product manipulations described above, and the equations giving the division of the agM into 3, 5 and 7 parts (in analogy with the division of the lemniscate). The last item is especially interesting because it relates to the second half of the 108th entry: "Moreover, in these same days, we have discovered the principles according to which the agM series should be interpolated, so as to enable us to exhibit by algebraic equations the terms in a given progression pertaining to any rational index" (see [12, X.1, p. 548]). There is no other record of this in 1800, yet here it is in 1808 resurfacing with other material (the infinite products) dating from 1800. Thus it is reasonable to assume that the rest of this material, including (2.7), also dates from 1800. Of course, to really check this conjecture, one would have to study the original documents in detail.

Given all of (2.6)-(2.9), it is still not clear where Gauss got the basic insight that $M(a, b)$ is a multiple valued function. One possible source of inspiration is the differential equation (1.12) whose solution (1.13) suggests linear combinations similar to those of Theorem 2.2. We get even closer to this theorem when we consiser the periods of $S(\phi)$:

$$m\varpi + in\varpi' = \pi \cos \nu \left(\frac{m}{M(1, \cos\nu)} + \frac{in}{M(1, \sin\nu)} \right)$$

where m, n are even integers. Gauss' struggles during May 1800 to understand the imaginary nature of these periods (see [12, X.2, pp. 70-71]) may have influenced his work on the agM. (We should point out that the above comments are related: Theorem 2.2 can be proved by analyzing the monodromy group—$\Gamma_2(4)$ in this case—of the differential equation (1.12).) On the other hand, Geppert suggests that Gauss may have taken a completely different route, involving the asymptotic formula (3.18), of arriving at Theorem 2.2 (see [14, pp. 173-175]). We will of course never really know how Gauss arrived at this theorem.

For many years, Gauss hoped to write up these results for publication. He mentions this in Disquisitiones Arithmeticae (see [11, § 335]) and in the research announcement to his 1818 article (see [12, III, p. 358]). Two manuscripts written in 1800 (one on the agM, the other on lemniscatic functions) show that Gauss made a good start on this project (see [12, III, pp. 360-371 and 413-415]). He also periodically returned to earlier work and rewrote it in more complete form (the 1808 notebook is an example of this). Aside from the many other projects Gauss had to distract him, it is clear why he never finished this one: it was simply too big. Given his predilection for completeness, the resulting work would have been enormous. Gauss finally gave up trying in 1827 when the first works of Abel and Jacobi appeared. As he wrote in 1828, "I shall most likely not soon prepare my investigations on transcendental functions which I have had for many years—since 1798—because I have many other matters which must be cleared up. Herr Abel has now, I see, anticipated me and relieved me of the burden in regard to one third of these matters, particularly since he carried out all developments with great concision and elegance" (see [12, X.1, p. 248]).

The other two thirds "of these matters" encompass Gauss' work on the agM and modular functions. The latter were studied vigorously in the nineteenth century and are still an active area of research today. The agM, on the other hand, has been relegated to the history books. This is not entirely wrong, for the history of this subject is wonderful. But at the same time the agM is also wonderful as mathematics, and this mathematics deserves to be better known.