Applications of Elliptic Functions in Classical and Algebraic Geometry

Jamie Snape
Collingwood College, University of Durham

Dissertation submitted for the degree of Master of Mathematics
at the University of Durham

It strikes me that mathematical writing is similar to using a language. To be understood you have to follow some grammatical rules. However, in our case, nobody has taken the trouble of writing down the grammar; we get it as a baby does from parents, by imitation of others. Some mathematicians have a good ear; some not... That's life.

JEAN-PIERRE SERRE

Jean-Pierre Serre (1926–). Quote taken from Serre (1991).

Contents

List of Figures

Part I

Background

Chapter 1

Elliptic Functions

While the elliptic functions can be defined in a variety of geometric and mechanical ways, we begin with an analytic definition.

1.1. Motivation [1]

A function $f(z)$ is defined to be a *simply periodic* function of z if there exists a non-zero constant ω_1 such that

$$f(z+\omega) = f(z)$$

holds for all values of z. This number ω is the period of $f(z)$. Clearly, if n is a non-zero integer then $n\omega$ is also a period. If no submultiple of ω is a period, then it is known as a *fundamental period.*

This is illustrated by the circular functions, for example

$$\sin(z+2\pi) = \sin z = \sin(z+2n\pi),$$

for all integer values of n. Therefore, $\sin z$ and also $\cos z$ are simply periodic functions with period 2π. Similarly, e^z is simply periodic with period $2\pi i$.

Jacobi wondered whether there exists a function, analytic except at its poles, with [2] two fundamental periods whose ratio is real. It turns out that such a function is constant. However, if the ratio is not real, this is not the case, and leads us to investigate further such functions with two fundamental periods.

1.2. Definition of an elliptic function [3]

Let ω_1 and ω_2 be two complex numbers whose ratio is not real. Then a function which satisfies

$$f(z+\omega_1) = f(z) = f(z+\omega_2),$$

for all values of z for which $f(z)$ is defined is known as a *doubly periodic* function of z with periods ω_1 and ω_2. A doubly periodic function that is analytic except at its poles,

[1] Copson (1935) 345.

[2] Carl Gustav Jacob Jacobi (1804–1851). It was in 1834 that Jacobi proved that if a one-valued function of one variable is doubly periodic, then the ratio of the periods is imaginary.

[3] Whittaker & Watson (1927) 429–430.

and which has no singularities other than these poles in a finite part of the complex plane is called an *elliptic function*.

Now, without loss of generality we take the imaginary part of the ratio of ω_1 and ω_2 to be real. Therefore, the points 0, ω_1, $\omega_1 + \omega_2$ and ω_2 when taken in order are the vertices of a parallelogram, known as the *fundamental parallelogram*. If we also consider the points of the *period lattice* defined as $\Omega = \{m\omega_1 + n\omega_2\}$, then the four points $m\omega_1 + n\omega_2$, $(m+1)\omega_1 + n\omega_2$, $(m+1)\omega_1 + (n+1)\omega_2$ and $m\omega_1 + (n+1)\omega_2$ are vertices of a similar parallelogram, obtained from the original parallelogram by a translation without rotation. This parallelogram is called a *period parallelogram*. Therefore, the entire complex plane is covered by a system of non-overlapping period parallelograms.

The points z and $z + m\omega_1 + n\omega_2$ lie in different period parallelograms. If we translate one such parallelogram so that it coincides with the other, then the points also coincide. Therefore, we say that z is *congruent* to $z + m\omega_1 + n\omega_2$ with respect to the period lattice Ω. As $m\omega_1 + n\omega_2$ is a period of $f(z)$, then it follows that $f(z)$ takes the same value at every one of a set of congruent points. Hence, the behaviour of an elliptic function is completely determined by its values in any fundamental parallelogram.

Suppose we wish to count the number of poles or zeroes in a given period parallelogram. We can simplify this calculation by translating the period parallelogram without rotation until no pole or zero lies on its boundary. This parallelogram is known as a *cell*, and the set of poles or zeroes within that cell is called an *irreducible set*.

1.3. Properties of an elliptic function

THEOREM 1.1. *The number of poles of an elliptic function $f(z)$ in any cell is finite.*

Proof (Copson, 1935). If there were an infinite number, then the set of these poles [4] would have a limit point. But the limit point of poles is an essential singularity, and so by definition the function would not be an elliptic function.

□

THEOREM 1.2. *The number of zeroes of an elliptic function $f(z)$ in any cell is finite.*

Proof (Whittaker & Watson, 1927). If there were an infinite number, then it would fol- [5] low that the reciprocal of the function would have an infinite number of poles. Therefore, it would have an essential singularity, and this would also be an essential singularity of the original function. Again, this would mean that the function was not an elliptic function.

□

THEOREM 1.3. *The sum of the residues of an elliptic function $f(z)$ at its poles in any cell is zero.*

Proof (Copson, 1935). If C is the contour formed by the edges of the cell, then, by the [6]

---

[4] Copson (1935) 350.

[5] Whittaker & Watson (1927) 431.

residue theorem, the sum of the residues of $f(z)$ at its poles within C is

$$\frac{1}{2\pi \mathrm{i}} \int_C f(z)\, \mathrm{d}z.$$

This is zero as the integrals along opposite sides of C cancel due to the periodicity of $f(z)$.

□

THEOREM 1.4. *Liouville's theorem for elliptic functions. An elliptic function $f(z)$ with no poles in a cell is constant.*

Proof (Whittaker & Watson, 1927). If $f(z)$ has no poles inside its cell, then it must be 7
analytic and its absolute value bounded inside or on the boundary of the cell. Hence, by the periodicity of $f(z)$, it is bounded on all z, and so by Liouville's theorem for
analytic functions it must be constant. 8

□

The number of poles of an elliptic function in any cell, counted with multiplicity, is called the *order* of the function. This is necessarily at least equal to 2, since an elliptic function of order 1 would have a single irreducible pole. If this were actually a pole, its residue would not be zero, and so contradict Theorem 1.3.

THEOREM 1.5. *An elliptic function $f(z)$ of order m has m zeroes in each cell.*

Proof (Copson, 1935). If $f(z)$ is of order m and has n zeroes in a cell, counted with 9
multiplicity, then $n - m$ is equal to the sum of residues of $f'(z)/f(z)$ at its poles in the cell. But $f'(z)$ is an elliptic function with the same periods as $f(z)$, so it follows that $f'(z)/f(z)$ is similarly an elliptic function. Therefore, $n - m = 0$ by Theorem 1.3.

□

THEOREM 1.6. *The sum of the set of irreducible zeroes, counted with multiplicity, of an elliptic function $f(z)$ is congruent to the sum of the set of irreducible poles, counted with multiplicity.*

Proof (Whittaker & Watson, 1927). For a cell C, it follows from the residue theorem, 10
that the difference between the sums is

$$\frac{1}{2\pi \mathrm{i}} \int_C \frac{z f'(z)}{f(z)}\, \mathrm{d}z.$$

[6] Copson (1935) 351.

[7] Whittaker & Watson (1927) 431–432.

[8] *Liouville's theorem for analytic functions. Let $f(z)$ be any analytic function and let its absolute value be bounded for all values of z, then $f(z)$ is constant.*

This theorem was actually discovered by Augustin Louis Cauchy (1789–1857), but was later falsely attributed to Joseph Liouville (1809–1882) after it appeared in his lectures of 1847.

[9] Copson (1935) 351.

[10] Whittaker & Watson (1927) 433.

If the vertices of C are t, $t+\omega_1$, $t+\omega_1+\omega_2$, $t+\omega_2$, then this is equal to

$$\begin{aligned}
&\frac{1}{2\pi \mathrm{i}}\int_t^{t+\omega_1}\left\{\frac{zf'(z)}{f(z)}-\frac{(z+\omega_2)f'(z+\omega_2)}{f(z+\omega_2)}\right\}\mathrm{d}z\\
&\qquad\qquad\qquad -\frac{1}{2\pi \mathrm{i}}\int_t^{t+\omega_2}\left\{\frac{zf'(z)}{f(z)}-\frac{(z+\omega_1)f'(z+\omega_1)}{f(z+\omega_1)}\right\}\mathrm{d}z\\
&\qquad =\frac{1}{2\pi \mathrm{i}}\left\{\omega_1\int_t^{t+\omega_2}\frac{f'(z)}{f(z)}\,\mathrm{d}z-\omega_2\int_t^{t+\omega_1}\frac{f'(z)}{f(z)}\,\mathrm{d}z\right\}\\
&\qquad =\frac{1}{2\pi \mathrm{i}}\left\{\omega_1\Big[\log f(z)\Big]_t^{t+\omega_2}-\omega_2\Big[\log f(z)\Big]_t^{t+\omega_1}\right\},
\end{aligned}$$

by using the periodic properties of $f(z)$ and $f'(z)$.

Now, $f(z)$ has the same values at $t+\omega_1$ and $t+\omega_2$ as at t, so it follows that the values of $\log f(z)$ at these points can only differ from the value of $\log f(z)$ at t by integer multiples of $2\pi\mathrm{i}$. Therefore, we have

$$\frac{1}{2\pi \mathrm{i}}\int_C \frac{zf'(z)}{f(z)}\,\mathrm{d}z = m\omega_1 + n\omega_2.$$

Hence, the difference between the sums is a period, and we have the result.

□

THEOREM 1.7. *If $f(z)$ and $g(z)$ are elliptic functions with poles at the same points, and with the same principal parts at these points, then $f(z)=g(z)+A$, for some constant A.*

Proof (Jones & Singerman, 1987). The function $f(z)-g(z)$ is an elliptic function of order zero, as it has no poles. Therefore, it is constant by Theorem 1.4. [11]

□

THEOREM 1.8. *If $f(z)$ and $g(z)$ are elliptic functions with zeroes and poles of the same order at the same points, then $f(z)=Ag(z)$, for some constant A.*

Proof (Jones & Singerman, 1987). By a similar argument to the previous proof, we have that the function $f(z)/g(z)$ is also an elliptic function of order zero. Therefore, by again applying Theorem 1.4, it is constant. [11]

□

[11] Jones & Singerman (1987) 75–76.

Chapter 2

Jacobi Elliptic Functions

In the previous chapter, we have seen that the order of an elliptic function is never less than 2, so in terms of singularities, the simplest elliptic functions are those of order 2. These can be divided into two classes: those which have a single irreducible double pole in each cell at which the residue is zero, and those which have two simple poles in each cell at which the two residues are equal in absolute value, but of opposite sign. The *Jacobi elliptic functions* are examples of the latter class.

2.1. Motivation [1]

Suppose we have the two integrals

$$u = \int_0^x \frac{\mathrm{d}t}{\sqrt{1-t^2}}, \tag{2.1}$$

$$\tfrac{1}{2}\pi = \int_0^1 = \frac{\mathrm{d}t}{\sqrt{1-t^2}},$$

where $-1 < x < 1$ is real.

If we take the square root to be positive for u between zero and π, then this defines u as an odd function of x. By inversion of the integral, we have defined z as an odd function of u. If denote this function by $\sin u$, then (2.1) reduces to the form

$$u = \sin^{-1} x.$$

We can define a second function $\cos u$ by

$$\cos u = \sqrt{1 - \sin^2 u}.$$

By taking the square root positive for u between $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$, we have u as an even function of x. It follows that we have the identity

$$\sin^2 u + \cos^2 u = 1. \tag{2.2}$$

We can also note that $\sin 0 = 0$ and $\cos 0 = 1$.

Suppose now we consider the derivative of (2.1) with respect to x, which is clearly

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{\sqrt{1-x^2}}.$$

[1] Bowman (1953) 7–8.

It follows that

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\sin u\right\} = \sqrt{1 - \sin^2 u} = \cos u,$$

as $x = \sin u$. Moreover, by differentiation of (2.2), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\cos u\right\} = -\sin u,$$

Now, we consider the equation

$$w = \sin u_1 \cos u_2 + \cos u_1 \sin u_2.$$

The partial derivatives of w with respect to u_1 and u_2 are equal, so it follows that $w = f(u_1 + u_2)$, where $f(u_1 + u_2)$ is a function of $u_1 + u_2$. Therefore,

$$f(u_1 + u_2) = \sin u_1 \cos u_2 + \cos u_1 \sin u_2.$$

If we set $u_2 = 0$, then $f(u_1) = \sin u_1$, and similarly $f(u_2) = \sin u_2$. Hence, $f(u_1 + u_2) = \sin(u_1 + u_2)$ and we obtain an addition formula

$$\sin(u_1 + u_2) = \sin u_1 \cos u_2 + \cos u_1 \sin u_2.$$

By (2.2), we also have

$$\cos(u_1 + u_2) = \cos u_1 \cos u_2 - \sin u_1 \sin u_2.$$

We can also use these two addition formulæ to see that both $\sin u$ and $\cos u$ are simply periodic functions with period 2π.

2.2. Definitions of the Jacobi elliptic functions [2]

The Jacobi elliptic function $\mathrm{sn}\, u$ is defined by means of the integral

$$u = \int_0^x \frac{\mathrm{d}t}{\sqrt{(1 - t^2)(1 - k^2 t^2)}}, \tag{2.3}$$

for some constant k. Therefore, by inversion of the integral, we have $x = \mathrm{sn}\, u$. It is clear that $\mathrm{sn}\, 0 = 0$.

The functions $\mathrm{cn}\, u$ and $\mathrm{dn}\, u$ are defined by the identities

$$\mathrm{sn}^2 u + \mathrm{cn}^2 = 1, \tag{2.4}$$

$$k^2 \,\mathrm{sn}^2 u + \mathrm{dn}^2 u = 1. \tag{2.5}$$

It follows that $\mathrm{cn}\, 0 = 1 = \mathrm{dn}\, 0$.

Each of the Jacobi elliptic functions depend on a parameter k, called the *modulus*. We also have the *complementary modulus* k' defined by

$$k^2 + k'^2 = 1.$$

[2] Bowman (1953) 8–9.

In the words of Arthur Cayley (1821–1895), $\mathrm{sn}\, u$ is *a sort of sine function, and* $\mathrm{cn}\, u$, $\mathrm{dn}\, u$ *are sorts of cosine functions.*

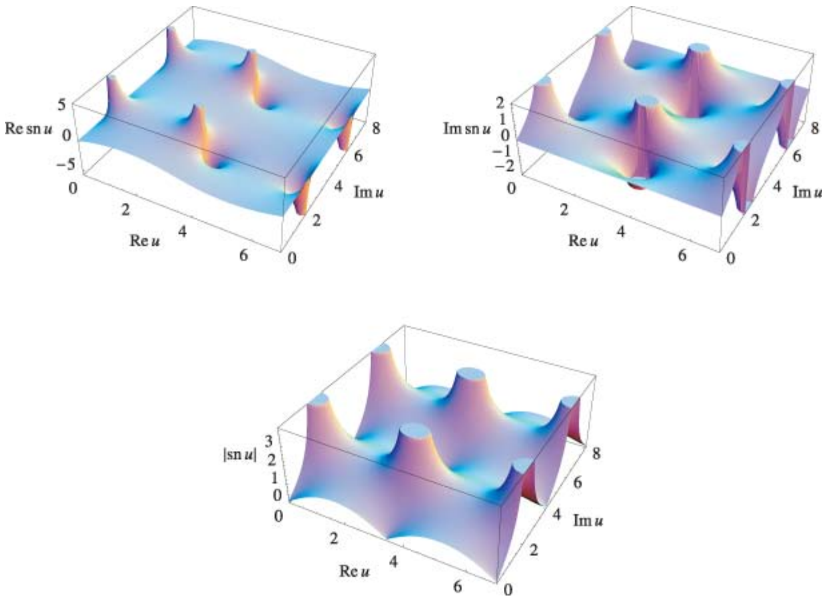

Figure 2.1. Graphs of the real part (a), imaginary part (b) and absolute value (c) of the Jacobi elliptic function $\mathrm{sn}(u, k)$ *with modulus* $k = 0.1$.

When emphasizing a particular modulus, we write the three functions as $\mathrm{sn}(u, k)$, $\mathrm{cn}(u, k)$ and $\mathrm{dn}(u, k)$. An alternative notation uses a parameter $m = k^2$, and so the functions are denoted $\mathrm{sn}(u|m)$, $\mathrm{cn}(u|m)$ and $\mathrm{dn}(u|m)$.

When $k = 0$, the functions $\mathrm{sn}\, u$ and $\mathrm{cn}\, u$ degenerate to the circular functions $\sin u$ and $\cos u$, respectively, while $\mathrm{dn}\, u$ degenerates to 1. When $k = 1$, we have $\mathrm{sn}\, u$ equal to the hyperbolic function $\tanh u$, while both $\mathrm{cn}\, u$ and $\mathrm{dn}\, u$ are equal to $\mathrm{sech}\, u$.

2.3. Properties of the Jacobi elliptic functions

THEOREM 2.1. *The function* $\mathrm{sn}\, u$ *is an odd function of* u*, while* $\mathrm{cn}\, u$ *and* $\mathrm{dn}\, u$ *are even functions of* u.

The notation $\mathrm{sn}\, u$, $\mathrm{cn}\, u$ and $\mathrm{dn}\, u$ was introduced by Christoph Gudermann (1798–1852) in lectures given in 1838. In his *Fundamenta nova theoriæ functionum ellipticarum*, Jacobi himself used $\sin \mathrm{am}\, u$, $\cos \mathrm{am}\, u$ and $\Delta\, \mathrm{am}\, u$. This follows from his definition of an additional function $\mathrm{am}\, u$ which follows by the inversion of the integral

$$u = \int_0^\theta \frac{\mathrm{d}\theta}{\sqrt{1 - k^2 \sin^2 \theta}}.$$

A shortened notation to express reciprocals and quotients was invented by James Whitbread Lee Glaisher (1848–1928) in which the reciprocals are denoted by reversing the orders of the letters of the function to obtain

$$\mathrm{ns}\, u = 1/\,\mathrm{sn}\, u, \quad \mathrm{nc}\, u = 1/\,\mathrm{cn}\, u, \quad \mathrm{nd}\, u = 1/\,\mathrm{dn}\, u.$$

Quotients are denoted by writing in order the first letters of the numerator and denominator functions. Hence,

$$\mathrm{sc}\, u = \mathrm{sn}\, u/\,\mathrm{cn}\, u, \quad \mathrm{sd}\, u = \mathrm{sn}\, u/\,\mathrm{dn}\, u, \quad \mathrm{cd}\, u = \mathrm{cn}\, u/\,\mathrm{dn}\, u,$$
$$\mathrm{cs}\, u = \mathrm{cn}\, u/\,\mathrm{sn}\, u, \quad \mathrm{ds}\, u = \mathrm{dn}\, u/\,\mathrm{sn}\, u, \quad \mathrm{dc}\, u = \mathrm{dn}\, u/\,\mathrm{cn}\, u.$$

Circular functions	*Jacobi elliptic functions*
$u = \int_0^x \frac{dt}{\sqrt{1-t^2}}$	$u = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$
$x = \sin u$	$x = \operatorname{sn} u$
$\frac{1}{2}\pi = \int_0^1 \frac{dt}{\sqrt{1-t^2}}$	$K = \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$
$\cos u = \sin(\frac{1}{2}\pi - u)$	$\operatorname{cd} u = \operatorname{sn}(K - u)$
$\sin \frac{1}{2}\pi = 1$	$\operatorname{sn} K = 1$
$\sin \pi = 0$	$\operatorname{sn} 2K = 0$
$\sin(\pi + u) = -\sin u$	$\operatorname{sn}(2K + u) = -\operatorname{sn} u$
$\sin(-u) = -\sin u$	$\operatorname{sn}(-u) = -\operatorname{sn} u$

Table 2.1. Comparison between the circular functions and the Jacobi elliptic functions.

Proof (Whittaker & Watson, 1927). If we substitute t for $-t$ in (2.3), then if the sign 3
of x is changed, the sign of u is similarly changed. Therefore, $\operatorname{sn}(-u) = -\operatorname{sn} u$.

By the identity (2.4), it follows that $\operatorname{cn}(-u) = \pm \operatorname{cn} u$. As $\operatorname{cn} u$ is a one-valued, then, by the theory of analytic continuation, either the upper sign, or else the lower sign must always be taken. When $u = 0$, the positive sign must be taken, so it is this that is always taken. It follows that $\operatorname{cn}(-u) = \operatorname{cn} u$.

By a similar argument, we obtain $\operatorname{dn}(-u) = -\operatorname{dn} u$.

□

THEOREM 2.2. *The derivatives of the Jacobi elliptic functions are* 4

$$\frac{d}{du}\left\{\operatorname{sn} u\right\} = \operatorname{cn} u \operatorname{dn} u,$$
$$\frac{d}{du}\left\{\operatorname{cn} u\right\} = -\operatorname{sn} u \operatorname{dn} u,$$
$$\frac{d}{du}\left\{\operatorname{dn} u\right\} = -k^2 \operatorname{sn} u \operatorname{cn} u.$$

Proof (Bowman, 1953). By differentiation of (2.3), we have 5

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

and, as $x = \operatorname{sn} u$, it follows that

$$\frac{d}{du}\left\{\operatorname{sn} u\right\} = \sqrt{(1-\operatorname{sn}^2 u)(1-k^2\operatorname{sn}^2 u)} = \operatorname{cn} u \operatorname{dn} u.$$

If we next differentiate (2.4), then we have

$$\frac{d}{du}\left\{\operatorname{cn} u\right\} = -\operatorname{sn} u \operatorname{dn} u,$$

[3] Whittaker & Watson (1927) 493.
[4] The derivative of $\operatorname{am} u$ is $\operatorname{dn} u$.
[5] Bowman (1953) 9–10.

and similarly, from (2.5), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\operatorname{dn} u\right\} = -k^2 \operatorname{sn} u \operatorname{cn} u.$$

$\square$

2.4. The addition formulæ for the Jacobi elliptic functions

A function $f(u)$ is said to possess an algebraic *addition formula* if there is an identity between $f(u_1)$, $f(u_2)$ and $f(u_1+u_2)$ of the form

$$R\left(f(u_1), f(u_2), f(u_1+u_2)\right) = 0,$$

for all u_1 and u_2, where R is a non-zero rational function of three variables. For example,

$$\tan(u_1+u_2) = \frac{\tan u_1 + \tan u_2}{1 - \tan u_1 \tan u_2}.$$

THEOREM 2.3. *The addition formulæ for the Jacobi elliptic functions are*

$$\operatorname{sn}(u_1+u_2) = \frac{\operatorname{sn} u_1 \operatorname{cn} u_2 \operatorname{dn} u_2 + \operatorname{sn} u_2 \operatorname{cn} u_1 \operatorname{dn} u_1}{1 - k^2 \operatorname{sn}^2 u_1 \operatorname{sn}^2 u_2}, \tag{2.6}$$

$$\operatorname{cn}(u_1+u_2) = \frac{\operatorname{cn} u_1 \operatorname{cn} u_2 - \operatorname{sn} u_1 \operatorname{sn} u_2 \operatorname{dn} u_1 \operatorname{dn} u_2}{1 - k^2 \operatorname{sn}^2 u_1 \operatorname{sn}^2 u_2},$$

$$\operatorname{dn}(u_1+u_2) = \frac{\operatorname{dn} u_1 \operatorname{dn} u_2 - k^2 \operatorname{sn} u_1 \operatorname{sn} u_2 \operatorname{cn} u_1 \operatorname{cn} u_2}{1 - k^2 \operatorname{sn}^2 u_1 \operatorname{sn}^2 u_2}.$$

Proof (Bowman, 1953). We denote $s_1 = \operatorname{sn} u_1$, $s_2 = \operatorname{sn} u_2$, $c_1 = \operatorname{cn} u_1$, $c_2 = \operatorname{cn} u_2$, [6]
$d_1 = \operatorname{dn} u_1$ and $d_2 = \operatorname{dn} u_2$. Let [7]

$$w = \frac{s_1 c_2 \operatorname{dn} u_2 + s_2 c_1 d_1}{1 - k^2 s_1^2 s_2^2}.$$

Then, by partial differentiation with respect to u_1, and after simplification using (2.4) and (2.3), we have

$$\frac{\mathrm{d}w}{\mathrm{d}u_1} = \frac{c_1 d_1 c_2 d_2 (1 + k^2 s_1^2 s_2^2) - s_1 s_2 (d_1^2 d_2^2 + k^2 c_1^2 c_2^2)}{(1 - k^2 s_1^2 s_2^2)^2}$$

Therefore, $\mathrm{d}w/\mathrm{d}u_1$ is symmetric in u_1 and u_2, and as w is symmetric, it follows that $\mathrm{d}w/\mathrm{d}u_2$ is equal to $\mathrm{d}w/\mathrm{d}u_1$. Hence, for a function $f(u_1+u_2)$ of u_1+u_2, we have $w = f(u_1+u_2)$, and it follows that

$$f(u_1+u_2) = \frac{s_1 c_2 d_2 + s_2 c_1 d_1}{1 - k^2 s_1^2 s_2^2}.$$

Putting $u_2 = 0$ gives $f(u_1) = s_1$, while $u_1 = 0$ gives $f(u_2) = s_2$. Therefore,

$$f(u_1+u_2) = \operatorname{sn}(u_1+u_2).$$

[6] Bowman (1953) 12–13. See Whittaker & Watson (1927) for an alternative proof.

[7] The notation $s_1, s_2, \ldots$ is also due to Glaisher.

Now, by (2.3) and (2.6), we have

$$\operatorname{cn}^2(u_1+u_2) = 1 - \operatorname{sn}^2(u_1+u_2) = \frac{(1-k^2s_1^2s_2^2)-(s_1c_2d_2+s_2c_1d_1)^2}{(1-k^2s_1^2s_2^2)^2}.$$

If we express $(1-k^2s_1^2s_2^2)^2$ in the form $(c_1^2+s_1^2d_2^2)(c_2^2+s_2^2d_1^2)$, then

$$\operatorname{cn}^2(u_1+u_2) = \frac{(c_1c_2-s_1s_2d_1d_2)^2}{(1-k^2s_1^2s_2^2)^2}.$$

We then take the square root, and to remove the ambiguity in sign we note that both of these expressions are one-valued functions of u_1 except at isolated poles, so, by the theory of analytic continuation, either the positive sign, or else the negative sign must always be taken. By setting $u_2 = 0$, it follows that the positive sign must be taken.

The formula for $\operatorname{dn}(u_1+u_2)$ follows by a similar argument.

□

2.5. The constants K and K' [8]

We can define the constant K by [9]

$$K = \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k^2t^2)}}.$$

It follows that

$$\operatorname{sn} K = 1, \quad \operatorname{cn} K = 0, \quad \operatorname{dn} K = k'.$$

Similarly, we denote the integral

$$\int_0^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k'^2t^2)}} = \int_1^{1/k} \frac{\mathrm{d}t}{\sqrt{(t^2-1)(1-k^2t^2)}}$$

by the symbol K'. Since

$$K + \mathrm{i}K' = \int_0^{1/k} \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k^2t^2)}},$$

we have

$$\operatorname{sn}(K+\mathrm{i}K') = 1/k, \quad \operatorname{cn}(K+\mathrm{i}K') = -\mathrm{i}k'/k, \quad \operatorname{dn}(K+\mathrm{i}K') = 0. \qquad (2.7)$$

2.6. Periodicity of the Jacobi elliptic functions

By definition of an elliptic function, the Jacobi elliptic functions must be doubly periodic. These periods can be expressed in terms of the constants K and K' introduced in the previous section.

---

[8] Whittaker & Watson (1927) 498–499, 501–502.

[9] The integral K is the complete elliptic integral of the first kind (see Section 4.1).

THEOREM 2.4. *The functions* $\operatorname{sn} u$ *and* $\operatorname{cn} u$ *each have a period* $4K$*, while* $\operatorname{dn} u$ *has a smaller period* $2K$*.*

Proof (Whittaker & Watson, 1927). By Theorem 2.3, we have 10

$$\operatorname{sn}(u+K) = \frac{\operatorname{sn} u \operatorname{cn} K \operatorname{dn} K + \operatorname{sn} K \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 K} = \operatorname{cd} u.$$

Similarly,

$$\operatorname{sn}(u+K) = -k' \operatorname{sd} u, \quad \operatorname{dn}(u+K) = k' \operatorname{nd} u.$$

Hence,

$$\operatorname{sn}(u+2K) = \frac{\operatorname{cn}(u+K)}{\operatorname{dn}(u+K)} = \frac{-k' \operatorname{sd} u}{k' \operatorname{nd} u} = -\operatorname{sn} u,$$

and it also follows that

$$\operatorname{cn}(u+2K) = -\operatorname{cn} u, \quad \operatorname{dn}(u+2K) = \operatorname{dn} u.$$

Finally,

$$\operatorname{sn}(u+4K) = \operatorname{sn} u, \quad \operatorname{cn}(u+4K) = \operatorname{cn} u.$$

□

THEOREM 2.5. *The functions* $\operatorname{sn} u$ *and* $\operatorname{dn} u$ *each have a period* $4K + 4\mathrm{i}K'$*, while* $\operatorname{cn} u$ *has a smaller period* $2K + 2\mathrm{i}K'$*.*

Proof (Whittaker & Watson, 1927). By (2.7), we have 11

$$\operatorname{sn}(K + \mathrm{i}K') = 1/k, \quad \operatorname{cn}(K + \mathrm{i}K') = -\mathrm{i}k'/k, \quad \operatorname{dn}(K + \mathrm{i}K') = 0.$$

Hence, by applying Theorem 2.3, we obtain

$$\begin{aligned} \operatorname{sn}(u + K + \mathrm{i}K') &= \frac{\operatorname{sn} u \operatorname{cn}(K + \mathrm{i}K') \operatorname{dn}(K + \mathrm{i}K') + \operatorname{sn}(K + \mathrm{i}K') \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2(K + \mathrm{i}K')} \\ &= (1/k) \operatorname{dc} u. \end{aligned} \tag{2.8}$$

By the same argument,

$$\operatorname{cn}(u + K + \mathrm{i}K') = -(\mathrm{i}k'/k) \operatorname{nc} u, \quad \operatorname{dn}(u + K + \mathrm{i}K') = \mathrm{i}k' \operatorname{sc} u.$$

Following further applications of the same formulæ, we have

$$\operatorname{sn}(u + 2K + 2\mathrm{i}K') = -\operatorname{sn} u, \quad \operatorname{cn}(u + 2K + 2\mathrm{i}K') = \operatorname{cn} u,$$
$$\operatorname{dn}(u + 2K + 2\mathrm{i}K') = -\operatorname{dn} u.$$

Hence,

$$\operatorname{sn}(u + 4K + 4\mathrm{i}K') = \operatorname{sn} u, \quad \operatorname{dn}(u + 4K + 4\mathrm{i}K') = \operatorname{dn} u.$$

□

THEOREM 2.6. *The functions* $\operatorname{cn} u$ *and* $\operatorname{dn} u$ *each have a period* $4\mathrm{i}K'$*, while* $\operatorname{sn} u$ *has a smaller period* $2\mathrm{i}K'$*.*

Proof (Whittaker & Watson, 1927). By using Theorem 2.3 and the result (2.8), we have 12

$$\operatorname{sn}(u + \mathrm{i}K') = \operatorname{sn}(u - K + K + \mathrm{i}K') = (1/k)\operatorname{dc}(u - K)$$

Hence,

$$\operatorname{sn}(u + \mathrm{i}K') = (1/k)\operatorname{ns} u,$$

and also

$$\operatorname{cn}(u + \mathrm{i}K') = -(\mathrm{i}/k)\operatorname{ds} u, \quad \operatorname{dn}(u + \mathrm{i}K') = -\mathrm{i}\operatorname{cs} u.$$

By repeated applications of these formulæ, we have

$$\operatorname{sn}(u + 2\mathrm{i}K') = \operatorname{sn} u, \quad \operatorname{cn}(u + 2\mathrm{i}K') = -\operatorname{cn} u, \quad \operatorname{dn}(u + 2\mathrm{i}K') = -\operatorname{dn} u,$$
$$\operatorname{cn}(u + 4\mathrm{i}K') = \operatorname{cn} u, \quad \operatorname{dn}(u + 4\mathrm{i}K') = \operatorname{dn} u.$$

□

As $4K$ and $4K'$ are periods of the Jacobi elliptic functions, we refer to K as the *quarter period* and K' as the *complementary quarter period*.

2.7. Poles and zeroes of the Jacobi elliptic functions 13

Each of the functions $\operatorname{sn} u$, $\operatorname{cn} u$ and $\operatorname{dn} u$ have two simple poles and two simple zeroes. In the case of $\operatorname{sn} u$, the poles are at points congruent to $\mathrm{i}K'$ or $2K + \mathrm{i}K'$ with residues $1/k$ and $-1/k$, respectively. It has simple zeroes at points congruent to zero and $2K$. For $\operatorname{cn} u$, we have poles at points congruent to $\mathrm{i}K'$ or $2K + \mathrm{i}K'$, while its zeroes are at points congruent to K and $-K$ with residues $-\mathrm{i}/k$ and $-\mathrm{i}/k$, respectively. Finally, $\operatorname{dn} u$ has poles congruent to $\mathrm{i}K'$ and $-\mathrm{i}K$ with residues $-\mathrm{i}$ and i, respectively, and zeroes congruent to $K + \mathrm{i}K'$ and $K - \mathrm{i}K'$.

2.8. The theta functions 14

It is possible to express the Jacobi elliptic functions in terms of quotients of non-elliptic functions known as *theta functions*.

[10] Whittaker & Watson (1927) 500.

[11] Whittaker & Watson (1927) 502–503.

[12] Whittaker & Watson (1927) 503.

[13] Whittaker & Watson (1927) 504–505.

[14] Whittaker & Watson (1927) 462–490.

The theta functions were first studied by Jacobi in the early nineteenth century. He developed these using elliptic functions and used purely algebraic methods to derive their properties. However, the first function considered to be like a theta function is due to Leonhard Euler (1707–1783), who introduced the partition function

$$\prod_{n=1}^{\infty} \frac{1}{1 - x^n z}$$

in the first volume of *Introductio in Analysin Infinitorum*, published in 1748. Euler also obtained properties of similar products, whose associated series had previously been investigated by Jakob Bernoulli (1654–1705). Theta functions also appear in *La Théorie Analytique de la Chaleur* by Jean-Baptiste Joseph Fourier (1768–1830).

