Lectures on Selected Topics in Mathematical Physics: Elliptic Functions and Elliptic Integrals

Lectures on Selected Topics in Mathematical Physics: Elliptic Functions and Elliptic Integrals

William A Schwalm

Department of Physics and Astrophysics,
University of North Dakota, USA

Morgan & Claypool Publishers

ISBN 978-1-6817-4230-4 (ebook)
ISBN 978-1-6817-4166-6 (print)
ISBN 978-1-6817-4102-4 (mobi)

DOI 10.1088/978-1-6817-4230-4

Version: 20151201

IOP Concise Physics
ISSN 2053-2571 (online)
ISSN 2054-7307 (print)

A Morgan & Claypool publication as part of IOP Concise Physics
Published by Morgan & Claypool Publishers, 40 Oak Drive, San Rafael, CA, 94903, USA

IOP Publishing, Temple Circus, Temple Way, Bristol BS1 6HG, UK

We dedicate these lectures to the late Professor Philip Gold of Portland State University, and to the memory of his stimulating lectures on continuous groups and other areas of applied mathematics and physics, which we attended as graduate students.

Contents

Preface

We provide lecture based units on topics in the mathematical methods of physics. The purpose is to accompany a set of good quality video lectures with text materials that correspond closely. Emphasis will be on the topics that, while important and interesting, are more apt not to appear these days in the standard courses, just because of time constraints. In this way the material is intended to augment and supplement rather than to replace course content in a modern course in theoretical methods. This volume is a basic introduction to certain aspects of elliptic functions and elliptic integrals. Primarily, the elliptic functions stand out as closed solutions to a class of physical and geometrical problems giving rise to nonlinear differential equations. While these nonlinear equations may not be the types of greatest interest currently, the fact that they are solvable exactly in terms of functions about which much is known makes up for this. The elliptic functions of Jacobi, or equivalently the Weierstrass elliptic functions, inhabit the literature on current problems in condensed matter and statistical physics, on solitons and conformal representations, and all sorts of famous problems in classical mechanics.

The lectures on elliptic functions have evolved as part of the first semester of a course on theoretical and mathematical methods given to first and second year graduate students in physics and chemistry at the University of North Dakota. They are for graduate students or for researchers who want an elementary introduction to the subject that nevertheless leaves them with enough of the details to address real problems. The style is supposed to be informal. The intention is to introduce the subject as a moderate extension of ordinary trigonometry in which the reference circle is replaced by an ellipse. This entre depends upon fewer tools and has seemed less intimidating than other typical introductions to the subject that depend on some knowledge of complex variables. The first three lectures assume only calculus, including the chain rule and elementary knowledge of differential equations. In the later lectures, the complex analytic properties are introduced naturally so that a more complete study becomes possible.

William A Schwalm
August, 2015

Bibliographical notes

The approach to elliptic functions as trigonometry of an ellipse is familiar to a few of my colleagues, but is difficult to find in print. The material presented here I learned from lectures by William M Kinnersley [1]. A recent, detailed exposition of the general development is *Elliptic Functions* by Armatage and Eberlein [2]. This is interesting, very readable and is recommended. A classical reference is *A Course of Modern Analysis* (CMA) by Whittaker and Watson [3], especially chapter XXII, where however the development depends heavily upon the material on theta functions in the previous chapter. The two-letter notations for the 12 Jacobi functions evolved as follows, according to Whittaker and Watson ([3] p 494). The two letter notations, *sn*, *cn* and *dn* were introduced by Gudermann [4] to replace *sinam*, *cosam* and used by Jacobi [5]. The other two-letter abbreviations, as well as the general ratio convention for defining them, was introduced by Glaisher [6]. In fact, Gudermann had used *tn* for *sc*. Two popular, concise books on the Jacobi elliptic functions are the ones by Milne-Thomson [7] and Bowman [8]. Useful material on the reduction of the general elliptic integral is found in the book by Baker [9]. Finally, it is more common nowadays to write the Jacobi functions as depending on the squared modulus m, where $m = k^2$ rather than on k itself. So, for instance one writes $\mathrm{sn}(u|m)$ rather than $\mathrm{sn}(u, k)$, where if $k > 0$, $\mathrm{sn}(u|k^2) = \mathrm{sn}(u, k)$.

Further reading

[1] Kinnersley W M 1972 Class notes for lectures (Montana State University)

[2] Armitage J A and Eberlein W F 2006 *Elliptic Functions* London Mathematical Society Student Texts 26 (Cambridge: Cambridge University Press)

[3] Whittaker E K and Watson G N 1958 *A Course of Modern Analysis* (Cambridge: Cambridge University Press)

[4] Gudermann C 1838 *Journal für Math* XVIII pp 12–20 (CMA p 494)

[5] Jacobi K 1827 Communication to H C Schumacher, who published extracts in *Astronomiche Nachtrichen* VI **123** (CMA p 512)

[6] Glaisher J W L 1882 *Messenger of Mathematics* XI p 86 (CMA p 494)

[7] Milne-Thomson L M 1950 *Jacobian Elliptic Function Tables* (New York: Dover)

[8] Bowman F 1953 *Introduction to Elliptic Functions* (Chichester: Wiley)

[9] Baker A L 1890 *Elliptic Functions: An Elementary Text-Book for Students of Mathematics* (New York: Wiley) pp 4–14

Acknowledgements

I am grateful to my teachers, in particular Dr Philip Gold and Dr William Kinnersley. It was in a lecture by Dr Kinnersley circa 1972 that I first saw the Jacobian elliptic functions developed as trigonometry. Thanks to Scott Crocket of LANL for suggesting the project in the first place providing motivation, encouragement and consultation throughout the process. I owe a great deal to my wife and collaborator, Mizuho Schwalm for organizing the lectures presented here, working through the material to correct errors in the calculations, and preparing the LaTeX manuscripts. Mizuho also recorded the lectures and edited the videos. Without her, the project would not have been possible at all. I must really thank some of my colleagues and graduate students for attending the lecture presentations, providing questions and corrections. In particular I want to thank my colleagues Yen Lee Loh, Timothy Young, and our students including Cody Rude, Albert Schmitz, Rajesh Dhakal, Pragalv Karki, Prasadh Kalawila Vinthanage, Madina Sultanova, Mohamed Radwan, Faisal Amir and Rasika Mohottige for active participation. We thank the University of North Dakota for providing space, including excellent blackboards, and for the use of recording equipment.

Author biography

William A Schwalm

Dr William A Schwalm was born in Portsmouth, New Hampshire, USA and spent his boyhood in the town of Greenland near the Atlantic coast. He received a Bachelor of Science degree in physics from The University of New Hampshire and PhD in condensed matter theory from Montana State University. Dr Schwalm is currently Professor of Physics in the Department of Physics and Astrophysics at The University of North Dakota, Grand Forks, where he has been employed since moving to North Dakota from Salt Lake City in 1980, where he held a postdoctoral appointment in the Department of Physics at The University of Utah. His research has specialized mathematical analysis of physical problems. Most recent publications pertain to groups and dynamical systems. He has held visiting professorships at Montana State University, The University of Minnesota, the University of Rome (Sapienza – Università di Roma) and Kyoto University. His teaching assignment normally includes a graduate two-semester course in methods of theoretical physics.

Lectures on Selected Topics in Mathematical Physics: Elliptic Functions and Elliptic Integrals

William A Schwalm

Chapter 1

Elliptic functions as trigonometry

This introduction to the Jacobi elliptic, *sn*, *cn*, *dn* and related functions is parallel to the usual development of trigonometric functions, except that the unit circle is replaced by an ellipse. These functions satisfy nonlinear differential equations that appear often in physical applications, for instance in particle mechanics. We will look at some applications and show methods for reducing certain differential equations to standard forms where they can be solved via elliptic functions or via their relatives, the elliptic integrals.

1.1 Definition of Jacobian elliptic functions and trigonometric identities

An ellipse is the locus of points P such that the sum of distances $r_1 + r_2$ from two foci has some fixed value. Another way to describe it is as a stretched circle. If I stretch a circle along the axes so that the semi major and semi minor axes are a and b respectively, it must somehow fix both the positions of the foci and the value of the sum $r_1 + r_2$ such that

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1.$$

So the distance $2c$ between the foci must be determined. To find c, imagine moving point P. When P is on the x axis to the right, say, then $r_1 = a + c$, $r_2 = a - c$, $r_1 + r_2 = 2a$.

Now moving P up to the y axis I see that $r_1 = r_2 = a$ so that $a^2 = b^2 + c^2$. The eccentricity is the ratio $k = c/a$, which is also called the *modulus* of the associated elliptic functions.

For normal trigonometry it is handy to define functions with respect to the unit circle. In the elliptic case one normalizes the y intercept to $b = 1$. Thus

doi:10.1088/978-1-6817-4230-4ch1

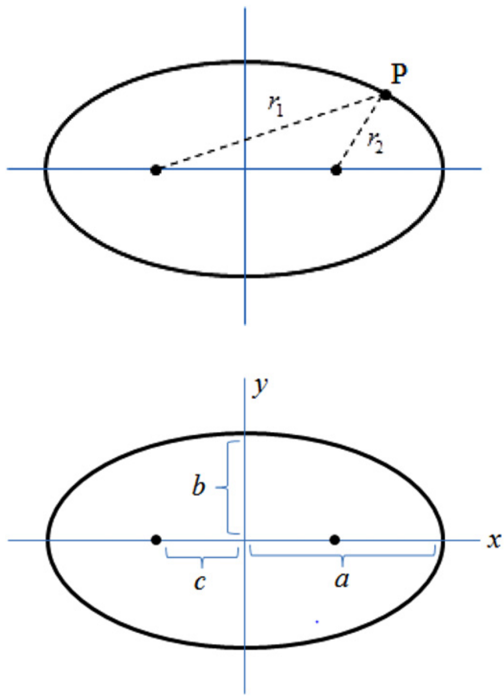


Figure 1.1.

$$\text{modulus} = k = \frac{\sqrt{a^2 - 1}}{a}. \tag{1.1}$$

Because the radius is not constant, the central equation of trigonometry on the unit circle, namely $x^2 + y^2 = 1$, is replaced in the elliptic case by two equations

$$\frac{x^2}{a^2} + y^2 = 1, \tag{1.2}$$

$$x^2 + y^2 = r^2. \tag{1.3}$$

Accordingly, there are three basic functions, denoted as *sn*, *cn* and *dn*, corresponding to the two functions, sine and cosine of the circular case. Putting off just for a moment the question of what exactly these are functions of (which turns out *not* to be the angle θ) the values are defined by the ratios

$$sn = y, \qquad cn = \frac{x}{a}, \qquad dn = \frac{r}{a}.$$

In fact, these are functions of two variables, since the each must depend on the modulus k that determines the shape of the ellipse.

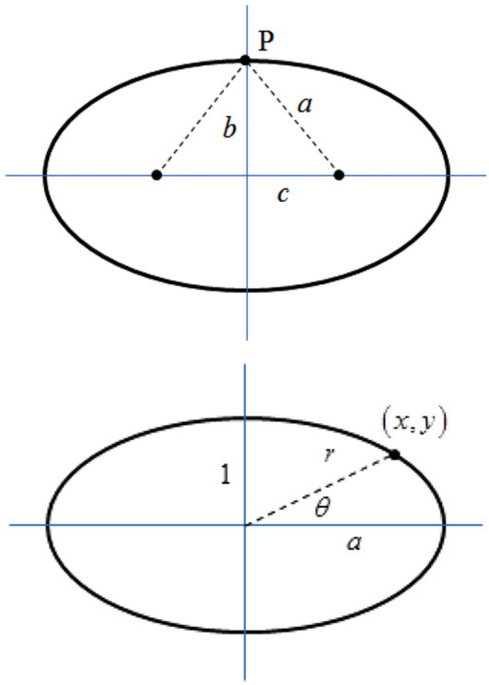


Figure 1.2.

The other argument u that determines the position of the point P, once the ellipse is determined, must correspond in some way to the angle θ. How it does so exactly is the main point of the development. It turns out that the argument of the elliptic functions is

$$u = \int_0^P r\mathrm{d}\theta, \qquad \text{so } \mathrm{d}u = r\mathrm{d}\theta. \tag{1.4}$$

For an angle in the first quadrant, the upper limit could be $\sin^{-1} y$, for instance. This is not intuitive, as far as I can see. It is not either the arc length or the area swept out. So including both arguments, the full notation should be

$$\mathrm{sn}(u, k) = y, \qquad \mathrm{cn}(u, k) = \frac{x}{a}, \qquad \mathrm{dn}(u, k) = \frac{r}{a}. \tag{1.5}$$

One often omits mention of the modulus, as long is the context makes clear what value k should have. Sometimes in lecture I even omit both arguments when deriving general relationships. So quite often I will use simply the italic notations *sn*, *cn* and *dn* rather than the full function notation.

From equation (1.2) we have

$$cn^2 + sn^2 = 1, \tag{1.6}$$

which generalizes $\cos^2\theta + \sin^2\theta = 1$. And then from equation (1.3)

$$dn^2 + k^2 sn^2 = 1. \tag{1.7}$$

These two algebraic relations take the palace of the single trigonometric $\sin^2 + \cos^2 = 1$ in the circular case.

The next step is to find derivatives. In order to get du start with

$$\theta = \tan^{-1}(y/x).$$

$$\frac{\mathrm{d}\left(\frac{y}{x}\right)}{\mathrm{d}\theta} = \frac{1}{x}\frac{\mathrm{d}y}{\mathrm{d}\theta} - y\frac{1}{x^2}\frac{\mathrm{d}x}{\mathrm{d}\theta}.$$

But

$$\frac{\mathrm{d}(\tan\theta)}{\mathrm{d}\theta} = 1 + \tan^2\theta.$$

Thus one has

$$\mathrm{d}\theta = \frac{1}{r^2}(x\mathrm{d}y - y\mathrm{d}x),$$

and

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

To differentiate sn one must replace dx, using equation (1.2)

$$\mathrm{d}x = -a^2\frac{y}{x}\,\mathrm{d}y, \quad \text{so } \mathrm{d}u = \frac{a^2}{xr}\,\mathrm{d}y.$$

From this

$$\frac{\mathrm{d}}{\mathrm{d}u}(sn) = cn\,dn. \tag{1.8}$$

Differentiating equation (1.6)

$$\frac{\mathrm{d}}{\mathrm{d}u}(cn) = -sn\,dn. \tag{1.9}$$

And differentiating equation (1.7)

$$\frac{\mathrm{d}}{\mathrm{d}u}(dn) = -k^2\,sn\,cn. \tag{1.10}$$

There are other Jacobi elliptic functions. There is also another development of elliptic functions, due to Weierstrass [1], which makes essential use of complex variable theory. That approach leads to a set of elliptic functions different from but related to the Jacobi functions.

There are other elliptic functions defined as ratios involving *sn*, *cn* and *dn*. You might think the ratio *sn*/*cn* would be called *tn* in analogy to the *tan* in the circular case. This is done sometime. But there is a nicer notation for the ratios introduced by Gudermann [2] and Glaisher [3] in which the reciprocals are indicated by reversing the letters, thus

$$ns = \frac{1}{sn}, \qquad nc = \frac{1}{cn}, \qquad nd = \frac{1}{dn}.$$

Also, ratios such as *sn*/*cn* are named by catenation of first letters, thus

$$sc = \frac{sn}{cn}, \qquad cs = \frac{cn}{sn}, \qquad sd = \frac{sn}{dn}, \qquad ds = \frac{dn}{sn}, \qquad cd = \frac{cn}{dn}, \qquad dc = \frac{dn}{cn}.$$

So there are twelve in all.

The functions are periodic. From looking at the reference ellipse it is fairly clear that *sn* and *cn* should be qualitatively like the sine and cosine functions, at least for small eccentricity k. The period of *sn* or *cn* is $4K$, and the period of *dn* is $2K$, where the quarter period $K = K(k)$ is the value of u at the top of the ellipse where $(x, y) = (0, 1)$. This function $K(k)$ will be identified as an elliptic integral and discussed further later on. Hence, as can be seen in figure 1.3

$$\mathrm{sn}(K(k), k) = 1, \quad \mathrm{cn}(K(k), k) = 0 \quad \text{and} \quad \mathrm{dn}(K(k), k) = \frac{1}{a} = \sqrt{1 - k^2}.$$

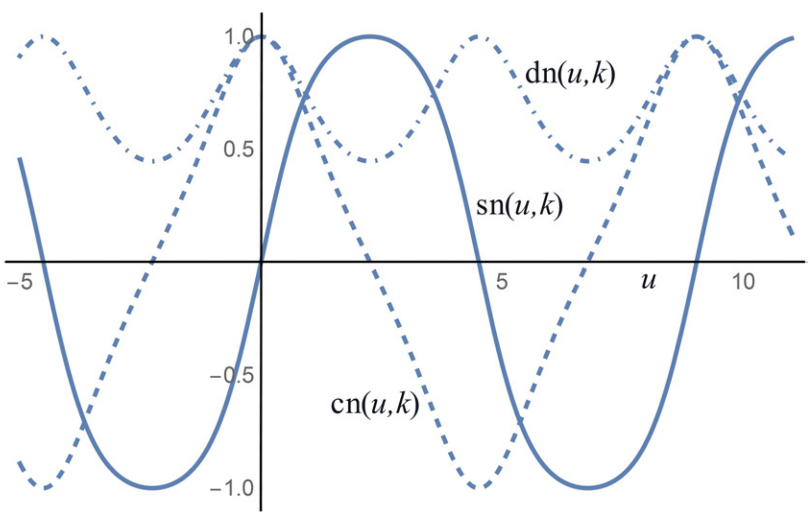


Figure 1.3. Qualitative features of *cn*, *sn*, *dn*. The ordinate is function value and abscissa is the argument u. Full period of *cn* or *sn* is $4K$ which defines K.

1.2 Differential equations

Each of the Jacobi functions satisfies a nonlinear differential equation of a certain form, namely

$$\left(\frac{\mathrm{d}}{\mathrm{d}u} zn\right)^2 = \alpha zn^4 + \beta zn^2 + \gamma \tag{1.11}$$

where $zn = \mathrm{zn}(u, k)$ is any one of the twelve elliptic functions and the coefficients α, β and γ are certain specific functions of the modulus k that characterize that particular elliptic function. For example,

$$\frac{\mathrm{d}}{\mathrm{d}u} sn = cn\, dn = \sqrt{1 - sn^2}\sqrt{1 - k^2 sn^2},$$

$$\left(\frac{\mathrm{d}}{\mathrm{d}u} sn\right)^2 = k^2 sn^4 - \left(1 + k^2\right) sn^2 + 1,$$

so for the case of sn one has that

$$\alpha = k^2, \qquad \beta = -\left(1 + k^2\right), \qquad \gamma = 1.$$

For nc,

$$\frac{\mathrm{d}}{\mathrm{d}u} nc = \frac{\mathrm{d}}{\mathrm{d}u}\frac{1}{cn} = \frac{sn\, dn}{cn^2} = \frac{1}{cn^2}\sqrt{1 - cn^2}\sqrt{1 - k^2\left(1 - cn^2\right)},$$

$$\left(\frac{\mathrm{d}}{\mathrm{d}u} nc\right)^2 = \left(1 - k^2\right) nc^4 - \left(1 - 2k^2\right) nc^2 - k^2,$$

and so for nc,

$$\alpha = \left(1 - k^2\right), \qquad \beta = -\left(1 - 2k^2\right), \qquad \gamma = -k^2.$$

One finds the results in table 1.1 by a similar calculation for each case.

Table 1.1. Coefficients appearing in equation (1.11) for each of the twelve Jacobi elliptic functions.

	α	β	γ
sn	k^2	$-(1 + k^2)$	1
cn	$-k^2$	$-(1 - 2k^2)$	$(1 - k^2)$
dn	-1	$(2 - k^2)$	$-(1 - k^2)$
ns	1	$-(1 + k^2)$	k^2
nc	$(1 - k^2)$	$-(1 - 2k^2)$	$-k^2$
nd	$-(1 - k^2)$	$(2 - k^2)$	-1
sc	$(1 - k^2)$	$(2 - k^2)$	1
cs	1	$(2 - k^2)$	$(1 - k^2)$
sd	$-k^2(1 - k^2)$	$-(1 - 2k^2)$	1
ds	1	$-(1 - 2k^2)$	$-k^2(1 - k^2)$
cd	k^2	$-(1 + k^2)$	1
dc	1	$-(1 + k^2)$	k^2

Notice that when

$$\left(\frac{\mathrm{d}}{\mathrm{d}u}zn\right)^2 = \alpha\, zn^4 + \beta\, zn^2 + \gamma,$$

then by differentiating,

$$2\left(\frac{\mathrm{d}^2}{\mathrm{d}u^2}zn\right)\left(\frac{\mathrm{d}}{\mathrm{d}u}zn\right) = \left(4\alpha\, zn^3 + 2\beta\, zn\right)\left(\frac{\mathrm{d}}{\mathrm{d}u}zn\right),$$

so

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2}zn = \beta\, zn + 2\alpha\, zn^3. \tag{1.12}$$

Each of elliptic functions also satisfies a second order DE of this type.

Now suppose in the process of solving a physical problem, the following differential equation comes up.

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = Ay^4 + By^2 + C. \tag{1.13}$$

It seems reasonable to seek a solution of the form

$$y(t) = a\,\mathrm{zn}(b\ t,\, k)$$

where zn is some Jacobi function. One would need to decide first which elliptic function would be appropriate, and then to figure out the value of k, and then finally get values for a and b.

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = a^2b^2\left(\alpha\, zn^4 + \beta\, zn^2 + \gamma\right)$$

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = a^2b^2\left(\alpha\left(\frac{y}{a}\right)^4 + \beta\left(\frac{y}{a}\right)^2 + \gamma\right) = \frac{b^2}{a^2}\alpha y^4 + b^2\beta y^2 + a^2b^2\gamma,$$

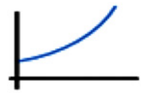
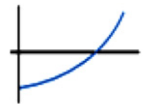
hence the coefficients of the physical differential equation are related to the elliptic differential equations by the following:

$$A = \frac{b^2}{a^2}\alpha, \qquad B = b^2\beta, \qquad C = a^2b^2\gamma.$$

Two points stand out. First the signs are preserved. The signs of A, B, C must match the signs of α, β, γ. This helps identify which elliptic function is appropriate. Then secondly,

$$\frac{AC}{B^2} = \frac{\alpha\gamma}{\beta^2}. \tag{1.14}$$

Table 1.2. Signs in equation (1.11) and sketch of right-hand side versus zn^2.

Signs (upper for $k^2 < 1/2$)	Functions possible	$\alpha\, zn^4 + \beta\, zn^2 + \gamma$ versus zn^2
$+ + +$	cs, sc	
$+ \mp -$	ds, nc	
$+ - +$	sn, ns, cd, dc	
		
$- \mp +$	cn, sd	
$- + -$	dn, nd	

So the first step in solving a differential equation with elliptic functions would be to compare qualitative behavior of the right-hand side when the equation is written in the form equation (1.13) to the cases shown in table 1.2. After identifying the appropriate elliptic function, values from table 1.2 can be substituted into the expression on the right side, and then this equation permits solving for k.

The physically applicable region of the solution in each case is where the polynomial shown in the right-hand column of table 1.2 takes on positive values. The combination $(-, -, -)$ does not appear, since zn squared cannot be negative. Once k is determined, the constants a and b are determined by

$$a = \pm\sqrt{\frac{C\beta}{B\gamma}} \qquad \text{and} \qquad b = \sqrt{\frac{B}{\beta}}\,.$$

I have chosen the positive value for b so that time (or independent variable) still runs in the same direction after rescaling.

1.3 Anharmonic oscillator

As a simple application to illustrate the general ideas, consider an oscillator comprising a block of mass m that slides without friction over a horizontal surface and which is attached to a spring with nonlinear elastic response given by the force law

$$F_x \approx -\kappa x + s x^3.$$

The potential energy is

$$U = \frac{1}{2}\kappa x^2 - \frac{1}{4} s x^4.$$

Thus

$$E = \text{constant} = \frac{1}{2} m \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \frac{1}{2}\kappa x^2 - \frac{1}{4} s x^4$$

so

$$\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 = \frac{2E}{m} - \frac{\kappa}{m} x^2 + \frac{1}{2}\frac{s}{m} x^4$$

$$C = +\frac{2E}{m}, \qquad B = -\frac{\kappa}{m}, \qquad A = +\frac{1}{2}\frac{s}{m}.$$

From the table, it appears that cd and sn are appropriate for periodic motion (they represent the same solution, shifted in time by a quarter period K) and ns or dc might represent what happens as the spring gives way. To see this, notice that there can be two allowed ranges for the squared function in the case (+, −, +). The inner permissable range includes 0 but not ∞. The outer range includes ∞ but not zero. Both ns and dc can have zero denominators so they must belong to the outside range. Taking for example sn, I find

$$\alpha = k^2, \qquad \beta = -\left(1 + k^2\right), \qquad \gamma = 1.$$

$$\frac{AC}{B^2} = \frac{Es}{\kappa^2}, \qquad \frac{\alpha\gamma}{\beta^2} = \frac{k^2}{\left(1 + k^2\right)^2} = \frac{Es}{\kappa^2},$$

$$k = \frac{1}{\sqrt{2}}\left(\frac{\kappa^2}{Es} - 2 - \sqrt{\left(\frac{\kappa^2}{Es}\right)^2 - 4\frac{\kappa^2}{Es}}\right)^{1/2}.$$

The minus sign is chosen since k should go to zero as the nonlinearity s goes to zero. So,

$$x(t) = a\,\mathrm{sn}(b\,t,\, k).$$

One sees there should be another constant of integration. Since t does not appear explicitly in the differential equation, time translation is a symmetry, so I write

$$x(t) = a\,\mathrm{sn}\big(b(t - t_0), k\big),$$

or

$$x(t) = \sqrt{\frac{2E}{\kappa}\Big(1 + k^2\Big)}\ \mathrm{sn}\left(\sqrt{\frac{\kappa}{m\Big(1 + k^2\Big)}}\,(t - t_0), k\right).$$

Using the translation by a quarter period yields the *cd* solution, or a cosine-like solution convenient for an initial condition in which the oscillator starts at rest, displaced by a certain amount from equilibrium. Finally, one should be careful to see that even though *cd* and *sn* are two independent solutions to the problem, it is *not* possible to combine them linearly, such as $x(t) = A\,\mathrm{sn}(b\,t, k) + B\,\mathrm{cd}(b\,t, k)$ since the differential equations is not linear homogeneous.

References

[1] Weierstrass K 1883 Zur Theorie der elliptischen Functionen *Sitzungberichte der Akadamie des Wissenschaften zu Berlin* pp 193–203 (1895 *Math. Werke* **2** 257–309) (Berlin: Mayer and Müler)

[2] Gudermann C 1838 *Journal für Math* XVIII pp 12–20 (CMA p 494)

[3] Glaisher J W L 1882 *Messenger of Mathematics* XI p 86 (CMA p 494)

IOP Concise Physics

William A Schwalm

Chapter 2

Differential equations satisfied by the Jacobi elliptic functions: pendula

Before going on to the general problem of solving differential equations with elliptic functions it is best to have another example. Here is one from mechanics, namely the motion of a pendulum through arbitrarily large angles, assuming no friction and no driving force.

I'll start with a brief review of the previous chapter. The three most basic of the Jacobi elliptic functions are

$$\mathrm{sn}(u, k) = y, \qquad \mathrm{cn}(u, k) = \frac{x}{a}, \qquad \mathrm{dn}(u, k) = \frac{r}{a}.$$

There are two geometric identities,

$$\mathrm{cn}(u, k)^2 + \mathrm{sn}(u, k)^2 = 1, \tag{2.1}$$

$$\mathrm{dn}(u, k)^2 + k^2\,\mathrm{sn}(u, k)^2 = 1, \tag{2.2}$$

and three differential identities

$$\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{sn}(u, k) = \mathrm{cn}(u, k)\,\mathrm{dn}(u, k), \tag{2.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{cn}(u, k) = -\mathrm{sn}(u, k)\,\mathrm{dn}(u, k), \tag{2.4}$$

$$\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{dn}(u, k) = -k^2\,\mathrm{sn}(u, k)\,\mathrm{cn}(u, k). \tag{2.5}$$

doi:10.1088/978-1-6817-4230-4ch2

The function $sn = \mathrm{sn}(u, k)$ satisfies the two differential equations. First

$$\left(\frac{\mathrm{d}}{\mathrm{d}u}\, sn\right)^2 = \left(1 - sn^2\right)\left(1 - k^2 sn^2\right) = k^2 sn^4 - \left(1 + k^2\right) sn^2 + 1, \tag{2.6}$$

and then, differentiating with respect to u,

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2}\, sn = -\left(1 + k^2\right) sn + 2k^2\, sn^3.$$

There are twelve elliptic functions defined as reciprocals or ratios of the basic three. It turns out that like sn, each one, $zn = \mathrm{zn}(u, k)$, satisfies two nonlinear DEs of these same two general forms. Thus the elliptic functions are all solutions of two nonlinear differential equations, one of first and one of second order. Thus

$$\left(\frac{\mathrm{d}}{\mathrm{d}u}\, zn\right)^2 = \alpha\, zn^4 + \beta\, zn^2 + \gamma$$

and thus by differentiating one finds also the second order DE

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2}\, zn = 2\alpha\, zn^3 + \beta\, zn.$$

With simple transformations one can solve general DEs in which the first derivative of the dependent variable, when squared, equals a general third or fourth degree polynomial in the independent variable, as we'll see.

2.1 Oscillatory motion of a pendulum at large amplitude

So, before going into some details of the general case, I will work out this standard example that illustrates some important points.

In figure 2.1 there is a pendulum comprising a small ball of mass m. Assume friction can be ignored somehow. The conserved energy is just translational kinetic energy and gravitational potential energy. If ℓ is the length, then

$$\frac{1}{2} m\ell^2 \left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + mg\ell(1 - \cos\theta) = \epsilon. \tag{2.7}$$

Suppose first the pendulum is displaced to an initial angle $\theta_o < \pi$ and released from rest. In this case the energy must be the initial potential energy,

$$\varepsilon = mg\ell\left(1 - \cos\theta_o\right),$$

and so

$$\frac{1}{2}\ell\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 = g\left(\cos\theta - \cos\theta_o\right).$$

So far this does not look awfully much like one of the forms satisfied by an elliptic function. Now comes a point worth noting. I will suggest the standard substitution

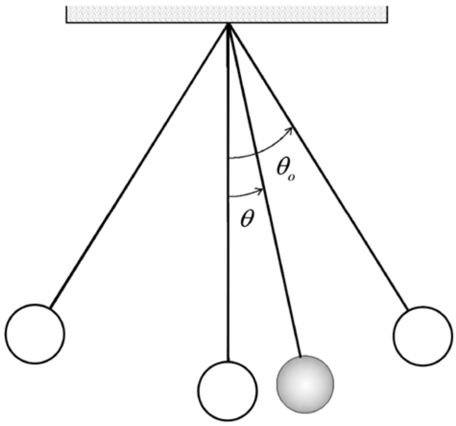


Figure 2.1. Pendulum. θ_o is the maximum displacement. The suspension is supposed to allow full circular motion.

I happen to know which transforms this to a polynomial form. I don't think it is obvious, and it is the kind of idiosyncratic assertion one would like to avoid if possible in favor of a more general and systematic procedure. But here it is.

Prompted by the fact that

$$\cos\theta = 1 - 2\sin^2\left(\frac{\theta}{2}\right),$$

let y be the sine of the half angle,

$$y = \sin\left(\frac{\theta}{2}\right).$$

For one thing, this trig substitution is such that $y \to 0$ linearly as $\theta \to 0$, and it turns out to clear the DE of trig functions.