Suppose τ is a constant complex number whose imaginary part is positive, and let $q = \mathrm{e}^{\pi \mathrm{i} \tau}$ for $|q| < 1$. Then we define the theta function $\vartheta_4(z, q)$ by

$$\vartheta_4(z, q) = 1 + 2\sum_{n=1}^{\infty}(-1)^n q^{n^2} \cos 2nz.$$

Clearly,

$$\vartheta_4(z + \pi, q) = \vartheta_4(z, q),$$

so $\vartheta_4(z, q)$ is a periodic function of z with period π. Moreover,

$$\vartheta_4(z + \pi\tau, q) = q^{-1}\mathrm{e}^{-2\mathrm{i}z}\vartheta_4(z, q).$$

It follows that we call $\vartheta_4(z, q)$ a *quasi doubly periodic* function of z. The effect of increasing z by π or $\pi\tau$ is the same as multiplying $\vartheta_4(z, q)$ by 1 or $-q^{-1}\mathrm{e}^{-2\mathrm{i}z}$. Therefore, 1 and $-q^{-1}\mathrm{e}^{-2\mathrm{i}z}$ are said to be the *multipliers* associated with the periods π and $\pi\tau$, respectively.

Three other theta functions are defined as follows:

$$\begin{aligned}
\vartheta_3(z, q) &= \vartheta_4(z + \tfrac{1}{2}\pi, q) = 1 + 2\sum_{n=1}^{\infty} q^{n^2} \cos 2nz, \\
\vartheta_1(z, q) &= \mathrm{i}\mathrm{e}^{\mathrm{i}z + \frac{1}{4}\pi\mathrm{i}\tau}\vartheta_4(z + \tfrac{1}{2}\pi\tau, q) = 2\sum_{n=0}^{\infty}(-1)^n q^{(n+\frac{1}{2})^2} \sin(2n+1)z, \\
\vartheta_2(z, q) &= \vartheta_1(z + \tfrac{1}{2}\pi, q) = 2\sum_{n=0}^{\infty} q^{(n+\frac{1}{2})^2} \cos(2n+1)z.
\end{aligned}$$

Clearly, $\vartheta_4(z, q)$ is an odd function of z, while $\vartheta_1(z, q)$, $\vartheta_2(z, q)$ and $\vartheta_3(z, q)$ are even functions of z.

If we do not need to emphasize the parameter q, we write $\vartheta(z)$ in place of $\vartheta(z, q)$. If we wish to alternatively show a dependence on τ, we write $\vartheta(z|\tau)$. Also we may replace $\vartheta(0)$ by ϑ and $\vartheta'(0)$ by ϑ'.

THEOREM 2.7. *A theta function $\vartheta(z)$ has exactly one zero inside each cell.*

Proof (Whittaker & Watson, 1927).[15] The function $\vartheta(z)$ is analytic in a finite part of the complex plane, so it follows, by the residue theorem, that if C is a cell with corners t, $t + \pi$, $t + \pi + \pi\tau$ and $t + \pi\tau$, then the number of zeroes in this cell is

$$\frac{1}{2\pi\mathrm{i}}\int_C \frac{\vartheta'(z)}{\vartheta(z)}\,\mathrm{d}z.$$

This integral reduces to

$$\frac{1}{2\pi\mathrm{i}}\int_t^{t+\pi} 2\mathrm{i}\,\mathrm{d}z,$$

and it follows that

$$\frac{1}{2\pi\mathrm{i}}\int_C \frac{\vartheta'(z)}{\vartheta(z)}\,\mathrm{d}z = 1.$$

Therefore, $\vartheta(z)$ has exactly one simple zero inside C.

□

[15] Whittaker & Watson (1927) 465–466.

Clearly, one zero of $\vartheta_1(z)$ is $z = 0$, so it follows that the zeroes of $\vartheta_2(z)$, $\vartheta_3(z)$ and $\vartheta_4(z)$ are the points congruent to $\frac{1}{2}\pi$, $\frac{1}{2}\pi + \frac{1}{2}\pi\tau$ and $\frac{1}{2}\pi\tau$, respectively.

We now define the Jacobi elliptic functions in terms of theta functions. We have

$$\begin{aligned} \mathrm{sn}(u, k) &= \frac{\vartheta_3\vartheta_1(u/\vartheta_3^2)}{\vartheta_2\vartheta_4(u/\vartheta_3^2)}, \\ \mathrm{cn}(u, k) &= \frac{\vartheta_4\vartheta_2(u/\vartheta_3^2)}{\vartheta_2\vartheta_4(u/\vartheta_3^2)}, \\ \mathrm{dn}(u, k) &= \frac{\vartheta_4\vartheta_3(u/\vartheta_3^2)}{\vartheta_3\vartheta_4(u/\vartheta_3^2)}, \end{aligned} \tag{2.9}$$

where $u = z\vartheta_3^2$ and $k = \vartheta_2^2/\vartheta_3^2$. As we have factors of $1/\vartheta_3^2$ in (2.9), we also consider the function

$$\Theta(u) = \vartheta_4(u/\vartheta_3^2|\tau).$$

This was Jacobi's original notation, and replaces the function $\vartheta_4(z)$. The periods associated with this function are $2K$ and $2\mathrm{i}K'$. The function $\Theta(u + K)$ therefore replaces $\vartheta_3(z)$, and in place of $\vartheta_1(z)$ we use the *eta function*, which is defined by

$$H(u) = -\mathrm{i}q^{-\frac{1}{4}}\mathrm{e}^{\mathrm{i}\pi u/2K}\Theta(u + \mathrm{i}K') = \vartheta_1(u/\vartheta_3^2|\tau).$$

Hence, instead of $\vartheta_2(z)$ we have $H(u + K)$.

A geometric application of the theta functions appears in *On the counting of colored tangles* (Zinn-Justin & Zuber, 2000). A *tangle* is a knotted structure from which four strings emerge, and it is possible to formulate the problem of counting coloured such tangles as one of theoretic physics, and then to use matrix integrals and Feynman diagrams. In the case of two-coloured tangles in which the string makes alternative under and over crossings with itself, the solution is expressed in terms of theta functions.

Chapter 3

Weierstrass Elliptic Functions

The *Weierstrass elliptic function* is a member of the second family of elliptic functions of order 2, those with a single irreducible double pole in each cell with residue equal to zero.

3.1. Motivation [1]

We first consider the properties of the circular function $\operatorname{cosec}^2 z$, which in some ways can be considered to be the circular analogue of the Weierstrass elliptic function.

We obtain the properties of $\operatorname{cosec}^2 z$ by the series

$$f(z) = \sum_{m=-\infty}^{\infty} \frac{1}{z - m\pi}.$$

This converges absolutely and uniformly except at the points $m\pi$ where it has double poles. Therefore, $f(z)$ is analytic throughout the whole complex plane except at these points. If we add a multiple of π to z, we have a series whose individual terms are identical to those occurs in the original series. Since the series is absolutely convergent, the sum of the series is unchanged. It follows that $f(z)$ is a simply periodic function with period π.

Now, consider $f(z)$ in the strip where the real part of z is between $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$. By periodicity, the value of $f(z)$ at any point on the plane is equal to its value at the corresponding point of the strip. In this strip, $f(z)$ has one singularity at $z = 0$, and is bounded as z tends to infinity.

In a domain including the point $z = 0$, the function

$$f(z) - \frac{1}{z^2}$$

It has been claimed that Karl Theodor Wilhelm Weierstrass (1815–1897) wasn't particularly geometrically minded. *Algebraic truths vs. geometric fantasies: Weierstrass' response to Riemann* (Bottazzini, 2002) recounts an argument in which Weierstrass accused Georg Friedrich Bernhard Riemann (1826–1866) of working with *geometric fantasies*, whereas he was only interested in *algebraic truths*. Whether this explains why there are fewer geometric applications of his elliptic function than that of Jacobi is debatable; a more obvious reason might be the visualization of the Jacobi elliptic functions as a generalization of the circular functions of plane trigonometry.

[1] Whittaker & Watson (1927) 438–439.

is analytic, and an even function. Therefore, it has the Taylor expansion

$$f(z) - \frac{1}{z^2} = \sum_{n=0}^{\infty} a_{2n} z^{2n},$$

when $|z| < \pi$. Clearly,

$$a_{2n} = 2\pi^{-2n-2}(2n+1)\sum_{m=1}^{\infty} m^{-2n-2},$$

and it follows that

$$a_0 = \tfrac{1}{3}, \quad a_2 = 6\pi^{-4}\sum_{m=1}^{\infty} m^{-4} = \tfrac{1}{15}.$$

Therefore, for small values of $|z|$, we have

$$f(z) = z^{-2} + \tfrac{1}{3} + \tfrac{1}{15}z^2 + O(z^4).$$

By taking the second derivative, it follows that

$$f''(z) = 6z^{-4} + \tfrac{2}{15} + O(z^2),$$

and, by squaring, we obtain

$$f^2(z) = z^{-4} + \tfrac{2}{3}z^{-2} + \tfrac{1}{20} + O(z^2).$$

Hence,

$$f''(z) - 6f^2(z) + 4f(z) = O(z^2).$$

So the function $f''(z) - 6f^2(z) + 4f(z)$ is analytic at the origin and is periodic. Moreover, it is bounded as z tends to infinity in the strip previously defined, as both $f(z)$ and $f'(z)$ are bounded. Hence, $f''(z) - 6f^2(z) + 4f(z)$ is bounded in the strip, and, as a consequence of its periodicity, it is bounded in the whole complex plane. Therefore, by Liouville's theorem, it is a constant. If we let z tend to zero, we have that the constant [2]
is zero. It follows that $\operatorname{cosec}^2 z$ satisfies the equation

$$f''(z) = 6f^2(z) - 4f(z).$$

3.2. Definition of the Weierstrass elliptic function [3]

The Weierstrass elliptic function $\wp(z)$ is defined by the infinite sum

$$\wp(z) = \frac{1}{z^2} + {\sum_{m,n}}' \left\{ \frac{1}{(z - m\omega_1 - n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right\}, \tag{3.1}$$

where ω_1 and ω_2 are periods, and the summation is taken over all integer values of m and n, except for when m and n are both equal to zero.

[2]See Section 1.3.

[3]Whittaker & Watson (1927) 433-434.

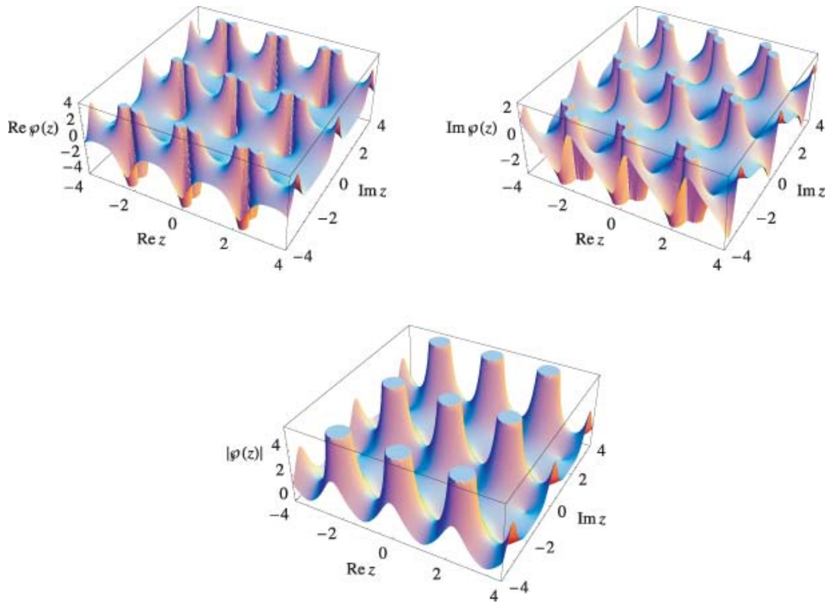

Figure 3.1. Graphs of the real part (a), imaginary part (b) and absolute value (c) of the Weierstrass function $\wp(z; g_2, g_3)$ with invariants $g_2 = 5$ and $g_3 = 2$.

The series converges absolutely and uniformly with respect to z apart from near its poles, which are points of $\Omega = \{m\omega_1 + n\omega_2\}$. It follows that $\wp(z)$ is analytic on the whole complex plane except at the points of Ω, where it has double poles.

Clearly, $\wp(z)$ is dependent on the values of its periods ω_1 and ω_2. When we wish to emphasize this, we can write $\wp(z|\omega_1, \omega_2)$, or the more commonly used notation $\wp(z; g_2, g_3)$, where g_2 and g_3 are constants defined in terms of the periods, and which will be defined precisely in Section 3.4.

3.3. Periodicity and other properties of the Weierstrass elliptic function

Since the series (3.1) is a uniformly convergent series of analytic functions, we can differentiate termwise to obtain

$$\wp'(z) = -2\sum_{m,n} \frac{1}{(z - m\omega_1 - n\omega_2)^3}.$$

Having established the definitions of $\wp(z)$ and $\wp'(z)$, we can now investigate two of their basic properties.

THEOREM 3.1. *The function $\wp'(z)$ is an odd function of z, while $\wp(z)$ is an even function of z.*

Proof (Whittaker & Watson, 1927). By the definition of $\wp'(z)$ we have 4

$$\wp'(-z) = 2\sum_{m,n} \frac{1}{(z + m\omega_1 + n\omega_2)^3}.$$

Now, the set of points $-\Omega$ is the same as those of Ω, so the terms of $\wp'(z)$ are identical as those of $-\wp'(z)$, but in a different order. However, as the series is absolutely convergent, we can rearrange terms to obtain

$$\wp'(-z) = -\wp'(z).$$

Similarly, we have that the terms of the absolutely convergent series

$$\sum_{m,n}{}' \left\{ \frac{1}{(z + m\omega_1 + n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right\}$$

are the terms of the series

$$\sum_{m,n}{}' \left\{ \frac{1}{(z - m\omega_1 - n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right\},$$

but rearranged. Hence, by uniform convergence it follows that

$$\wp(-z) = \wp(z).$$

□

THEOREM 3.2. *Both $\wp'(z)$ and $\wp(z)$ are elliptic functions with periods ω_1 and ω_2.*

Proof (Whittaker & Watson, 1927). We have [5]

$$\wp'(z + \omega_1) = -2\sum_{m,n} \frac{1}{(z - m\omega_1 - n\omega_2 + \omega_1)^3}.$$

The set of points $\Omega - \omega_1$ is the same as those of the set Ω so the terms of this series are the same as those of $\wp'(z)$, but in a different order. Once more using absolute convergence, we have

$$\wp'(z + \omega_1) = \wp'(z). \tag{3.2}$$

Therefore, $\wp'(z)$ has period ω_1. Similarly, it also has period ω_2. As $\wp'(z)$ is analytic except at its poles, it follows that it is an elliptic function.

Now, by integration of (3.2), we obtain

$$\wp(z + \omega_1) = \wp(z) + A,$$

for a constant A. Letting $z = -\omega_1$, then, as $\wp(z)$ is an even function, we have $A = 0$. Therefore,

$$\wp(z + \omega_1) = \wp(z),$$

and by an analogous argument $\wp(z + \omega_2) = \wp(z)$. Having seen in the previous section that it is analytic except at its poles, it follows that $\wp(z)$ is an elliptic function.

□

[4] Whittaker & Watson (1927) 434–435.

[5] Whittaker & Watson (1927) 435.

3.4. A differential equation satisfied by the Weierstrass elliptic function

THEOREM 3.3. *The function $\wp(z)$ satisfies the differential equation*

$$\wp'^2(z) = 4\wp^3(z) - g_2\wp(z) - g_3,$$

where the elliptic invariants g_2 and g_3 are given by

$$g_2 = 60\sum_{m,n}{}' \frac{1}{(z+m\omega_1+n\omega_2)^4}, \quad g_3 = 140\sum_{m,n}{}' \frac{1}{(z+m\omega_1+n\omega_2)^6}.$$

Proof (Whittaker & Watson, 1927). We have [6]

$$\wp(z) - \frac{1}{z^2} = \sum_{m,n}{}' \left\{ \frac{1}{(z-m\omega_1-n\omega_2)^2} - \frac{1}{(m\omega_1+n\omega_2)^2} \right\}$$

is analytic in a region around the origin, and as it is an even function of z, we obtain the Taylor expansion

$$\wp(z) - z^{-2} = \tfrac{1}{20}g_2z^2 + \tfrac{1}{28}g_3z^4 + O(z^6),$$

for sufficiently small values of $|z|$. Therefore,

$$\wp(z) = z^{-2} + \tfrac{1}{20}g_2z^2 + \tfrac{1}{28}g_3z^4 + O(z^6). \tag{3.3}$$

By differentiation, we have

$$\wp'(z) = -2z^{-3} + \tfrac{1}{10}g_2z + \tfrac{1}{7}g_3z^3 + O(z^5). \tag{3.4}$$

From (3.3) and (3.4), it follows that

$$\wp^3(z) = \frac{1}{z^6} + \tfrac{3}{20}g_2\frac{1}{z^2} + \tfrac{3}{28}g_3 + O(z^2),$$
$$\wp'^2(z) = \frac{4}{z^6} - \tfrac{2}{5}g_2\frac{1}{z^2} - \tfrac{4}{7}g_3 + O(z^2).$$

Hence,

$$\wp'^2(z) - 4\wp^3(z) + g_2\wp(z) + g_3 = O(z^2).$$

Now, the left hand side is an elliptic function, analytic at the origin and at all congruent points. But these points are the only possible singularities of the function, so it is an elliptic function with no singularities. By Theorem 1.4, it must be constant.

By letting z tend to zero, we have that this constant is zero.

□

We can also consider the converse statement.

THEOREM 3.4. *Given*

$$\left(\frac{\mathrm{d}y}{\mathrm{d}z}\right)^2 = 4y^3 - g_2y - g_3, \tag{3.5}$$

[6] Whittaker & Watson (1927) 436–437.

and if numbers ω_1 and ω_2 can be determined such that

$$g_2 = 60 \sum_{m,n}{}' \frac{1}{(z + m\omega_1 + n\omega_2)^4}, \quad g_3 = 140 \sum_{m,n}{}' \frac{1}{(z + m\omega_1 + n\omega_2)^6}.$$

then the general solution of the differential equation is

$$y = \wp(z + A),$$

for a constant A.

Proof (Whittaker & Watson, 1927). If we take a variable u defined by $y = \wp(u)$, an equation which always has solutions, then (3.5) reduces to [7]

$$\left(\frac{\mathrm{d}u}{\mathrm{d}z}\right)^2 = 1.$$

As $\wp(u)$ is an even function of u, we have $y = \wp(z \pm A)$. Without loss of generality, we can choose the choice of sign to be positive to obtain the result.

□

3.5. The addition formula for the Weierstrass elliptic function

An important property of the Weierstrass elliptic function is that it has an algebraic addition formula.

THEOREM 3.5. *The addition formula for the Weierstrass elliptic function is*

$$\begin{vmatrix} \wp(z_1) & \wp'(z_1) & 1 \\ \wp(z_2) & \wp'(z_2) & 1 \\ \wp(z_1 + z_2) & -\wp'(z_1 + z_2) & 1 \end{vmatrix} = 0.$$

Proof (Whittaker & Watson, 1927). We consider the equations [8]

$$\wp'(z_1) = A\wp(z_1) + B, \quad \wp'(z_2) = A\wp(z_2) + B. \tag{3.6}$$

These determine A and B in terms of z_1 and z_2 unless $\wp(z_1) = \wp(z_2)$. We also consider the function

$$\wp'(u) - A\wp(u) - B.$$

This has a triple pole at $u = 0$. By Theorem 1.5 it has exactly three irreducible zeroes. Moreover, by Theorem 1.6, as $u = z_1$ and $u = z_2$ are two zeroes, then the third irreducible zero must be congruent to $-z_1 - z_2$. Hence, we have $-z_1 - z_2$ is a zero of $\wp'(u) - A\wp(u) - B$, and it follows that

$$\wp'(-z_1 - z_2) = A\wp(-z_1 - z_2) + B.$$

The result follows by substituting from (3.6) to eliminate A and B.

□

[7] Whittaker & Watson (1927) 437.

A geometric proof by Niels Henrik Abel (1802–1829) of the addition formula for the Weierstrass elliptic function is described in Section 13.4.

[8] Whittaker & Watson (1927) 440.

If we introduce a further variable z_3 and set $z_1 + z_2 + z_3$, then Theorem 3.5 takes a symmetric form.

THEOREM 3.6. *If* $z_1 + z_2 + z_3 = 0$, *then the addition formula for the Weierstrass elliptic function is*

$$\begin{vmatrix} \wp(z_1) & \wp'(z_1) & 1 \\ \wp(z_2) & \wp'(z_2) & 1 \\ \wp(z_3) & \wp'(z_3) & 1 \end{vmatrix} = 0.$$

This leads us to consider the following:

THEOREM 3.7. *If* $\wp(z_1) = p_1$, $\wp(z_2) = p_2$ *and* $\wp(z_3) = p_3$ *and* $z_1 + z_2 + z_3 = 0$, *then the addition formula for the Weierstrass elliptic function is*

$$(p_1 + p_2 + p_3)(4p_1p_2p_3 - g_3) = (p_1p_2 + p_2p_3 + p_3p_1 + \tfrac{1}{4}g_2)^2. \tag{3.7}$$

Proof (Copson, 1935). From the proof of Theorem 3.5, it is clear that the values of u 9
which cause $\wp'(u) - A\wp(u) - B$ to vanish are congruent to one of the points z_1, z_2 or z_3. Hence, $\wp'^2(u) - (A\wp(u) - B)^2$ vanishes when u is congruent to any of z_1, z_2 or z_3, Therefore,

$$4\wp^3(u) - A^2\wp^2(u) - (2AB + g_2)\wp(u) - (B^2 + g_3)$$

vanishes when $\wp(u)$ is equal to each of $\wp(z_1)$, $\wp(z_2)$ and $\wp(z_3)$, or, in the notation of the theorem, p_1, p_2 and p_3.

For general values of z_1 and z_2, the variables p_1, p_2 and p_3 are not equal and so are roots of

$$4p^3 - A^2p^2 - (2AB + g_2)p - (B^2 + g_3).$$

Now, the left hand side of (3.7) contains the sum and product of these roots, so, from the formulæ that link the roots of equations with their coefficients, we have

$$p_1 + p_2 + p_3 = \tfrac{1}{4}A^2, \quad p_1p_2p_3 = \tfrac{1}{4}(g_3 + B^2).$$

It follows that

$$(p_1 + p_2 + p_3)(4p_1p_2p_3 - g_3) = \tfrac{1}{4}A^2B^2.$$

By a similar argument, we have that the right hand side of (3.7) is

$$(p_1p_2 + p_2p_3 + p_3p_1 + \tfrac{1}{4}g_2)^2 = \tfrac{1}{4}A^2B^2.$$

□

3.6. The constants e_1, e_2 and e_3

If $\omega_3 = -\omega_1 - \omega_2$, we have three constants e_1, e_2 and e_3 defined by

$$\wp(\tfrac{1}{2}\omega_1) = e_1, \quad \wp(\tfrac{1}{2}\omega_2) = e_2, \quad \wp(\tfrac{1}{2}\omega_3) = e_3.$$

THEOREM 3.8. *The constants* e_1, e_2 *and* e_3 *are mutually distinct and are roots of the equation*

$$4y^3 - g_2y - g_3 = 0.$$

Proof (Whittaker & Watson, 1927). We consider $\wp'(\frac{1}{2}\omega_1)$. Since $\wp'(z)$ is an odd, periodic function we have 10

$$\wp'(\tfrac{1}{2}\omega_1) = -\wp'(-\tfrac{1}{2}\omega_1) = -\wp'(\omega_1 - \tfrac{1}{2}\omega_1) = -\wp'(\tfrac{1}{2}\omega_1).$$

Therefore, we have $\wp'(\frac{1}{2}\omega_1) = 0$. Similarly,

$$\wp'(\tfrac{1}{2}\omega_2) = \wp'(\tfrac{1}{2}\omega_3) = 0.$$

Now, as $\wp'(z)$ is an elliptic function whose only singularities are triple poles at points congruent to the origin, it has exactly three irreducible zeroes by Theorem 1.5. Therefore, the only zeroes of $\wp'(z)$ are points congruent to $\frac{1}{2}\omega_1$, $\frac{1}{2}\omega_2$ and $\frac{1}{2}\omega_3$.

Next, we consider $\wp(z) - e_1$. By definition, this vanishes at $\frac{1}{2}\omega_1$, and since $\wp'(\frac{1}{2}\omega_1) = 0$, it has a double zero at $\frac{1}{2}\omega_1$. As $\wp(z)$ has only two irreducible poles, it follows that the only zeroes of $\wp(z) - e_1$ are congruent to ω_1. By the same argument, the only zeroes of $\wp(z) - e_2$ and $\wp(z) - e_3$ are double poles at points congruent to $\frac{1}{2}\omega_2$ and $\frac{1}{2}\omega_3$, respectively.

If $e_1 = e_2$, then $\wp(z) - e_1$ would have a zero at $\frac{1}{2}\omega_2$. But this would not be congruent to ω_1, so $e_1 \neq e_2 \neq e_3$.

Now, by Theorem 3.3,

$$\wp'^2(z) = 4\wp^3(z) - g_2\wp(z) - g_3,$$

and since $\wp'(z)$ vanishes at $\frac{1}{2}\omega_1$, $\frac{1}{2}\omega_2$ and $\frac{1}{2}\omega_3$, it follows that $4\wp^3(z) - g_2\wp(z) - g_3$ vanishes when $\wp(z)$ is equal to each of e_1, e_2 and e_3.

□

From the formulæ that link the roots of equations with their coefficients, we have also have the following formulæ:

$$\begin{aligned} e_1 + e_2 + e_3 &= 0, \\ e_1e_2 + e_2e_3 + e_3e_1 &= -\tfrac{1}{4}g_2, \\ e_1e_2e_3 &= \tfrac{1}{4}g_3. \end{aligned}$$

3.7. Connection with the Jacobi elliptic functions and the theta functions

If we write 11

$$y = e_3 + \frac{e_1 - e_3}{\operatorname{sn}(\lambda u)},$$

then it follows that

$$\begin{aligned} \left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)^2 &= 4(e_1 - e_3)^2\lambda^2 \operatorname{ns}^2 \lambda u \operatorname{cs}^2 \lambda u \operatorname{ds}^2 \lambda u \\ &= 4(e_1 - e_3)^2\lambda^2 \operatorname{ns}^2 \lambda u (\operatorname{ns}^2 \lambda u - 1)(\operatorname{ns}^2 \lambda u - k^2). \end{aligned}$$

[9] Copson (1935) 362–364.
[10] Whittaker & Watson (1927) 443–444.
[11] Whittaker & Watson (1927) 505.

Hence, if $\lambda^2 = e_1 - e_3$ and $k^2 = (e_2 - e_3)/(e_1 - e_3)$ then y satisfies the equation

$$\left(\frac{\mathrm{d}y}{\mathrm{d}u}\right)^2 = 4y^3 - g_2 y - g_3,$$

and it follows that

$$e_3 + (e_1 - e_3)\,\mathrm{ns}^2(u\sqrt{e_1 - e_3}) = \wp(u + A),$$

where A is a constant. Letting u tend to zero, we have that A is a period, and so

$$\wp(u) = e_3 + (e_1 - e_3)\,\mathrm{ns}^2(u\sqrt{e_1 - e_3}, k),$$

where the modulus is given by $k^2 = (e_2 - e_3)/(e_1 - e_3)$.

We can also express $\wp(z)$ in terms of theta functions in the following way: 12

$$\begin{aligned}\wp(z) &= e_1 + \left\{\frac{\vartheta_1'\vartheta_2(z)}{\vartheta_1(z)\vartheta_2}\right\}^2 \\ &= e_2 + \left\{\frac{\vartheta_1'\vartheta_3(z)}{\vartheta_1(z)\vartheta_3}\right\}^2 \\ &= e_3 + \left\{\frac{\vartheta_1'\vartheta_4(z)}{\vartheta_1(z)\vartheta_4}\right\}^2.\end{aligned}$$

This results from the double periodicity of quotients $\vartheta_2^2(z)/\vartheta_1^2(z)$, $\vartheta_2^3(z)/\vartheta_1^3(z)$ and $\vartheta_2^4(z)/\vartheta_1^4(z)$, and because they only have a single pole of order 2 within each cell.

3.8. The Weierstrass zeta and sigma functions

The *Weierstrass zeta function* is defined by the equation

$$\zeta'(z) = -\wp(z),$$

along with the condition $\lim_{z\to\infty}\{\zeta(z) - 1/z\} = 0$.

The series for $\wp(z) - 1/z^2$ converges uniformly throughout any domain from which the neighbourhood of the points of $\mathbf{\Omega}$, except the origin, are excluded. Therefore, we integrate termwise to obtain

$$\zeta(z) - \frac{1}{z} = -\sum_{m,n}{}' \int_0^z \left\{\frac{1}{(z - m\omega_1 - n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2}\right\} \mathrm{d}z,$$

and it follows that

$$\zeta(z) = \frac{1}{z} + \sum_{m,n}{}' \left\{\frac{1}{z - m\omega_1 - n\omega_2} + \frac{1}{m\omega_1 + n\omega_2} + \frac{z}{(m\omega_1 + n\omega_2)^2}\right\}.$$

The function $\zeta(z)$ is analytic over the whole complex plane except at simple poles at all the points of $\mathbf{\Omega}$.

THEOREM 3.9. *The Weierstrass zeta function $\zeta(z)$ is an odd function of z.*

Proof (*Whittaker & Watson, 1927*). It is clear that the series for $-\zeta(-z)$ is given by [13]

$$-\zeta(-z) = \frac{1}{z} + \sum_{m,n}{}' \left\{ \frac{1}{z + m\omega_1 + n\omega_2} - \frac{1}{m\omega_1 + n\omega_2} + \frac{z}{(m\omega_1 + n\omega_2)^2} \right\}.$$

This series consists of terms of the series for $\zeta(z)$ rearranged in the same way as in the proof of Theorem 3.1. Therefore,

$$\zeta(-z) = -\zeta(z).$$

$\square$

The *Weierstrass sigma function* is defined by the logarithmic derivative

$$\frac{\mathrm{d}}{\mathrm{d}z} \left\{ \log \sigma(z) \right\} = \zeta(z),$$

coupled with the condition $\lim_{z \to 0} \{\sigma(z)/z\} = 1$. As the series for $\zeta(z)$ converges uniformly except near the poles of $\zeta(z)$, we integrate the series termwise, and, after taking the exponential of each side, we have

$$\sigma(z) = z \prod_{m,n}{}' \left\{ \left(1 - \frac{z}{m\omega_1 + n\omega_2}\right) \exp\left(\frac{z}{m\omega_1 + n\omega_2} + \frac{z^2}{2(m\omega_1 + n\omega_2)^2} \right) \right\}.$$

By the absolute convergence of this series, it immediately follows that the product for $\sigma(z)$ converges uniformly and absolutely in any bounded domain of values of z.

THEOREM 3.10. *The Weierstrass sigma function $\sigma(z)$ is an odd function of z.*

Proof (*Whittaker & Watson, 1927*). By an analogous argument to the proof of Theorem 3.9, we have [14]

$$-\sigma(-z) = z \prod_{m,n}{}' \left\{ \left(1 + \frac{z}{m\omega_1 + n\omega_2}\right) \exp\left(\frac{z^2}{2(m\omega_1 + n\omega_2)^2} - \frac{z}{m\omega_1 + n\omega_2} \right) \right\}.$$

This series is absolutely convergent and so

$$\sigma(-z) = -\sigma(z).$$

$\square$

We can also define three further sigma functions $\sigma_1(z)$, $\sigma_2(z)$ and $\sigma_3(z)$ by

$$\sigma_1(z) = \mathrm{e}^{-\eta_1 z} \frac{\sigma(z + \frac{1}{2}\omega_1)}{\sigma(\frac{1}{2}\omega_1)},$$

$$\sigma_2(z) = \mathrm{e}^{-\eta_2 z} \frac{\sigma(z + \frac{1}{2}\omega_2)}{\sigma(\frac{1}{2}\omega_2)},$$

$$\sigma_3(z) = \mathrm{e}^{-\eta_3 z} \frac{\sigma(z + \frac{1}{2}\omega_3)}{\sigma(\frac{1}{2}\omega_3)},$$

[12] McKean & Moll (1999) 132.
[13] Whittaker & Watson (1927) 445.
[14] Whittaker & Watson (1927) 447.

where η_1, η_2 and η_3 are defined by $\zeta(z+\omega_i) = \eta_i + \zeta(z_i)$ for $i = 1, 2, 3$. Along with $\omega(z)$, these sigma functions can be considered to be analogous to the four theta functions.

If we continue with the analogy of Section 3.1, we can compare $\zeta(z)$ with the function $\cot z$ defined by the series

$$\frac{1}{z} + \sum_{m=-\infty}^{\infty}{}' \left\{ \frac{1}{z - m\pi} + \frac{1}{m\pi} \right\},$$

as the derivative of $-\cot z$ with respect to z is $\operatorname{cosec}^2 z$, the function analogous to $\wp(z)$. Similarly, we compare $\sigma(z)$ with the function $\sin z$ defined by

$$\prod_{m=-\infty}^{\infty}{}' \left\{ (1 - \frac{z}{m\pi}) \mathrm{e}^{z/m\pi} \right\}.$$

The logarithmic derivative of $\sin z$ with respect to z is $\cot z$.

Chapter 4

Elliptic Integrals

An *elliptic integral* takes the form

$$\int R(x, \sqrt{G(x)})\,\mathrm{d}x$$

where $G(x)$ is a either a cubic or quartic polynomial with no multiple root, and $R(x, y)$ is a rational function of two variables. An alternative definition of the elliptic functions is that they are the functions obtained by the inversion of an elliptic integral.