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{2}\cos\left(\frac{\theta}{2}\right)\frac{\mathrm{d}\theta}{\mathrm{d}t}.$$

Then,

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = \frac{1}{4}\cos^2\left(\frac{\theta}{2}\right)\frac{4g}{\ell}\left(\sin^2\left(\frac{\theta_o}{2}\right) - \sin^2\left(\frac{\theta}{2}\right)\right),$$

or

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = \frac{g}{\ell}\left(1 - y^2\right)\left(y_o^2 - y^2\right) = +\frac{g}{\ell}y^4 - \frac{g}{\ell}\left(1 + y_o^2\right)y^2 + \frac{g}{\ell}y_o^2. \tag{2.8}$$

At this point it may be worth mentioning a wrong turn that one could take. In order to match equation (2.6), it might be tempting to divide out y_o on the left side and to try to identify $1/y_o$ as the modulus. This is a bad idea, since the modulus k should be between 0 and 1. Instead, I proceed systematically as outlined in the first chapter.

Now compare equation (2.8) to the general form

$$\left(\frac{\mathrm{d}\,zn}{\mathrm{d}u}\right)^2 = \alpha\, zn^4 + \beta\, zn^2 + \gamma.$$

Because the signs of (α, β, γ) are $(+, -, +)$ in the differential equation for y, and because an oscillatory solution is anticipated, one substitutes

$$y = A\ \mathrm{sn}(bt, k)$$

Then,

$$\begin{aligned}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 &= A^2b^2\left(1 - \frac{y^2}{A^2}\right)\left(1 - k^2\frac{y^2}{A^2}\right)\\ &= \frac{b^2}{A^2}k^2y^4 - b^2\left(1 + k^2\right)y^2 + A^2b^2. \end{aligned} \tag{2.9}$$

Comparing like powers of y in equations (2.8) and (2.9)

$$\frac{b^2}{A^2}k^2 = \frac{g}{\ell}, \qquad b^2\left(1 + k^2\right) = \frac{g}{\ell}\left(1 + y_o^2\right) \qquad A^2b^2 = \frac{g}{\ell}y_o^2.$$

From the first chapter you know that the product of the first with the last of these three quantities divided by the middle one can be used to solve for k. In this case it turns out very nicely. One has

$$A = y_o, \qquad k = |y_o|, \qquad b = \sqrt{\frac{g}{\ell}}.$$

Of course, in this simple example it seems natural to change dependent variable from y to y/y_o, which amounts to the same thing. Anyway, the result is

$$y = \sin\left(\frac{\theta}{2}\right) = \sin\left(\frac{\theta_o}{2}\right)\ \mathrm{sn}\left(\sqrt{\frac{g}{\ell}}\,t,\ \left|\sin\left(\frac{\theta_o}{2}\right)\right|\right). \tag{2.10}$$

This is a possible motion for the pendulum. However, originally the Newton's law equation would have been second order. Starting with energy reduced the problem to first order, where the constant ϵ or equivalently the maximum displacement, is the first constant of integration. There is another constant, which corresponds to fixing the phase, or fixing the position at the time $t = 0$. There is no explicit time dependence in the differential equation, so $t \to t + \alpha$ is a symmetry. This choice can be made—as we began to assume for instance—in such a way as to make the pendulum start from rest at maximum displacement at $t = 0$.

The differential equation

$$\left(\frac{\mathrm{d}u}{\mathrm{d}v}\right)^2 = k^2u^4 - \left(1 + k^2\right)u^2 + 1 = \left(1 - u^2\right)\left(1 - k^2u^2\right)$$

satisfied by $u(v) = \mathrm{sn}(v, k)$ is separable. Hence

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

The maximum value of the *sn* function is 1. The function oscillates, going from 0 to +1 to 0 to −1 and back to 0 during a period. So the v value at a quarter cycle is

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

Thus, when the pendulum starts from rest at maximum displacement at $t = 0$,

$$y = \sin\left(\frac{\theta}{2}\right) = \sin\left(\frac{\theta_o}{2}\right)\ \mathrm{sn}\left(\sqrt{\frac{g}{\ell}}\,t + K,\ \left|\sin\left(\frac{\theta_o}{2}\right)\right|\right),$$

$K(k)$ is the quarter period, which corresponds in the case $k \to 0$ to $\pi/2$ for the circular functions. So it is seen that the quarter period $K(k)$ can be evaluated as an integral. Recall from the first chapter that the two functions $sn = \mathrm{sn}(u, k)$ and $cd = \mathrm{cd}(u, k)$ satisfy the same first order differential equation, so they differ by being different members of the general solution satisfying a different initial condition. By comparing initial conditions one sees

$$\mathrm{cd}(u, k) = \mathrm{sn}(u + K, k).$$

Therefore the previous solution for the pendulum starting at rest from θ_o is also

$$y = \sin\left(\frac{\theta}{2}\right) = \sin\left(\frac{\theta_o}{2}\right)\mathrm{cd}\left(\sqrt{\frac{g}{\ell}}\ t,\ \left|\sin\left(\frac{\theta_o}{2}\right)\right|\right).$$

Figure 2.2 shows displacement versus time qualitatively for several values of initial displacement. Notice the way the initially sinusoidal motion distorts and the period increases as the amplitude increases.

Figure 2.3 shows the motion for cases when the initial displacement puts the pendulum almost but not quite at the top of the vertical circle. Initial position closer and closer to the top of the circle distorts the motion further and further from simple sinusoidal and there is an increasing tendency to loiter at near the top of the circle.

2.2 Motion traversing the whole circle

The solution up to this point describes the motion when the energy is low enough so that the pendulum swings back and forth between limits $-\theta_o \leqslant \theta \leqslant \theta_o$. Now in contrast suppose it swings all the way around in a circular motion. Assume the pendulum is stiff so that it can swing around the whole circle and even loiter at the top where $\theta = \pi$. Then it is not useful to characterize the energy ϵ in terms of a maximum angle, as such an angle would not be real. Just let ϵ be any positive energy. In particular, one is interested in the case where $\epsilon \geqslant 2mg\ell$ so the pendulum swings all the way around.

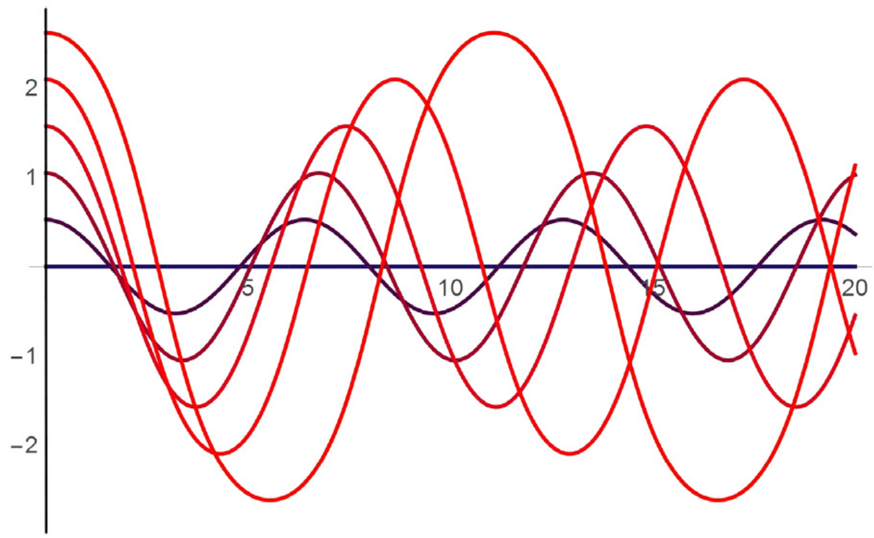


Figure 2.2. Displacement as a function of time, starting from maximum displacement. The ordinate is θ and the abscissa is bt. The intercepts at zero time show the initial displacement. The period of oscillation increases as the initial amplitudes increases. The shape also distorts.

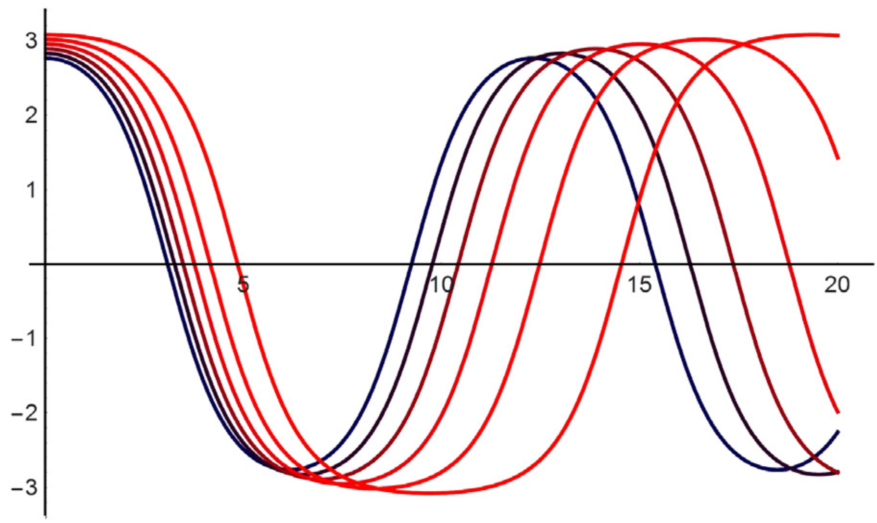


Figure 2.3. The angle versus bt for initial displacements of the pendulum close to the top of the circle at $\theta = \pi$. This shows the distortion of the curve and increasing tendency to loiter.

One has

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = \frac{g}{\ell}y^4 - \left(\frac{\epsilon}{2m\ell^2} + \frac{g}{\ell}\right)y^2 + \frac{\epsilon}{2m\ell^2},$$

and so the signs of (α, β, γ) are still $(+, -, +)$. The third column of table 1.2 in chapter 1 has two subcases. These subcases depend on whether the variable y is bounded or not. In the current problem, whether the energy is below or above the threshold above which the pendulum does complete loops the solution has to be periodic in θ so the solution for the underlying elliptic function u has to be bounded even when θ is not. When the pendulum swings back and forth, the *sn* or the *cd* solutions apply, since the allowable region connects to $u = 0$. And since these solutions are periodic, it is also possible to use *sn* or *cd* for u when θ goes all the way around. So again

$$y = A\,\mathrm{sn}(bt, k).$$

But since both *sn* and y must cover the entire range between -1 and 1, the coefficient A will turn out to have magnitude 1. There is no other way that both variables can span this same range.

Thus comparing

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = \frac{g}{\ell}y^4 - \left(1 + \frac{2mg\ell}{\epsilon}\right)\frac{\epsilon}{2m\ell^2}\,y^2 + \frac{\epsilon}{2m\ell^2},$$

with

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = \frac{b^2}{A^2}\,k^2y^4 - \left(1 + k^2\right)b^2y^2 + A^2b^2,$$

one finds indeed that

$$A = 1, \qquad b = \frac{1}{\ell}\sqrt{\frac{\epsilon}{2m}} \qquad k = \sqrt{\frac{2mg\ell}{\epsilon}}.$$

And so the solution for such cases where the pendulum traverses the whole circle is

$$y = \sin\left(\frac{\theta}{2}\right) = \mathrm{sn}\left(\sqrt{\frac{\epsilon}{2m}}\,\frac{t - t_o}{\ell},\ \sqrt{\frac{2mg\ell}{\epsilon}}\right)$$

As a check it is interesting to ask what happens as the energy ϵ increases without limit. The potential energy becomes negligible and so

$$\epsilon \approx \frac{1}{2}\,mv^2,$$

so

$$b = \sqrt{\frac{\epsilon}{2m}}\,\frac{1}{\ell} \approx \frac{v}{2\ell}.$$

That is consistent with

$$b\left(t - t_0\right) \approx \frac{v}{2\ell}\left(t - t_0\right) = \frac{\omega t + \theta_o}{2}$$

Meanwhile, as ϵ increases

$$k = \sqrt{\frac{2mg\ell}{\epsilon}} \to 0,$$

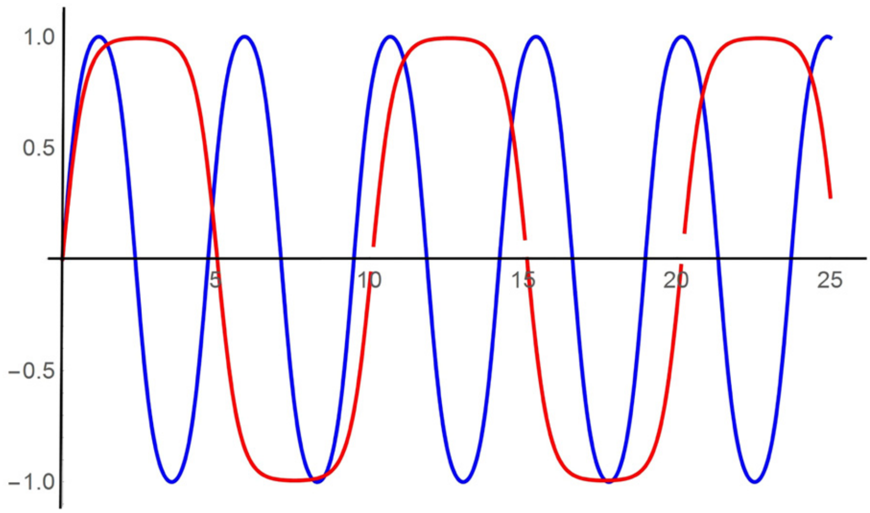


Figure 2.4. Displacement versus time when the pendulum travels around the full circle. Ordinate $y = \sin\theta/2$ versus abscissa bt. The red curve shows a case where ϵ is just slightly above $2mg\ell$, and the blue curve is a case when $\epsilon \gg 2mg\ell$.

so that the *sn* function goes over to the circular sine, and

$$\theta \approx 2n\pi + \omega t + \theta_o,$$

which is uniform circular motion. This is the expected limit.

Figure 2.4 shows motion in two cases with energy just above or considerably above $2mg\ell$, which is the energy needed for the pendulum to reach the unstable equilibrium at the top of the circle. For high energy, the motion is nearly uniform circular. So the solution to large amplitude motion of a simple pendulum, both for energy less than $2mg\ell$ when the pendulum swings back and forth, and for energy greater than $2mg\ell$ when the pendulum revolves all the way around the pivot, is expressed in terms of the elliptic function $\mathrm{sn}(u, k)$. This serves as motivation to study the general problem of equations with elliptic function solutions and the general reduction methods. This will be the topic of chapter 3.

2.3 The sine-Gordon equation: a series of pendula

From relativity, the energy of a classical particle is

$$c^2p^2 + m^2c^4 = E^2.$$

To make a quantum mechanical wave equation, replace physical quantities by operators acting on a function $u(x, t)$.

$$-\hbar^2c^2\frac{\partial^2 u}{\partial x^2} + m^2c^4u = -\hbar^2\frac{\partial^2 u}{\partial t^2},$$

or

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \mu^2 u = 0, \quad \text{where } \mu^2 = \frac{m^2c^2}{\hbar^2}.$$

This is the *Klein–Gordon equation*. Now consider a series of pendula, free to swing under the influence of gravity about a common axis, and coupled together harmonically by springs. Suppose the separation between them is a and the length of each one is ℓ and the mass is m.

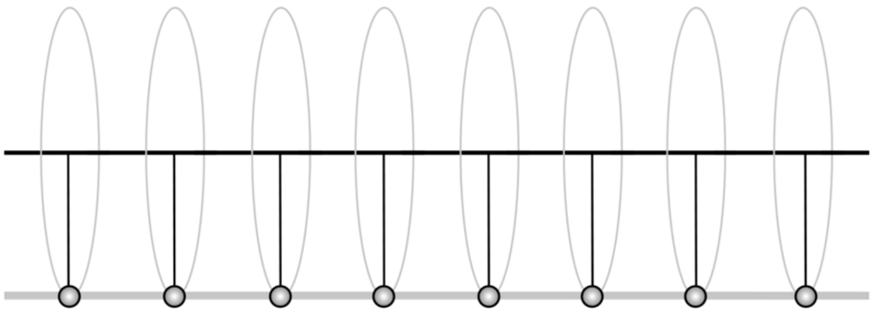


Figure 2.5.

The wide grey line represents the interactions. The Lagrangian function is

$$L = \sum_n \frac{1}{2}m\ell^2\left(\frac{\mathrm{d}\theta_n}{\mathrm{d}t}\right)^2 - \sum_n mg\ell\left(1 - \cos\theta_n\right) - \sum_n \frac{1}{2}\alpha\left(\theta_{n+1} - \theta_n\right)^2. \tag{2.11}$$

The Lagrange equations are found,

$$\frac{\partial L}{\partial \dot{\theta}_j} = m\ell^2\dot{\theta}_j, \tag{2.12}$$

$$\frac{\partial L}{\partial \theta_j} = -mg\ell\sin\theta_j - \sum_n \alpha\left(\theta_{n+1} - \theta_n\right)\left(\delta_{n+1,j} - \delta_{n,j}\right) \tag{2.13}$$

$$= -mg\ell\sin\theta_j + \alpha\left(\theta_{j+1} - 2\theta_j + \theta_{j-1}\right) \tag{2.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\theta}_j}\right) - \frac{\partial L}{\partial \theta_j} = m\ell^2\frac{\mathrm{d}^2\theta_j}{\mathrm{d}t^2} + mg\ell\sin\theta_j - \alpha\left(\theta_{j+1} - 2\theta_j + \theta_{j-1}\right) = 0. \tag{2.15}$$

Make this into a wave equation by introducing an envelope function and taking a limit where the distance a between pendula goes to zero keeping the envelope smooth.

$$\theta_j(t) \to \theta(x, t), \qquad \dot{\theta}_j(t) \to \frac{\partial \theta}{\partial t} \tag{2.16}$$

$$\theta_{j\pm1}(t) \to \theta(x \pm a,\, t) = \theta \pm a\frac{\partial\theta}{\partial x} + \frac{1}{2}a^2\frac{\partial^2\theta}{\partial x} + \cdots \tag{2.17}$$

$$\left(\theta_{j+1} - 2\theta_j + \theta_{j-1}\right) = a^2\frac{\partial^2\theta}{\partial x^2} + \cdots. \tag{2.18}$$

Thus

$$m\ell^2\frac{\mathrm{d}^2\theta_j}{\mathrm{d}t^2} + mg\ell\,\sin\theta_j - \alpha\left(\theta_{j+1} - 2\theta_j + \theta_{j-1}\right) = 0$$

becomes

$$\alpha a^2\frac{\partial^2\theta}{\partial x^2} - m\ell^2\frac{\partial^2\theta}{\partial t^2} - mg\ell\,\sin\theta = 0,$$

$$\frac{\partial^2\theta}{\partial x^2} - \frac{m\ell^2}{\alpha a^2}\frac{\partial^2\theta}{\partial t^2} - \frac{mg\ell}{\alpha a^2}\sin\theta = 0,$$

or

$$\frac{\partial^2\theta}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\theta}{\partial t^2} - \mu^2\sin\theta = 0, \quad \text{where } c = \frac{a}{\ell}\sqrt{\frac{\alpha}{m}}, \quad \mu = \frac{1}{a}\sqrt{\frac{mg\ell}{\alpha}}.$$

By analogy to the Klein–Gordon equation, this is called the ‘sine Gordon’ equation. So I want to use elliptic functions to construct a particular solution to the equation

$$\frac{\partial^2\theta}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\theta}{\partial t^2} - \mu^2\sin\theta = 0. \tag{2.19}$$

First I make a change of variables to

$$\theta(x,\, t) \to \theta(x - vt) = \theta(\xi).$$

$$\frac{\partial^2\theta}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\theta}{\partial t^2} - \mu^2\sin\theta = \left(1 - \frac{v^2}{c^2}\right)\frac{\mathrm{d}^2\theta}{\mathrm{d}\xi^2} - \mu^2\sin\theta = 0,$$

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}\xi^2} = \gamma^2\mu^2\sin\theta, \quad \text{where } \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

It is important to notice that, since this sine-Gordon wave equation did not really come from the context of relativity, the speed c is not a limit. The ratio v/c can take on any value. The quantity γ^2 could also be negative, so a better notation would be

$$f = \frac{1}{1 - v^2/c^2}. \tag{2.20}$$

The differential equation is of the type for which $\mathrm{d}\theta/\mathrm{d}\xi$ is an integrating factor,

$$\frac{\mathrm{d}\theta}{\mathrm{d}\xi}\frac{\mathrm{d}^2\theta}{\mathrm{d}\xi^2} = f\mu^2 \sin\theta\frac{\mathrm{d}\theta}{\mathrm{d}\xi},$$

$$\frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}\xi}\right)^2 = B - f\mu^2\cos\theta.$$

From experience with the simple pendulum, prompted by the fact that $\cos\theta = 1 - 2\sin^2\theta/2$, let $y = \sin\theta/2$. Then

$$\frac{\mathrm{d}y}{\mathrm{d}\xi} = \frac{1}{2}\cos\frac{\theta}{2}\frac{\mathrm{d}\theta}{\mathrm{d}\xi},$$

$$\left(\frac{\mathrm{d}y}{\mathrm{d}\xi}\right)^2 = \frac{1}{4}\left(1 - y^2\right)\left(\frac{\mathrm{d}\theta}{\mathrm{d}\xi}\right)^2 = \frac{1}{2}\left(1 - y^2\right)\left(B - f\mu^2\left(1 - 2y^2\right)\right), \tag{2.21}$$

$$\left(\frac{\mathrm{d}y}{\mathrm{d}\xi}\right)^2 = -f\mu^2y^4 - \frac{1}{2}\left(B - 3f\mu^2\right)y^2 + \frac{1}{2}\left(B - f\mu^2\right).$$

Referring to the signs, when B is in the range

$$f\mu^2 < B < 3f\mu^2, \tag{2.22}$$

one has the case $(-, +, +)$ in table 1.2, so the *cn* function or the *sd* function seem to match. Thus consider first the *cn* case and let

$$y = A\,\mathrm{cn}(b\xi, k).$$

Then

$$\left(\frac{\mathrm{d}y}{\mathrm{d}\xi}\right)^2 = -\frac{b^2}{A^2}k^2y^4 - b^2\left(1 - 2k^2\right)y^2 + A^2b^2\left(1 - k^2\right). \tag{2.23}$$

Equating the coefficients between equations (2.21) and (2.23),

$$\frac{b^2}{A^2}k^2 = f\mu^2, \quad b^2\left(1 - 2k^2\right) = \frac{1}{2}\left(B - 3f\mu^2\right), \quad A^2b^2\left(1 - k^2\right) = \frac{1}{2}\left(B - f\mu^2\right). \tag{2.24}$$

Eliminating k^2 and b^2 leads to the algebraic equation,

$$\left(A^2 - 1\right)\left(B + \left(2A^2 - 1\right)f\mu^2\right) = 0.$$

Glancing at the equation I see that either $A^2 = 1$ or $B = (1 - 2A^2)f\mu^2$. Since $A^2 \leqslant 1$ in any case, one sees that either $A^2 = 1$ or $-f\mu^2 \leqslant B \leqslant f\mu^2$ where we are assuming $f > 0$. But the latter contradicts equation (2.22), namely that $B > f\mu^2$ in order that the signs of the term in the DE match the *cn* case $(-, +, +)$. So we conclude that there is no *cn* solution for $A^2 \neq 1$.

On the other hand for $A = 1$ (solitons) the solution for real b does indeed exist [1],

$$b = \pm\left(\frac{B + f\mu^2}{2}\right)^{1/2}.$$

This means the solution covers the whole range of the angle $-\pi < \theta \leqslant \pi$, or indeed the angle variable is unlimited, so that the pendula circulate all the way around the axis. The variable $y = \sin(\theta/2)$ thus covers the range $-1 \leqslant y \leqslant 1$.

It is convenient to eliminate B in favor of k. The solution is

$$y = \sin\frac{\theta}{2} = \mathrm{cn}\left(\pm\frac{\mu}{k}\left(\frac{x - x_0 - v(t - t_0)}{\sqrt{1 - v^2/c^2}}\right), k\right).$$

Qualitatively, the solution is a series of twists in which the pendula go around the axis as shown in figure 2.6. The twist shown is right handed with respect to the forward propagation direction, and of course the twist can be either right or left handed depending on the sign of $b = \pm\mu/k$. One of the twists is shown in the figure and the whole soliton solution is a sequence of twists, periodic to reflect the periodicity of the *cn* function[1].

$$y = \sin\frac{\theta}{2} = \mathrm{cn}\left(\pm\frac{\mu}{k}\left(\frac{x - x_0 - v(t - t_0)}{\sqrt{1 - v^2/c^2}}\right), k\right)$$

$$\rightarrow \mathrm{sech}\left(\pm\frac{\mu}{k}\left(\frac{x - x_0 - v(t - t_0)}{\sqrt{1 - v^2/c^2}}\right)\right).$$

The single-kink limit is shown in figure 2.7. Thus in this limit the solution is no longer periodic and consists of individual kink or anti-kink solitons, which are topological

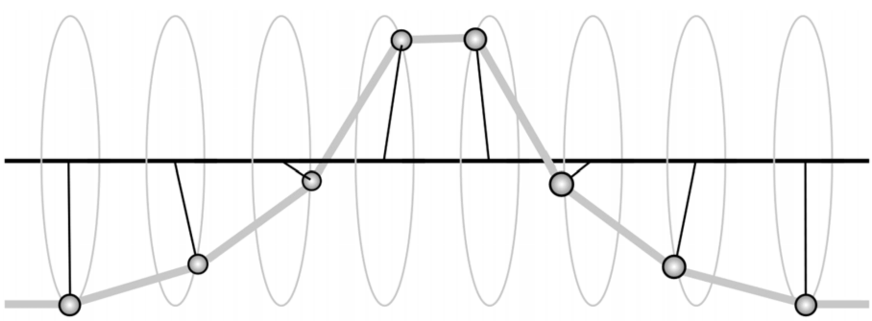


Figure 2.6.

[1] For visualization of certain soliton solutions of sine-Gordon equation see for example wikipedia http://en.wikipedia/wiki/Sine-Gordon_equation

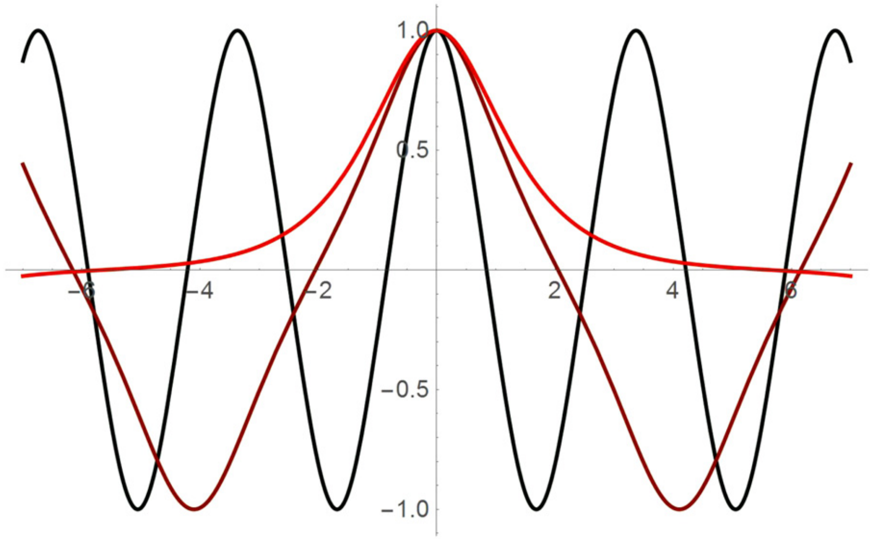


Figure 2.7. Soliton solution $y = \sin(\theta/2) = \mathrm{cn}(\mu\sqrt{f}x/k,\, k)$ versus $\tilde{x} = \mu\sqrt{f}x$ for k values 0.5, 0.9 and 0.99 (navy, dark red and red respectively). Each time y goes to ± 1 the angle θ goes past $\pm\pi$, increasing as $\tilde{x}$ increases. The modulus k is related to the energy per pendulum such that the energy diverges as k goes to zero. In the limit $k \to 1$, the *cn* becomes *sech*, and the train of pulses separates into widely spaced single-twist solitons.

excitations in the sense that they correspond to a fixed winding number where the spring connecting the pendula encircles the axis clockwise or anticlockwise.

The *cn* solution is thus a periodic string of such solutions. To find other solutions one could to go back and look at *sd* rather than *cn*. So one would consider $y = \sin(\theta/2) = \mathrm{sd}(b\xi,\, k)$. This works, but it turns out to be the same solution again, only phase shifted, because $\mathrm{sd}(u,\, k)$ is proportional to $\mathrm{cn}(u + K,\, k)$.

What seems to be missing is a solution that would reduce to the linearized limit, or in other words the limit as the displacement θ becomes small so the differential equation is approximately linear.

2.4 Series of pendula: 'super luminal' case

On thinking about this, it appears that the required solutions must be 'super luminal', meaning that the wave forms should move with a speed $v > c$. Thus assume

$$f < 0, \qquad f = -|f|, \qquad |f| = \frac{1}{v^2/c^2 - 1}.$$

So, the sine-Gordon equation becomes

$$\left(\frac{\mathrm{d}y}{\mathrm{d}\xi}\right)^2 = |f|\mu^2 y^4 - \frac{1}{2}\left(B + 3|f|\mu^2\right)y^2 + \frac{1}{2}\left(B + |f|\mu^2\right).$$

Then for $B > -|f|\mu^2$, the signs are $(+, -, +)$ and the periodic solution is *sn*.

$$y = \sin\theta/2 = A\,\mathrm{sn}[b\xi,\, k],$$

and from comparing the two differential equations

$$\frac{b^2}{A^2}k^2 = |f|\mu^2, \qquad b^2\left(1 + k^2\right) = \frac{1}{2}\left(B + 3|f|\mu^2\right), \qquad A^2b^2 = \frac{1}{2}\left(B + |f|\mu^2\right).$$

Eliminating k^2 and b^2 we obtain

$$\left(A^2 - 1\right)\left(B - 2A^2 - 1\right)|f|\mu^2 = 0,$$

so,

$$B = \left(2A^2 - 1\right)|f|\mu^2.$$

From this we see the value of B is in the valid range for $0 < A^2 \leqslant 1$. So, we have both the soliton solution ($A = 1$) and the solution for the arbitrary amplitude, as long as $|A| < 1$. By substituting the expression for B in the three equations above

$$k^2 = A^2, \qquad b^2 = |f|\mu^2.$$

In other words, $k = A$ and $b = \pm\sqrt{|f|}\mu$. Thus

$$y = \sin\frac{\theta}{2} = A\,\mathrm{sn}\left(\sqrt{|f|}\mu\xi,\, A\right) = A\,\mathrm{sn}\left(\frac{\mu\left(x - x_0 \pm v\left(t - t_0\right)\right)}{\sqrt{v^2/c^2 - 1}},\, A\right).$$

In the super luminal limit $v \to \infty$ the wavelength diverges and

$$y \to A\,\mathrm{sn}\left(c\mu\left(t - t_0\right),\, A\right).$$

The pendula are moving in unison. In the small amplitude limit, one has

$$y \to \frac{\theta}{2} \to A\sin\left(c\mu\left(t - t_0\right)\right) = A\sin\left(\sqrt{\frac{g}{\ell}}\left(t - t_0\right)\right).$$

For small amplitude as the parameter v goes to infinity, the pendula swing in phase with the normal small-amplitude frequency of a pendulum. Notice that in the case of the soliton solution the kink solution simply comes to rest as the parameter v goes to zero.

So the pendulum and the long chain of coupled pendula are examples of mechanical systems giving rise to elliptic functions. There are many others. The next step is to look at the reduction and solution of general differential equations of the forms

$$\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = c_0 + c_1y + c_2y^2 + c_3y^3 \quad \text{or} \quad \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = c_0 + c_1y + c_2y^2 + c_3y^3 + c_4y^4.$$

This is the subject of the next chapter.