4.1. The elliptic integral of the first kind [1]

We calculate the derivative of the inverse function $\mathrm{sn}^{-1}\,x$, where we take values of this function to be in the interval from zero to K. Therefore, by putting $u = \mathrm{sn}^{-1}\,x$ we have $x = \mathrm{sn}\,u$, and it follows that

$$\frac{\mathrm{d}x}{\mathrm{d}u} = \mathrm{cn}\,u\,\mathrm{dn}\,u = \sqrt{(1-x^2)(1-k^2x^2)}.$$

Hence, by integration, we obtain

$$u = \int_0^x \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k^2t^2)}}.$$

This is known as the *elliptic integral of the first kind*. As $\mathrm{sn}^{-1}\,1 = K$, we also have

$$K(k) = \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k^2t^2)}},$$

the *complete elliptic integral of the first kind*. Similarly as $K'(k) = K(k')$, it follows that

$$K'(k) = \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k'^2t^2)}},$$

The appellation *elliptic integral* is attributed to Giulio Carlo Fagnano (1682–1766). Despite never himself carrying out the inversion of such an integral, so important was Fagnano's work that Jacobi called 23 December 1751 the birthday of elliptic functions. It was on the date that Euler was asked to examine the collected papers of Fagnano as a referee to his proposed membership of the Berlin Academy.

[1] Lawden (1989) 50–52.

or, after a change of variables by $s = \sqrt{1 - k'^2 t^2}/k$, we obtain

$$K'(k) = \int_1^{1/k} \frac{\mathrm{d}s}{\sqrt{(s^2 - 1)(1 - k^2 s^2)}}.$$

If we let $t = \sin\varphi$, then the elliptic integral of the first kind takes the form

$$F(\varphi, k) = \int_0^k \frac{\mathrm{d}\varphi}{\sqrt{1 - k^2 \sin^2\varphi}}.$$

4.2. The elliptic integral of the second kind [2]

The *elliptic integral of the second kind* is defined by

$$\int_0^x \sqrt{\frac{1 - k^2 t^2}{1 - t^2}}\, \mathrm{d}t.$$

If we set $t = \sin\varphi$ then we have

$$E(k, \varphi) = \int_0^\varphi \sqrt{1 - k^2 \sin^2\varphi}\, \mathrm{d}\varphi.$$

Moreover, if we put $\operatorname{sn} u = t = \sin\varphi$, then we obtain

$$E(u) = \int_0^u \operatorname{dn}^2 u \, \mathrm{d}u.$$

Since $\operatorname{dn}^2 u$ is an even function of u with double poles at the points $2mK + (2n+1)\mathrm{i}K$, with residue zero, then it follows that $E(u)$ is an odd function of u with simple poles at the poles of $\operatorname{dn}^2 u$.

Finally, the complete elliptic of the second kind is given by

$$E = \int_0^1 \frac{\sqrt{1 - k^2 t^2}}{\sqrt{1 - t^2}}\, \mathrm{d}t,$$

or alternatively,

$$E(K) = \int_0^K \operatorname{dn}^2 u \, \mathrm{d}u.$$

---

[2] Whittaker & Watson (1927) 517.

For completeness, the *elliptic integral of the third kind* is given by

$$\int_0^x \frac{\mathrm{d}t}{(1 + nt^2)\sqrt{(1 - t^2)(1 - k^2 t^2)}},$$

and, with the substitution $\operatorname{sn} u = t = \sin\varphi$,

$$\Pi(\varphi, n, k) = \int_0^\varphi \frac{\mathrm{d}\varphi}{(1 + n\sin^2\varphi)\sqrt{1 - k^2 \sin\varphi}} = \int_0^u \frac{\mathrm{d}u}{1 + n \operatorname{sn}^2 u} = \Pi(u, n, k).$$

It was shown by Adrien-Marie Legendre (1752–1822) that any elliptic integral can, by suitable linear transformations and reduction formulæ, be expressed as the sum of a finite number of elliptic integrals of the first, second and third kinds. Hence, these three integrals are known as *Legendre forms*.

4.3. The addition formula for the elliptic integral of the second kind

In common with the Jacobi and Weierstrass elliptic functions, the elliptic integral of the second kind satisfies an algebraic addition formula.

THEOREM 4.1. *The addition formula for the elliptic integral of the second kind is*

$$E(u_1 + u_2) = E(u_1) + E(u_2) - k^2 \operatorname{sn} u_1 \operatorname{sn} u_2 \operatorname{sn}(u_1 + u_2).$$

Proof using Jacobi elliptic functions (Bowman, 1953). Suppose we have two variables 3
x and y. From the addition formula for the Jacobi elliptic function $\operatorname{dn} x$ (Theorem 2.3), and as $\operatorname{cn} x$ and $\operatorname{dn} x$ are even functions, we have

$$\operatorname{dn}(x + y) + \operatorname{dn}(x - y) = \frac{2 \operatorname{dn} x \operatorname{dn} y}{1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y},$$
$$\operatorname{dn}(x + y) - \operatorname{dn}(x - y) = \frac{-2k^2 \operatorname{sn} x \operatorname{sn} y \operatorname{cn} x \operatorname{cn} y}{1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y}.$$

Multiplying these gives us

$$\operatorname{dn}^2(x + y) - \operatorname{dn}^2(x - y) = \frac{-4k^2 \operatorname{sn} x \operatorname{sn} y \operatorname{cn} x \operatorname{cn} y \operatorname{dn} x \operatorname{dn} y}{(1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y)^2},$$

and, by integration with respect to y, it follows that

$$E(x + y) + E(x - y) = C_x - \frac{2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{\operatorname{sn}^2 x (1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y)}, \tag{4.1}$$

with C_x depending only on x. If $y = x$, we have

$$E(2x) = C_x - \frac{2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{\operatorname{sn}^2 x (1 - k^2 \operatorname{sn}^4 x)},$$

and so, by subtracting this from (4.1), we obtain

$$E(x + y) + E(x - y) - E(2x) = \left\{ \frac{2k^2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{1 - k^2 \operatorname{sn}^4 x} \right\} \left\{ \frac{\operatorname{sn}^2 x - \operatorname{sn}^2 y}{1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y} \right\}.$$

Now, once more using Theorem 2.3, and as $\operatorname{sn} x$ is an odd function, it follows that

$$\operatorname{sn}(x + y) \operatorname{sn}(x - y) = \frac{\operatorname{sn}^2 x - \operatorname{sn}^2 y}{1 - k^2 \operatorname{sn}^2 x \operatorname{sn}^2 y}. \tag{4.2}$$

Similarly,

$$\operatorname{sn} 2x = \frac{2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{1 - k^2 \operatorname{sn}^4 x}. \tag{4.3}$$

Hence, by (4.2) and (4.3), we have

$$E(x + y) + E(x - y) - E(2x) = k^2 \operatorname{sn} 2x \operatorname{sn}(x + y) \operatorname{sn}(x - y).$$

If we let $u_1 = x + y$ and $u_2 = x - y$, then the result follows.

□

Proof using theta functions (Whittaker & Watson, 1927). Consider the derivative [4] [5]

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\frac{\Theta'(u)}{\Theta(u)}\right\}. \tag{4.4}$$

It is a doubly periodic function of u with double poles at the zeroes of $\Theta(u)$. So, for a constant A, we have

$$\mathrm{dn}^2 - A\frac{\mathrm{d}}{\mathrm{d}u}\left\{\frac{\Theta'(u)}{\Theta(u)}\right\}$$

is a doubly periodic function of u with periods $2K$ and $2\mathrm{i}K$, and with only a single simple pole in any cell. Therefore, it is equal to a constant, which we can write as E/K.

To calculate A, we compare the principal parts of $\mathrm{dn}^2 u$ and the derivative in (4.4). At $\mathrm{i}K'$ the principal part of $\mathrm{dn}^2 u$ is $-(u-\mathrm{i}K')^{-2}$. The residue of $\Theta'(u)/\Theta(u)$ at this pole is 1, so the principal part of (4.4) is $-(u-\mathrm{i}K')^{-2}$. Hence, $A = 1$, and it follows that

$$\mathrm{dn}^2 u = \frac{\mathrm{d}}{\mathrm{d}u}\left\{\frac{\Theta'(u)}{\Theta(u)} + \frac{E}{K}\right\}.$$

By integration, we have

$$E(u) = \frac{\Theta'(u)}{\Theta(u)} + \frac{E}{K}u, \tag{4.5}$$

as $\Theta'(0) = 0$.

Now, consider the function

$$f(u_1) = \frac{\Theta'(u_1+u_2)}{\Theta(u_1+u_2)} - \frac{\Theta'(u_1)}{\Theta(u_1)} - \frac{\Theta'(u_2)}{\Theta(u_2)} + k^2 \,\mathrm{sn}\, u_1 \,\mathrm{sn}\, u_2 \,\mathrm{sn}(u_1+u_2)$$

as a function of u_1. It is doubly periodic with periods $2K$ and $2\mathrm{i}K$, and with simple poles congruent to $\mathrm{i}K'$ and $\mathrm{i}K' - u_2$. The residue of the first two terms of $f(u_1)$ at $\mathrm{i}K'$ is -1, while the residue of $\mathrm{sn}\, u_1 \,\mathrm{sn}\, u_2 \,\mathrm{sn}(u_1+u_2)$ is $(1/k)\,\mathrm{sn}\, u_2 \,\mathrm{sn}(\mathrm{i}K'+u_2) = 1/k^2$.

Therefore, the function $f(u_1)$ is doubly periodic and has no poles at points congruent to $\mathrm{i}K'$ or, by a similar argument, at points congruent to $\mathrm{i}K' - u_2$. So, by Liouville's theorem (Theorem 1.4), it is constant. If we put $u_1 = 0$, then the constant is zero, and by (4.5), we have the result.

□

4.4. The integral formula for the Weierstrass elliptic function [6]

We consider the equation

$$z = \int_u^\infty \frac{\mathrm{d}t}{\sqrt{4t^3 - g_2 t - g_3}}, \tag{4.6}$$

[3] Bowman (1953) 22–23.
[4] Whittaker & Watson (1927) 517–519.
[5] The function $\Theta(u)$ is defined in Section 2.8.
[6] Whittaker & Watson (1927) 437–438.

which determines z in terms of u. The path of integration may be chosen to be any curve not passing through a zero of $4t^3 - g_2t - g_3$. After differentiation, we have

$$\left(\frac{\mathrm{d}u}{\mathrm{d}z}\right)^2 = 4u^3 - g_2u - g_3,$$

and it follows, from Section 3.4, that

$$u = \wp(z + A),$$

for a constant A.

Now, if we let u tend to infinity, then z tends to zero, as the integral converges. Therefore, A is a pole of the function $\wp(z)$. It follows that $u = \wp(z)$, and that (4.6) is equivalent to $u = \wp(z)$. Therefore, we can write

$$\wp^{-1}(u) = \int_u^\infty \frac{\mathrm{d}t}{\sqrt{4t^3 - g_2t - g_3}}.$$

It can be proven that any elliptic integral can be reduced to the form

$$\int R(x, y)\,\mathrm{d}x,$$

where $y^2 = 4x^3 - g_2x - g_3$. Elliptic integrals written in this form are said to be in *Weierstrass form*.

Part II

Applications of the Jacobi Elliptic Functions

Chapter 5

Greenhill's Pendulums

> *When a pendulum swings through a finite angle about a horizontal axis, the determination of the motion introduces the Elliptic Functions in such an elementary and straightforward manner, that we may take the elliptic functions as defined by a pendulum motion, and begin the investigation of their use and theory by their application to this problem.*
>
> SIR GEORGE GREENHILL

A *simple pendulum* consists of particle suspended by a light wire moving freely in a vertical plane under the force of gravity. The interval of time for each complete oscillation, its period, is constant. This period may be increased by increasing the length of the string, but a change in the mass of the particle leaves the period unaffected. If the strength of the gravitational force increases then the pendulum swings faster, and so has a shorter period.

5.1. The simple pendulum [1]

We consider a simple pendulum OP oscillating about a horizontal axis OA, as described by Figure 5.1. We suppose that the particle is of mass m and is being acted upon by a gravitational force mg. We denote by ma^2 the moment of the pendulum about the horizontal axis through P, such that $m(h^2 + a^2)$ is the moment about the parallel axis through O.

Now, if at time t, the line OP makes an angle θ with OA, and if the pendulum is considered to be vertical at $t = 0$, then the equation of motion of the pendulum that

---

Sir Alfred George Greenhill (1847–1927) had somewhat of an obsession with pendulums. In his obituary in *The Times*, a visitor to his lodgings described *his walls festooned with every variety of pendulum, simple or compound.* Greenhill was an applied mathematician, with a strong interest in military applications, and so it is perhaps unsurprising that *The Applications of Elliptic Functions* (Greenhill, 1892) is so strongly based on this mechanical viewpoint. Srinivasa Aiyangar Ramanujan (1887–1920) is said to have learned much of his mathematics from Greenhill's book. In a review of his *Collected Papers* appearing in the *The Mathematical Gazette* (Littlewood, 1929), John Edensor Littlewood (1885–1977) wrote of Ramanujan

> *Above all, he was totally ignorant of Cauchy's theorem and complex function-theory. (This may seem difficult to reconcile with his complete knowledge of elliptic functions. A sufficient, and I think a necessary, explanation would be that Greenhill's very odd and individual Elliptic Functions was his text-book).*

[1] Greenhill (1892) 1–5.

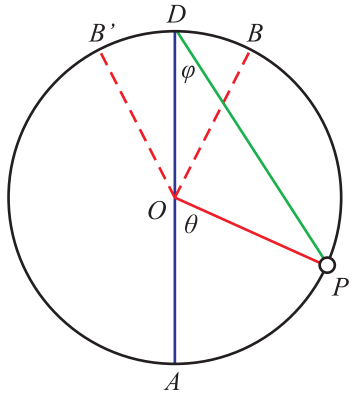

Figure 5.1. A simple pendulum.

follows by taking moments about O is

$$m(h^2 + a^2)\ddot{\theta} = -mgh\sin\theta,$$

where $\ddot{\theta}$ is the second derivative of θ with respect to time t. Therefore,

$$\left\{h + \frac{a^2}{h}\right\}\ddot{\theta} = -g\sin\theta,$$

and it follows, by making the substitution $h + a^2/h = l$, that

$$l\ddot{\theta} = -g\sin\theta.$$

Suppose we denote the angle of an oscillation between B and B' by 2α, and take this angle to be large. It follows that $\dot{\theta} = 0$ when $\theta = \alpha$, and by letting $g/l = \omega^2$, we have

$$\dot{\theta}^2 = 4\omega^2(\sin^2 \tfrac{1}{2}\alpha - \sin^2 \tfrac{1}{2}\theta). \tag{5.1}$$

Hence,

$$\omega t = \int_0^{\theta} \frac{d\frac{1}{2}\theta}{\sqrt{\sin^2 \frac{1}{2}\alpha - \sin^2 \frac{1}{2}\theta}}.$$

This is an elliptic integral of the first kind, and we reduce it to the standard form by letting

$$\sin\tfrac{1}{2}\theta = \sin\tfrac{1}{2}\alpha\sin\varphi,$$

where φ corresponds to the angle $\angle ADP$. Now,

$$\sin^2 \tfrac{1}{2}\alpha - \sin^2 \tfrac{1}{2}\theta = \sin^2 \tfrac{1}{2}\alpha\cos^2\varphi,$$

and

$$d\tfrac{1}{2}\theta = \frac{\sin\frac{1}{2}\alpha\cos\varphi\, d\varphi}{\sqrt{1 - \sin^2 \frac{1}{2}\alpha\sin^2\varphi}},$$

so it follows that

$$\omega t = \int_0^{\varphi} \frac{d\varphi}{\sqrt{1 - \sin^2 \frac{1}{2}\alpha\sin^2\varphi}}.$$

Now, if we replace $\sin \frac{1}{2}\alpha$ by k, then we have the elliptic integral of the first kind

$$F(\varphi, k) = \int_0^{\varphi} \frac{\mathrm{d}\varphi}{\sqrt{1 - k^2 \sin \varphi}},$$

where k is the modulus, and φ the amplitude. Therefore, we have

$$k = AD/AB.$$

It follows that we can define the Jacobi elliptic functions by

$$\varphi = \operatorname{am} \omega t, \quad \cos \varphi = \operatorname{cn} \omega t, \quad \sin \varphi = \operatorname{sn} \omega t, \quad \dot{\varphi} = \omega \operatorname{dn} \omega t.$$

5.2. The period of the pendulum [2]

The period T of the pendulum is the time taken by the pendulum to oscillate between B and B' and then back to B. It follows that the quarter period $\frac{1}{4}T$ is the time of motion of P from A to B. As t increases from zero to $\frac{1}{4}T$, then θ increases from zero to α, and φ from zero to $\frac{1}{2}\pi$, so that ωt increases from zero to K, where K is the complete elliptic integral of the first kind, defined by

$$K = \int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}.$$

Hence, this value K is the quarter period of the Jacobi elliptic functions.

Now, $k = \sin \frac{1}{2}\alpha$, so in a similar way we let $k' = \cos \frac{1}{2}\alpha$, thus defining the complementary modulus. It follows that we have

$$k^2 + k'^2 = 1,$$

and a complementary quarter period K'.

5.3. The pendulum just reaches its highest position [3]

As α increases from zero to π, the modulus k increases from zero to 1, while the quarter period increases from $\frac{1}{2}\pi$ to infinity.

In the case $k = 1$, the pendulum has just sufficient velocity to carry it to its highest position OD. However, this will take an infinite amount of time.

If we set $\alpha = \pi$, corresponding to $\angle ADP$, in (5.1), then we have

$$\dot{\theta}^2 = 4\omega^2 \cos \tfrac{1}{2}\theta,$$

and it follows that

$$\omega t = \int_0^{\theta} \sec \tfrac{1}{2}\theta \, \mathrm{d}\tfrac{1}{2}\theta.$$

[2] Greenhill (1892) 8–9.
[3] Greenhill (1892) 13–14.

If we let $\frac{1}{2}\theta = \varphi$, then

$$\operatorname{sn}\varphi = \tanh\omega t, \quad \cos\varphi = \operatorname{sech}\omega t, \quad \dot{\varphi} = 2\omega\operatorname{sech}\omega t.$$

Hence, when $k = 1$, the Jacobi elliptic functions degenerate to hyperbolic functions, namely

$$\sin\omega t = \tanh\omega t, \quad \operatorname{cn}\omega t = \operatorname{dn}\omega t = \operatorname{sech}\omega t.$$

5.4. The pendulum makes complete revolutions [4]

The situation in the previous section corresponds to the pendulum having an angular velocity $\omega = \sqrt{g/l}$ at its lowest position, and as it would just reach its highest position, it follows that if ω is increased further, then the pendulum will make complete revolutions.

If we again let $\frac{1}{2}\theta = \varphi$, equivalent to $\angle ADP$, then we have

$$l^2\dot{\varphi}^2 = g(R - l\sin^2\varphi).$$

where R is the radius of the circle traced out by the particle. It follows that

$$\dot{\varphi}^2 = \frac{gR}{l^2}\left\{1 - \frac{l}{R}\sin^2\varphi\right\}.$$

By putting $k^2 = l/R$ and $\omega^2 = g/l$, we have

$$\dot{\varphi}^2 = \frac{\omega^2}{k^2}(1 - k^2\sin^2\varphi),$$

and, by integration, we obtain

$$\frac{\omega t}{k} = \int_0^t \frac{\mathrm{d}t}{\sqrt{1 - k^2\sin^2\varphi}}.$$

Therefore,

$$\varphi = \operatorname{am}(\omega t/k), \quad \sin\varphi = \operatorname{sn}(\omega t/k), \quad \cos\varphi = \operatorname{cn}(\omega t/k), \quad \dot{\varphi} = 2(\omega/k)\operatorname{dn}(\omega t/k).$$

[4] Greenhill (1892) 18.

Chapter 6

Halphen's Circles and Poncelet's Polygons

As illustrated in Figure 6.1, we consider a circle S and a point P inside that circle. If we draw chords through that point and consider segments of these with length inversely proportional to the square root of the total length of the chord, then we obtain a closed, convex curve C that will allow us to define the Jacobi elliptic functions.

6.1. A circle and a closed curve [1]

We denote the endpoints of one of the chords by M and M', and let N and N' be the points of intersection of this chord with C. These are symmetric about P. We also have the maximum diameter N_0N_0' of C corresponding to the chord M_1M_1', and the minimum diameter of C corresponding to the chord M_1M_1'.

If we consider a moving line PN, then we define the angle between this and its initial position PN_0 to be increasingly positive as it moves anti-clockwise about the circle S, and negative if it moves in the opposite direction. Similarly, we take the area enclosed by the section of C limited by PN and PN_0 to be positive if the angles between the two lines is positive, and negative otherwise. If the position of PN as it moves around S completes a full circle, we continue to add to the area in the obvious way to obtain that the area is a function taking values on the whole real line.

Now, let R be the radius of S, and δ the distance OP from the centre of S to the

---

Georges-Henri Halphen (1844–1889) was a soldier in the French army, reaching the rank of Commander. He was conferred the title *Chevalier de la Legion d'Honneur* after the battle of Pont-Noyelles. Much of Halphen's work in mathematics was considered to be ahead of its time, and so he did not initially receive recognition from his peers. However, he was later elected to *l'Académie des Sciences de Paris*. Despite his relatively short life, and his primary career in the military, Halphen was a prolific author, his *Œuvres* being published in four volumes totaling almost 2500 pages. His treatment of Poncelet's poristic polygons appears predominantly in the second volume of his work *Traité des fonctions elliptiques et de applications* (Halphen, 1886), a series that was sadly left incomplete by his premature death. Greenhill was particularly influenced by Halphen. In an obituary of Greenhill in the *Journal of the London Mathematical Society* (Love, 1928) it was claimed that

> *At a later date, when Halphen's Traité des fonctions elliptiques was published, he devoured it avidly, and made it his constant companion, or, so to say, his bible.*

[1] Halphen (1886) **1** 1–4.

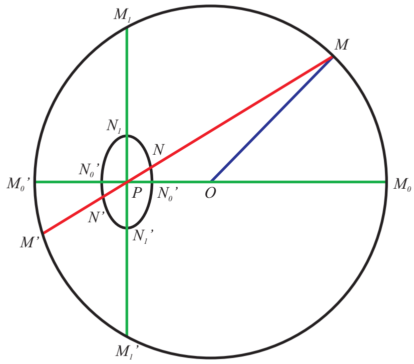

Figure 6.1. A circle and a closed curve.

point P. It follows that

$$PN = l\sqrt{\frac{2(R+\delta)}{MM'}},$$

for an arbitrary length l. Suppose we define the argument u by

$$u = \frac{\text{area } N_0PN}{l^2},$$

and define φ to be $\frac{1}{2}\angle M_0OM$. It follows that φ is a function of u, which we denote by

$$\tfrac{1}{2}\angle M_0OM = \varphi = \operatorname{am} u.$$

This is the amplitude of u.

Suppose we denote the ratio of the total area enclosed by the C and l^2 by $2K$, then it follows from the definition of the amplitude that

$$\operatorname{am} 0 = 0, \quad \operatorname{am} K = \tfrac{1}{2}\pi, \quad \operatorname{am} 2K = \pi,$$
$$\operatorname{am}(2K+u) = \pi + \operatorname{am} u, \quad \operatorname{am}(-u) = -\operatorname{am} u.$$

6.2. Eccentricity of the curve [2]

When the point P is chosen such that it coincides with the centre of the circle S, then the curve C reduces to a circle with radius l. Therefore, the function $\operatorname{am} u$ is simply equal to u. In general $\operatorname{am} u$ differs from u in proportion to the eccentricity of C. This eccentricity is the modulus k defined by

$$k = \frac{2\sqrt{R\delta}}{R+\delta}.$$

Similarly, we have the complementary modulus

$$k' = \frac{R-\delta}{R+\delta}.$$

Clearly, it follows that $k^2 + k'^2 = 1$, and that k and k' take values in the interval from zero to 1.

[2] Halphen (1886) **1** 3–4.

6.3. The Jacobi elliptic functions 3

We can define the Jacobi elliptic functions by the sine and cosine of $\operatorname{am} u$, and by the distance of the point P from M which corresponds to the argument u on the circle S. It follows that we have

$$\operatorname{sn} u = \sin \operatorname{am} u, \quad \operatorname{cn} u = \cos \operatorname{am} u, \quad \operatorname{dn} u = \frac{PM}{PM_0} = \frac{PM}{R+\delta}.$$

From the properties of the circular functions $\sin\theta$ and $\cos\theta$, it follows that $\operatorname{sn} u$ and $\operatorname{cn} u$ clearly take values between -1 and 1 and satisfy the identity

$$\operatorname{sn}^2 u + \operatorname{cn}^2 u = 1. \tag{6.1}$$

The function $\operatorname{dn} u$ is, by definition, always positive, and its maximum value is 1, corresponding to the line PM_0, while its minimum corresponds to PM_0' and is equal to k'.

Now, the angle $\angle PM_0M$ is equal to $\frac{1}{2}\pi - \varphi$, and its cosine is equal to $\operatorname{sn} u$. Moreover, the length of the chord M_0M can be expressed as $2R\sin\varphi$, or equivalently $2R \operatorname{sn} u$.

By the cosine formula of plane trigonometry we have 4

$$PM^2 = PM_0^2 + MM_0 - 2PM_0.MM_0 \cos \angle PM_0M.$$

Hence,

$$\operatorname{dn}^2 u = 1 - \frac{4R\delta}{(R+\delta)^2} \operatorname{sn}^2 u,$$

and, from the definition of the modulus, we have

$$\operatorname{dn}^2 u + k^2 \operatorname{sn}^2 u = 1. \tag{6.2}$$

Eliminating $\operatorname{sn}^2 u$ from (6.1) and (6.2), we have

$$\operatorname{dn}^2 u - k^2 \operatorname{cn} u = k'^2.$$

6.4. Quarter and half periods 5

From Figure 6.1, we immediately have

$$\operatorname{dn} \tfrac{1}{2}K = \frac{CM_1}{CM_0} = \frac{\sqrt{R^2-\delta^2}}{R+\delta} = \sqrt{\frac{R-\delta}{R+\delta}} = \sqrt{k'},$$

and it follows from (6.2) and (6.2) that

$$\operatorname{sn} \tfrac{1}{2}K = \frac{1}{\sqrt{1+k'}}, \quad \operatorname{cn} \tfrac{1}{2}K = \sqrt{\frac{k'}{1+k'}}.$$

[3] Halphen (1886) **1** 4–5.

[4] *The cosine formula. If a, b and c are the sides of a plane triangle, and α, β and γ are their opposite angles, then*

$$c^2 = a^2 + b^2 - 2ab\cos\alpha.$$

A further discussion of this can be found in Section 9.1.

[5] Halphen (1886) **1** 6–7.

Now, by the symmetry of the curve C, all lines passing through its centre P divide the area enclosed by this curve into two equal parts. If u is the argument corresponding to a point N on C. then the point N' corresponds to the argument $u+K$. As $PM.PM' = R^2 - \delta^2$, we have

$$(R+\delta)^2 \operatorname{dn} u \operatorname{dn}(u+K) = R^2 - \delta^2,$$

and from the definition of k', we obtain

$$\operatorname{dn}(u+K) = k' \operatorname{nd} u.$$

On the triangle PM_0M, we have that the angle at M_0 is $\frac{1}{2}\pi - \varphi$ and its sine is $\operatorname{cn} u$. Moreover, the angle M is equal to the half angle measured by the arc M_0M' on the circle. Therefore, the amplitude corresponding to the point N' is $\operatorname{am}(u+K)$. Now, the sine of the angle at M is $\operatorname{sn}(u+K)$, and so we have

$$\frac{PM}{PM_0} = \frac{\operatorname{cn} u}{\operatorname{sn}(u+K)}.$$

By definition, the left hand side is equal to $\operatorname{dn} u$, and it simply follows that

$$\operatorname{sn}(u+K) = \operatorname{cd} u.$$

Replacing u by $u+K$, we obtain

$$\operatorname{cn}(u+K) = -k' \operatorname{sd} u.$$

Similarly, by substituting $-u$ in place of u, we have [6]

$$\operatorname{dn}(K-u) = k' \operatorname{nd} u, \quad \operatorname{sn}(K-u) = \operatorname{cd} u, \quad \operatorname{cn}(K-u) = k' \operatorname{sd} u.$$

6.5. Derivatives of the Jacobi elliptic functions [7]

Suppose θ is the angle of the variable line PN with the initial line PN_0. The derivative with respect to θ of the area bounded by these lines and the curve C is $\frac{1}{2}PN^2$. From the definitions of u and PN, we have

$$\frac{\mathrm{d}u}{\mathrm{d}\theta} = \frac{R+\delta}{MM'}.$$

Now, given φ and φ', the half angles $\frac{1}{2}\angle M_0OM$ and $\frac{1}{2}\angle M_0OM'$. We obtain from the properties of the circle,

$$\theta = \tfrac{1}{2}(2\varphi + 2\varphi' - \pi).$$

Therefore, it follows that

$$\mathrm{d}\theta = \mathrm{d}\varphi + \mathrm{d}\varphi'.$$

Suppose that we have a second line PN_1 which intersects S at M_1 and M_1', then we obtain

$$\frac{MM_1}{M'M_1'} = \frac{PM}{PM_1'}.$$

---

[6] Jacobi referred to $K-u$ as the *complementary argument*, and its amplitude as the *coamplitude* of u, denoted $\operatorname{coam} u$. Similarly he adopted the notation $\operatorname{sn}(K-u) = \sin\operatorname{coam} u$, $\operatorname{cn}(K-u) = \cos\operatorname{coam} u$ and $\operatorname{dn}(K-u) = \Delta\operatorname{coam} u$.

[7] Halphen (1886) **1** 7–9.

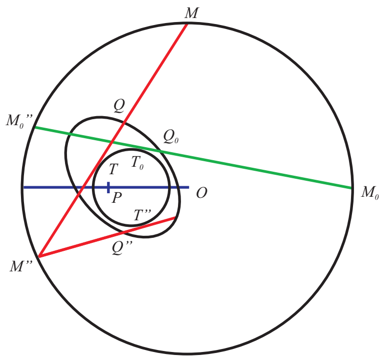

Figure 6.2. A theorem of Jacobi.

If M_1 is infinitely close to M, then it follows that

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\varphi'} = \frac{PM}{PM'},$$

$$\frac{\mathrm{d}\varphi + \mathrm{d}\varphi'}{\mathrm{d}\varphi} = \frac{MM'}{PM} = \frac{MM'}{(R+\delta)\,\mathrm{dn}\,u}.$$

But

$$\frac{\mathrm{d}u}{\mathrm{d}\varphi + \mathrm{d}\varphi'} = \frac{R+\delta}{MM'},$$

so by the chain rule, and by replacing φ with $\mathrm{am}\,u$, we have

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\mathrm{am}\,u\right\} = \mathrm{dn}\,u.$$

It follows from the derivatives of $\sin\theta$ and $\cos\theta$ that

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\mathrm{sn}\,u\right\} = \mathrm{cn}\,u\,\mathrm{dn}\,u, \qquad \frac{\mathrm{d}}{\mathrm{d}u}\left\{\mathrm{cn}\,u\right\} = -\,\mathrm{sn}\,u\,\mathrm{dn}\,u,$$

and by application of (6.2) that

$$\frac{\mathrm{d}}{\mathrm{d}u}\left\{\mathrm{dn}\,u\right\} = -k^2\,\mathrm{sn}\,u\,\mathrm{cn}\,u.$$

6.6. A theorem of Jacobi [8]

We consider a circle U inside a second circle S, with radius passing through the point P. On U, we take the tangent from a variable point T. This moving tangent intersects S at a point M corresponding to T, as shown in Figure 6.2.

Now, let T_0 be the initial point of T, and corresponding to M_0. Then we have

$$\frac{TM}{T_0M_0} = \frac{PM}{PM_0} = \mathrm{dn}\,u.$$

[8] Halphen (1886) **1** 10–13.

If we construct a curve C in the same way as in the preceding sections, then we have the length

$$PN = l\sqrt{\frac{PM_0}{MM'}}.$$

Denoting by M'' the second point of intersection of TM with S, and taking the length TQ on the tangent TM given by

$$TQ = l\sqrt{\frac{2T_0M_0}{MM''}},$$

then the point Q traces out a closed convex curve D exterior to the circle U.

We now consider the area between D and U, limited by the initial tangent T_0Q_0 and the tangent TQ moving around the circle. Using the same conventions as before, this new area takes values on the whole real line. We can show that this area is equivalent to that enclosed by the earlier curve C and the lines PNM and PN_0M_0.

We define the argument u_1 by

$$u_1 = \frac{\text{area } Q_0T_0TQ}{l^2},$$

and let θ_1 be the angle between TM and PM_0. The derivative of this new area with respect to θ_1 is $\frac{1}{2}TQ^2$. Therefore,

$$\frac{\mathrm{d}u_1}{\mathrm{d}\theta_1} = \frac{TM_0}{MM''}.$$

If we denote the angles $\angle MOM_0$ and $\angle M''OM_0$ by 2φ and $2\varphi''$, then we have

$$\mathrm{d}\theta_1 = \mathrm{d}\varphi + \mathrm{d}\varphi'', \quad \frac{\mathrm{d}\varphi}{\mathrm{d}\varphi''} = \frac{TM}{TM''}, \quad \frac{\mathrm{d}\theta_1}{\mathrm{d}\varphi} = \frac{MM''}{TM}.$$

It follows that

$$\frac{\mathrm{d}\varphi}{\mathrm{d}u_1} = \frac{TM}{T_0M_0} = \operatorname{dn} u.$$

As we have

$$\frac{\mathrm{d}\varphi}{\mathrm{d}u} = \operatorname{dn} u,$$

then

$$\frac{\mathrm{d}u_1}{\mathrm{d}u} = 1,$$

and $u_1 = u$.

Now, let $M''T''$ be a second tangent to the circle U. Its point of contact T'' corresponds to M'' on the outer circle S. But the tangents $M''T$ and $M''T''$ are equal. Therefore, as

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\varphi''} = \frac{TM}{TM''},$$

we obtain

$$\frac{\mathrm{d}\varphi}{TM} = \frac{\mathrm{d}\varphi''}{T''M''}.$$

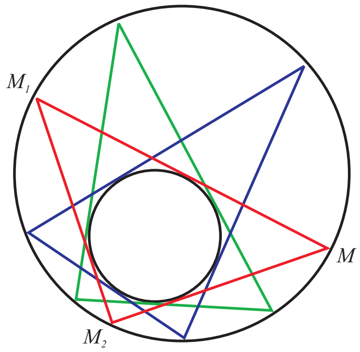

Figure 6.3. Poristic triangles.

Letting u'' be the argument corresponding to M'' in an analogous way to u corresponding to M, it follows that

$$\frac{\mathrm{d}\varphi}{TM} = \frac{\mathrm{d}u}{T_0 M_0}, \qquad \frac{\mathrm{d}\varphi''}{T''M''} = \frac{\mathrm{d}u''}{T_0 M_0}.$$

Hence, $\mathrm{d}u = \mathrm{d}u''$.

In the case of the circle reducing to the point P, then M'' is replaced by M', and so we have $\mathrm{d}u = \mathrm{d}u'$. As a result of the symmetry of C, we obtain $u' = u + K$.

The difference $u'' - u$ is also constant, but it is not equal to K. If we consider the position of the initial tangent $M_0 T_0 M_0''$, then the constant is the argument u_0'' corresponding to the point M_0''. Therefore, we have the following result of Jacobi:

THEOREM 6.1. *Let U be a circle in the interior of a second circle S, and let P be a point not lying on the radical axis. If we draw a variable chord MM'' tangent to the* 9
circle U, then the difference between the arguments u and u'' corresponding to M and M'' is constant.

6.7. Poncelet's poristic polygons 10

We consider the triangle MM_1M_2 shown in Figure 6.3, which is inscribed in a circle S and circumscribed around a circle U. The arguments of these three vertices, taken successively, are u, $u + u_0''$ and $u + 2u_0''$. As the chord M_2M is tangent to the circle U, the area enclosed by the curve is completely described, and so we have that the argument of the point M corresponds to the arguments

$$u + 2K, \quad u + 3u_0''.$$

Therefore,

$$u_0'' = \tfrac{2}{3}K.$$

[9]If O is the centre of a circle of radius r, and if Q is any point, then $OP^2 - r^2$ is called the *power* of Q with respect to the circle. Moreover, the locus of a point Q which moves so that its powers with respect to two circles are equal is called the *radical axis* of the circle.

[10]Halphen (1886) **1** 3–4 and Halphen (1886) **2** 367–412. See also Poncelet (1865). Greenhill (1892) derives Poncelet's porism using the properties of a simple pendulum.

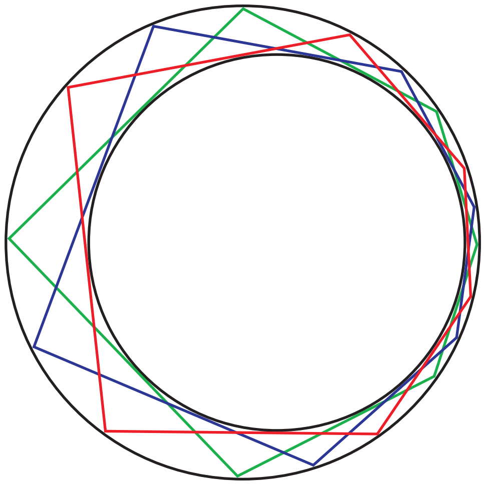

Figure 6.4. Poristic hexagons.

It follows that the circle U is determined by the condition that the point M_0'' corresponds to the argument $\frac{2}{3}K$. Hence, each point M on S is a vertex of a triangle inscribed in S and circumscribed around U.

Now, instead of a triangle, suppose we have a polygon with p sides, inscribed in S and circumscribed around U. In this case, we may also have that the polygon is a star polygon formed by connecting with straight lines every qth point out of the p. It follows that for a convex polygon we would have $q = 1$. Therefore, if u is the argument of the first vertex, then in a similar way to the triangle, we have equivalent expressions

$$u + 2qK, \quad u + pu_0'',$$

and we obtain

$$u_0'' = \frac{2qK}{p}.$$

This allows us to state a result discovered by Poncelet.

THEOREM 6.2. *Poncelet's porism. Given two circles, one inscribed in and the other circumscribed around a closed polygon, then there exists an infinite number of polygons with the same number of sides, circumscribed around the first circle, and inscribed in the other.*

Like Halphen, Jean-Victor Poncelet (1788–1867), fought in the French army, though as an engineer, and was part of Napoléon's ill-fated march on Russia in 1812. Having been abandoned as dead after the Battle of Krasnoy, he was imprisoned for two years. During this time, Poncelet studied projective geometry, and wrote his first book *Applications d'analyse et de la géométrie*. This work was originally intended as an introduction for his most famous work *Traité des propriétés projectives des figures* (Poncelet, 1865), however, in common with all his works, it soon expanded to fill two full volumes with numerous diagrams. Although one of the founders of modern projective geometry, Poncelet spent much of his life concerned with mechanical and, naturally, military applications of mathematics, and proposed improvements to turbines and water wheels.

Chapter 7

Fagnano's Ellipses

An *ellipse* is a closed curve formed from the intersection of a circular cone and a plane not parallel to the base of the cone. It can also be considered to be the path of a point moving in a plane such that the ratio of its distances from a fixed point, the *focus*, and a fixed line, the *directrix*, is a constant less than one. As such a path has the same properties with respect to a second fixed point and a second fixed line, ellipses are said to have two foci and two directrices. The ratio of the distances is known as the *eccentricity*. We take the ellipse to be represented in general by the equation

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

where a and b are its axes. An ellipse is symmetric about both its axes, and, of course, if the axes are equal, we have a circle.

7.1. Fagnano's theorem on arcs of an ellipse

In the case of the circle it is easy to construct an arc whose length is equal to the sum of the lengths of two other arcs of the same circle. This is linked to the ability to express $\sin(\theta_1 + \theta_2)$ in terms of $\sin\theta_1$ and $\sin\theta_2$ by

$$\sin(\theta_1 + \theta_2) = \sin\theta_1\sqrt{1 - \sin^2\theta_2} + \sin\theta_2\sqrt{\sin^2\theta_1}.$$

In Section 4.3, we have seen that there exists an addition formula for the elliptic integral of the second kind, and it is in terms of this that the arc length of an ellipse is expressed.

Fagnano was considered a gifted child, and by 14 was studying theology and philosophy at the college of Clementine in Rome. While he was there he assiduously avoided mathematics despite the encouragement of mathematician Quateroni. However, while studying the work *De la recherche de la vérité* of the philosopher and mathematician Nicolas Malebranche (1638–1715), Fagnano recognized the need to study mathematics, and abandoned philosophy.

In 1743, when it was feared that the dome of St. Peter's Basilica in Rome was in danger of collapse, Fagnano was appointed as the engineer and architect tasked with stabilizing its structure. In reward of his efforts, Pope Benedict XIV commissioned the publication of Fagnano's works, which first appeared in 1750.

Fagnano is buried in Senigallia on the Adriatic coast of Italy, north of Ancona. His tombstone begins with the words

Veritas Deo ∞ *gloria...*

In recognition of his most famous works, there is a portrait of Fagnano in the town hall of Senigallia. In one hand he holds a lemniscate.

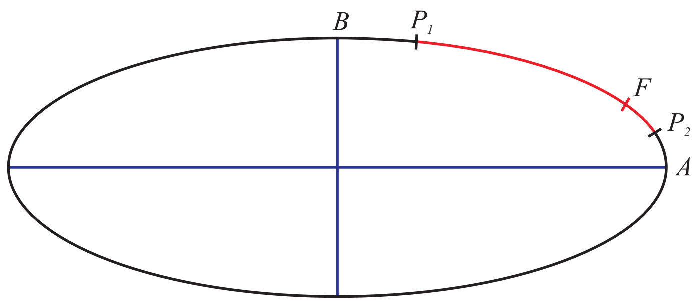

Figure 7.1. Fagnano's theorem on arcs of an ellipse.

THEOREM 7.1. *Fagnano's theorem. If $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ are two points on the ellipse*

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

whose eccentric angles φ_1, φ_2 are such that

$$\tan\varphi_1 \tan\varphi_2 = \frac{b}{a}, \tag{7.1}$$

then

$$\operatorname{arc} BP_1 + \operatorname{arc} BP_2 - \operatorname{arc} BA = \frac{k^2 x_1 x_2}{a}.$$

Proof (Bowman, 1953). We consider an ellipse with eccentricity k and parameterize the points P_1, P_2 by [1]

$$x_1 = a \operatorname{sn} u, \quad y_1 = b \operatorname{cn} u, \quad x_2 = a \operatorname{sn} v, \quad y_2 = b \operatorname{cn} v, \tag{7.2}$$

where

$$\operatorname{sn} u = \sin\varphi_1, \quad \operatorname{sn} v = \sin\varphi_2. \tag{7.3}$$

Now,

$$\operatorname{sn}(u + K) = \operatorname{cd} u, \quad \operatorname{cn}(u + K) = -k' \operatorname{sd} u.$$

Hence, if we have $u + v = K$, then

$$\operatorname{sn} u = \operatorname{cd} v, \quad \operatorname{cn} u = k' \operatorname{sd} v.$$

Now, $k' = b/a$, so it follows that

$$bu = a \operatorname{cs} v.$$

Hence,

$$\operatorname{sc} u \operatorname{sc} v = a/b.$$

But, by applying (7.3), it follows that

$$\tan\varphi_1 \tan\varphi_2 = a/b.$$

Therefore, the case $u + v = K$ is equivalent to (7.1).

[1] Bowman (1953) 27.

Now, by (7.2), we have

$$\frac{\mathrm{d}x}{\mathrm{d}u} = a \operatorname{cn} u \operatorname{dn} u, \qquad \frac{\mathrm{d}y}{\mathrm{d}u} = -b \operatorname{sn} u \operatorname{dn} u.$$

Therefore,

$$\mathrm{d}x^2 + \mathrm{d}y^2 = (a^2 \operatorname{cn}^2 u \operatorname{dn}^2 u + b^2 \operatorname{sn}^2 u \operatorname{dn}^2 u)\, \mathrm{d}u = a^2 \operatorname{dn}^2 u \, \mathrm{d}u,$$

and, after integration, it follows that

$$\operatorname{arc} BP_1 = \int_0^u \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = a \int_0^u \operatorname{dn}^2 u \, \mathrm{d}u = aE(u). \tag{7.4}$$

Similarly,

$$\operatorname{arc} BP_2 = aE(v), \tag{7.5}$$

and as $\sin \frac{\pi}{2} = 1 = \operatorname{sn} K$,

$$\operatorname{arc} BA = aE(K). \tag{7.6}$$

Hence,

$$\operatorname{arc} BP_1 + \operatorname{arc} BP_2 - \operatorname{arc} BA = aE(u) + aE(v) - aE(K).$$

We have $u + v = K$, so

$$\operatorname{arc} BP_1 + \operatorname{arc} BP_2 - \operatorname{arc} BA = a\left\{E(u) + E(v) - E(u+v)\right\}.$$

Applying the addition formula for $E(u)$, it follows that

$$\operatorname{arc} BP_1 + \operatorname{arc} BP_2 - \operatorname{arc} BA = ak^2 \operatorname{sn} u \operatorname{sn} v \operatorname{sn}(u+v).$$

But $u + v = K$ and $\operatorname{sn} K = 1$, so

$$\operatorname{arc} BP + \operatorname{arc} BP' - \operatorname{arc} BA' = ak^2 \operatorname{sn} u \operatorname{sn} v.$$

By the parameterizations in (7.2) we have the result.

□

7.2. Fagnano's point

THEOREM 7.2. *If the points P_1 and P_2 coincide in Fagnano's point $F = (x, y)$, then*

$$\operatorname{arc} BF - \operatorname{arc} AF = a - b.$$

Proof (Lawden, 1989). Considering the same ellipse, and parameterizing F by (7.2), then clearly this point satisfies [2]

$$x = a \operatorname{sn} \tfrac{1}{2}K, \qquad y = b \operatorname{cn} \tfrac{1}{2}K. \tag{7.7}$$

By the duplication formulæ for $\operatorname{cn} u$, $\operatorname{dn} u$, we have

$$\frac{1 - \operatorname{cn} 2u}{1 + \operatorname{dn} 2u} = \operatorname{sn}^2 u.$$

[2] Lawden (1989) 99.

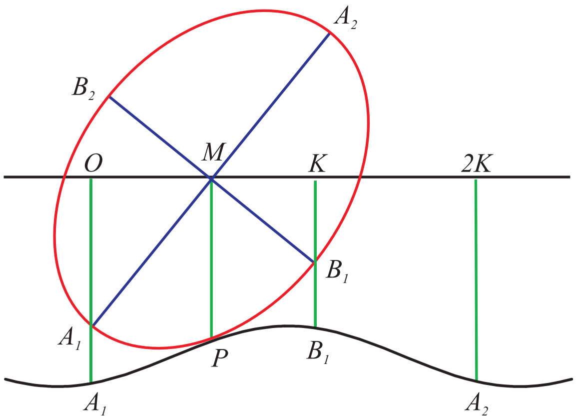

Figure 7.2. Rolling an ellipse.

It follows that

$$\operatorname{sn}\tfrac{1}{2}K = \sqrt{\frac{1-\operatorname{cn}K}{1+\operatorname{dn}K}} = \sqrt{\frac{1}{1+k'}}, \quad \operatorname{cn}\tfrac{1}{2}K = \sqrt{\frac{k'}{1+k'}}. \tag{7.8}$$

Hence, the coordinates of F are

$$x = \sqrt{\frac{a^3}{a+b}}, \quad y = \sqrt{\frac{b^3}{a+b}}.$$

We now apply Theorem 7.1 and (7.7), and obtain

$$\operatorname{arc} BF - \operatorname{arc} AF = \operatorname{arc} BF + \operatorname{arc} BF - \operatorname{arc} BA = \frac{k^2x^2}{a} = k^2 a \operatorname{sn}^2 \tfrac{1}{2}K.$$

As $k^2 + k'^2 = 1$ and $k' = b/a$, then, by applying (7.8), this reduces to the result.

□

7.3. Rolling an ellipse on a curve

THEOREM 7.3. *The curve on which the ellipse*

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

can roll such that its centre describes a straight line is

$$y = a\operatorname{dn}(x/b, k),$$

where $k = \sqrt{a^2 - b^2}/a$.

Proof (Greenhill, 1892). Consider the ellipse illustrated in Figure 7.2. The semiaxes [3]
of this ellipse are $A_1A_2 = a$ and $B_1B_2 = b$. Let M be the centre of the ellipse and

[3] Greenhill (1892) 71–73.

denote the straight line that it describes to be Ox. If P is the point of contact of the ellipse with the curve, then we must have that M lies vertically above P. Let $MP = r$ be the line joining these two points. If we also define θ to be the angle between MP and MA as shown, then we have that the polar equation of the ellipse is

$$\frac{1}{r^2} = \frac{\cos^2\theta}{a^2} + \frac{\sin^2\theta}{b^2}, \tag{7.9}$$

and, moreover, that

$$MG = -\frac{\mathrm{d}r}{\mathrm{d}\theta} = -y\frac{\mathrm{d}y}{\mathrm{d}x}.$$

After differentiation of (7.9), we have

$$-\frac{2}{r^3}\frac{\mathrm{d}r}{\mathrm{d}\theta} = \left\{\frac{1}{b^2} - \frac{1}{a^2}\right\} 2\sin\theta\cos\theta. \tag{7.10}$$

It also follows that

$$\frac{1}{r^2} - \frac{1}{a^2} = \left\{\frac{1}{b^2} - \frac{1}{a^2}\right\}\sin^2\theta,$$
$$\frac{1}{b^2} - \frac{1}{r^2} = \left\{\frac{1}{b^2} - \frac{1}{a^2}\right\}\cos^2\theta.$$

Therefore, (7.10) reduces to

$$-\frac{2}{r^3}\frac{\mathrm{d}r}{\mathrm{d}\theta} = 2\sqrt{\left\{\frac{1}{r^2} - \frac{1}{a^2}\right\}\left\{\frac{1}{b^2} - \frac{1}{r^2}\right\}},$$

and we have

$$-\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{r\sqrt{(a^2 - r^2)(r^2 - b^2)}}{ab}.$$

Hence, as $MP = r = y$,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\sqrt{(a^2 - y^2)(y^2 - b^2)}}{ab},$$
$$x = \int_y^a \frac{ab\,\mathrm{d}y}{\sqrt{(a^2 - y^2)(y^2 - b^2)}}.$$

If we now make the substitution $s^2 = (a^2 - y^2)/(a^2 - b^2)$, then we have an elliptic integral of the first kind, thus obtaining

$$x = b\,\mathrm{dn}^{-1}(y/a),$$

where $k = \sqrt{a^2 - b^2}/a$. The constant of integration vanishes as initially $x = 0$ and $y = a$. By inversion of the integral, we obtain the result.

□

Chapter 8

Bernoulli's Lemniscate

Men will fight long and hard for a bit of coloured ribbon

NAPOLÉON BONAPARTE 1

The *lemniscate* is a figure of eight shaped curve whose equation in polar coordinates is

$$r^2 = a^2 \cos 2\theta,$$

or in Cartesian coordinates,

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

It is a special case of the Cassinian ovals illustrated in Figure 8.1. A Cassinian oval is the locus of the intersection of a tangent to a conic and the perpendicular taken from the origin to the tangent. If the curve is a regular hyperbola, then we have a lemniscate.

8.1. Rectification of the lemniscate

The problem of calculating the arc length of the lemniscate provided one the first motivations for the study of elliptic integrals.

THEOREM 8.1. *The length of the arc connecting the origin and a point r on the lemniscate*

$$r^2 = \cos 2\theta \tag{8.1}$$

is given by the lemniscatic integral

$$s = \int_0^r \frac{\mathrm{d}t}{\sqrt{1 - t^4}}.$$

Proof (Prasolov & Solovyev, 1997). We define polar coordinates by 2

[1]Napoléon Bonaparte (1769–1821).

The appellation lemniscate stems from the Latin *lemniscatus* meaning decorated with ribbons. Its properties were first studied by the astronomer Giovanni Domenico Cassini (1625–1712) who considered more general curves, that now take his name, to describe the orbits of the planets more precisely than with ellipses. However, Cassini's work was only published, posthumously, in 1749, and so the lemniscate is more often associated with the 1694 papers of Jakob Bernoulli, and from which follows the name *Bernoulli's lemniscate*.

The lemniscatic integral first appeared in connection with the rectification of a curve known as the *curva elastica*. The term curva elastica has historical significance due to its appearance in the diary of Johann Carl Friedrich Gauss (1777–1855). In entry 51, Gauss starts to write *curvam elasticam*, but crosses through *elasticam* and replaces it with *lemniscatam*.

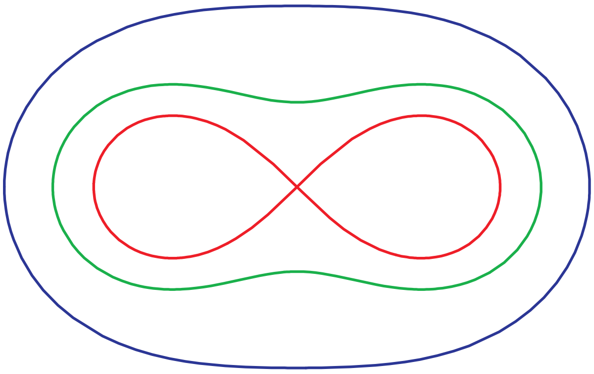

Figure 8.1. Cassinian ovals. The lemniscate is a Cassinian oval with a self intersection.

$$x = r\cos\theta, \qquad y = r\sin\theta, \tag{8.2}$$

and it follows that

$$\mathrm{d}x^2 + \mathrm{d}y^2 = (\cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta)^2 + (\sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta)^2 = \mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2.$$

Now, by (8.1), we have

$$2r\,\mathrm{d}r = -2\sin 2\theta\,\mathrm{d}\theta.$$

Hence,

$$\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 = \mathrm{d}r^2 + \frac{r^4\mathrm{d}r^2}{1-\cos^2 2\theta} = \frac{\mathrm{d}r^2}{1-r^4},$$

and, after integration, we obtain the result.

□

8.2. The lemniscate functions

If we denote the lemniscatic integral by

$$\varphi_1 = \int_0^x \frac{\mathrm{d}t}{\sqrt{1-t^4}}, \tag{8.3}$$

then we can express the relationship between φ_1 and x as

$$x = \mathrm{sl}\,\varphi_1$$

[2] Prasolov & Solovyev (1997) 78–79.

The notation $\mathrm{sl}\,\varphi$, $\mathrm{cl}\,\varphi$ and ω is that which appears in Gauss's notebooks of 1801 and 1808. The most significant part of his work in relation to the lemniscate is that he worked with an inverse function, a concept that had not previous been employed, and one that was rediscovered by Jacobi years later.

Gauss did not publish his on results on the lemniscate functions, and, consequently, elliptic functions, though it was mentioned by Jacobi in a letter to Legendre that he believed that Gauss already had in his possession many of the results that Jacobi himself published in 1827. Legendre, who was said to have had a particular dislike of Gauss, expressed serious doubt and incredulity that anyone could have such results, but leave them unpublished. It has been suggested that his decision not to publish stems from the disdainful reception that his *Disquisitiones Arithmeticæ* of 1801 received from *l'Académie des Sciences de Paris*, of whom Legendre was, of course, a member.

Circular functions	*Lemniscate functions*
$\varphi = \int_0^x \frac{\mathrm{d}t}{\sqrt{1-t^2}}$	$\varphi = \int_0^x \frac{\mathrm{d}t}{\sqrt{1-t^4}}$
$x = \sin\varphi$	$x = \mathrm{sl}\,\varphi$
$\frac{1}{2}\pi = \int_0^1 \frac{\mathrm{d}t}{\sqrt{1-t^2}}$	$\frac{1}{2}\omega = \int_0^1 \frac{\mathrm{d}t}{\sqrt{1-t^4}}$
$\cos\varphi = \sin(\frac{1}{2}\pi - \varphi)$	$\mathrm{cl}\,\varphi = \mathrm{sl}(\frac{1}{2}\omega - \varphi)$
$\sin\frac{1}{2}\pi = 1$	$\mathrm{sl}\,\frac{1}{2}\omega = 1$
$\sin\pi = 0$	$\mathrm{sl}\,\omega = 0$
$\sin(\pi + \varphi) = -\sin\varphi$	$\mathrm{sl}(\omega + \varphi) = -\,\mathrm{sl}\,\varphi$
$\sin(-\varphi) = -\sin\varphi$	$\mathrm{sl}(-\varphi) = -\,\mathrm{sl}\,\varphi$

Table 8.1. Comparison between the circular functions and the lemniscate functions.

where $\mathrm{sl}\,\varphi_1$ is known as the *lemniscate sine*. Suppose we also have

$$\begin{aligned}\varphi_2 &= \int_x^1 \frac{\mathrm{d}t}{\sqrt{1-t^4}},\\ \tfrac{1}{2}\omega &= \int_0^1 \frac{\mathrm{d}t}{\sqrt{1-t^4}},\end{aligned} \tag{8.4}$$

then we denote the *lemniscate cosine* by

$$x = \mathrm{cl}\,\varphi_2,$$

and it follows that we have the identity

$$\mathrm{sl}\,\varphi = \mathrm{cl}(\tfrac{1}{2}\omega - \varphi).$$

The constant ω is equal to the length of arc of one petal of the lemniscate, and has the same significance for the lemniscate as π does for the circle. The similarities between the circular and lemniscate functions are further illustrated in Table 8.1.

The lemniscate functions can also be expressed in term of elliptic functions by the equations

$$\begin{aligned}\mathrm{sl}\,\varphi &= \tfrac{1}{2}\sqrt{2}\,\mathrm{sd}(\sqrt{2}\varphi, k),\\ \mathrm{cl}\,\varphi &= \mathrm{cn}(\sqrt{2}\varphi, k),\end{aligned}$$

where $k = \frac{1}{2}\sqrt{2}$.

Moreover, $\frac{1}{2}\omega$ is the smallest positive value of φ for which

$$\mathrm{cn}(\sqrt{2}, k) = 0.$$

Therefore, it follows that

$$\omega = \sqrt{2}K(k), \tag{8.5}$$

and from this result we can express $K(k)$ in terms of gamma functions. [3]

[3]The *gamma function* $\Gamma(z)$ was first defined by Euler using an infinite integral, namely

$$\Gamma(z) = \int_0^\infty \mathrm{e}^{-t} t^{z-1}\,\mathrm{d}t,$$

when the real part of z is positive. The notation $\Gamma(z)$ was introduced by Legendre in 1814.

THEOREM 8.2. *The complete elliptic integral of the first kind with modulus* $k = \frac{1}{2}\sqrt{2}$ *can be expressed as*

$$K(k) = \frac{1}{4\sqrt{\pi}}\Gamma^2(\tfrac{1}{4}),$$

where $\Gamma(z)$ *is the gamma function.*

Proof (Whittaker & Watson, 1927). From (8.3) and (8.4) it follows that [4]

$$K(k) = \sqrt{2}\int_0^1 \frac{\mathrm{d}t}{\sqrt{1-t^4}}.$$

Making the substitution $s = t^4$, then we have

$$K(k) = \tfrac{1}{4}\sqrt{2}\int_0^1 \frac{\mathrm{d}s}{s^{\frac{3}{4}}\sqrt{1-s}},$$

or, in terms of gamma functions,

$$K(k) = \tfrac{1}{4}\sqrt{2}\frac{\Gamma(\frac{1}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{4})}.$$

Moreover, as $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(z)\Gamma(1-z) = \pi/\sin \pi z$, we have the result.

□

8.3. Fagnano's doubling of the lemniscate

The earliest instance of an addition formula related to elliptic integrals is attributed to Fagnano for his discovery of a formula to double an arc of the lemniscate. This formula also has the geometric context that we can double the arc of the lemniscate using ruler and compasses alone.

THEOREM 8.3. *An arc of the lemniscate can be doubled using the formula*

$$\int_0^a \frac{\mathrm{d}t}{\sqrt{1-t^4}} = 2\int_0^c \frac{\mathrm{d}t}{\sqrt{1-t^4}},$$

where

$$c = \frac{2a\sqrt{1-a^4}}{1+a^4}.$$

Proof (Nekovář, 2004). We first make the substitution $a = (1+\mathrm{i})b/\sqrt{1-b^4}$ in (8.3), [5]
and it follows that

$$\int_0^a \frac{\mathrm{d}a}{\sqrt{1-a^4}} = (1+\mathrm{i})\int_0^b \frac{\mathrm{d}b}{\sqrt{1-b^4}}.$$

Next, if we let $b = (1-\mathrm{i})c/\sqrt{1-c^4}$, then we have

$$\int_0^b \frac{\mathrm{d}b}{\sqrt{1-b^4}} = (1-\mathrm{i})\int_0^c \frac{\mathrm{d}c}{\sqrt{1-c^4}}.$$

[4] Whittaker & Watson (1927) 524.

[5] Nekovář (2004) 16–18.

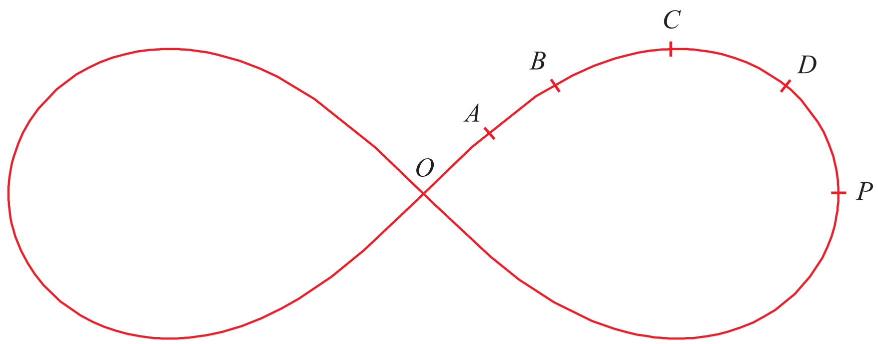

Figure 8.2. Division of the lemniscate into five equal parts.

By combining our two substitutions, it follows that

$$a = \frac{(1+\mathrm{i})(1-\mathrm{i})c}{\sqrt{(1-b^4)(1-c^4)}} = \frac{2c\sqrt{1-c^4}}{1+c^4},$$

and hence,

$$\int_0^a \frac{\mathrm{d}a}{\sqrt{1-a^4}} = 2\int_0^c \frac{\mathrm{d}c}{\sqrt{1-c^4}}.$$

□

We can reformulate Theorem 8.3 in terms of the lemniscate functions of Section 8.2.

THEOREM 8.4. *The doubling formula for the lemniscate sine is*

$$\mathrm{sl}\,2\varphi = \frac{2\,\mathrm{sl}\,\varphi\sqrt{1-\mathrm{sl}^4\varphi}}{1+\mathrm{sl}^4\varphi}.$$

8.4. Division of the lemniscate 6

One application of Theorem 8.3 made by Fagnano is in the division of a quadrant of the lemniscate. We consider the division into five parts, illustrated in Figure 8.2.

Let $OA = z$, $OB = u$, $OC = v$ and $OD = w$. If

$$u = \frac{2z\sqrt{1-z^4}}{1+z^4},$$

then arc $OB = 2\,\mathrm{arc}\,OA$. If

$$v = \frac{2u\sqrt{1-u^4}}{1-u^4},$$

then, similarly, arc $OC = 2\,\mathrm{arc}\,OB$. Finally, if

$$w = \sqrt{\frac{1-z^2}{1+z^2}},$$

then arc $OA = \mathrm{arc}\,PD$. Setting $v = w$ and eliminating u and z, we have 7

[6] Ayoub (1984) 144–145.

$$w^{24} + 50w^{20} - 125w^{16} + 300w^{12} - 105w^8 - 62w^4 + 5 = 0.$$

The polynomial $w^8 - 2w^4 + 5$ is a factor, and leaves

$$w^{16} + 52w^{12} - 26w^8 - 12w^4 + 1 = 0$$

which factors as

$$(w^8 + (26 - 12\sqrt{5})w^4 + 9 - 4\sqrt{5})(w^8 + (26 + 12\sqrt{5})w^4 + 9 + 4\sqrt{5}) = 0.$$

It follows that the real roots of this can be constructed solely with ruler and compasses.

8.5. Euler's addition formula

Theorem 8.4 was later extended by Euler to give an addition theorem for the lemniscate sine, which was then generalized further to give an addition formula for elliptic integrals.

THEOREM 8.5. *The addition formula for the lemniscate sine is*

$$\mathrm{sl}(\varphi + \psi) = \frac{\mathrm{sl}\,\varphi\sqrt{1 - \mathrm{sl}^4\,\psi} + \mathrm{sl}\,\psi\sqrt{1 - \mathrm{sl}^4\,\varphi}}{1 + \mathrm{sl}^2\,\varphi\,\mathrm{sl}^2\,\psi}.$$

Proof (Markushevich, 1992). Consider an equation of the form [8]

$$u^2 + v^2 + Au^2v^2 + 2Buv - C^2 = 0, \tag{8.6}$$

where A, B, C, are constants. By differentiation, we have

$$\Big\{u(1 + Av^2) + Bv\Big\}\,\mathrm{d}u + \Big\{v(1 + Au^2) + Bu\Big\}\,\mathrm{d}v = 0. \tag{8.7}$$

But

$$(1 + Av^2)u^2 + 2Buv + v^2 - C^2 = 0,$$

and it follows that

$$u(1 + Av^2) + Bv = \sqrt{C^2 + (AC^2 + B^2 - 1)v^2 - Av^4}. \tag{8.8}$$

Moreover, as (8.6) is symmetric in u and v, we can interchange these variables to similarly obtain

$$v(1 + Au^2) + Bu = \sqrt{C^2 + (AC^2 + B^2 - 1)u^2 - Au^4}. \tag{8.9}$$

[7] One of the many relations proven by Fagnano was that if $t = \sqrt{1 + z^2/1 - z^2}$, then

$$\int \frac{\mathrm{d}z}{\sqrt{1 - z^4}} = \int \sqrt{\frac{1 + z^2}{1 - z^2}}\,\mathrm{d}z + \int \frac{t^2}{\sqrt{t^4 - 1}}\,\mathrm{d}t - zt.$$

Geometrically, the integral on the left hand side is the arc length of a lemniscate, the first of the right hand side is the arc length of an ellipse, while the second is that of a rectangular hyperbola.

[8] Markushevich (1992) 2–4.

Therefore, by applying both (8.8) and (8.9) we have that (8.7) reduces to

$$\frac{\mathrm{d}u}{\sqrt{C^2 + (AC^2 + B^2 - 1)u^2 - Au^4}} + \frac{\mathrm{d}v}{\sqrt{C^2 + (AC^2 + B^2 - 1)v^2 - Av^4}} = 0. \tag{8.10}$$

It follows that (8.6) is a solution of this differential equation.