Reference

[1] Drazin P G and Johnson R S 1989 *Solitons: An Introduction* (Cambidge: Cambidge University Press)

Chapter 3

General reduction of the DE in terms of Jacobi functions

The previous chapter covered some examples of physical problems for which the differential equations reduced to the general form

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \alpha y^4 + \beta y^2 + \gamma.$$

Each of the Jacobi elliptic functions satisfies such an equation, where each of the constants α β and γ is a certain characteristic function of the modulus k. These functions are tabulated in table 1.1 of chapter 1. This chapter is about reduction of differential equations of the more general form

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = c_0 + c_1 y + \cdots + c_4 y^4$$

where $c_4 \neq 0$

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = c_0 + c_1 y + \cdots + c_3 y^3.$$

where $c_3 \neq 0$.

3.1 Linear fractional transformation and cross ratio

First consider the quartic case. I assume initially that the roots $y_1, y_2, \cdots y_4$ of the polynomial $c_0 + c_1 y + \cdots + c_4 y$ are real and in increasing order. However, the methods used are chosen in such a way as to apply to the case of complex conjugate pairs of roots, or even to the case of complex coefficients, when the roots

doi:10.1088/978-1-6817-4230-4ch3

become arbitrary complex number. The strategy is to use a linear fractional transformation

$$w = \frac{py + q}{ry + s}, \tag{3.1}$$

to express w in terms of y, or y in terms of w, where w is an elliptic function that satisfies

$$\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 = b^2\left(\alpha w^4 + \beta w^2 + \gamma\right).$$

This is elaborated on below. An example would be

$$w(x) = \mathrm{sn}(bx, k),$$

in which case,

$$\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 = b^2\left(k^2 w^4 - (k^2 + 1)w^2 + 1\right).$$

I calculate first the left side and then the right side of this equation in terms of y. Then by comparing like powers of y it will be possible to determine the parameters of the transformation, the modulus k and the scale factor b of the independent variable. Notice that the left side, when calculated from the transformation itself, will not contain b, while the right side evidently contains b^2 from the chain rule. That will make it possible to determine b.

For the left side, begin by differentiating the linear fractional expression, no b appears, since y is a function of x not of bx.

$$\frac{\mathrm{d}w}{\mathrm{d}x} = \frac{ps - qr}{(ry + s)^2}\frac{\mathrm{d}y}{\mathrm{d}x},$$

so, using the original differential equation,

$$\begin{aligned}\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 &= \frac{(ps - qr)^2}{(ry + s)^4}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\\ &= \frac{(ps - qr)^2}{(ry + s)^4}\left(c_0 + c_1 y + \cdots + c_4 y^4\right).\end{aligned}$$

The polynomial in y factors completely over the complex numbers. So

$$\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 = \frac{(ps - qr)^2}{(ry + s)^4}c_4(y - y_1)(y - y_2)(y - y_3)(y - y_4), \tag{3.2}$$

where for the time being the roots are assumed to be real.

And then for the right side, which does contain b since $w(x) = \mathrm{sn}(bx, k)$,

$$\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 = b^2\left(\alpha w^4 + \beta w^2 + \gamma\right) = b^2\alpha\left(w^2 - w_a^2\right)\left(w^2 - w_b^2\right). \tag{3.3}$$

Substituting equation (3.1) in equation (3.3)

$$\begin{aligned}\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 &= b^2\alpha\left(w^2 - w_a^2\right)\left(w^2 - w_b^2\right)\\ &= \frac{b^2\alpha}{(ry+s)^4}\left((py+q)^2 - w_a^2(ry+s)^2\right)\left((py+q)^2 - w_b^2(ry+s)^2\right).\end{aligned}$$

Collecting powers of y,

$$\begin{aligned}\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 &= \frac{b^2\alpha}{(ry+s)^4}\left(\left(p^2 - w_a^2 r^2\right)\left(p^2 - w_b^2 r^2\right)y^4 + \cdots\right)\\ &= \frac{b^2\alpha}{(ry+s)^4}\left(p^2 - w_a^2 r^2\right)\left(p^2 - w_b^2 r^2\right)(y - y_1)(y - y_2)(y - y_3)(y - y_4).\end{aligned} \tag{3.4}$$

Comparing equations (3.2) and (3.4)

$$\begin{aligned}&\frac{(ps - qr)^2}{(ry+s)^4}c_4(y - y_1)(y - y_2)(y - y_3)(y - y_4)\\ &= \frac{b^2\alpha}{(ry+s)^4}\left(p^2 - w_a^2 r^2\right)\left(p^2 - w_b^2 r^2\right)(y - y_1)(y - y_2)(y - y_3)(y - y_4).\end{aligned} \tag{3.5}$$

So

$$(ps - qr)^2 c_4 = b^2\alpha\left(p^2 - w_a^2 r^2\right)\left(p^2 - w_b^2 r^2\right). \tag{3.6}$$

This gives the scale factor b once all the other coefficients are known. The principal objective will be to obtain the other parameters by tracing how the linear fractional transformation has to map roots.

In order to find the other parameters go back and equate the right sides of equations (3.3) and (3.4),

$$\begin{aligned}&b^2\alpha\left(w^2 - w_a^2\right)\left(w^2 - w_b^2\right)\\ &= \frac{b^2\alpha}{(ry+s)^4}\left(p^2 - w_a^2 r^2\right)\left(p^2 - w_b^2 r^2\right)(y - y_1)(y - y_2)(y - y_3)(y - y_4).\end{aligned} \tag{3.7}$$

The right side expressed in terms of y can only be zero when y is one of the roots y_1, y_2, y_3, or y_4. On the other hand, when it is expressed in terms of w it appears that it can only be zero when w is one of the four roots $-w_a$, $+w_a$, $-w_b$ or $+w_b$. There would seem to be 4! or 24 different ways to map each of the four roots in y into one of the four roots in w. Does it matter which is mapped into which? Yes, certainly. To see how all this works, it is appropriate to make a slight excursion.

For any four complex numbers, z_1, z_2, z_3 and z_4 there is a cross ratio defined thus,

$$\textit{cross ratio}[z_1, z_2; z_3, z_4] = \frac{z_1 - z_3}{z_1 - z_4}\Big/\frac{z_2 - z_3}{z_2 - z_4} = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

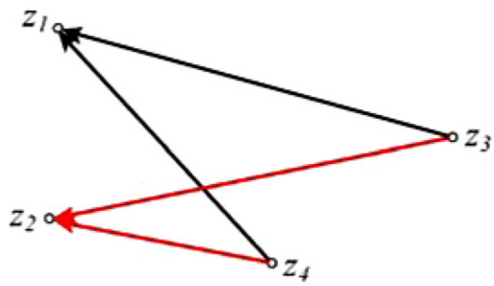


Figure 3.1.

Apparently there are several different cross ratios for any given tetrad of numbers. There are indeed $4! = 24$ different permutations of four numbers, but the subgroup (a Klein 4-group) of four permutations generated by the two substitutions

$$\{z_1 \to z_2, z_2 \to z_1, z_3 \to z_4, z_4 \to z_3\}$$
$$\{z_1 \to z_3, z_2 \to z_4, z_3 \to z_1, z_4 \to z_2\}$$

leaves the cross ratio invariant.

So that means there are 24/4 = 6 different cross ratios in general for any four complex numbers. The linear fractional transformation preserves the cross ratio: cross ratio, invariant under linear fractional transformation

$$\text{If} \quad w = \frac{pz + q}{rz + s} \quad \text{then} \quad \frac{(w_1 - w_3)(w_2 - w_4)}{(w_1 - w_4)(w_2 - w_3)} = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

I will use a different notation for the cross ratio, which I prefer because the subscripts in the top row of the fractions occur in increasing order. Let

$$\rho(z_1, z_2, z_3, z_4) \equiv \textit{cross ratio}[z_1, z_3, z_2, z_4]$$

so that

$$\rho(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_4)(z_3 - z_2)}. \tag{3.8}$$

Then

$$\rho(z_1, z_2, z_3, z_4) = \rho(w_1, w_2, w_3, w_4). \tag{3.9}$$

Now I can make a linear fractional transformation that takes any three points (including a point at infinity) into any other three points. How does one do this? The best way to see how it works is to do it in two steps. First, notice the linear fractional transformation

$$u = \rho(z, a, b, c) = \frac{(z-a)(b-c)}{(z-c)(b-a)} = \frac{(b-c)z - a(b-c)}{(b-a)z - c(b-a)}$$

maps z values $\{a, b, c\}$ into u values $\{0, 1, \infty\}$ respectively. Or if you solve for z, you find the inverse transformation

$$z = \frac{c(a-b)u + a(b-c)}{(a-b)u + b - c},$$

which takes u values $\{0, 1, \infty\}$ back into the z value $\{a, b, c\}$. The equation $u = \rho(z, a, b, c)$ thus defines both the transformation and its inverse.

Now to make a linear fractional transformation

$$w = \frac{pz + q}{rz + s}$$

that takes z values $\{a, b, c\}$ into w values $\{d, e, f\}$ one can make a two-step composite mapping. First $u = \rho(z, a, b, c)$ takes z values $\{a, b, c\}$ to u values $\{0, 1, \infty\}$. Then, the inverse transformation you get by solving $u = \rho(w, d, e, f)$ for w gives a transformation that takes the u values $\{0, 1, \infty\}$ to the w values $\{d, e, f\}$. Thus solving

$$\rho(z, a, b, c) = \rho(w, d, e, f)$$

for w gives the desired formula taking z values $\{a, b, c\}$ to the w values $\{d, e, f\}$. At the same time, if you solve the equations for z instead of for w, you get a linear fractional transformation of the form

$$z = \frac{rw - q}{s - pw}$$

that takes w values $\{d, e, f\}$ into z values $\{a, b, c\}$. Both the transformation and its inverse are determined by setting the two cross ratios equal.

3.2 Reduction of general quartic case

Now return to the elliptic functions and reducing the general quartic case. I want to use the transformation defined above to substitute the new variable w for y in such a way that

$$\begin{aligned}\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 &= \frac{b^2\alpha}{(ry+s)^4}\left(p^2 - w_a^2 r^2\right)\left(p^2 - w_b^2 r^2\right)(y-y_1)(y-y_2)(y-y_3)(y-y_4) \\ &= b^2\alpha\left(w^2 - w_a^2\right)\left(w^2 - w_b^2\right),\end{aligned}$$

where the parameters have to be selected so that each of the four roots $y_1, \ldots, y_4$ corresponds to a unique one of the four roots $\pm w_a$, $\pm w_b$. At first this doesn't seem possible, since the linear fractional transformation takes three points onto three points. But the reason that it is possible is that there is one more degree of freedom, namely the choice of the modulus k. (Remember the scale factor b has been set aside, having been expressed already in terms of the four mapping parameters and k.)

The roots of equation (3.3) are

$$w_a = \sqrt{-\frac{\beta}{2\alpha} + \sqrt{\left(\frac{\beta}{2\alpha}\right)^2 - \frac{\gamma}{\alpha}}}, \qquad w_b = \sqrt{-\frac{\beta}{2\alpha} - \sqrt{\left(\frac{\beta}{2\alpha}\right)^2 - \frac{\gamma}{\alpha}}}.$$

Since, from table 1.1 of chapter 1, the coefficients α, β, γ are functions only of k and depend on which elliptic function one is trying to fit to the problem, the determination of k, and also of the other parameters, will depend on several choices. One can choose any ordering for the arguments $\{-w_a, -w_b, +w_b, +w_a\}$ of ρ. Each value of ρ depends only on the ratio w_b/w_a of the roots. Of the 24 choices, six are distinct. So the question is, which of these is most useful?

Consider for instance

$$\rho(-w_a, w_a, -w_b, w_b) = \frac{4w_a w_b}{(w_a + w_b)^2} = \rho_0, \tag{3.10}$$

where $\rho_0 \equiv \rho(y_1, y_3, y_2, y_4)$, and $y_1 < y_2 < y_3 < y_4$. For this ordering $0 \leqslant \rho_0 \leqslant 1$. This is quite convenient. It turns out to be possible also, and often sufficient, to select $w = \mathrm{sn}(bx, k)$. In that case, $\alpha = k^2$, $\beta = -(1 + k^2)$, $\gamma = 1$ so $\alpha > 0$, $\beta < 0$, $\gamma > 0$ and $\beta^2 - 4\alpha\gamma = (1 - k^2) \geqslant 0$.

$$\rho_0 = \rho(-w_a, w_a, -w_b, w_b) = \frac{4w_a w_b}{(w_a + w_b)^2} = -\frac{4\sqrt{\alpha\gamma}}{\beta - 2\sqrt{\alpha\gamma}} = \frac{4\sqrt{\alpha\gamma}}{|\beta| + 2\sqrt{\alpha\gamma}}$$

and hence $0 \leqslant \rho_0 \leqslant 1$. So the left-hand side of the transformation equation (3.10) is

$$\rho_0 = \frac{4\sqrt{\alpha\gamma}}{|\beta| + 2\sqrt{\alpha\gamma}} = \frac{2\sqrt{t}}{1 + \sqrt{t}}, \quad \text{where } t = 4\frac{\alpha\gamma}{\beta^2}.$$

It was seen in previous chapters that this ratio is the same for $A\ \mathrm{zn}(bx, k)$ as it is for $\mathrm{zn}(x, k)$, so it can be used to select a particular elliptic function *zn* and then to determine k.

For the *sn* case,

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

So

$$t = \left(\frac{\rho_0}{2 - \rho_0}\right)^2 = \left(\frac{2k}{1 + k^2}\right)^2.$$

Thus in the case of sn, the modulus is one of four roots

$$k = \pm\left(\frac{2\left(1 \pm \sqrt{1 - \rho_0}\right) - \rho_0}{\rho_0}\right).$$

If $0 \leqslant \rho \leqslant 1$ then the root that leads to $0 \leqslant k \leqslant 1$ is

$$k = \frac{2\left(1 - \sqrt{1 - \rho_0}\right) - \rho_0}{\rho_0}. \tag{3.11}$$

So in the case where the roots of $c_4 y^4 + \cdots + c_1 y + c_0 = 0$ are all real, one strategy will be to assume a solution involving sn, sort the roots so that $y_1 < y_2 < y_3 < y_4$ and then use $\rho_0 = \rho(y_1, y_4, y_2, y_3)$ (which is also $\rho(y_2, y_3, y_1, y_4)$). Then, since the modulus k that this choice determines makes $\rho(y_1, y_4, y_2, y_3) = \rho(-w_a, w_a, -w_b, w_b)$, an appropriate linear fractional transformation is found, for example, as

$$\rho(y, y_4, y_2, y_3) = \rho(w, w_a, -w_b, w_b).$$

3.3 Finding the coefficients of the linear fractional transformation

One can solve this either for y or for w. In the end y is needed as a function of w, however I also need to have the time scale factor b. So for that I need k and the transformation coefficients. I could solve first for w. This determines q/p, r/p, s/p, and so together with the fact that k is known, I have all information needed to build the solution. So I continue to build the solution using the sn function.

$$\begin{aligned}
\rho(w, w_a, -w_b, w_b) &= \frac{(w - w_a)(-w_b - w_b)}{(w - w_b)(-w_b - w_a)} \\
&= \frac{2w_b}{(w_a + w_b)}\frac{(w - w_a)}{(w - w_b)} \\
&= \frac{(2/k)}{(1 + 1/k)}\frac{(w - 1)}{(w - 1/k)} \\
\rho(y, y_4, y_2, y_3) &= \frac{(y - y_4)(y_2 - y_3)}{(y - y_3)(y_2 - y_4)},
\end{aligned}$$

so

$$w = \frac{\left(2ky_4 + (1 - k)y_2 - (k + 1)y_3\right)y + (y_3 - y_2)y_4 + 2ky_2y_3 - k(y_2 + y_3)y_4}{k\left((k - 1)y_2 - (k + 1)y_3 + 2y_4\right)y + 2ky_2y_3 - ky_4\left((k + 1)y_2 + (1 - k)y_3\right)},$$

Thus it appears that

$$\begin{aligned}
p &= 2ky_4 + (1 - k)y_2 - (k + 1)y_3, \\
q &= (y_3 - y_2)y_4 + 2ky_2\, y_3 - k(y_2 + y_3)y_4 \\
r &= k\big((k - 1)y_2 - (k + 1)y_3 + 2y_4\big), \\
s &= 2ky_2\, y_3 - ky_4\big((k + 1)y_2 + (1 - k)y_3\big).
\end{aligned}$$

Now to find the scale factor b use equation (3.6),

$$b^2 = \frac{(ps - qr)^2 c_4}{k^2\big(p^2 - w_a^2 r^2\big)\big(p^2 - w_b^2 t^2\big)}, \qquad b = \pm\frac{(ps - qr)\sqrt{c_4}}{\sqrt{p^2 - r^2}\sqrt{k^2p^2 - r^2}}$$

Notice that b is a function only of the ratios q/p, r/p, s/p, and k as it should be. These ratios are known now, and so together with the fact that k is known, I have all information needed to build the solution. I can go back and solve the other way for y in terms of w, where most generally $w = \mathrm{sn}(\pm b(x - x_0), k)$,

$$y = \frac{sw - q}{p - rw} = \frac{s\,\mathrm{sn}\big(\pm b(x - x_0), k\big) - q}{p - r\,\mathrm{sn}\big(\pm b(x - x_0), k\big)}.$$

This solves the differential equation

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^4 = c_0 y^4 + \cdots + c_1 y + c_0,$$

the general quartic case.