We now rewrite (8.10) as

$$\frac{\mathrm{d}u}{\sqrt{1 + mu^2 + nu^4}} + \frac{\mathrm{d}v}{\sqrt{1 + mv^2 + nv^4}} = 0, \tag{8.11}$$

by simply setting $B^2 = 1 + mC^2 + nC^4$. Therefore, (8.6) becomes

$$u^2 + v^2 - nC^2u^2v^2 + 2\sqrt{1 + mC^2 + nC^4}uv - C^2 = 0,$$

and it follows that

$$4(1 + mC^2 + nC^4)u^2v^2 = \left\{(u^2 + v^2) - (nu^2v^2 + 1)C^2\right\}^2.$$

If we let $1 + mt^2 + nt^4 = P(t)$, then

$$(1 - nu^2v^2)^2C^4 - 2\left\{u^2P(v) + v^2P(u)\right\}C^2 + (u^2 - v^2) = 0.$$

As $(1 - nu^2v^2)(u^2 - v^2) = u^2P(v) - v^2P(u)$, we have

$$C^2 = \frac{u^2P(v) + v^2P(u) + 2uv\sqrt{P(u)P(v)}}{(1 - nu^2v^2)^2},$$

from which we obtain

$$C = \frac{u\sqrt{P(v)} + v\sqrt{P(u)}}{1 - nu^2v^2}. \tag{8.12}$$

This an integral of (8.11) expressed in algebraic form. However, this integral can also be written in the transcendental form [9]

$$\int_0^u \frac{\mathrm{d}t}{\sqrt{P(t)}} + \int_0^v \frac{\mathrm{d}t}{\sqrt{P(t)}} = \alpha + \beta = \gamma, \tag{8.13}$$

where γ is a constant. Now, (8.12) must follow from (8.13). If we set $u = 0$ in (8.13), then

$$\int_0^v \frac{\mathrm{d}t}{\sqrt{P(t)}} = \gamma.$$

Moreover, (8.12) implies that $C = v$. Hence,

$$\int_0^u \frac{\mathrm{d}t}{\sqrt{P(t)}} + \int_0^v \frac{\mathrm{d}t}{\sqrt{P(t)}} = \int_0^C \frac{\mathrm{d}t}{\sqrt{P(t)}},$$

If we set $P(t) = 1 - t^4$, then we have the result.

□

[9] A *transcendental form* is one which cannot be expressed in terms of algebra.

Chapter 9

Spherical Trigonometry

Spherical trigonometry is the study of *spherical triangles*. These are formed by the intersection of three great circles on a sphere. The angles of a spherical triangle are [1] defined by the angle of intersection of the corresponding tangent lines to each vertex.

9.1. Formulæ of spherical trigonometry

Suppose ABC is a spherical triangle on a unit sphere. Let the sides BC, CA and AB be denoted by a, b and c, respectively, and their corresponding opposite angles by α, β and γ. Then we can state the following formula:

THEOREM 9.1. *The spherical cosine formula is*

$$\cos a = \cos b \cos c + \sin b \sin c \cos \alpha. \tag{9.1}$$

Proof (Smart, 1977). Let AD be the tangent at A to the great circle AB, and let AE be [2] the tangent at A to the great circle AC. The radius OA is perpendicular to AD and AE. As AD lies in the plane of the great circle AB, the radius OB intersects the tangent AD at D. Similarly, the radius OC meets the tangent AE at E. Now, the spherical angle $\alpha = \angle BAC$ is defined to be the angle between the tangents at A to the two great circles AB and AC. Therefore, $\alpha = \angle DAE$.

In the plane triangle OAD, we have $\angle BAC = \pi$ and $\angle AOD = \angle AOB = c$. It follows that

$$AD = OA \tan c, \quad OD = OA \sec c.$$

Similarly, from the plane triangle OAE, we have

$$AE = OA \tan b, \quad OE = OA \sec b.$$

Now, from the plane triangle DAE, we obtain

$$\begin{aligned} DE^2 &= AD^2 + AE^2 - 2AD.AE \cos \angle DAE \\ &= OA^2(\tan^2 c + \tan^2 b - 2 \tan b \tan c \cos \alpha), \end{aligned} \tag{9.2}$$

Until the sixteenth century, it was primarily spherical trigonometry that interested academics studying this branch of geometry, as a result of the popularity of astronomy within the natural sciences. The first reference to spherical triangles appeared in *Spherærica* in which Menelaus of Alexandria (*c* 70–130 AD) developed spherical analogues of the results of plane triangles found in *The Elements* of Euclid of Alexandria (*c* 325–265 BC).

[1] A *great circle* is the circle cut out on the surface of the sphere by a plane through the centre of the sphere.

[2] Smart (1977) 6–8.

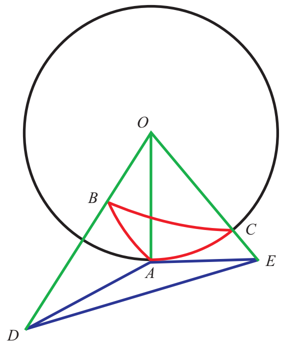

Figure 9.1. A spherical triangle.

and, from DOE, it follows that

$$DE^2 = OD^2 + OE^2 - 2OD.OE\cos\angle DOE.$$

But $\angle DOE = \angle BOC = a$, and so

$$DE^2 = OA^2(\sec^2 c + \sec^2 b - 2\sec b - 2\sec b\sec c\cos a). \tag{9.3}$$

Hence, by combining (9.2) and (9.3), we have

$$\sec^2 c + \sec^2 b - 2\sec b\sec c\cos a = \tan^2 c + \tan^2 b - 2\tan b\tan c\cos\alpha.$$

By application of the identity $\sec^2\theta = 1 + \tan^2\theta$, and after simplification, we have the result.

□

The spherical cosine formula can be seen as the fundamental formula of spherical trigonometry as from it we can derive each of the other formulæ of spherical trigonometry.

THEOREM 9.2. *The spherical sine formula is*

$$\frac{\sin a}{\sin\alpha} = \frac{\sin b}{\sin\beta} = \frac{\sin c}{\sin\gamma} = k, \tag{9.4}$$

where $k < 1$ is constant.

Proof (Smart, 1977). By (9.1), we have [3]

$$\sin b\sin c\cos\alpha = \cos a - \cos b\cos c.$$

If we square each side, and write the left hand side in the form

$$\sin^2 b\sin^2 c - \sin^2 b\sin^2 c\sin^2\alpha,$$

or, alternatively, in the form

$$1 - \cos^2 b - \cos^2 c + \cos^2 b\cos^2 c - \sin^2 b\sin^2 c\sin^2\alpha,$$

[3] Smart (1977) 9–10.

it follows that

$$\sin^2 b \sin^2 c \sin^2 \alpha = 1 - \cos^2 a - \cos^2 b - \cos^2 c + 2 \cos a \cos b \cos c. \qquad (9.5)$$

Let k be a positive constant defined by

$$\sin^2 a \sin^2 b \sin^2 c = k^2(1 - \cos^2 a - \cos^2 b - \cos^2 c + 2 \cos a \cos b \cos c).$$

Then, by (9.5), we have

$$\frac{\sin^2 a}{\sin^2 \alpha} = k^2.$$

Hence,

$$k = \pm\frac{\sin a}{\sin \alpha}.$$

But in a spherical triangle the lengths of each side are less than π, as are the angles. Therefore, we take the positive sign.

By a similar argument, we have

$$k = \frac{\sin b}{\sin \beta} = \frac{\sin c}{\sin \gamma}.$$

□

THEOREM 9.3. *The analogue formula is*

$$\sin a \cos \beta = \cos b \sin c - \sin b \cos c \cos \alpha. \qquad (9.6)$$

Proof (Smart, 1977). Applying formulæ analogous to (9.1), we have [4]

$$\begin{aligned} \sin c \sin a \sin \beta &= \cos b - \cos c \cos a \\ &= \cos b - \cos c(\cos b \cos c + \sin b \sin c \cos \alpha) \\ &= \sin^2 c \cos b - \sin b \sin c \cos c \cos \alpha. \end{aligned}$$

The result follows by dividing by $\sin c$.

□

THEOREM 9.4. *The four parts formula is*

$$\cos a \cos \gamma = \sin a \cot b - \sin \gamma \cot \beta. \qquad (9.7)$$

Proof (Smart, 1977). Again using analogues of (9.1), we have [5]

$$\begin{aligned} \cos b &= \cos a \cos c + \sin a \sin c \cos \beta \\ &= \cos a(\cos b \cos a + \sin b \sin a \cos \gamma) + \sin a \sin c \cos \beta. \end{aligned}$$

Therefore,

$$\cos b \sin^2 a = \cos a \sin b \sin a \cos \gamma + \sin a \sin c \cos \beta.$$

Dividing by $\sin a$ and $\sin b$, it follows that

$$\cot b \sin a = \cos a \cos \gamma + \frac{\sin c}{\sin b} \cos \beta,$$

and by application of (9.4), we have the result.

□

[4] Smart (1977) 10–11.
[5] Smart (1977) 12.

We consider the polar triangle of ABC. If we denote this by $A'B'C'$, the sides by [6]
a', b' and c', and the opposite angles by α', β' and γ', then [7]

$$\alpha' = \pi - a, \quad \beta' = \pi - b, \quad \gamma' = \pi - c,$$

and similarly

$$a' = \pi - \alpha, \quad b' = \pi - \beta, \quad c' = \pi - \gamma.$$

Then we have the polar formulæ corresponding to (9.1) and (9.6).

$$\begin{aligned} \cos\alpha' &= -\cos\beta'\cos\gamma' + \sin\beta'\sin\gamma'\cos a', \\ \sin\alpha'\cos b' &= \cos\beta'\sin\gamma' + \sin\beta'\cos\gamma'\cos a'. \end{aligned} \tag{9.8}$$

Suppose we have now have a sphere of radius r. The formulæ of spherical trigonometry still hold, but we replace each of a, b and c by a/r, b/r and c/r, respectively. If we let $1/r = l$, and denote the angles of this spherical triangle by α_l, β_l and γ_l, then we have

$$\cos la = \cos lb\cos lc + \sin lb\sin lc\cos\alpha_l, \tag{9.9}$$

$$\frac{\sin la}{\sin\alpha_l} = \frac{\sin lb}{\sin\beta_l} = \frac{\sin lc}{\sin\gamma_l}. \tag{9.10}$$

If we subtract 1 from each side of (9.9) and then divide by l^2, it follows that by letting l tend to zero, we have the cosine formula of plane trigonometry, namely

$$c^2 = a^2 + b^2 - 2ab\cos\alpha_0.$$

By dividing both sides of (9.10) by k and similarly letting l tend to zero, then we have the sine formula of plane trigonometry, which can be stated as

$$\frac{a}{\sin\alpha_0} = \frac{b}{\sin\beta_0} = \frac{c}{\sin\gamma_0}.$$

9.2. Elliptic measures of the angles of a spherical triangle

We introduce elliptic functions to spherical trigonometry by defining the *elliptic measures* of the angles α, β and γ. If we denote these by u, v and w, then they are given by the equations

$$\operatorname{sn} u = \sin\alpha, \quad \operatorname{sn} v = \sin\beta, \quad \operatorname{sn} w = \sin\gamma.$$

The modulus of each of the elliptic functions is taken to be k, as defined in (9.4).

THEOREM 9.5. *The elliptic measures of the angles of a spherical triangle always sum to* $2K$.

Proof (Lawden, 1989). We can choose each measure to be in the interval from zero to [8]

[6] Smart (1977) 15.

[7] Suppose ABC is a spherical triangle. The great circle of which BC is an arc has two poles, one in each of the hemispheres into which it divides the sphere. If we denote by A' the pole in the hemisphere in which A lies, and similarly denote by B' and C' the appropriate poles of CA and AB, then $A'B'C'$ is the *polar triangle* of ABC.

K. Then we have

$$\alpha = \operatorname{am} u, \quad \beta = \operatorname{am} v, \quad \gamma = \operatorname{am} w,$$

and it follows that

$$\operatorname{cn} u = \cos\alpha, \quad \operatorname{cn} v = \cos\beta, \quad \operatorname{cn} w = \cos\gamma.$$

By (9.4), we obtain

$$\sin a = k \operatorname{sn} u, \quad \sin b = k \operatorname{sn} v, \quad \sin c = k \operatorname{sn} w,$$

from which it follows that

$$\cos a = \operatorname{dn} u, \quad \cos b = \operatorname{dn} v, \quad \cos c = \operatorname{dn} w.$$

Now, by substituting in (9.1) and (9.8) we have

$$\operatorname{dn} w = \operatorname{dn} u \operatorname{dn} v + k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} w, \tag{9.11}$$

$$\operatorname{cn} w = -\operatorname{cn} u \operatorname{cn} v + \operatorname{sn} u \operatorname{sn} v \operatorname{dn} w. \tag{9.12}$$

Solving for $\operatorname{cn} w$ and $\operatorname{dn} w$, and applying the relevant addition formulæ (see below), we have

$$\operatorname{cn} w = -\operatorname{cn}(u+v), \quad \operatorname{dn} w = \operatorname{dn}(u+v).$$

As we have chosen that each of u, v and w lie in the interval from zero to K, these equations have a unique solution, namely

$$w = 2K - (u+v).$$

Therefore,

$$u + v + w = 2K. \tag{9.13}$$

□

9.3. A further derivation of the Jacobi addition formulæ

Spherical trigonometry also provides an alternative method to deriving the addition formulæ for the Jacobi elliptic functions.

THEOREM 9.6. *The addition formulæ for the Jacobi elliptic functions are*

$$\operatorname{sn}(u+v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v},$$

$$\operatorname{cn}(u+v) = \frac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{sn} v \operatorname{dn} u \operatorname{dn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v},$$

$$\operatorname{dn}(u+v) = \frac{\operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} u \operatorname{cn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}.$$

Proof (Lawden, 1989). If we keep c and γ constant, and vary a, b, c, α and β, then, by [9]

[8] Lawden (1989) 103–105.

applying (9.4), we clearly see that k is constant. After differentiation of (9.8), we have

$$(\sin\alpha\cos\beta + \cos\alpha\sin\beta\cos c)\,\mathrm{d}\alpha + (\cos\alpha\sin\beta + \sin\alpha\cos\beta\cos c)\,\mathrm{d}\beta = 0.$$

By (9.9) this can be simplified to obtain

$$\cos b\,\mathrm{d}\alpha + \cos a\,\mathrm{d}\beta = 0. \tag{9.14}$$

Now, the elliptic measure of the angle α was given by the equation $\operatorname{sn} u = \sin A$, and so, by differentiation, it follows that

$$\operatorname{cn} u \operatorname{dn} u\,\mathrm{d}u = \cos\alpha\,\mathrm{d}\alpha = \operatorname{cn} u\,\mathrm{d}\alpha,$$

and hence,

$$\mathrm{d}\alpha = \operatorname{dn} u\,\mathrm{d}u = \cos a\,\mathrm{d}u.$$

Similarly,

$$\mathrm{d}\beta = \operatorname{dn} v\,\mathrm{d}v = \cos b\,\mathrm{d}v.$$

It follows that (9.14) is equivalent to

$$\mathrm{d}u + \mathrm{d}v = 0.$$

Therefore, for fixed c and γ, we have that $u + v$ is equal to a constant. To calculate this constant, we put $a = 0$ and $b = c$. Then $\alpha = 0$ and $\beta = \pi - \gamma$, and it follows that $u = 0$ and $v = 2K - w$. Therefore, the constant is equal to $2K - w$, and we have (9.13). By (9.1) and (9.8), we obtain (9.11) and (9.12), and by the substitution $w = 2K - (u + v)$, we have

$$\operatorname{dn}(u+v) = \operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn}(u+v),$$

$$\operatorname{cn}(u+v) = -\operatorname{cn} u \operatorname{cn} v + \operatorname{sn} u \operatorname{sn} v \operatorname{dn}(u+v).$$

The result follows by solving for $\operatorname{cn}(u+v)$ and $\operatorname{dn}(u+v)$, from which it is easy to deduce the formula for $\operatorname{sn}(u+v)$.

□

[9] Lawden (1989) 105.

Chapter 10

Surface Area of an Ellipsoid

An *ellipsoid* is a closed surface of which all plane sections are either ellipses or circles. If a, b and c are its axes, then the equation of a general ellipsoid in Cartesian coordinates is

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

When $a=b=c$, the surface is a sphere, while if two axes are equal then the ellipsoid is a *spheroid*, formed by revolving an ellipsoid about one of its axes. These can be seen in Figure 10.1.

10.1. Surface area in terms of elliptic integrals of the second kind

While it is easy to calculate that the surface area of a sphere of radius r is $4\pi r^2$, the equivalent problem for the ellipsoid is altogether more complicated, and requires elliptic integrals of the second kind.

THEOREM 10.1. *The surface area of the ellipsoid*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1 \tag{10.1}$$

is given by

$$S=2\pi c^2+\frac{2\pi b}{\sqrt{a^2-c^2}}\left\{(a^2-c^2)E(u)+c^2u\right\},$$

where a, b, c *are constant and* $\operatorname{sn}^2 u=(a^2-c^2)/a^2$.

Proof (Bowman, 1953). Let p be the perpendicular from the centre of the ellipsoid on [1]
the tangent plane at the point (x, y, z), and let $\cos\alpha$, $\cos\beta$ and $\cos\gamma$ be the direction cosines of the normal at that point. Then

$$\cos\alpha=\frac{px}{a^2},\quad \cos\beta=\frac{py}{b^2},\quad \cos\gamma=\frac{pz}{c^2},$$

and

$$\frac{1}{p^2}=\frac{x^2}{a^4}+\frac{y^2}{b^4}+\frac{z^2}{c^4}.$$

Although mentioned in a letter written by Sir Isaac Newton (1643–1727) in 1672, the term *ellipsoid* was subsequently displaced by Euler's *elliptoid*. Its modern usage is attributed to Sylvestre François Lacroix (1765–1843) and Jean-Baptiste Biot (1774–1862).

[1] Bowman (1953) 31–32.

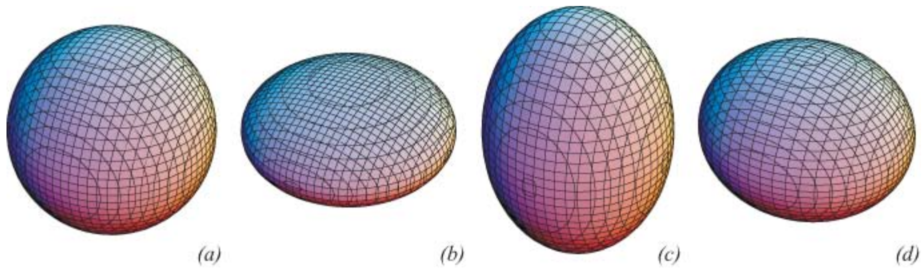

Figure 10.1. A sphere, two spheroids and an ellipsoid. The oblate spheroid (b) is formed by revolving an ellipse about its minor axis, while the prolate spheroid (c) results from a revolution about its major axis.

Therefore, the points at which the normals make a constant angle with the z-axis lie on the cone given by the equation

$$\left\{\frac{x^2}{a^4}+\frac{y^2}{b^4}+\frac{z^2}{c^4}\right\}\cos^2\alpha=\frac{z^2}{c^4}. \tag{10.2}$$

Moreover, by eliminating z from (10.1) and (10.2) we have

$$\left\{\frac{\cos^2\gamma}{a^2}+\frac{\sin^2\gamma}{c^2}\right\}\frac{x^2}{a^2}+\left\{\frac{\cos^2\gamma}{b^2}+\frac{\sin^2\gamma}{b^2}\right\}\frac{y^2}{b^2}=\frac{\sin^2\gamma}{c^2},$$

and so these points also lie on an elliptic cylinder. If A is the area of the cross-section of this cylinder and S is the area of the surface of the ellipsoid intercepted by the cylinder where $z>0$, then

$$\mathrm{d}S=\sec\gamma\ \mathrm{d}A. \tag{10.3}$$

Now, the semiaxes of the cross-section of the cylinder have lengths

$$\frac{a^2\sin\gamma}{\sqrt{c^2\cos^2\gamma\ +a^2\sin^2\gamma}},\qquad\frac{b^2\sin\gamma}{\sqrt{c^2\cos^2\gamma\ +b^2\sin^2\gamma}},$$

and it follows that its area is

$$A=\frac{\pi a^2b^2\sin^2\gamma}{\sqrt{(c^2\cos^2\gamma\ +a^2\sin^2\gamma\,)(c^2\cos^2\gamma\ +a^2\sin^2\gamma\,)}}.$$

Suppose we define e_1 and e_2 by $e_1^2=(a^2-c^2)/a^2$ and $e_2^2=(b^2-c^2)/b^2$, then

$$A=\frac{\pi ab\sin^2\gamma}{\sqrt{(1-e_1^2\cos^2\gamma\,)(1-e_2^2\cos^2\gamma\,)}}.$$

Now, $e_1^2>e_2^2$ if $a^2>b^2>c^2$, so if we let $t=e_1\cos\gamma$ and $k=e_2/e_1$, we have

$$A=\frac{\pi ab(e_1^2-t^2)}{e_1^2\sqrt{(1-t^2)(1-k^2t^2)}}.$$

Moreover, if we put $e_1 = \operatorname{sn} u$ and $t = \operatorname{sn} v$, it follows that

$$\sec\gamma = \frac{\operatorname{sn} u}{\operatorname{sn} v}, \qquad A = \frac{\pi ab}{\operatorname{sn}^2 u}\left\{\frac{\operatorname{sn}^2 u - \operatorname{sn}^2 v}{\operatorname{cn} v \operatorname{dn} v}\right\}.$$

Hence, after simplification, we have

$$\frac{\operatorname{sn} u}{\pi ab}\,\mathrm{d}A \sec\gamma = -\left\{\frac{\operatorname{dn}^2 u}{\operatorname{dn}^2 v} + \frac{\operatorname{cn}^2 u}{\operatorname{cn}^2 v}\right\}\mathrm{d}v.$$

As γ varies from zero to $\frac{1}{2}\pi$, t varies from e_1 to zero, while v varies from u to zero. If S now denotes the surface area of the whole ellipsoid, then, by (10.3),

$$\frac{\operatorname{sn} u}{\pi ab}\frac{S}{2} = \int_0^u \left\{\frac{\operatorname{dn}^2 u}{\operatorname{dn}^2 v} + \frac{\operatorname{cn}^2 u}{\operatorname{cn}^2 v}\right\}\mathrm{d}v. \tag{10.4}$$

Now, consider the derivative

$$\frac{\mathrm{d}}{\mathrm{d}v}\left\{\frac{\operatorname{sn} v \operatorname{cn} v}{\operatorname{dn} v}\right\} = \frac{1}{k^2}\left\{\operatorname{dn}^2 v - \frac{k'^2}{\operatorname{dn}^2 v}\right\}.$$

By integration, we have

$$\int \frac{\mathrm{d}v}{\operatorname{dn}^2 v} = \frac{1}{k'^2}\left\{E(v) - \frac{k^2 \operatorname{sn} v \operatorname{cn} v}{\operatorname{dn} v}\right\}. \tag{10.5}$$

Similarly,

$$\frac{\mathrm{d}}{\mathrm{d}v}\left\{\frac{\operatorname{sn} v \operatorname{dn} v}{\operatorname{cn} v}\right\} = \frac{k'^2}{\operatorname{cn}^2 v} - k'^2 - \operatorname{dn}^2 v,$$

and

$$\int \frac{\mathrm{d}v}{\operatorname{cn}^2 v} = v + \frac{1}{k'^2}\left\{\frac{\operatorname{sn} v \operatorname{dn} v}{\operatorname{cn} v} - E(v)\right\}. \tag{10.6}$$

By (10.5) and (10.6),

$$\int_0^u \left\{\frac{\operatorname{dn}^2 u}{\operatorname{dn}^2 v} + \frac{\operatorname{cn}^2 u}{\operatorname{cn}^2 v}\right\}\mathrm{d}v = \operatorname{sn} u \operatorname{cn} u \operatorname{dn} u + \frac{\operatorname{dn}^2 u - \operatorname{cn}^2 u}{k'^2}E(u) + u\operatorname{cn}^2 u.$$

Therefore, (10.4) becomes

$$\frac{\operatorname{sn} u}{2\pi ab}S = \operatorname{sn} u \operatorname{cn} u \operatorname{dn} u + \frac{\operatorname{dn}^2 u - \operatorname{cn}^2 u}{k'^2}E(v) + u\operatorname{cn}^2 u,$$

and it follows that

$$\frac{\sqrt{a^2 - c^2}}{2\pi a^2 b}S = \frac{c^2\sqrt{a^2 - c^2}}{a^2 b} + \frac{a^2 - c^2}{a^2}E(u) + \frac{c^2}{a^2}u.$$

□

Chapter 11

Seiffert's Spherical Spiral

Suppose we have a particle moving on the surface of a unit sphere with a constant speed v and a constant angular velocity a. The curve traced out by this motion is known as *Seiffert's spherical spiral.*

If φ is the longitude of the position of the particle, then it follows, from the definition, that

$$\varphi = at.$$

By setting $t = s/v$ we can eliminate t to obtain

$$\varphi = ks, \quad k = a/v, \tag{11.1}$$

where s is the length of the curve measured from the north pole.

The constant k is the only parameter that determines the behaviour of a Seiffert's spiral. As the angular velocity is constant, it follows that if the surface velocity is high, then k has to be small, and vice versa.

11.1. Parametric equations of a Seiffert's spiral in cylindrical coordinates [1]

To investigate the properties of a Seiffert's spiral, we introduce spherical coordinates (ρ, φ, z) of the particle at a position P. Referring to Figure 11.1 and denoting the north and south poles by N and S, respectively, we define ρ to be the distance from the axis NS, φ the longitude of P, and z to be the height above the equatorial plane. We also define θ to be the angle between rays from the centre of the sphere to P and to N. [2] If ρ and z are considered as Cartesian coordinates in the meridian plane of P, then ρ changes sign when the path passes through either of N or S. The coordinate z will change sign as the particle crosses the equator. We need to further define that when the particle passes through S, our value of φ remains unchanged. If we had adopted the conventional rule, then passing through a pole would cause φ to suddenly increase (or decrease) by π, therefore contradicting (11.1). This modification is consistent with the

The article *Spiraling the earth with C. G. J. Jacobi* (Erdös, 2000) uses Seiffert's spiral to introduce the Jacobi functions in a completely geometric way. Without needing any previous knowledge of elliptic functions, the definitions, identities, periodicity and reciprocity relations can also be derived from the spiral, much in the same way that Halphen and Greenhill use circles (Section 6.1) and simple pendulums (Section 5.1), respectively.

[1] Erdös (2000) 888-889.

[2] It turns out that $\theta = \operatorname{am} s$.

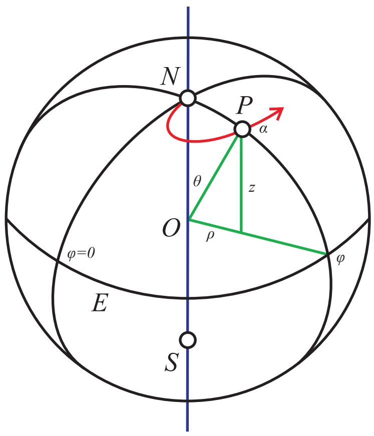

Figure 11.1. The beginning of a Seiffert's spiral on the sphere.

intuitive view that when the particle passes through S it remains in the same meridian plane.

Using these coordinates we can consider the following result:

THEOREM 11.1. *The parametric equations of a Seiffert's spiral in cylindrical coordinates are*

$$\begin{cases} \rho = \operatorname{sn}(s|m), \\ \varphi = \sqrt{m}s, \\ z = \operatorname{cn}(s|m), \end{cases} \tag{11.2}$$

where $m = k^2$.

Proof (Erdös, 2000). Consider the infinitesimal line element $\mathrm{d}s$ on the sphere, illustrated in Figure 11.2. Clearly, we can express $\mathrm{d}s^2$ for any curve on the sphere by

$$\mathrm{d}s^2 = \rho^2 \mathrm{d}\varphi^2 + \mathrm{d}\rho^2 + \mathrm{d}z^2. \tag{11.3}$$

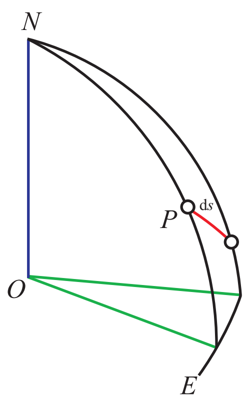

Figure 11.2. An infinitely small line element $\mathrm{d}s$ *on the surface of the sphere.*

As we have a unit sphere, then $\rho^2 + z^2 = 1$, and we eliminate $\mathrm{d}z^2$. Hence,

$$\mathrm{d}s^2 = \rho^2 \mathrm{d}\varphi^2 + \frac{1}{1-\rho^2}\mathrm{d}\rho^2,$$

for $\rho \neq 1$. Applying (11.1), we can replace $\mathrm{d}\varphi^2$ by $k^2\mathrm{d}s$ to obtain

$$\mathrm{d}s = \frac{\mathrm{d}\rho}{\sqrt{(1-\rho^2)(1-k^2\rho^2)}}, \tag{11.4}$$

where $|\rho| < 1/k$. Therefore, the total distance s travelled along the curve from N to P can be expressed in terms of the distance ρ of that point from the axis NS by the integral

$$s(\rho, k) = \int_0^\rho \frac{\mathrm{d}\rho}{\sqrt{(1-\rho^2)(1-k^2\rho^2)}}. \tag{11.5}$$

This defines the elliptic integral of the first kind with modulus k. As the square of k occurs in this integral, then, when describing Seiffert's spirals, we use the notation $m = k^2$, which we similarly use to represent the parameter of the elliptic integral. Therefore, by inversion of (11.5), we immediately have the Jacobi elliptic function $\mathrm{sn}(s|m)$, and it follows that

$$\rho = \mathrm{sn}(s|m). \tag{11.6}$$

Suppose instead of eliminating $\mathrm{d}z$ from (11.3), we use the property $\rho^2 + z^2 = 1$ to eliminate $\mathrm{d}\rho$. It follows that we obtain an expression of the length of the path travelled as a function of the distance z of P from the equatorial plane, namely

$$s(z, k) = \int_1^z \frac{\mathrm{d}z}{\sqrt{(1-z^2)(1-k^2+k^2z^2)}}, \tag{11.7}$$

where the positive square root $\sqrt{1-z^2} < 1/k$. Hence, by inversion of this integral we have the Jacobi elliptic function $\mathrm{cn}(s|m)$, and it follows that

$$z = \mathrm{cn}(s|m). \tag{11.8}$$

□

To summarize, we take (11.2) as our definition of a Seiffert's spiral on the surface of a unit sphere, with the length s measured from the north pole N (at which $\varphi = 0$, $\rho = 0$ and $z = 1$). We distinguish different Seiffert's spirals by a parameter m.

We can finally note that by (11.6) and (11.8), and as $\rho^2 + z^2 = 1$ we have the identity

$$\mathrm{sn}^2(s|m) + \mathrm{cn}^2(s|m) = 1. \tag{11.9}$$

11.2. Properties of a Seiffert's spiral [3]

[3] Erdös (2000) 889–891.

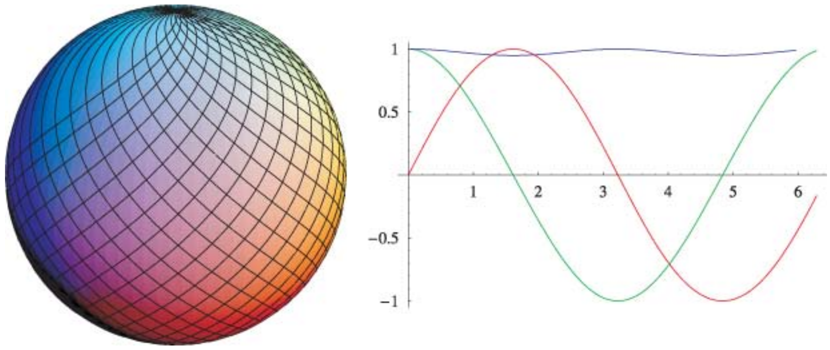

Figure 11.3. A Seiffert's spiral with parameter $m = 0.1$ of length 50π. The graph shows the three elliptic functions in this case, with $\mathrm{sn}(0|0.1)$ starting at zero, $\mathrm{cn}(s|0.1)$ starting at 1 and $\mathrm{dn}(s|0.1)$ staying close to 1.

Figure 11.2 shows part of a spiral traced out for some parameter m. At an arbitrary point P, the curve crosses the meridian at an angle α. If we consider the illustrated infinitely small spherical triangle at P, we have

$$\cos\alpha = \frac{\sqrt{\mathrm{d}z^2 + \mathrm{d}\rho^2}}{\mathrm{d}s} = \sqrt{1 - m\,\mathrm{sn}^2(s|m)}, \tag{11.10}$$

by applying (11.4). The right hand side of this equation therefore defines the third Jacobi elliptic function

$$\mathrm{dn}(s|m) = \cos\alpha. \tag{11.11}$$

From our geometric point of view, this means that if we travel distance s from N along a Seiffert's spiral, then the cosine of the angle formed at that point with the meridian arc NS equals $\mathrm{dn}(s|m)$.

Now, suppose we set $m = 0$. Even though this corresponds to an infinite-valued v in (11.1), (11.2) still remains valid, and we have φ remaining zero for any s. As we travel along the meridian arc NS, then as s increases from zero to π, we can see that $s = \theta$, $\rho = \sin s$, $z = \cos s$ and $\alpha = 0$. Therefore, from (11.2) and (11.11) it follows that

$$\mathrm{sn}(s|0) = \sin s, \quad \mathrm{cn}(s|0) = \cos s, \quad \mathrm{dn}(s|0) = 1.$$

Hence, we have shown that the circular functions can to considered to be special cases of the Jacobi elliptic functions.

To ensure that we do not have cusps in the spiral for $m > 0$, when the particle reaches S we do not reverse direction and travel along the meridian $\varphi = 0$, but instead continue smoothly. Therefore, by the convention previously adopted, φ remains zero and we complete a full meridian circle.

We now suppose that $m \neq 0$, but is still much less than 1. The curve defined by (11.2) will make a small angle α with the meridians it crosses. It follows that $\mathrm{dn}(s|m)$ remains close to 1 for all s and, as illustrated in Figure 11.3, the Jacobi elliptic functions are still similar to their circular counterparts. As we can begin to see in this figure, except when it is a closed curve, a Seiffert's spiral crosses each point on the sphere, other than its poles N and S, twice. The poles are crossed an infinite number of times, and it follows that the length of the spiral is itself infinite.

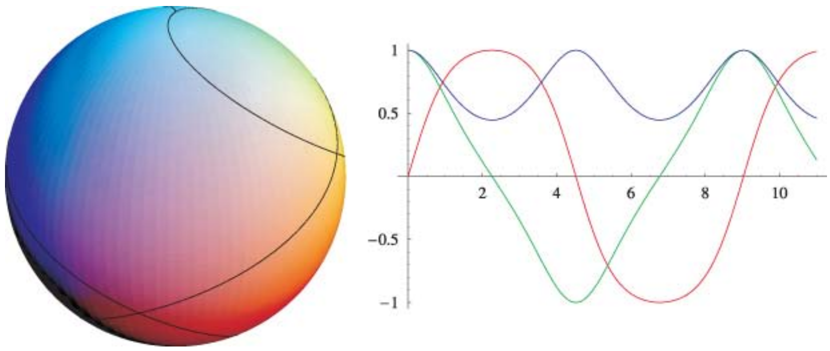

Figure 11.4. A Seiffert's spiral with parameter $m = 0.8$ of length $3\frac{1}{2}\pi$. We have $\mathrm{sn}(s|0.8)$ *starting at zero,* $\mathrm{cn}(s|0.8)$ *starting at 1 and* $\mathrm{dn}(s|0.8)$ *always greater than 1.*

Figure 11.4 shows a Seiffert's spiral for $m = 0.8$. In this case the Jacobi elliptic functions differ considerably from the circular functions.

When m increases towards 1, the angle α at which the spiral crosses the meridian at the equator approaches $\frac{1}{2}\pi$. This follows from the fact that on the equator $\rho = \mathrm{sn}(s|m) = 1$, and, by (11.10), setting $m = 1$ implies that $\mathrm{dn}(s|m) = \cos\alpha = 0$. Therefore, this spiral (shown in Figure 11.5) never crosses the equator. Moreover, it winds around the northern hemisphere, asymptotically approaching the equator, because its length s, given by (11.7) becomes infinite for $z = 0$ when $m = 1$. So for $m = 1$, (11.1) reduces to $\varphi = s$, and we have

$$\sin\alpha = \rho\frac{\mathrm{d}\varphi}{\mathrm{d}s} = \rho,$$

and, since $\sin\theta = \rho$, it follows that $\alpha = \theta$. Hence, at every point on the spiral, the angle α that it makes with the meridian equals the latitude θ. Moreover, we can express (11.5) and (11.7) in terms of inverse hyperbolic functions and it follows that

$$\mathrm{sn}(s|1) = \tanh s, \quad \mathrm{sn}(s|1) = \mathrm{dn}(s|1) = \mathrm{sech}\, s.$$

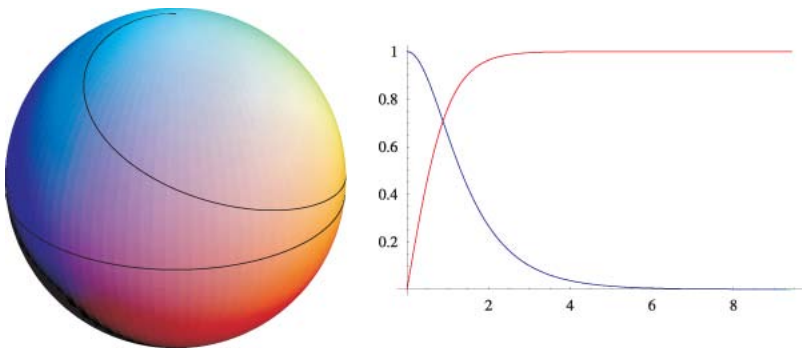

Figure 11.5. A Seiffert's spiral with parameter $m = 1$ of length 3π. The three elliptic functions are illustrated by two curves on the graph, with $\mathrm{sn}(s|1) = \tanh s$ *starting at zero and* $\mathrm{cn}(s|1) = \mathrm{dn}(s|1) = \mathrm{sech}\, s$ *starting at 1.*

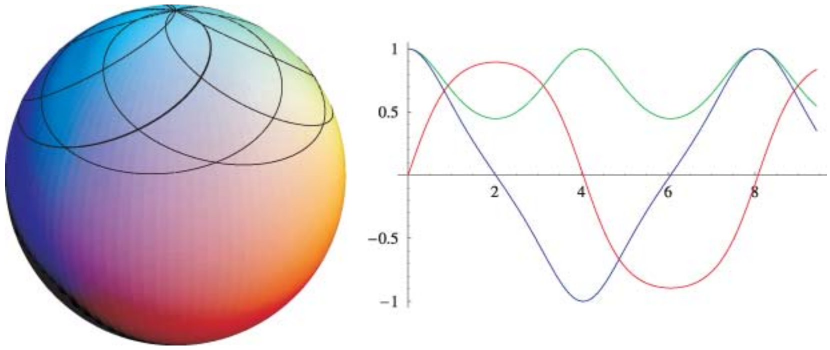

Figure 11.6. A Seiffert's spiral with parameter $m = 1.25$ *of length* 32*. The elliptic function* $\mathrm{sn}(s|1.25)$ *can be scaled to coincide with* $\mathrm{sn}(s|0.8)$*, while* $\mathrm{cn}(s|1.25)$ *and* $\mathrm{dn}(s|1.25)$ *can be scaled to coincide with* $\mathrm{dn}(s|0.8)$ *and* $\mathrm{cn}(s|0.8)$ *(see Figure 11.4).*

11.3. Low speed travel 4

We now consider a Seiffert's spiral with $m > 1$. Such a spiral can be seen in Figure 11.6, and corresponds to the case when the surface speed is low. In order for the angular velocity to remain the same as at high surface speed, the spiral must stay close to N, so as to circle the sphere in a short time. As with case $m = 1$, for $m > 1$ the spiral stays entirely in the northern hemisphere. When m is much greater than 1, we have that ρ remains small for all s, and it follows that $\mathrm{sn}(s|m)$ will be small in absolute value. Contrasting with this, values of z, and hence, $\mathrm{cn}(s|m)$ will be close to 1 for all s.

Suppose that we denote the lowest point of the spiral by L. At this point, the spiral is tangent to a circle of latitude, and we have $\cos\alpha_L = 0$. By (11.10) we also have

$$\rho_L = \max\{\mathrm{sn}(s|m)\} = 1/\sqrt{m},$$

for $m > 1$. The value ρ_L is the maximum distance of the spiral from the axis NS. From (11.9), we have that the latitude of the lowest point of the spiral is

$$z_L = \min\{\mathrm{cn}(s|m)\} = \sqrt{1 - (1/m)},$$

for $m > 1$.

11.4. Periodicity of the Jacobi elliptic functions 5

To demonstrate the periodicity of the Jacobi elliptic functions, we return to considering a Seiffert's spiral with m much less than 1. This spiral is close to the meridian circle, and its segments lying between any two successive passages of the same pole are congruent in that they, and their directions, can be mapped onto each other by a rotation around the axis NS. As a result of this congruence, we consider such a segment to be the period of the spiral.

[4] Erdös (2000) 889.
[5] Erdös (2000) 891–892.

Now, the length s_E of the arc between the pole and equator depends on m, and, by (11.5), is equal to $K(m)$, the complete elliptic integral of the first kind. The total length of the period is therefore $4s_E = 4K(m)$. Similarly, this is the period of $\rho = \mathrm{sn}(s|m)$ and $z = \mathrm{cn}(s|m)$. As the cosine of the angle between the spiral and the meridian is equal to 1 at each pole crossing, the period of $\cos\alpha = \mathrm{dn}(s|m)$ is $2s_E = 2K(m)$.

11.5. Seiffert's spiral as a closed curve

Suppose we wish to find under what condition Seiffert's spiral is a periodic function on the sphere, or, from a geometric point of view, that it is a closed curve. This problem is different from the discussion of the previous section as an elliptic function is always periodic, except at $m = 1$, while we will see that for a Seiffert's spiral it may not be the case.

THEOREM 11.2. *Seiffert's spiral forms a closed curve if its parameter $m < 1$ is chosen such that*

$$f(m) = \frac{2}{\pi}\sqrt{m}K(m)$$

is a rational number.

Proof (Erdös, 2000). We can express the condition for the spiral to be a closed curve by [6]
requiring that the particle, starting at N at longitude φ_0 and having travelled a distance s_n, returns to N at the angle

$$\varphi_n = \varphi_0 + 2\pi n, \tag{11.12}$$

for n a positive integer. While this passage may not be the first return to N, we can assume that it is the first return at angle φ_n. In this case, we clearly have that the returning branch of the spiral merges smoothly with the starting branch, and it follows that the curve is periodic. Without loss of generality, we set $\varphi_0 = 0$. At N we have $\rho = 0$, so by (11.2) we require

$$\mathrm{sn}(s_n|m) = 0.$$

Therefore, again by (11.2), the condition of the return angle φ_n is

$$s_n = \frac{2\pi}{\sqrt{m}}n, \tag{11.13}$$

where n is again a positive integer, and s_n denotes the length of the length of the closed spiral.

Now, combining (11.12) and (11.13), it follows that

$$\mathrm{sn}\left(\frac{2\pi}{\sqrt{m}}n \middle| m\right) = 0.$$

Since $\mathrm{sn}(s|m)$ is periodic with period $4K(m)$ and $\mathrm{sn}(0|m) = 0$, this reduces to

$$\frac{2\pi}{\sqrt{m}}n = 4K(m)p, \tag{11.14}$$

for an integer p. Our condition on m follows trivially.

□

[6] Erdös (2000) 892–894.

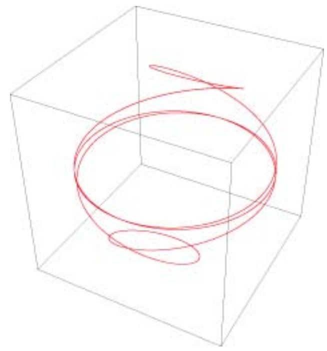

Figure 11.7. A closed Seiffert's spiral with $n = 4$ and $p = 1$, from which it follows that $m = 0.999944$.

Now, suppose we express (11.14) as $f(m) = n/p$, for positive integers p and n, and assume that the common factors between these two numbers have been cancelled out. As the function $f(m)$ varies between zero and infinity when m varies between zero and 1, the rational numbers n/p form a dense set over every interval from zero [7] to infinity. Therefore, there are infinitely many closed spirals for any interval of m. Unfortunately, the value of m which corresponds to a particular n/p can only be found by a numerical, and therefore approximate, solution of the transcendental equation (11.14). However, the integers p and n have the following geometric interpretation: Since p is the number of periods of $\operatorname{sn}(s|m)$ completed before the spiral closes for the last time, and in every period $\rho = \operatorname{sn}(s|m)$ has two zeroes, then p equals the number of times the closed curve passes each pole. Similarly, n represents the number of times the curve circles the axis NS.

An example of a closed Seiffert's spiral is shown in Figure 11.7.

11.6. Projecting a Seiffert's spiral

Suppose we project a Seiffert's spiral with $m < 1$ on the unit sphere onto the northern hemisphere of a second concentric sphere of radius

$$R = 1/\sqrt{m} \tag{11.15}$$

placed around our original sphere. This projection is parallel to the axis NS. The new spiral is not actually a Seiffert's spiral as there is no linear proportionality between its arc length s_2 and the meridian angle φ_2. In fact, we can show that the arc lengths s_1 of the first spiral, and s_2 of the second spiral are related by

$$s_2 = \int_0^{s_1} \sqrt{1 - (1 - m)\operatorname{sn}^2(s|m)}\, \mathrm{d}s.$$

Since s_1 is linearly proportional to φ_1, then s_2 is not. It follows that if we wish to maintain constant angular velocity $a = \dot{\varphi_1}$ along the projected spiral we must vary $\dot{s_2}$.

[7]A subset of a set is *dense* if its closure is equal to the whole set.

Despite not having a Seiffert's spiral after the projection, we can still use the projected spiral to derive more properties of Jacobi elliptic functions. These are known as the reciprocity relations.

THEOREM 11.3. *The reciprocity relations of the Jacobi elliptic functions, which connect functions with parameter $m < 1$ to functions with parameter $1/m$, are*

$$\operatorname{sn}\left(\sqrt{m}s \,\middle|\, \frac{1}{m}\right) = \sqrt{m}\,\operatorname{sn}(s|m),$$
$$\operatorname{cn}\left(\sqrt{m}s \,\middle|\, \frac{1}{m}\right) = \operatorname{dn}(s|m),$$
$$\operatorname{dn}\left(\sqrt{m}s \,\middle|\, \frac{1}{m}\right) = \operatorname{cn}(s|m).$$

Proof (Erdös, 2000). By comparison with (11.2), the equations of original spiral, and [8]
from the method of our projection, we have

$$\varphi_2 = \varphi_1, \quad \rho_2 = \rho_1 = \sin\theta_1 = R\sin\theta_2. \tag{11.16}$$

We then eliminate ρ_1 from (11.4), by using (11.15) and (11.16), to obtain

$$\sqrt{m}\mathrm{d}s_1 = \frac{\mathrm{d}\theta_2}{\sqrt{1-(1/m)\sin^2\theta_2}}.$$

By inversion of this integral, we have

$$\sin\theta_2 = \operatorname{sn}\left(\sqrt{m}s_1 \,\middle|\, \frac{1}{m}\right).$$

Applying (11.2) and (11.16), we have the relation for $\operatorname{sn}(s|m)$. By (11.9) and (11.10), we obtain relations for $\operatorname{cn}(s|m)$ and $\operatorname{dn}(s|m)$.

□

In each case reciprocity relations s is the length of the arc of the original Seiffert's spiral, and not that of its projected equivalent.

[8] Erdös (2000) 894.

Part III

Applications of the Weierstrass Elliptic Functions

Chapter 12

The Spherical Pendulum

A *spherical pendulum* is a pendulum that is suspended from a pivot mounting, which allows it to swing in any of an infinite number of vertical planes through the point of suspension. [1]

12.1. Equations of motion of the spherical pendulum in cylindrical coordinates

We consider a particle of mass m, suspended from a fixed point O by a light wire of length l. The forces acting on the particle are its weight, a constant force due to gravity of magnitude mg vertically downward, and by the tension in a wire.

If we let AOA' be the vertical diameter of the sphere on which the particle moves, and let PN be the perpendicular from the particle to this diameter, then we can define cylindrical coordinates (ρ, θ, z) as illustrated in Figure 12.1. Therefore, we have that $\rho = PN, \theta$ is the angle between the meridian APA' and a datum meridian AMA', and $z = ON$, taking positive values for N below O. Also we have that the velocity of the particle is given by $(\dot{\rho}, \rho\dot{\theta}, \dot{z})$ in this coordinate system.

THEOREM 12.1. *The equations of motion of the spherical pendulum in cylindrical coordinates are*

$$
\begin{aligned}
r^2 &= -l^2\{\wp(u) - \wp(\alpha)\}\{\wp(u) - \wp(\beta)\},\\
\mathrm{e}^{2\mathrm{i}\theta} &= -E^2\frac{\sigma(u+\beta)\sigma(u-\alpha)}{\sigma(u+\alpha)\sigma(u-\beta)}\mathrm{e}^{\{\zeta(\alpha)-\zeta(\beta)\}u},\\
z &= l\wp\left(\sqrt{\frac{g}{2l}}t + \tfrac{1}{2}\omega_2\right) + \frac{1}{6g}(v_0^2 + 2gz_0),
\end{aligned}
$$

where v_0 is the initial velocity of the pendulum, z_0 is the initial value of z, $t = \lambda u$ and α, β, λ, E and l are constants. The periods of $\wp(u)$ are ω_1 and ω_2.

Proof (Dutta & Debnath, 1965). Since both forces acting on the particle have zero moment about the z-axis, the particle's momentum about this axis is conserved. Hence, for a constant h, it follows that [2]

$$m\rho^2\dot{\theta} = mh. \tag{12.1}$$

Energy is also conserved, so

$$\tfrac{1}{2}m(\dot{\rho}^2 + \rho^2\dot{\theta}^2 + z^2) - mgz = \tfrac{1}{2}mv_0^2 - mgz_0,$$

[1] The *spherical pendulum* can be considered a generalization of the simple pendulum of Section 5.1.

[2] Dutta & Debnath (1965) 88–91.

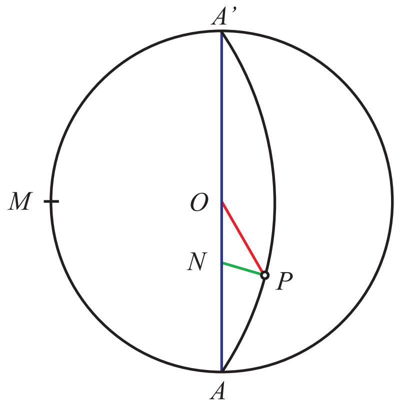

Figure 12.1. A spherical pendulum.

with z_0 being the initial position of the particle along the z-axis and v_0 the initial velocity of the particle. If we define a further constant $c = v_0^2 + 2gz_o/2g$, then

$$\dot{\rho}^2 + \rho^2\dot{\theta}^2 + \dot{z}^2 = 2g(c - z). \tag{12.2}$$

By eliminating $\dot{\theta}$ from (12.1) and (12.2), we have

$$\dot{\rho}^2 + \dot{z}^2 = 2g(c - z) - \frac{h^2}{r^2}.$$

But as the trajectory of the particle lies on the sphere, we have $\rho^2 + z^2 = l^2$, which implies $\rho\dot{\rho} + z\dot{z} = 0$. Therefore, we can also eliminate ρ and $\dot{\rho}$ to obtain

$$l^2\dot{z}^2 = 2g(c - z)(l^2 - z^2) - h^2. \tag{12.3}$$

This determines the variation of z with time t. If we now consider the cubic

$$\varphi(z) = 2g(c - z)(l^2 - z^2) - h^2,$$

then (12.3) reduces to

$$l^2\dot{z}^2 = \varphi(z). \tag{12.4}$$

For the particle to be in motion we must have that the component of its velocity along the z-axis is real, and that the corresponding initial velocity is non-zero. Therefore, $\varphi(z) > 0$. Moreover, it is easy to see that

$$\varphi(-l) = -h^2, \quad \varphi(l) = -h^2, \quad \varphi(\infty) = \infty.$$

Denoting these zeroes by z_1, z_2 and z_3, respectively, then it follows that

$$-l < z_1 < z_0 < z_2 < l < z_3 < \infty.$$

The zero z_3 does not lie on the sphere, so the particle moves between an upper horizontal plane $z = z_1$ and a lower horizontal plane $z = z_2$, touching these planes in turn when $z = 0$.

By (12.4), we have that $\dot{z}^2$ is equal to a cubic in z, so z must be an elliptic function of t. We now wish to bring (12.3) into the form

$$\{\wp'(u)\}^2 = 4\wp(u)^3 - g_2\wp(u) - g_3, \tag{12.5}$$

for a variable u. Introducing z' and c' defined by $z = lz'$ and $c = ac'$, then (12.3) becomes

$$l^2\dot{z}'^2 = 2gl(c'-z')(1-z'^2) - \frac{h^2}{l^2}. \tag{12.6}$$

Next, we define t' by $t = \lambda t'$. It follows that (12.6) reduces to the form

$$\begin{aligned}\left(\frac{\mathrm{d}z'}{\mathrm{d}t'}\right)^2 &= \frac{\lambda^2}{l^2}2gl(c'-z')(1-z'^2) - \frac{\lambda^2 h^2}{l^2} \\ &= 4(c'-z')(1-z'^2) - h'^2,\end{aligned} \tag{12.7}$$

where $\lambda^2 = 2l/g$ and $h' = \sqrt{(2/lg)}h$. If we take $z' = z'' + n$ then (12.7) takes the form

$$\left(\frac{\mathrm{d}z''}{\mathrm{d}t'}\right)^2 = 4z''^3 - g_2 z'' - g_3, \tag{12.8}$$

where $n = \frac{1}{3}c'$, and the invariants g_2, g_3 are given by

$$g_2 = \tfrac{4}{3}(c'^2+3), \quad g_3 = h'^2 + \tfrac{8}{27}c'^3 - \tfrac{8}{3}c'.$$

By putting $z'' = \wp(u)$, we obtain the form (12.5), and it follows that

$$\left(\frac{\mathrm{d}z''}{\mathrm{d}t'}\right)^2 = \pm 1,$$

where $u = \pm(t'+\alpha)$ for a constant of integration α.

Now, $\wp(u)$ is an even function, so taking the case $u = t' + \alpha$, we have

$$z'' = \wp(u) = \wp(t'+\alpha). \tag{12.9}$$

The zeroes of the cubic are all real, so if we take ω_1 and ω_2 to be the periods of $\wp(u)$, we have that ω_1 is real, while the ω_2 is purely imaginary. If e_1, e_2 and e_3 are the roots of $4z''^3 - g_2z'' - g_3 = 0$ and $e_1 > e_3 > e_2$, then $z'' = e_2$ when $z = z_1$, $z'' = e_3$ when $z = z_2$, and $z'' = e_1$ when $z = z_3$.

Choosing $t = 0$ to be the instant at which the particle is at its highest level, we have $z = z_1$, and hence, $z'' = e_2$. If we now substitute in (12.9), we have $\wp(\alpha) = e_2$ and $\alpha = \frac{1}{2}\omega_2$. Therefore, it follows that

$$\begin{aligned}z = lz' &= l\wp\left(\sqrt{\frac{g}{2l}}t + \tfrac{1}{2}\omega_2\right) + \frac{l}{3} \\ &= l\wp\left(\sqrt{\frac{g}{2l}}t + \tfrac{1}{2}\omega_2\right) + \frac{1}{6g}(v_0^2 + 2gz_0).\end{aligned}$$

Now, by (12.1), we have

$$\dot{\theta} = \frac{h}{r^2} = \frac{h}{l^2 - z^2},$$

as $\rho^2 + z^2 = l^2$. Hence,

$$\dot{\theta} = \frac{h}{2a}\left(\frac{1}{a-z} + \frac{1}{a+z}\right) = \frac{h}{2a^2}\left(\frac{1}{1-z'} + \frac{1}{1+z'}\right),$$

with z' defined as before.

Now, $u = t' + \alpha = (t/\lambda) + \alpha$ so $\dot{u} = 1/\lambda$. Therefore,

$$\frac{\mathrm{d}\theta}{\mathrm{d}u} = \frac{\mathrm{d}\theta}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}u} = \frac{\lambda h}{2l^2}\left\{\frac{1}{1-z'} + \frac{1}{1+z'}\right\}.$$

If we define $\wp(\alpha) = -(1 + \frac{1}{3}c')$ and $\wp(\beta) = 1 - \frac{1}{3}c'$, and as

$$z' = z'' + \tfrac{1}{3}c' = \wp(u) + \tfrac{1}{3}l',$$

then

$$\frac{\mathrm{d}\theta}{\mathrm{d}u} = \frac{\mathrm{d}\theta}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}u} = \frac{\lambda h}{2l^2}\left\{\frac{1}{\wp(u)-\wp(\alpha)} - \frac{1}{\wp(u)-\wp(\beta)}\right\}. \tag{12.10}$$

When $z' = \pm 1$ or c', then, from the definition of z', and by (12.7), we have

$$\left(\frac{\mathrm{d}z'}{\mathrm{d}t'}\right)^2 = \wp'^2(u) = 4(c'-z')(1-z'^2) - h'^2 = -h'^2.$$

Hence,

$$\wp'^2(\alpha) = \wp'^2(\beta) = -h^2,$$

and it follows that

$$\wp'(\alpha) = \wp'(\beta) = \mathrm{i}h'.$$

Now, by (12.10)

$$\begin{aligned} 2\mathrm{i}\frac{\mathrm{d}\theta}{\mathrm{d}u} &= \frac{\wp'(\alpha)}{\wp(u)-\wp(\alpha)} - \frac{\wp'(\beta)}{\wp(u)-\wp(\beta)} \\ &= \zeta(u+\beta) - \zeta(u-\beta) - 2\zeta(\beta) - \zeta(u+\alpha) + \zeta(u-\alpha) + 2\zeta(\alpha). \end{aligned}$$

After integration, we have

$$\mathrm{e}^{2\mathrm{i}\theta} = -E^2\frac{\sigma(u+\beta)\sigma(u-\alpha)}{\sigma(u+\alpha)\sigma(u-\beta)}\mathrm{e}^{2\{\zeta(\alpha)-\zeta(\beta)\}u}, \tag{12.11}$$

where $-E^2$ is a constant of integration that can be determined by the initial conditions.

Finally, from the equation of the sphere

$$r^2 = l^2 - z^2 = (l+z)(l-z) = -l^2\{\wp(u)-\wp(\alpha)\}\{\wp(u)-\wp(\beta)\}. \tag{12.12}$$

□

12.2. Equations of motion of the spherical pendulum in Cartesian coordinates

THEOREM 12.2. *The equations of motion of the spherical pendulum in Cartesian coordinates are*

$$x + \mathrm{i}y = El\frac{\sigma(u-\alpha)\sigma(u+\beta)}{\sigma^2(u)\sigma(\alpha)\sigma(\beta)}\mathrm{e}^{\{\zeta(\alpha)-\zeta(\beta)\}u},$$
$$x + \mathrm{i}y = -\frac{l}{E}\frac{\sigma(u+\alpha)\sigma(u-\beta)}{\sigma^2(u)\sigma(\alpha)\sigma(\beta)}\mathrm{e}^{\{\zeta(\alpha)-\zeta(\beta)\}u},$$

where $t = \lambda u$ and α, β, λ, E and l are constants.

Proof (Dutta & Debnath, 1965). We define Cartesian coordinates by $x = \rho\cos\theta$ and 3
$y = \rho\sin\theta$. From (12.11), we have

$$\mathrm{e}^{2\mathrm{i}\theta} = \frac{x+\mathrm{i}y}{x-\mathrm{i}y} = -E^2\frac{\sigma(u+\beta)\sigma(u-\alpha)}{\sigma(u+\alpha)\sigma(u-\beta)}\mathrm{e}^{2\{\zeta(\alpha)-\zeta(\beta)\}u}.$$

Also, from (12.12), we obtain

$$\rho^2 = (x+\mathrm{i}y)(x-\mathrm{i}y) = -a^2\{\wp(u)-\wp(\alpha)\}\{\wp(u)-\wp(\beta)\}.$$

□

[3]Dutta & Debnath (1965) 91–92.

Chapter 13

Elliptic Curves and Cryptography

He, first of men, with awful wing pursu'd the comet thro' the long elliptic curve

JAMES THOMSON, *A Poem Sacred to the Memory of Sir Isaac Newton* 1

If we have a field K with characteristic not equal to 2 or 3, then we define an *elliptic curve* over this field to be a non-singular projective cubic C of the form

$$Y^2 = X^3 + aX + b, \tag{13.1}$$

where a and b are in K. If the characteristic of the field is equal to 2 then we have that C is instead defined by

$$Y^2 + cXY + dY = X^3 + aX + b, \tag{13.2}$$

with c and d also in K.

13.1. Addition on the cubic 2

Let P_1 and P_2 be rational points on a non-singular cubic C over an arbitrary field and choose a point O also on C. Let P_3 be the third point of intersection with C of the line through P_1 and P_2. Let P_4 be the third point of intersection of the line through O and P_3. Then we define the addition of P_1 and P_2 by

$$P_1 + P_2 = P_4.$$

If two or more points coincide, for example $P_1 = P_2$, then to construct the line analogous to P_1P_2 above we take the tangent at P_1.

13.2. The group law

THEOREM 13.1. *The construction in Section 13.1 defines an Abelian group law on C, with O as its neutral element.*

Proof (Cassels, 1991). The difficult part of this proof is associativity. However, we 3

[1] James Thomson (1700–1748).
[2] Cassels (1991) 27.

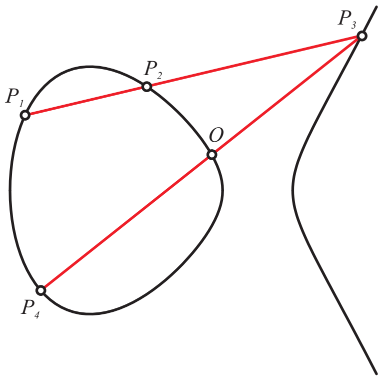

Figure 13.1. The group law.

also need a neutral element, an inverse, and, as we seek to prove to prove that the construction is an Abelian group, commutativity.

It is clear that

$$P_1 + P_2 = P_2 + P_1,$$

and

$$O + P_1 = P_1,$$

for all P_1.

Now, consider the tangent at O. Let the third point of intersection of this tangent at O to be P_1. Let P_2^- to be the third point of intersection of the line through P_1, and P_2. It then simply follows that

$$P_2 + P_2^- = O,$$

thus constructing an inverse P_1^- for any P_1.

We now need to show

$$(P_1 + P_2) + P_3 = P_1 + (P_2 + P_3).$$

Geometric argument. Let O, P_1, P_2 and P_3 be given and consider the situation in Figure 13.2. We have that $L_1, L_2, \ldots, L_6$ are lines and $P_1, P_2, \ldots, P_6$ are points on C. All of these except P_6 and P_9 are intersections of two lines.

It follows that $P_1 + P_2 = P_5$, and so $(P_1 + P_2) + P_3$ is the third point of intersection of the line through O and P_6. Similarly, $P_1 + (P_2 + P_3)$ is the third point of intersection of the line through O and P_9. To prove associativity, we therefore need that P_6 and P_9 are not as in Figure 13.2, but instead coincide in Q, the intersection of L_2 and L_4.

Next, we need the following result of algebraic geometry:

THEOREM 13.2. *Let $P_1, P_2, \ldots, P_8$ be eight points of the plane in general position.* 4
Then there is a ninth point P_9 such that every cubic curve through $P_1, P_2, \ldots, P_8$ also passes through P_9.

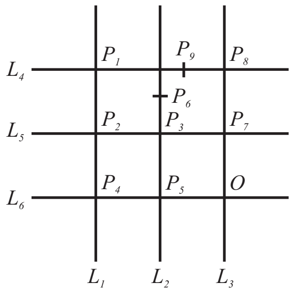

Figure 13.2. Proving the associativity of the group law.

Proof (Cassels, 1991). [5] A cubic form $F(X)$ has ten coefficients. The equation $F(X) = 0$ imposes a linear condition on the coefficients, and passing through the eight points $P_1, P_2, \ldots, P_8$ imposes eight conditions. Hence, if $F_1(X)$ and $F_2(X)$ are linearly independent forms through the eight points, then we have that any third equation, which we denote $F_3(X)$, is of the shape

$$F_3(X) = \lambda F_1(X) + \mu F_2(X).$$

Now, as $F_1(X) = 0$ and $F_2(X) = 0$ have nine points in common, so $F_3(X)$ also passes through all of them.

□

To apply Theorem 13.2, we let the equation for line L_1 to be $L_1(X) = 0$ and consider the two reducible cubics

$$F_1(X) = L_1(X)L_2(X)L_3(X) = 0,$$
$$F_2(X) = L_4(X)L_5(X)L_6(X) = 0.$$

The cubic C passes through eight of the points of intersection of $F_1(X)$ and $F_2(X)$ and so, by Theorem 13.2, must pass through the ninth. Hence, $P_6 = P_9$, as required.

Algebraic argument. Suppose we have a linear form $L_1(X)$. This goes not give a meaningful function on the curve C as the coefficients X are homogeneous. However, if we also have another linear form, $L_6(X)$, then the quotient

$$G(X) = \frac{L_1(X)}{L_6(X)}$$

does give something meaningful.

Now, again referring to Figure 13.2, we have $L_1(X) = 0$ passes through P_1, P_2 and P_4, and $L_6(X) = 0$ passes through O, P_4 and P_5 with all being points on C. Therefore, $G(X)$ has a zero at P_1 and P_2 and a pole at O and P_5. As both $L_1(X) = 0$ and $L_6(X) = 0$ pass through P_4, then these zeroes of the linear form cancel, and we have neither a zero nor a pole. The notion of the order of a zero or pole at a non-singular

[3] Cassels (1991) 28–30.

[4] An arrangement of points is in *general position* if no three are collinear. Three points are said to be *collinear* if they lie on a single straight line.

[5] Cassels (1991) 29.

point of an algebraic curve generalizes in an obvious way to that of the order of zero or pole of a rational function of a single variable. Clearly, we have simple zeroes at P_1 and P_2, and simple poles at O and P_5 and the equation

$$P_5 = P_1 + P_2$$

is equivalent to the existence of such a function. Similarly, we have

$$R = (P_1 + P_2) + P_3$$

is equivalent to the existence of a function with simple poles at P_1, P_2 and P_3, a double zero at O and a simple zero at a point R. It follows that

$$(P_1 + P_2) + P_3 = P_1 + (P_2 + P_3).$$

□

13.3. The simplified group law 6

Suppose we have the cubic (13.2). If we take the point with homogeneous coordinates [0, 1, 0] as the identity, then the group law simplifies, and Theorem 13.1 can be restated as follows:

THEOREM 13.3. *There is a unique Abelian group law on C, with $O = [0, 1, 0]$ as its neutral element, the map from (x, y) to $(x, -y)$ as its inverse, and $P_1 + P_2 + P_4 = O$ if and only if P_1, P_2 and P_4 are collinear.*

The group law in this form corresponds to the situation in Figure 13.3. If the field is the real numbers, we have a simple description of the point P_3: If we draw a line through P_1 and P_2 (or the tangent line to the curve at P_1 if $P_1 = P_2$) and denote the third point of intersection with C, then we have P_4 is the negative of P_3.

We can derive algebraic formulæ from this description which can be applied over 7

[6] Reid (1988) 39–40.