William A Schwalm

Chapter 4

Elliptic integrals

An elliptic integral is one of the form

$$\int R\left(y,\, \sqrt{u}\right) \mathrm{d}y, \tag{4.1}$$

where R is a rational function, which is to say it is a ratio of two polynomials in the two quantities y and $\sqrt{u}$, and u is a polynomial $u(y)$ of degree three or four in y. Thus

$$u(y) = c_4 y^4 + c_3 y^3 + c_2 y^2 + c_1 y + c_0,$$

where c_4 and c_3 are not both zero. For the time being, I suppose the coefficients are real. Further down this requirement will be relaxed, since in many applications they will need to be complex. These integrals are connected logically and historically to the elliptic functions. Historically, they appear early on in problems such as pendulum motion and finding arc length of an ellipse. The latter problem is responsible for the name.

In previous chapters it turned out that the Jacobi elliptic functions afford solutions to

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = c_4 y^4 + c_3 y^3 + c_2 y^2 + c_1 y + c_0.$$

Solving for the derivative, I see that this is really two separable differential equations

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm\sqrt{c_4 y^4 + c_3 y^3 + c_2 y^2 + c_1 y + c_0},$$

each of which has either one or zero real solutions through a general point $(x,\, y)$ where the polynomial has either positive or negative value respectively. There are generally also singular solutions, or in other words envelope solutions, where

doi:10.1088/978-1-6817-4230-4ch4

$y = y_c =$ constant for all x. For a constant solution, the left hand side must be zero, so the possible y_c values are the real zeros of the polynomial, if there are such. Thus one has

$$x - x_0 = \pm \int_{y_c}^{y} \frac{\mathrm{d}z}{\sqrt{c_4 z^4 + c_3 z^3 + c_2 z^2 + c_1 z + c_0}}.$$

So in this case it appears that if y can be expressed in terms of an elliptic function of x, then x can be expressed as an elliptic integral of y. The relation between the elliptic integrals and the elliptic functions may be more complicated generally, but this example gives some idea of how it works.

In the most general case,

$$R\left(y,\ \sqrt{u}\right) = \frac{f_1(y) + f_2(y)\sqrt{u}}{f_3(y) + f_4(y)\sqrt{u}},$$

when $f_1, f_2 \cdots f_4$ are rational functions. Hence,

$$\begin{aligned} R\left(y,\ \sqrt{u}\right) &= \left(\frac{f_1(y) + f_2(y)\sqrt{u}}{f_3(y) + f_4(y)\sqrt{u}}\right)\left(\frac{f_3(y) - f_4(y)\sqrt{u}}{f_3(y) - f_4(y)\sqrt{u}}\right) \\ &= \frac{f_1 f_3 - f_2 f_4 u}{f_3^2 - f_4^2 u} + \frac{f_2 f_3 - f_1 f_4}{f_3^2 - f_4^2 u}\sqrt{u}, \end{aligned}$$

and so, finally

$$R\left(y,\ \sqrt{u}\right) = \frac{f_1 f_3 - f_2 f_4 u}{f_3^2 - f_4^2 u} + \frac{f_2 f_3 - f_1 f_4}{f_1^2 - f_4^2 u}\frac{u}{\sqrt{u}} = R_1(y) + \frac{R_2(y)}{\sqrt{u(y)}}, \tag{4.2}$$

where R_1 and R_2 are two rational functions of only y. Of course, the polynomial coefficients may contain other parameters, such as for instance the modulus of some elliptic function. But this shows the most general functional dependence on y.

Because the rational function R is general, the problems with solutions as elliptic integrals is a larger class than the problems solvable directly via elliptic functions. These integrals arise routinely in classical mechanics, conformal mappings, and in the context of thermostatistical and other problems on lattices. They form an important class of special functions. Another point about these integrals is that, due to the presence of the square roots, an understanding of branch points, cuts and Riemann surfaces is sometimes unavoidable. This remains somewhat true even when the integrations are performed numerically.

In the following we look first as an example at the sort of inverse relation between the Jacobi elliptic function *sn* and the first Legendre elliptic integral *F*. The periodicity of *sn* is intimately connected to the multivaluedness of *F*. Extending the arguments to complex variables, it is easy to show that *sn* actually has two independent periodicities in the complex plane of its argument. In fact an elliptic function can be defined, in general terms, as a complex function having two independent periods and having no singularities other than poles in the finite plane of its argument. Then coming back to the general problem at the end of the section, a procedure is shown for reducing the most general form of elliptic integral to a closed expression in terms of three special forms, namely the Legendre elliptic integrals of the first, second and third kind. It should be stressed again that the reason for doing this reduction is that if the branch cuts of a solution—even a numerical solution of a physical problem—are not understood, then the solution is not really under control and subsequent analysis based on it cannot proceed in a satisfactory way.

To proceed from here it is useful to review some things about complex variables. A reader not familiar with the topic should consult any standard text.

4.1 Review of complex variables up through residues

If complex analysis up through the residue theorem is quite familiar, the reader may want to move on to the next section, which is a careful but informal discussion of branching and multivaluedness. The material covered there is essential to the central development, and the preliminary review here is provided for reference. The following is just a restatement of some basic things in order to fix ideas and some standard vocabulary. I will point out some logical connections where it seems instructive and is easy to do so, but will offer no rigorous development at this point. First, suppose z is a complex number. If I write $z = x + \mathrm{i}y$, that usually means x and y are both real numbers. The complex conjugate $z^* = x - \mathrm{i}y$ taken together with z makes possible to transform back and forth between (x, y) and (z, z^*). Another way to represent z would be in polar coordinates, letting $x = r\cos\theta$, $y = r\sin\theta$, so that

$$z = (\cos\theta + \mathrm{i}\sin\theta)r.$$

Then by solving a differential equation initial value problem, or by other means,

$$\cos\theta + \mathrm{i}\sin\theta = \mathrm{e}^{\mathrm{i}\theta}, \qquad \text{so} \qquad z = r\mathrm{e}^{\mathrm{i}\theta}.$$

A functional relation $w = f(z)$ between z and $w = u + \mathrm{i}v$ implies a transformation between the two-dimensional (x, y) and (u, v) planes. The function is *analytic* or *holomorphic* in a certain domain if its complex derivative exists in a neighborhood of every point in the domain. The derivative,

$$f'(z) = \lim_{\delta z \to 0} \frac{f(z + \delta z) - f(z)}{\delta z},$$

exists only if the limit is independent of how δz goes to zero, which leads to the Cauchy Riemann equations. That is to say, suppose $w = f(z)$ defines the transformation equations

$$u = u(x, y), \qquad v = v(x, y).$$

Then, if f is analytic inside some domain, then inside that domain,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

So these are the *Cauchy–Riemann conditions* required by analyticity of f. One of several consequences of analyticity is the *Cauchy–Goursat theorem*:

> If a function f is analytic at every point interior to and on a simple closed curve C then $\int_C f(z)\,\mathrm{d}z = 0$.

If in addition to existing, $f'(z)$ is also continuous at each point inside and on C, then this is the same as *Green's theorem* in the plane. The important use of the Cauchy–Goursat theorem is for moving integration contours.

The function f can fail to be analytic in a number of ways. I am mostly concerned with cases where $f'(z)$ fails to exist at isolated points. If $f(z)$ is analytic at z_0 then it has not only a first derivative but all its derivatives, and it turns out that the Taylor series at that point converges and it converges to $f(z)$, so that close enough to z_0,

$$f(z) = f(z_0) + (z - z_0)f'(z) + \frac{1}{2!}(z - z_0)^2 f''(z_0) + \cdots.$$

So consider a case where $f'(z)$ fails to exist at a single isolated point z_0. This point is an isolated singularity of f. Isolated singularities can be branching or non-branching. In the vicinity of a non-branching singularity, the function is single valued. Near a branching singularity, or a branch point, it is multivalued. An example of a branching singularity is a square root function. Branch points and particularly square root branch points are taken up in more detail below. So first, consider a non-branching singularity.

There can be a pole of order n at z_0. At such a point $f(z)$ is not analytic, but for some positive integer n, $(z - z_0)^n f(z)$ is analytic. If n is the smallest such integer, z_0 is a pole of order n. Then close enough to z_0,

$$f(z) = \frac{a_{-n}}{(z - z_0)^n} + \frac{a_{-n+1}}{(z - z_0)^{n+1}} + \cdots + \frac{a_{-1}}{z - z_0} + a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \cdots$$

This is a Laurent expansion. The coefficient a_{-1} is called the residue of f at z_0 and it has special meaning. When $n = 1$ the point z_0 is a simple pole. In the case there is no maximum negative power n, z_0 is called an (a non-branching) *essential singularity*.

The residue theorem is an important result. Consider the special case when the function f has only one, isolated, non-branching singularity located at z_0 in a certain region R. Suppose I integrate counter clockwise around a simple, closed contour C that remains in R and goes around z_0. Then using the Cauchy–Goursat theorem to deform the contour, the integration can be taken around a small circle centered at z_0. That is to say, the integration can be performed around the contour C_0 such that $|z - z_0| = \epsilon$. Assuming the validity of the Laurent expansion this integration gives the result

$$I_0 = \int_C f(z)\,\mathrm{d}z = \lim_{\epsilon\to 0}\int_{C_0} f(z)\,\mathrm{d}z = 2\pi i a_{-1}.$$

And so, if the function f is analytic except for a finite set of isolated non-branching singularities $\{z_k\}$ in the region R, and it is integrated counter clockwise around a simple, closed curve C that remains in R,

$$I = \int_C f(z)\,\mathrm{d}z = 2\pi i \sum_{z_k \in C} \mathrm{res}(z_k).$$

The summation is over all poles interior to the contour C. This is *the residue theorem*. Here $\mathrm{res}(z_k)$ denotes the residue of f at the pole z_k. It ends my sketchy reminder of basic material up to some point short of the topic of branching singularities or branch points and branch cuts. Now let's slow down and look at this matter of branching and multi-valued functions in somewhat more detail.

4.2 Branching and multi-valued functions in complex planes

A *branch point* z_0 of function $f(z)$ is a singularity such that the function is not single valued and cannot be made continuously single valued in any deleted neighborhood z_0. For example, the log function has a branch point at the origin, which is to say that in any neighborhood of $z = 0$, even with the point $z = 0$ removed, it isn't possible to make $\ln(z)$ both continuous and single valued. Consider z in polar coordinates,

$$\ln(z) = \ln\left(re^{\mathrm{i}\theta}\right).$$

When z is real and positive, the log is just the real, positive log of r. But when the angle increases to give z a positive imaginary part—if one assumes the basic property of the log function—the result is

$$\ln(z) = \ln\left(re^{\mathrm{i}\theta}\right) = \ln(r) + \mathrm{i}\theta.$$

Keeping the magnitude r constant, if z moves once around the circle counter-clockwise, the number z comes back to exactly the same place, but it would seem that the log value does not, since

$$\mathrm{e}^{2\pi i}z = z \quad \text{but} \quad \ln\left(\mathrm{e}^{2\pi i}\, z\right) = \ln(r) + \mathrm{i}\theta + 2\pi i.$$

So, if I put a certain number in for $z = r\mathrm{e}^{\mathrm{i}\theta}$, this recipe gives an ambiguous result. As θ increases from 0 to 2π the log changes smoothly from $\ln(r)$ to $\ln(r) + 2\pi i$. Should the log of z be $\ln(r) + \mathrm{i}\theta$, or $\ln(r) + \mathrm{i}\theta + 2\pi i$ or even $\ln(r) + \mathrm{i}\theta + 2n\pi i$ for any integer n? The logarithm is not really defined unless the ambiguity is resolved. It appears that in this sense the logarithm is a multi-valued function, so it is really more than one function.

Here is a way to proceed. Start by defining the log of a positive real number in the usual way. Then define a particular logarithm function in steps. First, select a range of θ

$$\theta_0 \leqslant \theta < \theta_0 + 2\pi \quad \text{or else} \quad \theta_0 < \theta \leqslant \theta_0 + 2\pi,$$

and say that in this range, the log is defined by a function

$$f_0(z) = \ln(r) + \mathrm{i}\theta,$$

and then notice that this log function is defined uniquely at all points in the z plane, but that it is discontinuous along the ray $\theta = \theta_0$. For example, one often takes $-\pi < \theta \leqslant +\pi$, so the discontinuity is along the negative real z axis. But this is only one possible choice. There are also infinitely many other possible values for the log function, namely

$$f_n(z) = \ln(r) + \mathrm{i}\theta + 2n\pi i.$$

One can think of these different functions as different *branches* of the log function. As θ increases, the branch

$$f_0(z) = \ln(r) + \mathrm{i}\theta \quad \text{for } \theta_0 < \theta \leqslant \theta_0 + 2\pi$$

is discontinuous at the *branch cut* where $\theta = \theta_0$. But it can be sutured onto the next branch,

$$f_1(z) = \ln(r) + \mathrm{i}\theta + 2\pi i \quad \text{for } \theta_0 < \theta \leqslant \theta_0 + 2\pi,$$

or, what is the same thing,

$$f_1(z) = \ln(r) + \mathrm{i}\theta, \quad \text{for } \theta_0 < \theta + 2\pi \leqslant \theta_0 + 4\pi.$$

To do this requires that z at $\theta + 2\pi$ is considered to be a different value from z at θ. In other words, there needs to be two copies of the z plane, one representing points for $\theta_0 < \theta \leqslant \theta_0 + 2\pi$ and another, different one for $\theta_0 + 2\pi < \theta \leqslant \theta_0 + 4\pi$. In fact, in order to define a log function that is continuous and single valued everywhere except at $z = 0$, there has to be an infinite collection of copies of the z plane, so that one is associated with each branch of the function. These copies are glued together at the branch cuts where $\theta = \theta_0 + 2n\pi$. Thus the log becomes a continuous, single valued function defined not on the z plane but on a domain consisting of copies of the z plane glued together to form a multi sheeted surface, which in this case is a helicoid.

So in general when it is useful near a branch point z_0 of a function $f(z)$—where the function has multiple values—these different values can define different component functions or different *branches* $f_n(z)$ of the function $f(z)$. Then one can think of the domain of definition differently. Rather than thinking of z as a point on the z plane, I can suppose that there are multiple *sheets* or different copies of the z plane sewn together along some branch cut to make a multi-sheeted surface, a *Riemann surface*. The multiple branches $f_n(z)$ are supposed to be sewn together in such a way that the function is continuous on the Riemann surface, except at the branch point itself. Then as z varies along the Riemann surface, the function definition shifts continuously, smoothly from one branch to another as z crosses a branch cut from one sheet to another.

Another example is $\sqrt{z - z_0}$. If I start by assuming the root is real and positive when $z - z_0$ is real and positive, then the primary branch of the square root is

$$f_1(z) = \sqrt{e^{\mathrm{i}\theta}|z - z_0|} \equiv e^{\mathrm{i}\theta/2}\sqrt{|z - z_0|}.$$

But there are two major points to consider. The first point is that the second equality ($\equiv$) is partly a definition. It is not just an identity, since there is a choice involved. The second point is that the branch cut has yet to be defined. Again there is a choice. One can have either $\theta_0 \leqslant \theta < \theta_0 + 2\pi$ or $\theta_0 < \theta \leqslant \theta_0 + 2\pi$ for any θ_0. In fact, the cut need not even be a straight line. One choice that is very common for the primary branch of the square root is to put the cut along the negative real axis of $z - z_0$. Thus

$$f_1(z) = e^{i\theta/2}\sqrt{|z - z_0|} \quad \text{for } -\pi < \theta \leqslant \pi.$$

Very often a computer language will assume this cut by default. In this case, the second branch of the function is

$$f_2(z) = -e^{i\theta/2}\sqrt{|z - z_0|}, \quad \text{for } -\pi < \theta \leqslant \pi,$$

or

$$e^{i\theta/2}\sqrt{|z - z_0|} \quad \text{for } \pi < \theta \leqslant 3\pi.$$

There are other useful choices. Sometimes it is handy to make the cut along the positive real axis of $z - z_0$, so that

$$f_1(z) = e^{i\theta/2}\sqrt{|z - z_0|}, \qquad f_2(z) = e^{-i\theta/2}\sqrt{|z - z_0|} \quad \text{for } 0 \leqslant \theta < 2\pi.$$

In any case, for the square root function, there are only two branches and the Riemann surface for the argument comprises two copies of the complex plane, sutured together in such a way that the function returns to the same value after traversing a circuit twice around the branch point, once on the first sheet where the branch is $f_1(z)$ and once on the second sheet where the branch is $f_2(z)$.

In exploring the elliptic integrals, one is often confronted with an expression of the form $\sqrt{u(z)}$ where $u(z)$ is a polynomial in z. There are important branch cut considerations here. Consider first the simple case $u(z) = 2z^2 - 3z + 1$. There will be two branching singularities for $\sqrt{u(z)}$, each of the square root type, since there are two roots of $u(z) = 0$.

$$\sqrt{u(z)} = \sqrt{2z^2 - 3\,z + 1} = \sqrt{2(z - 1)\left(z - \frac{1}{2}\right)}$$

Each root requires a choice of branch cut. Each one requires attention. So one must look at each factor separately. I am going to write

$$\sqrt{2(z - 1)\left(z - \frac{1}{2}\right)} = \sqrt{2}\,\sqrt{z - 1}\,\sqrt{z - \frac{1}{2}}\,.$$

This also is not an identity. It is not true, for example, that

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\,\sqrt{-1} = i\,i = -1.$$

For complex numbers it is not generally the case that $\sqrt{a}\,\sqrt{b}$ is the same as $\sqrt{ab}$ unless the branch cuts are taken in a way that allows it to be true. So in factoring the root expression, I am already starting to define the branch cuts. Next, on the primary sheet (the one for which $\sqrt{1} = +1$) I want to have a real positive number when z is a positive real number greater than 1. Defining the primary branch of each factor will now fix the primary branch of $\sqrt{u(z)}$. Let both factors be cut along the negative real axis,

$$\sqrt{2(z-1)\left(z-\frac{1}{2}\right)} = \sqrt{2}\sqrt{z-1}\sqrt{z-\frac{1}{2}},$$

let

$$(z-1) = |z-1|e^{i\theta_1}, \qquad \left(z-\frac{1}{2}\right) = \left|z-\frac{1}{2}\right| e^{i\theta_2},$$

and make two cuts such that

$$\sqrt{z-1} = e^{i\theta_1/2}\sqrt{|z-1|} \quad \text{for } -\pi < \theta_1 \leqslant \pi,$$

$$\sqrt{z-\tfrac{1}{2}} = e^{i\theta_2/2}\sqrt{\left|z-\tfrac{1}{2}\right|} \quad \text{for } -\pi < \theta_2 \leqslant \pi.$$

The discontinuity in the first root is along the ray from $+1$ to negative real infinity, and the discontinuity in the second is along the ray from $+\frac{1}{2}$ to negative real infinity. When these factors are multiplied together, the sign changes of the two factors cancel each other along the ray from $+\frac{1}{2}$ to negative real infinity, and so the final branch cut exists only along the segment of the real z axis from $+\frac{1}{2}$ to $+1$.

Sometimes it is useful to express the primary sheet in terms of the Cartesian real coordinates of z, namely (x, y) such that $z = x + iy$. Then if

$$w = \sqrt{z} = e^{i\theta/2}\sqrt{|z|}, \quad \text{with } -\pi < \theta \leqslant \pi \quad \text{and} \quad w = u + iv,$$

some elementary algebra shows on the primary sheet thus defined,

$$u = \frac{1}{\sqrt{2}}\sqrt{x + \sqrt{x^2+y^2}}, \qquad v = \frac{1}{\sqrt{2}}\frac{y}{\sqrt{x+\sqrt{x^2+y^2}}}.$$

Similar expressions are found in cases where the branch cut is placed differently.

Often one thinks of a functional relation between two complex quantities as a mapping between two plains, or between two Riemann surfaces. Let's think about that. The complex square root will serve as a good example.

Generally, a functional relationship, $w = f(z)$ where $z = x + iy$, $w = u + iv$ defines a mapping between the z plane with coordinates (x, y) and the w plane with coordinates (u, v). And the general notation will be to indicate which plane is which by naming the complex variable in the upper right as shown in figure 4.1.

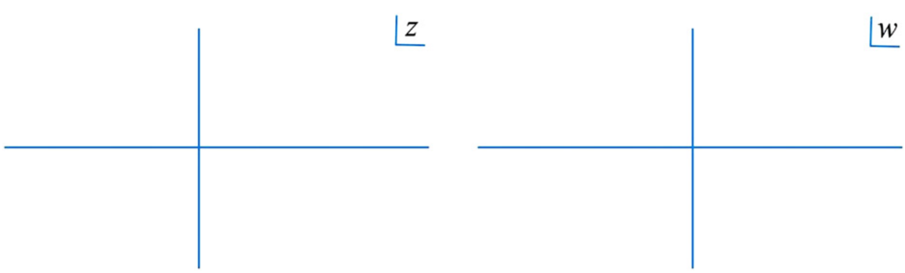


Figure 4.1.

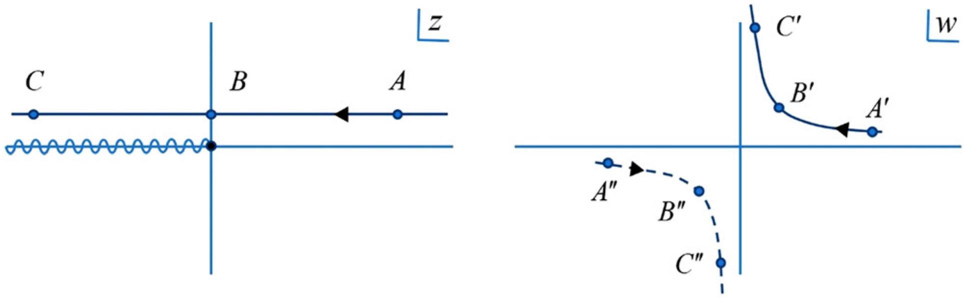


Figure 4.2.

To a point in the z plane there may correspond one or more points in the w plane and vice versa. Consider the case $z = w^2$. To each point in z correspond two points in w. If you imagine z as moving along the straight line A to B to C, it corresponds over in the w plane to a point moving either A' to B' to C' or A'' to B'' to C'' depending on which branch of the root you take. I have chosen to cut the z plane along the negative real axis as indicated by the wavy line. If point z at A were to move down onto the positive real axis, becoming the positive real number x, then the point A' would move onto $\sqrt{x}$ and point A'' would move onto $-\sqrt{x}$. So, as long as z does not cross the branch cut (remains on the same sheet of the z plane) I can call the first choice $\sqrt{z}$ and the second choice $-\sqrt{z}$. In polar form for this sheet of z, $\sqrt{z} = \sqrt{r}\,\mathrm{e}^{\mathrm{i}\theta/2}$, $-\pi < \theta \leqslant \pi$.

Since there are two branches of the square root, the z plane has two sheets (figure 4.2). At the branch cut, if z stays on the same sheet (let's call it the upper sheet, imagining the two sheets one on top of the other) then both branches of the function are discontinuous, because for example just above the cut, $\sqrt{-5 + .02\mathrm{i}} \approx +\mathrm{i}\sqrt{5}$, while just below it, $\sqrt{-5 - .02\mathrm{i}} \approx -\mathrm{i}\sqrt{5}$ so there is a discontinuity. However, the root function can be extended to the second sheet in such a way that the function on the connected surface is continuous. The branch $\sqrt{z}$ on the first sheet is sutured to $\sqrt{z}$ on the second and vice versa.

What happens as z moves on a contour that crosses a branch cut? Often one avoids doing this, but what happens depends on the meaning of 'crossing the branch cut.' The line A to B to C in figure 4.3 illustrates what happens if z is supposed to stay on the same sheet. Then the function is discontinuous, and so w moves discontinuously on A' to B' to C' where there is a jump. On the other hand, when z moves along D to E to F, the dashed segment indicates that the line goes under the cut onto the other sheet, as if disappearing into a pocket. F is thus actually on the other sheet of the z plane. Then the function is continuous, and w moves smoothly, without a jump, along the arc D' to E' to F''.

This concludes the brief but important recap of some things about complex variables. Complex variables will be a subject covered later on in more detail. But for now, these few facts are more or less sufficient for a preliminary encounter with elliptic integrals and the relations between them and the elliptic functions dealt with previously.

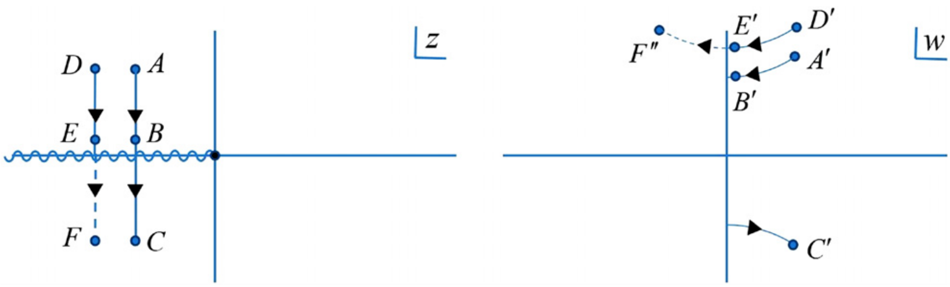


Figure 4.3.

4.3 Elliptic integrals and elliptic functions in complex planes

Returning to elliptic integrals, if $y(x) = \mathrm{sn}(x, k)$ then

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \left(1 - y^2\right)\left(1 - k^2y^2\right).$$

Normally, the variables can all be taken as real, and if y is positive, then for the real valued range of the *sn* function I can choose $0 \leqslant y \leqslant 1 \leqslant 1/k$. For this range, each radicand in the factorization

$$\mathrm{d}x = \frac{\mathrm{d}y}{\sqrt{(1 - y^2)(1 - k^2y^2)}} = \frac{\mathrm{d}y}{\sqrt{1 - y}\sqrt{1 + y}\sqrt{1 - ky}\sqrt{1 + ky}},$$

is positive and the equality holds. With the initial condition $y(0) = 0$, and so for $0 \leqslant y < 1$, one can write the real integral

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

This is the Legendre elliptic integral of the first kind, one of the standard elliptic integrals. It is often written as a function of ϕ such that $y = \sin\phi$. Thus ϕ is the polar angle in the construction of the ellipse that was used in chapter 1 (*q.v.*) to define the Jacobi elliptic functions. Then the *Legendre elliptic integral of the first kind* is usually defined as

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

Thus with a trig substitution, $t = \sin\theta$,

$$F(\phi, k) = \int_0^{\sin\phi} \frac{\mathrm{d}t}{\sqrt{(1 - t^2)}\sqrt{(1 - k^2t^2)}}. \tag{4.3}$$

So the elliptic integral of the first kind is related simply to the Jacobi *sn* function. If $y = \mathrm{sn}(x, k)$, then $x = F(\sin^{-1} y, k)$. In other words the elliptic integral of the first kind is an inverse of the *sn* elliptic function. However, the relationship is multi valued, as can be anticipated by the existence of branch points in the integrand, or by the fact that the inverse trigonometric functions are multi valued for the same reason, namely that the functions themselves are periodic. Thus the function values continually reoccur as the argument increases monotonically. This makes the inverse a multivalued function.

Some essential properties of elliptic integrals are uncovered when the arguments become complex. If x and y are both complex numbers, the integration is extended into the complex plane. It becomes an integral over a curve, a contour integral,

$$x = \int_{0 \text{ along } C}^{\text{to } y} \frac{\mathrm{d}z}{\sqrt{1 - z}\sqrt{1 + z}\sqrt{1 - kz}\sqrt{1 + kz}},$$

where indeed not only the limits of integration but also the contour C of integration needs to be specified. This lets one to explore complex values the function $x = F(\sin^{-1} y, k)$ considered as a relationship between x and y. Factoring the radicand and interpreting the square root as the square root of each factor separately permits continuing the function F analytically. When x is between 0 and 1, and the integration is along the real z axis, then the integral reduces back to the standard form.

$$x = \int_{0 \text{ along } C}^{\text{to } y} \frac{\mathrm{d}z}{\sqrt{1 - z}\sqrt{1 + z}\sqrt{1 - kz}\sqrt{1 + kz}}$$

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

But in general x, y and z can be complex variables. Because of the inverse relation between *sn* and *F*, this continuation also exposes several interesting properties of *sn*. Look at the integral as a mapping between x and y, both complex. In general,

$$x = F(\sin^{-1} y,\ k), \qquad y = \mathrm{sn}(x, k).$$

The z plane is the plane of the integrand. Since $y = \mathrm{sn}(x, k)$, one has that $x = 0$ when $y = 0$ and so the lower limit of integration, or the initial point on the integration contour, is $z = 0$. The upper limit is where $z = y$ and the integral along the contour specified in the z plane gives a value for x. The fact that choosing different contours connecting 0 to y will generally give different x values shows that there are different x values with the same value of y when $y = \mathrm{sn}(x, k)$.

So consider now the function defined by the complex contour integral, with the integration contour as shown in the z plane at the left. I take the branch cut for each root factor to be back along the negative real direction of its argument.