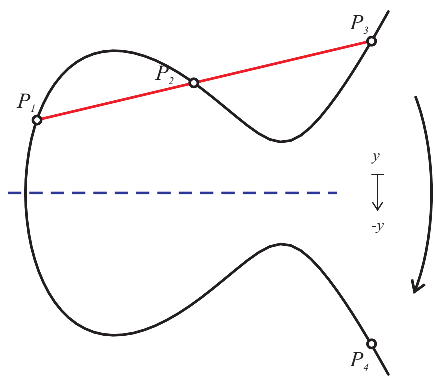

Figure 13.3. The simplified group law.

any field K. If C has the equation (13.2), and if we let

$$P_1 = (x_1, y_1), \quad P_2 = (x_2, y_2), \quad P_4 = (x_4, y_4),$$

then

$$\begin{cases} x_4 = -x_1 - x_2 + \alpha^2 + c\alpha, \\ y_4 = -cx_4 - d - y_1 + \alpha(x_1 - x_4), \end{cases} \tag{13.3}$$

where

$$\alpha = \begin{cases} (y_2 - y_1)/(x_2 - x_1) & \text{if} P_1 \neq P_2, \\ (3x_1^2 + a - cy_1)/(2y_1 + cx_1 + d) & \text{if} P_1 = P_2. \end{cases}$$

If C is an elliptic curve described by (13.1) then we can set $c = d = 0$, and the formulæ simplify further.

13.4. Abel's method of proving addition formulæ

The method of addition on curves in the previous section was used by Abel to establish addition formulæ for the elliptic functions. We consider the case of the addition formula for the Weierstrass elliptic function: [8]

THEOREM 13.4. *If $z_1 + z_2 + z_3 = 0$, then the addition formula for the Weierstrass elliptic function is*

$$\begin{vmatrix} \wp(z_1) & \wp'(z_1) & 1 \\ \wp(z_2) & \wp'(z_2) & 1 \\ \wp(z_3) & \wp'(z_3) & 1 \end{vmatrix} = 0.$$

Geometric proof (Whittaker & Watson, 1927). We consider the intersections of the cubic curve [9]

$$y^2 = 4x^3 - g_2 x - g_3,$$

with a variable line

$$y = mx + n.$$

If we take a point (x_1, y_1) on the cubic, then the equation

$$\wp(z) - x_1 = 0$$

has two solutions, $-z_1$ and z_1, parameters of (x_1, y_1) on the cubic. All other solutions are congruent to these two.

Now, since $\wp'^2(z) = 4\wp^3(z) - g_2\wp(z) - g_3$, we have

$$\wp'^2(z_1) = y_1^2,$$

---

[7] Koblitz (1987) 204.

[8] Analogous geometric proofs of the addition formulæ for the Jacobi elliptic functions use either the intersections of a twisted curve defined by the equations $x^2 + y^2 = 1$ and $z^2 + k^2x^2 = 1$ with a variable plane $lx + my + nz = 1$, or alternatively the intersections of the curve $y^2 = (1 - x^2)(1 - k^2x^2)$ with a variable curve $y = 1 + mx + nx^2$.

[9] Whittaker & Watson (1927) 442–443. The original proof can be found in Abel (1881).

and choose z_1 to be the solution for which $\wp'(z_1) = y_1$. We denote the intersections of the cubic with the variable line by x_1, x_2 and x_3. These are the roots of

$$\varphi(x) = 4x^3 - g_2x - g_3 - (mx + n^2)^2 = 4(x - x_1)(x - x_2)(x - x_3) = 0.$$

The variations in δx_1, δx_2 and δx_3 of x_1, x_2 and x_3 due to the position of the line are a consequence of small changes δm and δn in the coefficients m and n. These can be described by the equation

$$\varphi'(x_i)\delta x_i + \frac{\partial \varphi}{\partial m}\delta n = 0,$$

where $i = 1, 2, 3$. Hence,

$$\varphi'(x_i)\delta x_i = 2(mx_i + n)(x_i\delta m + \delta n),$$

and it follows that

$$\sum_{i=1}^{3} \frac{\delta x_1}{mx_i + n} = 2\sum_{i=1}^{3} x_i \frac{\delta m + \delta n}{\varphi'(x_i)},$$

provided that $\varphi'(x_i) \neq 0$.

Now, we consider the expression

$$\frac{x(x\delta m + \delta n)}{\varphi(x)}.$$

By putting this into partial fractions, we have

$$\frac{x(\delta m + \delta n)}{\varphi(x)} = \sum_{i=1}^{3} \frac{A_i}{(x - x_i)},$$

where

$$\begin{aligned} A_i &= \lim_{x \to x_i} x(x\delta m + \delta n)\frac{x - x_i}{\varphi(x)} \\ &= x_i(x_i\delta m + \delta n) \lim_{x \to x_i} \frac{x - x_i}{\varphi(x)}, \end{aligned}$$

for $i = 1, 2, 3$. By applying Taylor's theorem, we have

$$A_i = \frac{x_i(x\delta m + \delta n)}{\varphi'(x_i)}.$$

If we put $x = 0$, it follows that

$$\sum_{i=0}^{3} \frac{\delta x_i}{y_i} = 0.$$

Therefore, we have that the sum of the parameters z_1, z_2 and z_3 is constant and independent of the position of the line.

By varying the line so that all the points of intersection move off to infinity, we can see that $z_1 + z_2 + z_3$ is equal to the sum of the parameters when the line is at infinity. However, when the line is at infinity, each of z_1, z_2 and z_3 is a period of $\wp(z)$ and it clearly follows $z_1 + z_2 + z_3$ is also a period of $\wp(z)$.

Hence, the sum of the parameters of three collinear points on the cubic is congruent to zero.

□

Cryptosystem	*Mathematical Problem*	*Method for Solving*
Integer factorization	Given a number n, find its prime factors	Number field sieve
Discrete logarithm	Given a prime n, and numbers g and h, find x such that $h = g^x \pmod{n}$	Number field sieve
Elliptic curve discrete logarithm	Given an elliptic curve C, and points P and Q on C, find x such that $Q = xP$	Pollard-rho algorithm

Table 13.1. Comparison of various cryptosystems.

13.5. Elliptic curve cryptography [10]

The idea of cryptography using elliptic curves was first suggested in mid-1980s as an alternative to the more traditional systems based on the factoring of primes (see Table 13.1). The main advantage of elliptic curve cryptography lies in the difficulty [11] of the underlying mathematical problem, that of the factoring of elliptic curves. The best known way to solve this problem is fully exponential and so, compared to other systems, substantially smaller key sizes can be used to obtain equivalent strengths. Moreover, the elliptic curve cryptosystem is considered to require smaller bandwidth requirements and so may be the system of choice in an emerging trend towards mobile computing.

The essence of elliptic curve cryptography is that the plain text to be encrypted is embedded in points P_m on an elliptic curve over a field $GF(q)$, for large q. Repeated applications of the group law, or more specifically the formulæ (13.3), are used to calculate multiples of these points which are then transmitted. There are various ways that this process is carried out, we describe two of these.

Massey-Omura system. Suppose that a user A would like to send a message m, embedded as points of an elliptic curve, to a user B. Given our curve C of order N, and our embedding P_m, user A chooses a random integer c satisfying $0 < c < N$ and $\gcd(c, N) = 1$. User A then calculates cP_m as described above and transmits this to B. Next, B chooses a random integer d satisfying the same properties as c, and similarly calculates and transmits $d(cP_m)$ back to A. Finally, A transmits $c'(dcP_m) = dP_m$ where $c'c \equiv 1 \pmod{N}$ back to B. To decrypt the message, B computes $d'dP_m$ for $d'd \equiv 1 \pmod{N}$ to recover P_m.

ElGamal system. This is similar to that of Massey-Omura except that we do not need to know N. We assume the same construction as before, except that we need an additional publicly known point Q on C. Now, user B chooses any integer a and publishes aG. To transmit the message m, user A chooses a further random integer k, and sends to B the pair of points $(kG, P_m + k(\{aG\})$. In this case, to decrypt the message, user B multiplies the first point of the pair by a, and then subtracts the result from the second point of the pair, thus recovering P.

To break either system requires the solution of the elliptic curve analogue of the discrete logarithm problem:

[10] Koblitz (1987) 203–209.

[11] *Elliptic curve cryptography* is more correctly termed *elliptic curve discrete logarithm cryptography*, though it is usually abbreviated *ECC* or *EC*^{2}.

Given an elliptic curve C defined over $GF(q)$ and two points P and Q in C, find an integer x such that $Q = xP$, if such x exists.

As the existing methods for solving the classical discrete logarithm problem depend on a finite Abelian group, they can also be applied to the elliptic curve analogue. However, these methods are generally much slower because of the added complexity of the addition operation, and the fact that only those attacks aimed at a general group have so far proven to be of any use.

Chapter 14

The Nine Circles Theorem

A lesser known elementary gem

SERGE TABACHNIKOV

A *chain of circles* is a sequence $S_1, S_2, \ldots, S_n$ of circles in which each circle touches both its neighbours. To be precise, S_1 touches S_2, S_2 touches $S_3, \ldots, S_{n-1}$ touches S_n. A chain is said to be *closed* if the last circle S_n touches the first circle D_1. From this definition it may seem surprising that a geometric problem so simple and easily visualized could ever require the use of elliptic functions to obtain proofs, but this is indeed the case. Perhaps even more surprising, it is the Weierstrass elliptic function that is used.

14.1. A triangle and six circles

THEOREM 14.1. *The Money-Coutts theorem. Let $A_1A_2A_3$ be a triangle in the plane and let S_1 be any circle which touches the sides A_3A_1 and A_1A_2 of the triangle. Consider then the following triangular chain of circles: S_2 is a circle touching A_1A_2, A_2A_3 and S_1, S_3 is a circle touching A_2A_3, A_3A_1 and S_2, S_4 is a circle touching A_1A_2, A_2A_3 and S_3, S_5 is a circle touching A_1A_2, A_2A_3 and S_4, S_6 is a circle touching A_2A_3, A_3A_1 and S_5, S_7 is a circle touching A_3A_1, A_1A_2 and S_6. There are various choices available at each stage, but if the choice at each stage is appropriately made, then the last circle S_7 coincides with the first circle S_1 and we have a closed chain.*

To reduce the number of choices at each stage, we will first add the further constraint that each of the circles should lie within the triangle. This gives us at most one choice for each successive circle.

Proof (Evelyn et al.*, 1974).* Let a, b and c denote the lengths of sides A_2A_3, A_3A_1 and [1]
A_1A_2 of the triangle and let l, m, n, l', m', n' and l'' denote the lengths of the tangents to S_1, S_2, S_3, S_4, S_5, S_6 and S_7 from A_1, A_2, A_3, A_1, A_2, A_3 and A_1, respectively.

We begin by relating l and m and then extend this relation to link each of the other tangents.

Figure 14.2 shows that if X and Y are the centres of S_1 and S_2, and J and K are their points of contact with A_1A_2, we have $A_1J = l$ and $A_2K = m$. Since X is on the

[1] Evelyn *et al.* (1974) 49–54. An alternative proof can be found in Maxwell (1971).

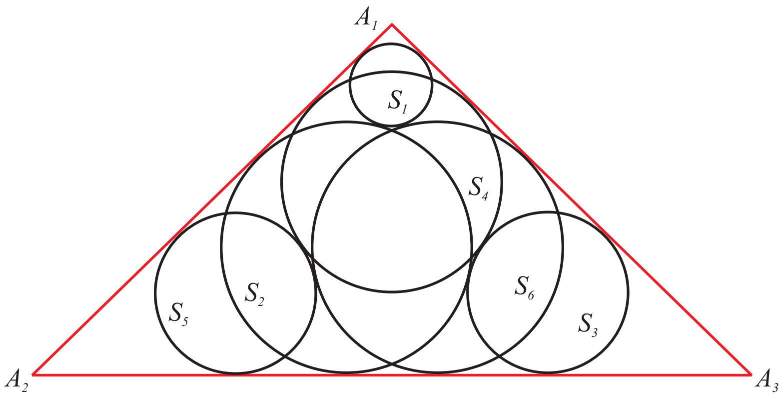

Figure 14.1. The Money-Coutts theorem.

internal bisector of the angle A_1, we have that the radius of S_1 is

$$JX = l \tan \tfrac{1}{2} A_1.$$

Similarly, the radius of S_2 is

$$KY = m \tan \tfrac{1}{2} A_2.$$

Now, from elementary geometry, if two circles of radii r_1 and r_2 touch externally, then the length of each common tangent is $2\sqrt{r_1 r_2}$, so we have

$$JK = 2\sqrt{lm \tan \tfrac{1}{2} A_1 \tan \tfrac{1}{2} A_2}.$$

As $A_1A_2 = A_1J + A_2K + JK$, it follows that

$$c = l + m + 2\sqrt{lm \tan \tfrac{1}{2} A_1 \tan \tfrac{1}{2} A_2}. \tag{14.1}$$

If we define $s = \frac{1}{2}(a + b + c)$, then by the half angle formulæ for a triangle we have

$$\tan \tfrac{1}{2} A_1 \tan \tfrac{1}{2} A_2 = \sqrt{\frac{(s-b)(s-c)}{s(s-a)}} \sqrt{\frac{(s-c)(s-a)}{s(s-b)}} = \frac{s-c}{s}.$$

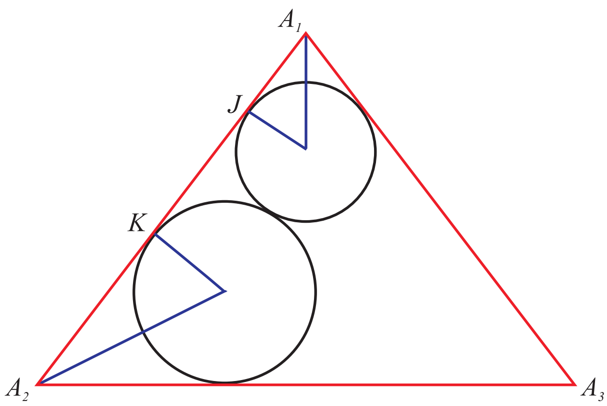

Figure 14.2. Obtaining a relation between tangents to the first two circles.

T_1	T_2	T_3	T_4	T_5	T_6
h	f	g	h	f	g
p	q	r	p'	q'	r'
q	r	p'	q'	r'	p''
ψ	θ	φ	ψ	θ	φ

Table 14.1. Relations between lengths and angles in the proof of the Money-Coutts theorem.

Then (14.1) reduces to

$$c = l + m + 2\sqrt{mn(s-c)/s}. \tag{14.2}$$

Thus, we obtain a relation between l and m. The same argument can be applied to the other circles to give relations connecting m and $n, \dots, n'$ and l'' and we have

$$a = m + n + 2\sqrt{mn(s-c)/s},$$

and so on, to give six relations in all.

We now write p, q, r, p', q', r' and p'' for the positive square roots of l, m, n, l', m', n' and l'', and, similarly, f, g, h and t for the square roots of a, b, c, s. We then define obtuse angles θ, φ and ψ by

$$\begin{aligned}
\cos\theta &= -\sqrt{1-\frac{a}{s}} = -\sqrt{1-\frac{f^2}{t^2}}, \quad \sin\theta = \frac{f}{t},\\
\cos\varphi &= -\sqrt{1-\frac{b}{s}} = -\sqrt{1-\frac{g^2}{t^2}}, \quad \sin\varphi = \frac{g}{t},\\
\cos\psi &= -\sqrt{1-\frac{c}{s}} = -\sqrt{1-\frac{h^2}{t^2}}, \quad \sin\psi = \frac{h}{t}.
\end{aligned} \tag{14.3}$$

This is perfectly valid as the quantities under the square root are positive and between zero and 1. Therefore, (14.2) reduces to

$$h^2 = p^2 + q^2 - 2pq\cos\psi. \tag{14.4}$$

The five other equations can be constructed by substituting for h, p, q and ψ as indicated in Table 14.1.

Now, (14.1), describes the cosine rule so we can construct a triangle T_1 with sides p, q, h and an angle ψ opposite to h. Similarly, we have T_2, T_3, T_4, T_5 and T_6 corresponding to the other relations in Table 14.1.

Once more using elementary geometry, we have that the diameter of the circumcircle of a triangle is equal to the length of any side divided by the sine of the opposite angle. Therefore, for T_1 we have diameter $h/\sin\varphi$, which by (14.3) is equal to t. Moreover, it is similarly easy to see that the circumradii of the other triangles is also t.

From this it follows that if we fit the triangles T_1, T_2 so that their sides of length q coincide, then their circumcircles also coincide. If we proceed in a similar fashion we can fit all six triangles together into the same circumcircle. By removing certain lines we can see in Figure 14.3 that we have an open heptagon $X_0X_1\dots X_7$ inscribed in a circle, with angles ψ at X_2 and X_5, angles θ at X_2 and X_4, and angles φ at X_3 and X_6.

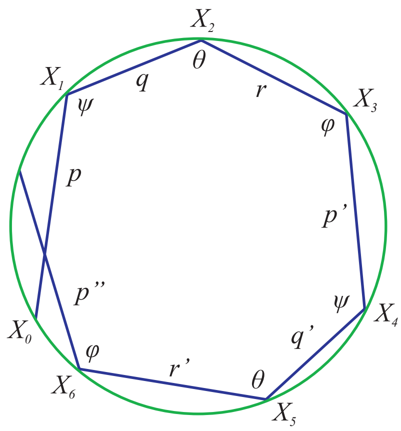

Figure 14.3. Fitting all six triangles into the same circumcircle.

Now, the triangles $X_0X_2X_4$ and $X_3X_5X_7$ are congruent, so the angles $X_0X_2X_4$ and $X_3X_5X_7$ are equal. Subtracting these angles from the equal angles $X_1X_2X_3$ and $X_4X_5X_6$ we have

$$\angle X_0X_2X_1 + \angle X_3X_2X_4 = \angle X_3X_5X_4 + \angle X_6X_5X_7.$$

But $X_3X_2X_4$ and $X_3X_5X_4$ are also equal so it follows that

$$\angle X_0X_2X_1 = \angle X_6X_5X_7.$$

Therefore, X_0X_1 and X_6X_7 subtend angles at the circumference of the circle, so these chords are equal in length.

Hence, $p = p''$, and it follows that $l = l''$.

□

Now, we relax the condition that circles lie inside the triangle, and instead only require that their centres lie on the internal bisectors of the respective angles of the triangle. In this case, for any given choice of S_1, there are three possible choices for S_2, two of which may not be real.

Proof (Evelyn et al., 1974). We first consider the case that we choose S_2 such that it 2
touches A_1A_2 at the same point as S_1, and so touches S_1 internally. If we make this special choice, then we can also choose S_3 in a similar way so that it touches A_2A_3 at the same point as S_2. Building up a complete chain by making the analogous choice for each successive circle S_4, S_5, S_6 and S_7, it follows that it necessarily closes up (see Figure 14.4).

Using the same notation as in the earlier proof, we immediately have

$$\begin{aligned} l + m &= c, & l + m' &= c, \\ m + n &= a, & m' + n &= a, \\ n + l' &= m, & n' + l'' &= b. \end{aligned}$$

[2] Evelyn *et al.* (1974) 54–58.

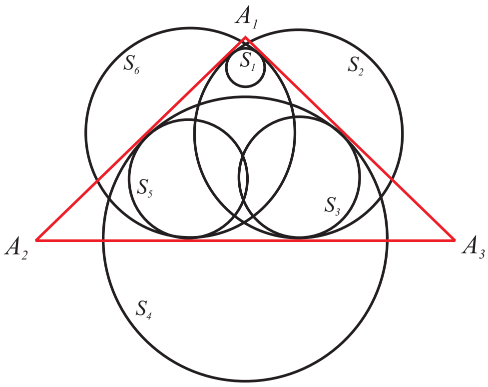

Figure 14.4. A chain constructed by taking special choices at each stage.

Hence, $l = l''$.

Now, the other two choices for S_2 are real if and only if S_2 does not lie entirely outside the triangle. With the preceding notation, we have that l and m still satisfy (14.2), though we must now consider both the positive and negative square roots. These correspond to the two possibilities for S_2. Writing (14.1) in the form

$$s(c - l - m)^2 = 4(s - c)lm, \tag{14.5}$$

then, as l is given, this is a quadratic for m and has real roots lying between zero and s if and only if l lies between zero and s.

If we construct a triangular chain, taking one of these two choices for the following circle, then after six steps we appear to have 64 possibilities for S_7 for any given S_1. However, not all of these possibilities will coincide with S_1. To determine which choices will give a closed chain, we can use the method of the previous proof, or alternatively consider the following argument.

We replace l, m, n, l', m', n' and l'' by $\lambda, \mu, \nu, \lambda', \mu', \nu'$ and λ'' defined by

$$l = s\sin^2\lambda, \quad m = s\sin^2\mu, \ldots, l'' = s\sin^2\lambda'',$$

where $\lambda, \mu, \ldots, \lambda''$ are determined modulo π. In terms of these new variables, (14.5) reduces to

$$\cos 2\mu - \cos(2\lambda \pm 2\psi),$$

where $\cos\psi = -\sqrt{1 - (c/s)}$.

Therefore,

$$\mu = \pm\lambda \pm \psi \pmod{\pi},$$

and similarly,

$$\nu = \pm\mu \pm \theta \pmod{\pi}, \ldots, \lambda'' = \pm\nu' \pm \varphi \pmod{\pi}.$$

Hence,

$$\lambda'' = \pm\lambda + (\pm\psi \pm \theta \pm \varphi \pm \psi \pm \theta \pm \varphi) \pmod{\pi},$$

where successive choices of sign correspond with the successive choices of circle. In order for the chain to close up, we can make the first three choices arbitrarily, but must then choose the final three so that the terms cancel in pairs to give $\lambda'' = \pm\lambda$.

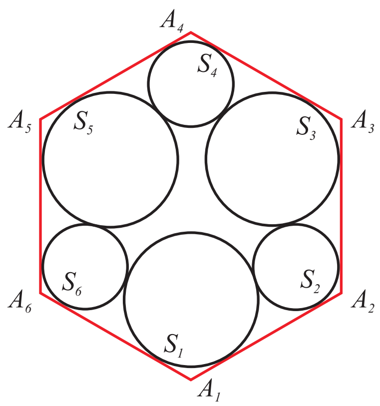

Figure 14.5. The Money-Coutts theorem for pentagons.

It follows that of the 64 possible chains in this case, we have eight which will close up. If we have our original constraint that the chain of circles lies inside the triangle then that chain must be one of these eight.

□

14.2. Polygons and circles

We next consider whether Theorem 14.1 extends to polygons other than triangles. An example of this would be a regular n-gon with vertices $A_1, A_2, \ldots, A_n$. By symmetry, the circles S_{i-1} and S_{i+1} are congruent for all $i = 1, 2, \ldots, n$. It follows that for such a regular n-gon we have S_1 coincides with S_{2n+1} if n is odd, and S_1 coincides with S_{n+1} if n is even. This behaviour is shown in Figure 14.5.

Unfortunately, Figure 14.6 illustrates that this periodicity is destroyed for certain perturbations of a regular n-gon. However, there is a subclass of irregular n-gons for which periodicity holds. We let $A_1, A_2, \ldots, A_n$ be the vertices of a convex n-gon P, and let 2α be the interior angle at A_i. We also let $a_i = |A_i A_{i+1}|$. As in the first

Figure 14.6. Perturbation of a regular pentagon.

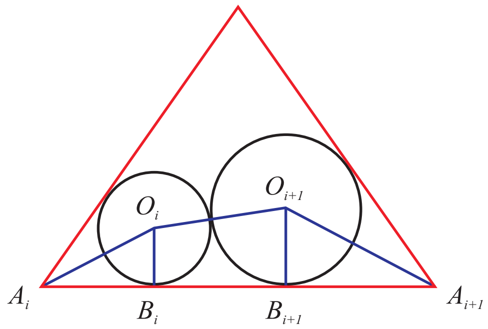

Figure 14.7. Fixing the choice of circles at each stage.

part of our proof of Theorem 14.2, we inscribe a circle S_1 in the angle of the vertex A_i for $i = 1, 2, \ldots, n$, and inscribe a further circle S_1 in A_2, tangent to S_1, and then continuing cyclically. If we let O_i be the centre of S_i and r_i its radius, we fix the choice of circles at each step by assuming that the orthogonal projections B_i and B_{i+1} of O_i and O_{i+1} on the line A_iA_{i+1} fall on the segment A_iA_{i+1}, and that B_i is closer to A_i than B_{i+1}. This is illustrated in Figure 14.7.

We can now define the class of n-gons for which Theorem 14.1 extends. Assume that $n \geq 5$ and $\alpha_i + \alpha_{i+1} > \frac{1}{2}\pi$ for all $i = 1, 2, \ldots, n$. Let D_i be the intersection of the
lines $A_{i-1}A_i$ and $A_{i+1}A_{i+2}$. Referring to Figure 14.8, we consider the escribed circles 3
of the triangles $A_{i-1}A_iD_{i-1}$ and $A_iA_{i-1}D_i$ that are tangent to the sides A_iD_{i-1} and A_iD_i, respectively. If these circles coincide for all i, then we have a *nice n-gon*.

THEOREM 14.2. *Let P be a nice n-gon. If n is odd then the sequence of circles S_i is $2n$-periodic: $S_1 = S_{2n+1}$. If n is even assume that*

$$\prod_{i=1}^{n}(1 + \sqrt{1 - \cot\alpha_i \cot\alpha_{i+1}})^{(-1)^i} = 1, \tag{14.6}$$

then the sequence of circles is n-periodic: $S_1 = S_{n+1}$.

Proof (Tabachnikov, 2000). In Figure 14.7, we have 4

$$B_iB_{i+1} = 2\sqrt{r_ir_{i+1}}, \quad |A_iB_i| = r_i \cot\alpha_i, \quad |A_{i+1}B_{i+1}| = r_{i+1}\cot\alpha_{i+1},$$

and it follows that

$$r_i \cot\alpha_i + 2\sqrt{r_ir_{i+1}} + r_{i+1}\cot\alpha_{i+1} = a_i. \tag{14.7}$$

If we now introduce the variables

$$u_i = \sqrt{r_i \cos\alpha_i}, \quad e_i = \sqrt{\tan\alpha_i \tan\alpha_{i+1}}, \quad c_i = \sqrt{a_i}, \tag{14.8}$$

where $e_i > 1$, then (14.7) becomes

$$u_i^2 + 2e_iu_iu_{i+1} + u_{i+1}^2 = c_i^2.$$

[3]The circle tangent to two lines extended from non-adjacent sides of a given triangle, and also to the other side of the triangle is known as an *escribed circle*.

[4]Tabachnikov (2000) 202–206.

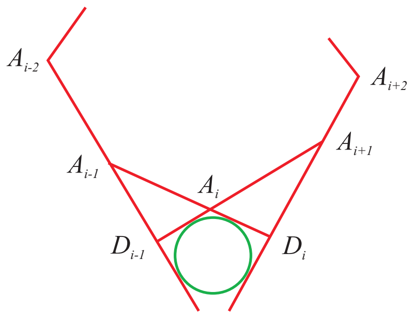

Figure 14.8. Defining nice n-gons.

Solving this in hyperbolic functions, we have

$$u_i\sqrt{c_i^2 + (e_i^2 - 1)u_{i+1}^2} + u_{i+1}\sqrt{c_i^2 + (e_i^2 - 1)u_i^2} = c_i. \tag{14.9}$$

We introduce *niceness* by the following result:

THEOREM 14.3. *An n-gon P is a nice if and only if there exists a constant $\rho > 0$ such that*

$$a_i = \rho^2(\tan\alpha_i \tan\alpha_{i+1} - 1),$$

for all $i - 1, 2, \ldots, n$.

Proof (Tabachnikov, 2000). We consider the escribed circle, of radius r_1, of the triangle $A_i A_{i+1} D_i$ shown in Figure 14.9. By elementary geometry, we have [5]

$$|EA_i| = r_i \cot\alpha_i, \quad |EA_i + 1| = r_i \tan\alpha_{i+1}.$$

It follows that

$$r_i \tan\alpha_{i+1} = r_i \cot\alpha_i + a_i,$$

and we obtain

$$\frac{a_i}{\tan\alpha_i \tan\alpha_{i+1} - 1} = \frac{r_1}{\tan\alpha_i}. \tag{14.10}$$

[5] Tabachnikov (2000) 204.

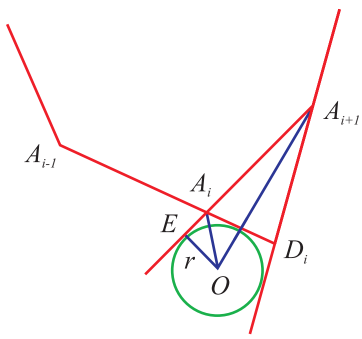

Figure 14.9. Proving a theorem about nice n-gons.

Similarly, let r_2 be the radius of the escribed circle of the triangle $A_{i-1}A_iD_{i-1}$, then

$$\frac{a_{i-1}}{\tan\alpha_{i-1}\tan\alpha_i - 1} = \frac{r_2}{\tan\alpha_i}. \tag{14.11}$$

The n-gon is nice if and only if $r_1 = r_2$, so, by (14.10) and (14.11), this corresponds to

$$\frac{a_i}{\tan\alpha_i\tan\alpha_{i+1}}$$

being independent of i, as required.

□

By application of this theorem, and by (14.8), we have

$$a_i = \rho^2(\tan\alpha_i\tan\alpha_{i+1} - 1), \quad c_i^2 = \rho^2(e_i^2 - 1).$$

Therefore, (14.9) can be written as

$$\frac{u_i}{\rho}\sqrt{1 + \frac{u_{i+1}^2}{\rho^2}} + \frac{u_{i+1}}{\rho}\sqrt{1 + \frac{u_i^2}{\rho^2}} = \frac{c_i}{\rho}.$$

If we let $x_i = \sinh^{-1}(u_i/\rho)$, then we have

$$\sinh(x_i + x_{i+1}) = c_i/\rho. \tag{14.12}$$

Now, we denote the family of circles inscribed in the ith angle of P by F_i. Then we have a map $T_i : F_i \longrightarrow F_{i+1}$ that takes S_i to S_{i+1}. Using x_i as a coordinate in F_i, then by (14.12) we have that T_i is the reflection

$$T_i(x_i) = x_{i+1} = \sinh^{-1}(c_i/\rho) - x_i.$$

If n is odd, the map $T_nT_{n-1}\dots T_1 : F_1 \longrightarrow F_1$ is a reflection, and its second iteration is the identity. If n is even, then the same map is the translation taking x_1 to

$$x_1 + \sum(-1)^i\sinh^{-1}(c_i/\rho).$$

By Theorem 14.3, the vanishing of this alternating sum is equivalent to (14.6).

□

If $e_i = 1$ in (14.8), then we have the case that the $(i-1)$st and $(i+1)$st sides of the n-gon are parallel. Therefore, we have

$$u_i + u_{u+1} = c_i.$$

It follows that if P is a parallelogram, then the sequence of circles is not periodic, unless P is a rhombus, in which case $S_1 = S_5$. In fact, we have the following result:

THEOREM 14.4. *Suppose P is a parallelogram, then, for an initial circle S_1, the sequence of circles S_i is preperiodic with eventual period 4.* [6]

[6] Troubetzkoy (2000) 289.

If we instead consider an open quadrilateral, then we may instead have that the behaviour of sequence of circles becomes chaotic. We can use the concept of the topological entropy of the map between circles in the sequence to define this. This is a topological invariant that measures the dynamical complexity of a map. A definition of chaos that follows is positivity of topological entropy. Therefore, we can state the following result:

THEOREM 14.5. *There is an open set of quadrilaterals for which the circle map has* 7
positive topological entropy.

14.3. The nine circles theorem

Suppose that the sides of the triangle in Theorem 14.1 are made into arcs of circles. This chain, illustrated in Figure 14.10, is still periodic, which we can state as follows:

THEOREM 14.6. *The nine circles theorem. Let C_1, C_2 and C_3 be three circles in general position in the plane and let S_1 be any circle which touches C_1 and C_2. Consider then the following chain of circles: S_2 is a circle touching C_2, C_3 and S_1; S_3 is a circle touching C_3, C_1 and S_2; S_4 is a circle touching C_1, C_2 and S_3; S_5 is a circle touching C_2, C_3 and S_4; S_6 is a circle touching C_3, C_1 and S_5; S_7 is a circle touching C_1, C_2 and S_6. There are a finite number of choices available at each stage, but if the choice at each stage is appropriately made, then the last circle S_7 coincides with the first circle S_1 and we have a closed chain.*

To summarize, we have a symmetric system where each circle is touched by four others. We can represent this by the following square array, in which two circles touch if and only if they do not appear in the same column:

$$\begin{array}{ccc} C_1 & C_2 & C_3 \\ S_5 & S_3 & S_1 \\ S_2 & S_6 & S_4 \end{array}$$

To describe the appropriate choices mentioned in the theorem, we first need the fact that given three circles in general position, eight circles can be described to touch all three. However, in this case, two of the three given circles already touch, and so this is reduced to six, two counting doubly and four simply. For example, there are two circles of the coaxal family determined by C_2 and S_1 which also touch C_3. These are 8
the two choices which count doubly, and we can describe them as special choices for S_2. The four other possible choices for S_2 will be referred to as general choices. We can make this distinction at each stage in the construction of the chain.

---

[7] Troubetzkoy (2000) 290.

The nine circles theorem was discovered by a group of friends, including amateur mathematicians G. B. Money-Coutts and C. J. A. Evelyn, who used to meet in the cafés of Piccadilly in London to discuss geometry. They first conjectured the special case of a triangle and six circles given in Theorem 14.1, and then, by careful drawing, found that by replacing each side of the triangle with a circle, the more general result of Theorem 14.6 appeared also to be true. As a result of this, this arrangement of circles has been named a Money-Coutts configuration. The proof using elliptic functions was formulated by a third member of the party, professional mathematician J. A. Tyrrell and his research student M. T. Powell.

[8] Circles which share a radical axis with a given circle are said to be *coaxal*. The centres of coaxal circles are collinear.

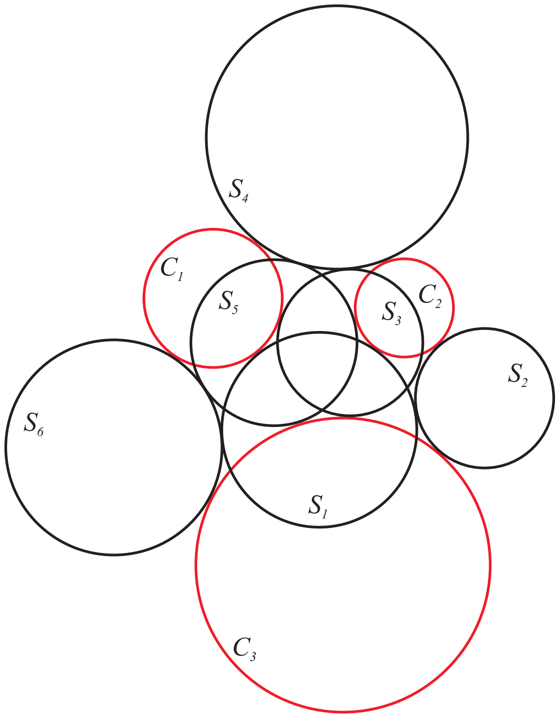

Figure 14.10. The nine circles theorem.

Now, if a circle S is drawn to touch two given circles, then the line joining the
two points of contact necessarily passes through one of the centres of similitude of the 9
given circles, and so we may say that S belongs to that centre of similitude. Of the six possible choices for S_2, there are three choices (one special and two general) belonging to each centre of similitude.

Finally, we know that the centres of similitude of three circles, taken in pairs, are the six vertices of a complete quadrilateral. Using this and the earlier facts, we can complete the statement of Theorem 14.6 as follows:

Select three collinear centres of similitude (one for each pair of circles C_1, C_2 and C_3), insisting that at every stage in the construction of the chain, the circle S_i belongs to the appropriate one of these fixed centres of similitude. If at each stage we make the special choice then the chain will close up. If we make the general choice, and if S_2, S_3 and S_4 are arbitrarily chosen, then it is always possible to choose S_5 and S_6 so that the chain closes up.

The consequence of this is that, for a given position of S_1, there are eighteen chains which close up, of which two are special and sixteen general.

Proof (Tyrrell & Powell, 1971). Let Q be a projective model of the system of circles 10

[9]The point of intersection of the two external common tangents is called the *centre of similitude* of the two circles.

in the plane. This is a non-singular quadric surface, and we have a birational correspondence between Q and the plane, in which plane sections of Q correspond to the [11] circles in the plane. Importantly, if two plane sections of Q touch then the corresponding circles in the plane also touch. Therefore, if we formulate the theorem in terms of plane sections of Q, then by proving this reformulation as a theorem of complex projective geometry, then the original theorem about circles automatically follows. Now, the rule of choice for the plane sections, corresponding to the rule in the plane covering the three fixed collinear centres of similitude, is to select three of the size cones with collinear vertices, and, every stage in the construction of the chain, to choose a section which touches the appropriate one of these cones.

We next consider a system of homogeneous coordinates $[x, y, z, w]$ in Q defined such that the planes of C_1, C_2 and C_3 are given by $x = 0$, $y = 0$ and $z = 0$, respectively, and in which $w = 0$ is the polar plane of the point of intersection of the planes C_1, C_2 and C_3. By the hypothesis of generality, we can assume that the planes of C_1, C_2 and C_3 are linearly independent and that the coefficients of x^2, y^2 and z^2 are non-zero. Therefore, by an appropriate choice of unit point, the equation of Q reduces to the form

$$x^2 + y^2 + z^2 - 2yz\cos\alpha - 2zx\cos\beta - 2xy\cos\gamma - w^2 = 0,$$

where α, β, γ are constants. It follows that the equations of the pairs of cones through (C_1, C_2), (C_2, C_3) and (C_3, C_1) are respectively

$$\begin{aligned}
&x^2 + y^2 + z^2 - 2yz\cos\alpha - 2zx\cos\beta + 2xy\cos(\alpha \pm \beta) - w^2 = 0,\\
&x^2 + y^2 + z^2 + 2yz\cos(\beta \pm \gamma) - 2zx\cos\beta - 2xy\cos\gamma - w^2 = 0,\\
&x^2 + y^2 + z^2 - 2yz\cos\alpha + 2zx\cos(\gamma \pm \alpha) - 2xy\cos\gamma - w^2 = 0.
\end{aligned} \tag{14.13}$$

These are collinear if we take all three choices of sign as negative, or if we take two as positive and one as negative. Since changing the sign of one of the constants is equivalent to changing two of the three sign alternatives, and does not effect the equation of Q, then without loss of generality, we can take all three to be negative.

[10] Tyrrell & Powell (1971) 70–74.

[11] It is easier to visualize the *birational correspondence* as being similar to the stereographic projection in Figure 14.11, but in a higher dimension (see also Baker (1925)). The *stereographic projection* is a map that projects each point on the sphere onto a plane tangent to the south pole along a straight line from the north pole.

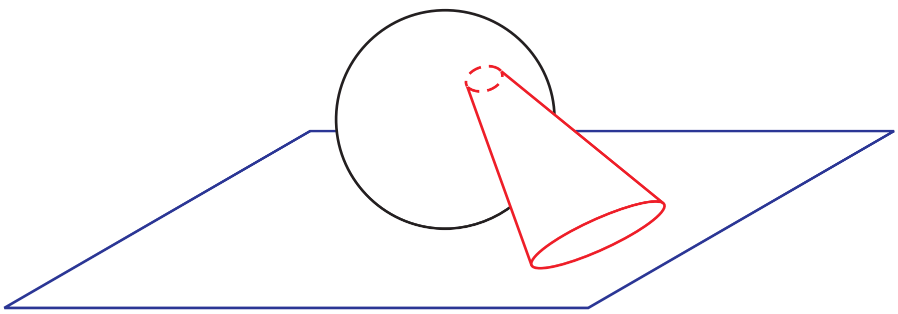

Figure 14.11. The projective model.

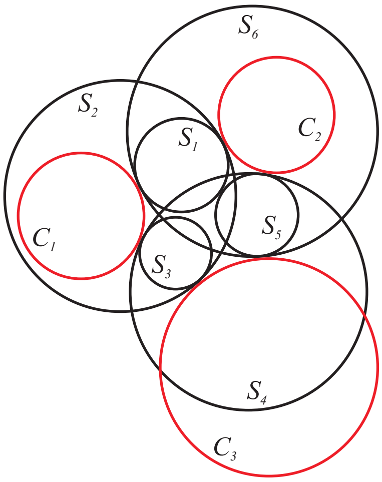

Figure 14.12. A special chain.

We now parameterize the sections C_1, C_2 and C_3 in the circular forms

$$\begin{aligned} C_1 &: (0, \cos\theta, \cos(\theta-\alpha), \sin\alpha), \\ C_2 &: (\cos(\varphi-\beta), 0, \cos\varphi, \sin\beta), \\ C_3 &: (\cos\psi, \cos(\psi-\gamma), 0, \sin\gamma), \end{aligned} \tag{14.14}$$

identifying the sections $S_1, S_2, \ldots, S_7$ by the parameters θ, φ, ψ, θ', φ', ψ' and θ'' of the points at which they touch C_1, C_2, C_3, C_1, C_2, C_3 and C_1, respectively. Hence, the tangent plane to the cone (14.13) at the point (14.14) is given by the equation

$$x\sin(\theta-\beta) + y\sin(\theta-\alpha) - z\sin\theta + w = 0, \tag{14.15}$$

and represents the plane of a variable section S_1 touching C_1 at the point (14.14) and C_2 at the point with parameter $\varphi = \beta - \theta$. Similarly, we have that any section touching C_2 and C_3 is cut by a plane of the form

$$-x\sin\varphi + y\sin(\varphi-\alpha) + z\sin(\varphi-\beta) + w^2 = 0. \tag{14.16}$$

This is the plane of a possible position for S_2 if we choose φ so that the sections given by (14.15) and (14.16) touch.

Now, the condition on φ resulting for our wish that these sections touch is reducible. One factor gives $\varphi = \beta - \theta$ and corresponds to the special choice for S_2, while from the other factor we obtain

$$\cos(\theta+\varphi-\beta) - 2(\cos\beta + \rho\sin\beta)\cos\theta\cos\varphi = 1, \tag{14.17}$$

where $\rho = \tan\frac{1}{2}(\alpha+\beta+\gamma)$. For a given θ, this gives the values of φ which correspond to the general choices for S_2.

We first consider the case in which the special choice is taken at each stage, an example of which is Figure 14.12. Permuting the parameters cyclically we have $\theta' = \alpha - \psi$, and, therefore, that $\theta' = \alpha - \gamma + \beta - \theta$. A further three steps in the chain gives us a position of S_7 that coincides with S_1.

For the general case, we use rational projective parameters instead of the circular θ, φ and ψ previously employed. We can use any bilinear function of $\tan \frac{1}{2}\theta$ as a rational [12] parameter on C_1, and in this case we choose a parameter t on C_1 related to θ by

$$\tan \tfrac{1}{2}\theta = \frac{t + 2\rho - 3}{t + 2\rho + 3}. \tag{14.18}$$

Similarly, we choose rational parameters u on C_2 and v on C_3 given in terms of φ and ψ, respectively, by formulæ analogous to (14.18). Therefore, (14.5) reduces to the form

$$(t + u + b)(4btu - g_3) = (tu + bt + bu + \tfrac{1}{4}g_2)^2, \tag{14.19}$$

where $g_2 = 48\rho^2 + 36$, $g_3 = 64\rho^3 + 72\rho$ and $b = 3\tan \frac{1}{2}\beta - 2\rho$.

We now need the following result:

THEOREM 14.7. *If $\wp(z_1) = p_1$, $\wp(z_2) = p_2$ and $\wp(z_3) = p_3$ and $z_1 + z_2 + z_3 = 0$, then the addition formula for the Weierstrass elliptic function is*

$$(p_1 + p_2 + p_3)(4p_1p_2p_3 - g_3) = (p_1p_2 + p_2p_3 + p_3p_1 + \tfrac{1}{4}g_2)^2. \tag{14.20}$$

Applying this, we therefore have that (14.19) is a form of the addition formula for the Weierstrass elliptic function and so can be written in the transcendental form

$$\wp^{-1}(u) \equiv \wp^{-1}(t) \pm \wp^{-1}(b) \pmod{\Omega}, \tag{14.21}$$

where Ω is the period lattice of $\wp(u)$.

As $\wp(u)$ is an even function of u of order two, $\wp^{-1}(u)$ is a two-valued function with equal and opposite values. Therefore, for a given t, the right hand side of (14.21) has four possible values modulo Ω. These are two equal and opposite pairs, and so u is determined as a two-valued function of t.

Now, the reduction to the form in (14.21) gives us that the constants g_2 and g_3 on which $\wp(u)$ depends are functions of ρ only. As ρ is also an expression symmetric in α, β and γ then it follows, by cyclically permuting the letters, that the parameters v and t' of the sections S_3 and S_4 are given by

$$\wp^{-1}(v) \equiv \pm\wp^{-1}(u) \pm \wp^{-1}(c),$$
$$\wp^{-1}(t) \equiv \pm\wp^{-1}(v) \pm \wp^{-1}(a),$$

where $\wp(u)$ is the same elliptic function as before, and where $c = 3\tan \frac{1}{2}\gamma - 2\rho$ and $a = 3\tan \frac{1}{2}\alpha - 2\rho$. Therefore, we have

$$\wp^{-1}(t') \equiv \pm\wp^{-1}(t) \pm \wp^{-1}(a) \pm \wp^{-1}(b) \pm \wp^{-1}(c), \tag{14.22}$$

with the eight equal and opposite pairs of values of the right hand side giving us t' as an eight-valued function of t. Similarly, the parameter t'' is an eight-valued function of the parameter t' of S_4, and is given by

$$\wp^{-1}(t'') \equiv \pm\wp^{-1}(t') \pm \wp^{-1}(a) \pm \wp^{-1}(b) \pm \wp^{-1}(c). \tag{14.23}$$

[12] A function of two variables is *bilinear* if it is linear with respect to each variable.

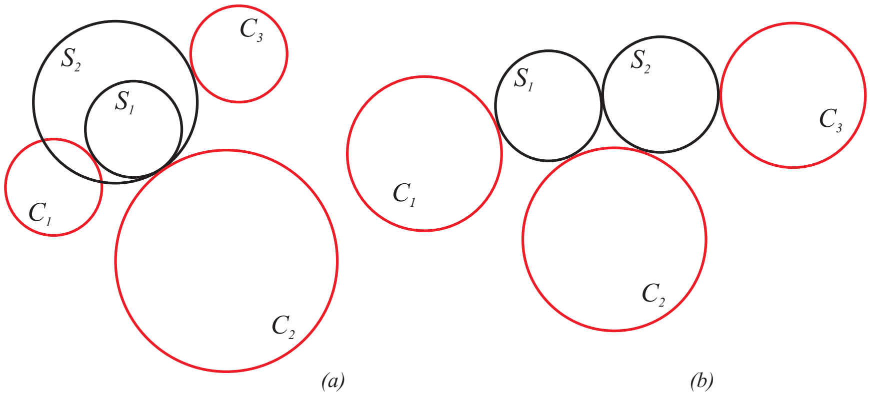

Figure 14.13. Irreducible components of the curve consisting of pairs of circles (S_1, S_2)*. These are rational curves (a) and an elliptic curve (b).*

If we choose signs arbitrarily in (14.22), it is possible to choose signs in (14.23) to obtain $\wp^{-1}(t'') \equiv \wp^{-1}(t)$, from which it follows that $t'' = t$. This tells us that, for a given S_1, if we choose S_2, S_3 and S_4 arbitrarily at each stage, then it is possible to choose S_5, S_6 and S_7 so that S_7 coincides with S_1.

□

Now, suppose we have three families of circles F_1, F_2, F_3, which respectively touch the pairs of circles (C_1, C_2), (C_2, C_3) and (C_3, C_1). If we consider the curve consisting of pairs of circles (S_1, S_2) in $F_1 \times F_1$ such that S_1 is tangent to S_2, then it has two irreducible components that are rational curves of bidegree $(1, 1)$ and another [13] irreducible component that is an elliptic curve of bidegree $(2, 2)$. These are illustrated in Figure 14.13. Therefore, we have elliptic curves E_1 in $F_1 \times F_2$, E_2 in $F_2 \times F_3$ and E_3 in $F_3 \times F_1$. This allows us to restate Theorem 14.6 as follows:

THEOREM 14.8. *Suppose that circles S_1 in F_1, S_2 in F_2, S_3 in F_3 and S_4 in F_4 are chosen such that (S_1, S_2) is in E_1, (S_2, S_3) is in E_2, and (S_3, S_1) is in E_3. Then it is possible to choose circles S_5 in F_2, S_6 in F_3, and S_7 in F_1 in such a way that (S_4, S_5) is in E_1, (S_5, S_6) is in E_2, and (S_6, S_7) is in E_3, and such that S_7 coincides with S_1.*

Proof (Barth & Bauer, 1996). We denote by π_1 and π_1' the projections of E_1 onto F_1 [14] and F_2, respectively, and similarly denote the projections of E_2 onto F_2 and F_3, and E_3 onto F_3 and F_1 by π_2, π_2', π_3 and π_3', respectively, having used the following result:

THEOREM 14.9. *Given F_1 and F_2, the family F_3 can be chosen in such a way that the projections π_1', π_2 as well as π_2', π_3 and π_3', π_1 have the same branch points F_2, F_3 and F_1, respectively.*

Hence, there are isomorphisms $\varphi_1 : E_1 \longrightarrow E_2$, $\varphi_2 : E_2 \longrightarrow E_3$ and $\varphi_3 : E_3 \longrightarrow E_1$ such that $\pi_2\varphi_1 = \pi_1'$, $\pi_3\varphi_2 = \pi_2'$ and $\pi_1\varphi_3 = \pi_3'$.

We define the curve E by identifying $E_1 = E_2 = E_3$ using φ_1 and φ_2. It follows that φ_3 becomes an automorphism τ of E such that $\pi_1\tau = \pi_3'$. Since the elliptic curve E is determined by the intersection points of C_1, C_2 and two Apollonian circles of C_1, [15]

[13] The *bidegree* of a non-zero polynomial $f(x, y) = \sum_{m,n} a_{ij} x_i y_j$ is the maximal (i, j) such that $a_{ij} \neq 0$.

[14] Barth & Bauer (1996) 16–20.

C_2 and C_3, then, for general C_1, C_2 and C_3, the curve E is also general. Therefore, the automorphism τ is either a translation or an involution. By an appropriate choice of φ_1 [16] and φ_2, we have that τ is a translation.

Now, we let σ_1, σ_2 and σ_3 be the involutions of E that interchange the preimages of π_1, π_2 and π_3, respectively. We have $S_1 = \pi_1(e)$ for some e in E, and then $S_2 = \pi_1'(\alpha_1 e)$ where α_1 is either the identity or σ_1. Continuing in a similar way, we have

$$S_4 = \pi_3'(\alpha_3\alpha_2\alpha_1 e) = \pi_1(\tau\alpha_3\alpha_2\alpha_1 e),$$

where α_2 and α_3 are defined in an analogous way to α_1.

Now, we determine the circles S_5, S_6 and S_7 such that (S_4, S_5), (S_5, S_6) and (S_6, S_7) are in E. It follows that

$$S_7 = \pi_3'(\beta_3\beta_2\beta_1\tau\alpha_3\alpha_2\alpha_1 e) = \pi_1(\tau\beta_3\beta_2\beta_1\tau\alpha_3\alpha_2\alpha_1 e),$$

where $\beta_2 = \alpha_2$, $\beta_3 = \alpha_3$ and we choose β_1 such that $\beta_3\beta_2\beta_1$ is an involution. Therefore, we obtain

$$\tau\beta_3\beta_2\beta_1\tau\alpha_3\alpha_2\alpha_1 e = \beta_3\beta_2\beta_1\alpha_3\alpha_2\alpha_1 e = (\alpha_3\alpha_2\beta_1)^2\beta_1\alpha_1 e = \beta_1\alpha_1 e.$$

Hence,

$$S_7 = \pi_1(\beta_1\alpha_1 e) = \pi_1(e) = S_1.$$

□

14.4. A biquadratic six cycle

In *The nine circles theorem* (Lyness, 1973), an attempt is made find an elementary proof of Theorem 14.6 in the case that radii of the circles C_1, C_2 and C_3. It relies upon the existence of a biquadratic six cycle

$$T_i = (1+k)t_i^2 - 2kt_i \cot\alpha_i + (1-k), \quad T_i T_{i+1} = (t_i + t_{i+1})^2,$$

in which given t_0, there are two values of t_1, four values of t_2 and eight values of t_3. Similarly, given t_0, the cycle leads to two values of t_{-1} from which follows eight values of t_{-3} that are equal to those of t_3. Theorem 14.1 is then formulated with an

---

[15] *Apollonian circles theorem. Let A and B be two distinct points in the plane, and let k be a positive real number not equal to 1. Then the locus of points P that satisfies the ratio $PA/PB = k$ is a circle whose centre lies on the line through A and B.*

If we include the point at infinity in the locus, then for every positive value k we have a generalized circle known as an *Apollonian circle*, named after the geometer Apollonius of Perga (*c* 262–190 BC).

[16] An *involution* is a map that is its own inverse.

Perhaps evident from his frequent contributions to the *Mathematical Gazette*, Lyness was a school teacher and chief inspector of schools for HMI. He was cited as an influence in the early academic life of Sir John Anthony Pople (1925–2004), the 1998 Nobel Laureate in Chemistry. In an autobiographical piece in *Les Prix Nobel* (Frängsmyr, 1998), Pople writes

> *My grades outside of mathematics and science were undistinguished so I usually ended up several places down in the monthly class order. This all changed suddenly three years later when the new senior mathematics teacher, R. C. Lyness, decided to challenge the class with an unusually difficult test. I succumbed to temptation and turned in a perfect paper, with multiple solutions to many of the problems.*

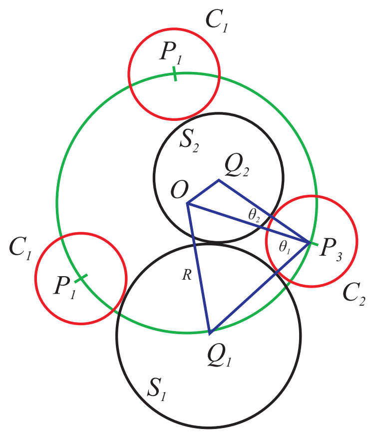

Figure 14.14. A biquadratic six cycle.

initial circle S_0 touching C_3 and C_1 which in turn touches S_1 in the chain. It follows there are a set of eight possible circles S_3, and if we have a similarly constructed set S_{-3}, then by proving that if the two sets are we prove the theorem.

Unfortunately, while partly verified by computer, the biquadratic six cycle in question has not been proven. However, referring to Figure 14.14, we consider the first two positive terms of the cycle. 17

Let the centres of C_1, C_2 and C_3 be P_1, P_2 and P_3, respectively, and the common value of their radii be r. Further, suppose that the centre and radius of the circle $P_1P_2P_3$ are O and R, respectively. If denote by Q_1 and Q_2 the centres of S_1 and S_2, then Q_1 lies on the perpendicular bisector of P_1P_2, and Q_2 on the perpendicular bisector of P_3P_1. Letting $\angle OP_1Q_1 = \theta_1$ and $\angle Q_2P_1O = \theta_2$ so that $t_1 = \tan\frac{1}{2}\theta_1$ and $t_2 = \tan\frac{1}{2}\theta_2$, then we can apply elementary geometry to obtain

$$T_1T_2 = (t_1 + t_2)^2,$$

where $T_i = (1+k)t_i^2 - 2kt_i \cot\alpha_i + (1-k)$ and $k = r/R$.

14.5. Tangent and anti-tangent cycles

Suppose that we now have oriented circles. We refer to these as *cycles*. In this context, we consider two cycles to be *tangent* if they are oriented in the same direction at their point of tangency, while if they are oriented in opposite directions, then they are *anti-tangent*.

In the case of tangent cycles, we have the following result:

THEOREM 14.10. *Let H_1 and H_2 be two cycles tangent at P, and let H_3 be a cycle not tangent to either H_1 or H_2. Then there exists just one cycle tangent to H_1 and H_2 (at P) and H_3.*

This is easily proven by inversion, mapping the point P to the point at infinity. 18

[17] Lyness (1974).

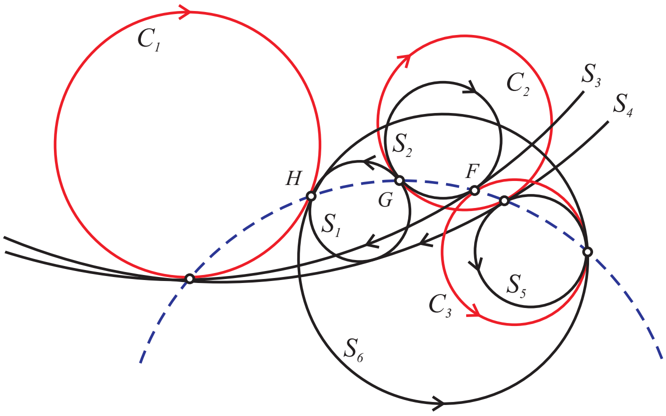

Figure 14.15. Nine tangent cycles.

It allows us to state a special case of Theorem 14.1 in terms of tangent cycles (Figure 14.15).

THEOREM 14.11. *Let C_1, C_2 and C_3 be three cycles in general position in the plane and let S_1 be any cycle tangent to C_1 and C_2. Consider then the following uniquely defined chain of cycles: S_2 is a cycle tangent to C_2, C_3 and S_1; S_3 is a cycle tangent to C_3, C_1 and S_2; S_4 is a cycle tangent to C_1, C_2 and S_3; S_5 is a cycle tangent to C_2 C_3 and S_4; S_6 is a cycle tangent to C_3, C_1 and S_5; S_7 is a cycle tangent to C_1, C_2 and S_6. Then the last cycle S_7 coincides with the first cycle S_1.*

The uniqueness of the choices of cycle at each step clearly follows from Theorem 14.10. The proof of this result is simplified by the following observation:

THEOREM 14.12. *The six points of tangency of the circles in Theorem 14.11 are concyclic.* 19

Proof of Theorems 14.11 and 14.12 (Rigby, 1981a). We invert Figure 14.15 so that the 20
points labelled F, G and H become collinear to obtain Figure 14.16. If we let C_1, C_2 and C_3 meet the line FGH again at K, L and M then clearly all of the angles θ are equal. Hence, we have points of tangency of S_3 at H and K; S_4 at K and L; S_5 at L and M; and S_6 at M and F.

□

The restatement in terms of anti-tangent cycles, as shown in Figure 14.17, is as follows:

THEOREM 14.13. *Let C_1, C_2 and C_3 be three cycles in general position in the plane* 21
and let S_1 be any cycle anti-tangent to C_1 and C_2. Consider then the following chain

[18] Suppose S is a circle with centre O and radius r, and let P be any point except O. If P' is the point on a line OP that lies on the same side of O as P and satisfies the equation $OP.OP = r^2$, then P' is known as the *inverse* of P with respect to S. The point O is the *centre of inversion*, while S is the *circle of inversion*. The map $P \mapsto P'$, for all P not equal to O, is known as *inversion*.

[19] A set of points which all lie on the same circle are said to be *concyclic*.

[20] Rigby (1981a) 361–363.

[21] Rigby (1981b) 110–132.

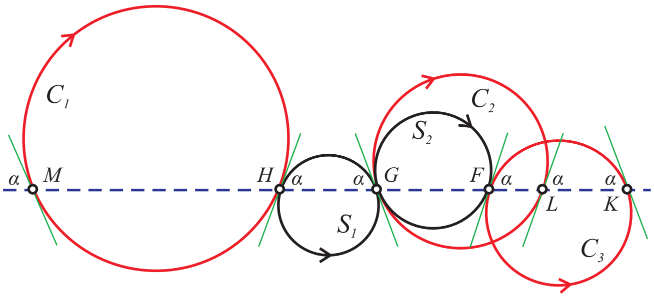

Figure 14.16. Inversion of nine tangent cycles.

of cycles: S_2 is a cycle anti-tangent to C_2, C_3 and S_1; S_3 is a cycle anti-tangent to C_3, C_1 and S_2; S_4 is a cycle anti-tangent to C_1, C_2 and S_3; S_5 is a cycle anti-tangent to C_2 C_3 and S_4; S_6 is a cycle anti-tangent to C_3, C_1 and S_5; S_7 is a cycle anti-tangent to C_1, C_2 and S_6. Then the choices for the last cycle S_7 coincide with the choices for the first cycle S_1.

In this case we do not have a unique chain, and the points of tangency are not concyclic, so we have the general case given in Theorem 14.1. It follows that the proof is considerably more complex than the case of tangent cycles. However, while not entirely elementary, appealing to small amount of complex analysis, it is the only complete proof that does not use elliptic functions or elliptic curves.

We can conclude with a related problem involving 45 circles. If we have a chain of nine circles chosen in the same way as Theorem 14.1, then it is possible, in certain circumstances, to combine this with nine others to form a chain of 45 circles touching in threes at 60 points.

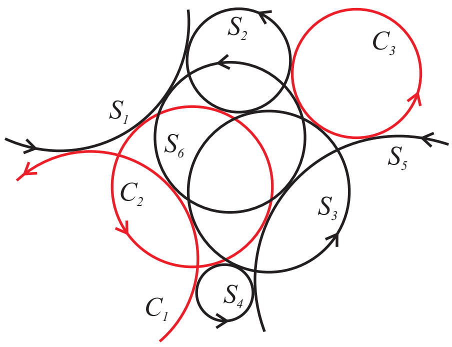

Figure 14.17. Nine anti-tangent cycles.

Bibliography

Articles

Ayoub, R. G. (1984). The lemniscate and Fagnano's contributions to elliptic integrals. *Archive for History of Exact Sciences* **29** (2) 131–149.

Barth, W. & Bauer, T. (1996). Poncelet theorems. *Expositiones Mathematicæ* **14** (2) 125–144.

Bottazzini, U. (2002). 'Algebraic truths' vs. 'geometric fantasies': Weierstrass' response to Riemann. *Proceedings of the International Congress of Mathematicians* **3** 923–934.

Caelli, W. J., Dawson, E. D. & Rea, S. A. (1999). PKI, elliptic curve cryptography, and digital signatures. *Computers and Security* **18** (1) 47–66.

Erdös, P. (2000). Spiraling the earth with C. G. J. Jacobi. *American Journal of Physics* **68** (10) 888–895.

Koblitz, N. (1987). Elliptic curve cryptosystems. *Mathematics of Computation* **48** (177) 203–209.

Littlewood, J. E. (1929). Collected papers of Srinivasa Ramanujan. *The Mathematical Gazette* **14** (200) 425–428.

Love, A. E. H. (1928). Alfred George Greenhill. *Journal of the London Mathematical Society* **3** (1) 27–32.

Lyness, R. C. (1974). A biquadratic 6-cycle. *The Mathematical Gazette* **58** (404) 142. Mathematical Note 3355.

Maxwell, E. A. (1971). Circles modulo 6. *The Mathematical Gazette* **55** (391) 61–62. Mathematical Note 3293.

Rigby, J. F. (1981a). The geometry of cycles, and generalized Laguerre inversion. C. Davis, B. Grübaum & F. A. Sherk (eds.) *The Geometric Vein: The Coxeter Festschrift*, 355–379. Springer-Verlag, New York, United States.

Rigby, J. F. (1981b). On the Money-Coutts configuration of 9 anti-tangent cycles. *Proceedings of the London Mathematical Society* **43** (1) 110–132.

Tabachnikov, S. (2000). Going in circles: Variations on the Money-Coutts theorem. *Geometriæ Dedicata* **80** (1–3) 201–209.

Troubetzkoy, S. (2000). Circles and polygons. *Geometriæ Dedicata* **80** (1–3) 289–296.

Tyrrell, J. A. & Powell, M. T. (1971). A theorem in circle geometry. *Bulletin of the London Mathematical Society* **3** (1) 70–74.

Vanstone, S. A. (2003). Next generation security for wireless: Elliptic curve cryptography. *Computers and Security* **22** (5) 412–415.

Zinn-Justin, P. & Zuber, J.-B. (2000). On the counting of colored tangles. *Journal of Knot Theory and Its Ramifications* **9** (8) 1127–1141.

Books

Abel, N. H. (1881). *Œuvres complètes*. Nouvelle édition. Grøndahl & Søn, Christiana (Oslo), Norway.

Baker, H. F. (1925). *Principles of Geometry: Higher Geometry*. Cambridge University Press, Cambridge, United Kingdom.

Bowman, F. (1953). *Introduction to Elliptic Functions with Applications*. English Universities Press, London, United Kingdom.

Brannan, D. A., Esplen, M. F. & Gray, J. J. (1999). *Geometry*. Cambridge University Press, Cambridge, United Kingdom.

Britannica (2003). *Encyclopædia Britannica*. Britannica, Chicago, Illinois, United States.

Cassels, J. W. S. (1991). *Lectures on Elliptic Curves*. Number 24 in London Mathematical Society Student Texts. Cambridge University Press, Cambridge, United Kingdom.

Copson, E. T. (1935). *An Introduction to the Theory of Functions of a Complex Variable*. Clarendon Press, Oxford, United Kingdom.

Du Val, P. (1973). *Elliptic Functions and Elliptic Curves*. Number 9 in London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, United Kingdom.

Durell, C. V. (1928). *Modern Geometry: The Straight Line and Circle*. Macmillan, London, United Kingdom.

Dutta, M. & Debnath, L. (1965). *Elements of the Theory of Elliptic and Associated Functions with Applications*. The World Press Private, Calcutta (Kolkata), India.

Evelyn, C. J. A., Money-Coutts, C. B. & Tyrrell, J. A. (1974). *The Seven Circles Theorem and Other New Theorems*. Stacey International, London, United Kingdom.

Frängsmyr, T. (1998). *Les Prix Nobel*. Almqvist & Wiksell International, Stockholm, Sweden.

Greenhill, A. G. (1892). *The Applications of Elliptic Functions*. Macmillan, London, United Kingdom.

Halphen, G.-H. (1886). *Traité des fonctions elliptiques et de leurs applications*. Gauthier-Villars, Paris, France.

Jones, G. A. & Singerman, D. (1987). *Complex Functions: An Algebraic and Geometric Viewpoint*. Cambridge University Press, Cambridge, United Kingdom.

Lawden, D. F. (1989). *Elliptic Functions and Applications*. Number 80 in Applied Mathematical Sciences. Springer-Verlag, New York, United States.

Markushevich, A. I. (1992). *Introduction to the Classical Theory of Abelian Functions*. Number 96 in Translations of Mathematical Monograms. American Mathematical Society, Providence, Rhode Island, United States.

Maxwell, E. A. (1949). *Geometry for Advanced Pupils*. Clarendon Press, Oxford, United Kingdom.

McKean, H. & Moll, V. (1999). *Elliptic Curves: Function Theory, Geometry, Arithmetic*. Cambridge University Press, Cambridge, United Kingdom.

Poncelet, J.-V. (1865). *Traité des propriétés projectives des figures*. Gauthier-Villars, Paris, France.

Prasolov, V. & Solovyev, Y. (1997). *Elliptic Functions and Elliptic Integrals*. Number 170 in Translations of Mathematical Monograms. American Mathematical Society, Providence, Rhode Island, United States.

Reid, M. A. (1988). *Undergraduate Algebraic Geometry*. Number 12 in London Mathematical Society Student Texts. Cambridge University Press, Cambridge, United Kingdom.

Smart, W. M. (1977). *Textbook on Spherical Astronomy*. Sixth edition. Cambridge University Press, Cambridge, United Kingdom.

Whittaker, E. T. & Watson, G. N. (1927). *A Course of Modern Analysis*. Fourth edition. Cambridge University Press, Cambridge, United Kingdom.

Unpublished

Lyness, R. C. (1973). *The Nine Circles Theorem*. Letter to J. V. Armitage, University of Nottingham, Nottingham, United Kingdom.

Nekovář, J. (2004). *Elliptic Curves and Modular Forms: A Classical Introduction*. Lecture course at Université Pierre et Marie Curie, Paris, France.

Serre, J.-P. (1991). *Mathematical Writing*. Letter to D. Goss, Ohio State University, Columbus, Ohio, United States.