Thus the sign changes cancel in such a way that there are two remaining segments of branch cut, like twin button holes, one from $-1/k$ to -1 and the other from 1 to $1/k$. The value of the integral starting at 0, integrating along the contour C as shown on the left, is thus a function of the endpoint y. To any given value of y along C in the z plane corresponds a value of x. As the integration proceeds in the z plane from 0 to y in the range $0 < y < 1$, the x value of the integral is a point along the real x axis, corresponding to the real-valued elliptic integral function $x = F(\sin^{-1} y, k)$. Viewing this relation as $y = \sin(x, k)$ and remembering symmetries of the reference ellipse it is clear that the value $y = 1$ should occur first at the quarter period point of the argument. This is similar to the way in which the sine function $y = \sin(x)$ first reaches its maximum value $y = 1$ at the quarter period $x = \pi/2$ of its argument. Thus

$$\mathrm{sn}(K, k) = 1 \quad \text{where } K = K(k) \equiv \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1 - t^2)(1 - k^2 t^2)}} = F\left(\frac{\pi}{2}, k\right).$$

This function $K(k)$ is the *complete elliptic integral of the first kind*, and again from symmetry of the reference ellipse it is seen that for any real x,

$$\mathrm{sn}(x + 4K, k) = \mathrm{sn}(x, k).$$

Now what happens if the integration continues to a y value greater than 1, by means of 'hopping over' the branch point at $z = 1$? The phase of the radicand $1 - z$ changes. In polar coordinates centered at $z = 1$ one has,

$$\sqrt{1 - t} \to \sqrt{\mathrm{e}^{-\mathrm{i}\pi}|1 - t|} = \mathrm{e}^{-\mathrm{i}\pi/2}\sqrt{t - 1}\,.$$

The change in phase is $\mathrm{e}^{+\mathrm{i}\pi}$ because upon hopping over the branch point from left to right the polar angle decreases by π. (For ducking under the branch point the angle would increase by π.) Thus on hopping over $z = 1$,

$$\sqrt{1 - t} \to -\,\mathrm{i}\sqrt{t - 1}\,.$$

So I see that when y is a point on the real z axis such that $1 < y < 1/k$, then if the contour humps over the branch point and continues,

$$\begin{aligned} x = F(\sin^{-1} y, k) &= \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1 - t^2)(1 - k^2 t^2)}} + \mathrm{i}\int_1^y \frac{\mathrm{d}t}{\sqrt{(t^2 - 1)(1 - k^2 t^2)}} \\ &= K + \mathrm{i}\int_1^y \frac{\mathrm{d}t}{\sqrt{(t^2 - 1)(1 - k^2 t^2)}}. \end{aligned}$$

So such values of y in the z plane correspond to points on the vertical segment of image curve shown in the x plane. The x point continues to move upward as t approaches $1/k$. Then if the y value hops over the next branch point at $1/k$,

$$\sqrt{1 - kt} \to \sqrt{\mathrm{e}^{-\mathrm{i}\pi}|1 - kt|} = \mathrm{e}^{-\mathrm{i}\pi/2}\sqrt{kt - 1} = -\mathrm{i}\sqrt{kt - 1}$$

and so

$$x = F(\sin^{-1} y, k) = K + \mathrm{i}\int_1^{1/k} \frac{\mathrm{d}t}{\sqrt{(t^2-1)(1-k^2t^2)}} - \int_{1/k}^{y} \frac{\mathrm{d}t}{\sqrt{(t^2-1)(k^2t^2-1)}},$$
$$= K + \mathrm{i}K' - \int_{1/k}^{y} \frac{\mathrm{d}t}{\sqrt{(t^2-1)(k^2t^2-1)}},$$

where

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

This shows that as the end point y moves along the positive real axis, $z = t$, with $1/k < t$ as long as the integration contour humps over each of the branch points, then the corresponding point x is somewhere along the upper horizontal line shown. As y increases toward ∞, x must be moving to the left.

The constant K' depends on k, but it also can be seen as another complete elliptic integral of the first kind, but as a function of a different variable $k' = \sqrt{1-k^2}$, the *complementary modulus*. Here is how it works. First change integration variables from t to u on such a way that $t = \pm 1$ maps to $u = \pm 1$ and $t = \pm 1/k$ maps to $u = 0$. This can be done easily by assuming a linear relationship between t^2 and u^2. So let $t^2 = Au^2 + B$. Applying the correspondences $B = 1/k^2$, $A = 1 - 1/k^2$. Solving for u and for t, since these are both real positive,

$$u = \frac{\sqrt{k^2t^2-1}}{\sqrt{k^2-1}}, \qquad t = \frac{\sqrt{(k^2-1)u^2+1}}{k}.$$

Then

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

Finally, the result is

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

Similarly, suppose z (which is the same as t in this case, since it is real) continues out along the real z axis, and as it goes to $t \to \infty$,

$$y \to K + \mathrm{i}K' - \int_\infty^{1/k} \frac{\mathrm{d}t}{\sqrt{(t^2-1)(k^2t^2-1)}},$$

and in the integrand, letting $t = 1/(ks)$, the s integral becomes $K(k)$ so that

$$y \to \mathrm{i}K',$$

as indicated in figure 4.4.

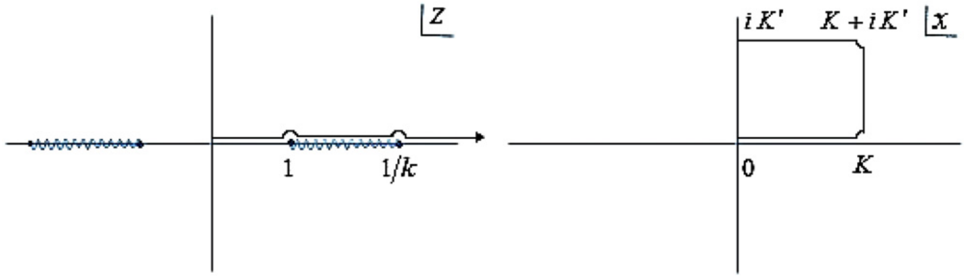


Figure 4.4.

Now suppose the integration is performed again, this time exploring the negative real axis. As y moves out along the negative real z axis, again hopping over the branch points, the image point x goes around a similar rectangle, from 0 to $-K$ to $-K + iK'$ to $+iK'$. If I let the z point have a small positive imaginary part near the origin, then the integrand and thus also the integral x also has a small positive imaginary part. If the end point y of the integration contour moves around in the upper half or the image point x moves around inside the rectangle. Therefore, the whole upper z plane is mapped by the integration to the interior of a rectangle, half of which is shown in figure 4.4.

Continuing on, if the integration is performed again along the positive real axis, this time ducking under the branch points, then the image point moves along a rectangular path from 0 to K to $K - iK'$. And if the same is done with integration along the negative real axis, ducking under the singularities, the image point traces out a rectangular path from 0 out to $-K$ to $-K - iK'$ to $-iK'$. So it turns out that the lower half z plane is mapped into the lower half rectangle.

So one sees that an image of all of the whole z plane, excluding the singularities, maps to the inside of the rectangle with corners at $K + iK'$, $-K + iK'$, $-K - iK'$ and $K - iK'$ in the x plane. And so now one wonders, what about the rest of the x plane? I know that, at least for real x, the function $y = \mathrm{sn}(x, k)$ is a periodic function defined for all x, which corresponds to the fact that its inverse must be multivalued.

Consider two other integration paths, C_1 and C_2 as in figure 4.5. Using the Cauchy–Goursat theorem to deform the contours, the two integration paths C_1 and C_2 can both be changed into a straight line along the real axis from 0 to 1, and then another straight line from 1 to y. The contours appear the same, then, except that the contour C_2 makes a tiny 270° turn around the branch point and thus winds down onto the lower sheet of the z plane. So the integrand on the second leg of the C_2 journey, the line from 1 to y, has an extra minus sign. The value of the integral from 0 to 1 is K in either case, and so if the integral from 0 to y along C_1 is $K + J$ then the integral from 0 to y along C_2 is $K - J$. One has

$$x_1 = K + J, \qquad x_2 = K - J, \qquad x_2 = 2K - x_1,$$

so

$$\mathrm{sn}(2K - x, k) = \mathrm{sn}(x, k).$$

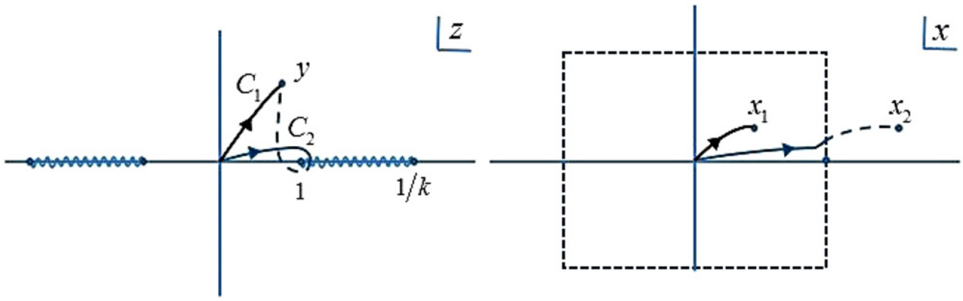


Figure 4.5.

This again could be seen from symmetry of the reference ellipse and the fact that K is the quarter period. It is analogous to $\sin(\pi - \theta) = \sin\theta$. However, the interesting thing about the elliptic functions of complex x is that they have two independent periods in the x plane. One of the periods is real, and in the case of the *sn* and *cn* that period is $4K$.

But to see the other periodicity, consider two more integration paths, both in the z plane as shown as shown in figure 4.6, each still ending at the same value y. The curve C_1 is the same in each case. In the left part of figure 4.6, C_2 consists of three steps. Step one goes from 0 to just short of 1, step two goes around the branch cut counter clockwise, and step three goes from the crossing point just short of 1 up to y. Thus, in the figure on the left, the contour C_1 and the contour C_2 remain on the first sheet of the integrand variable z. The first and last steps in C_2 taken together are equivalent via Cauchy–Goursat, to the curve C_1—meaning that these two parts of C_2 spliced together could be deformed into C_1 and give an integral with the same value as integrating on C_1 would give, the difference between integrating on C_2 and integrating on C_1 comes just from going around the branch cut. When z hops over the branch point at 1, the phase of the integrand increases from 1 to i. Integrating along the top from 1 to $1/k$ gives a contribution $+iK'$. The trip all the way around the branch point $1/k$ counter clockwise, gives an additional phase change of -1, but then integrating back along the bottom of the contour, since the direction of integration is also reversed, giving another -1, the

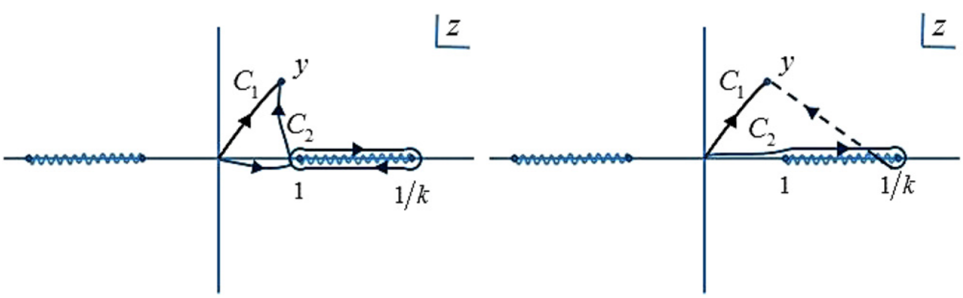


Figure 4.6.

contribution is again $+iK'$. Thus, if integrating along C_1 gives, x_1 then integrating along C_2 in the figure on the left gives $x_2 = x_1 + 2iK'$. Thus

$$\mathrm{sn}(x + 2iK', k) = y = \mathrm{sn}(x, k).$$

This demonstrates that *sn* has two different periods in the complex plane of its argument, namely a real period $4K$ and an imaginary period $2iK'$. Because the elliptic functions are related to each other algebraically, it seems clear that the others all have two periods as well. A function of complex argument x cannot have more than two periods, because if α and β are two complex numbers representing periods, then when seen as vectors in two dimensions, they need to be independent of one another, meaning that α/β cannot be real. But two independent vectors already span the complex plane, so there cannot be more than two independent periods. Any other period has to be a combination of these two.

Consider next figure 4.6. Along the curve C_2 the integration variable hops over the branch point at 1 so that the integrand acquires a phase *i*. Then, as before, the integration above the cut from 1 to $1/k$ gives a contribution iK'. Then going again all the way around the branch point $1/k$ changes the sign. The integration contour ducks under the branch cut, and continues on the second sheet. At this point, let's say point p, z is arbitrarily close to $1/k$ and the phase of the integrand is *minus* what it was at the corresponding point p on the upper edge of the cut before going around the branch point. So there is point p on the upper sheet and point p on the lower sheet. The integration contour C_1 can be deformed over to touch point p on the upper sheet. So the corresponding integral has the value $x_1 = K + iK' + J$, where J is the integral from p on the upper sheet over to y, which one could take along a straight line as shown. The difference in going to y via contour C_2 as opposed to going via C_1 is only in the last step. Along C_2 the contributions are first K to get from 0 to 1, then hopping over the branch point at 1 and integrating to $1/k$ gives a contribution iK', all the same, and the integration variable z arrives again at p on the top sheet. But the difference is that now z has gone around the branch point and the integrand has changed changes sign. So, starting from p on the lower sheet and integrating over to y (which is y on the lower sheet) gives a contribution of $-J$ rather than $+J$. Hence

$$x_1 = K + iK' + J, \qquad x_2 = K + iK' - J,$$

and thus $\mathrm{sn}(K + iK' + J, k) = y = \mathrm{sn}(K + iK' - J, k)$, where of course J is arbitrary since y is arbitrary. So one finds generally, by means of the right part of figure 4.6, that for any x,

$$\mathrm{sn}(x, k) = y = \mathrm{sn}(2K + 2iK' - x, k).$$

This is a combination of two previous results. It demonstrates again the method of reasoning that relates periodicity of the elliptic functions to multi valuedness of the inverse functions, which are elliptic integrals.

4.4 Example

Here is an example calculation. A finite difference model Schrödinger equation for a free particle in one dimension would have a Hamiltonian matrix

$$H_{m,n} = -\nu\left(\delta_{m,n+1} + \delta_{m,n-1}\right),$$

The density of states, or in other words the energy eigenvalue density defined by

$$D(E) = \frac{1}{N}\sum_k \delta\left(E - E_k\right),$$

where the sum is over the eigenvalues E_k of H and N is the size of H turns out as $N \to \infty$ to be

$$D(E) = \frac{1}{\pi}\frac{1}{\sqrt{4\nu^2 - E^2}}\,\Theta\left(1 - \frac{E^2}{4\nu^2}\right).$$

Here Θ is the unit step function. Now consider a two dimensional model defined on a square lattice via the new Hamiltonian

$$K_{pq,rs} = -\nu\left(\delta_{p,r+1} + \delta_{p,r-1}\right)\delta_{qs} - \nu\delta_{p,r}\left(\delta_{q,s+1} + \delta_{q,s-1}\right).$$

The Schrödinger difference equation based on K separates, and the energy eigenvalues of the two dimensional problem are just sums of pairs of eigenvalues of H,

$$\tilde{E}_{k_x k_y} = E_{k_x} + E_{k_y}$$

So,

$$\tilde{D}(E) = \frac{1}{N^2}\sum_{k_x k_y} \delta\left(E - E_{k_x} - E_{k_y}\right).$$

Thus $\tilde{D}(E)$ is a convolution.

$$\begin{aligned}
\tilde{D}(E) &= \frac{1}{N^2}\sum_{k_x k_y}\int_{-\infty}^{\infty} \delta\left(E - E_{k_x} - x\right)\delta\left(x - E_{k_y}\right)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{N}\sum_k \delta\left(E - E_k - x\right)\frac{1}{N}\sum_k \delta\left(x - E_k\right)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} D(E - x)D(x)\,\mathrm{d}x.
\end{aligned}$$

This process gives an elliptic integral for the density of states for an electron on a two dimensional square lattice, namely

$$\tilde{D}(E) = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{4\nu^2 - (E - x)^2}}\frac{1}{\sqrt{4\nu^2 - x^2}}\Theta\left(1 - \frac{(E - x)^2}{4\nu^2}\right)\Theta\left(1 - \frac{x^2}{4\nu^2}\right)\mathrm{d}x.$$

The range of integration is limited to within

$$-2\nu < x < 2\nu \quad \text{and also within} \quad E - 2\nu < x < E + 2\nu.$$

Thus it is zero when $|E| > 4\nu$. For $0 < E < 4\nu$, the range is

$$E - 2\nu < x < 2\nu.$$

For the energy range $-4\nu < E < 0$, the range of integration is

$$-2\nu < x < E + 2\nu.$$

$\tilde{D}(-E) = \tilde{D}(E)$ in this case, so one only needs to consider the positive energy case.

$$\tilde{D}(E) = \frac{1}{\pi^2} \int_{E-2\nu}^{2\nu} \frac{1}{\sqrt{4\nu^2 - (E-x)^2}} \frac{1}{\sqrt{4\nu^2 - x^2}} \mathrm{d}x.$$

Roots give branch points at $x_1 = -2\nu$, $x_2 = E - 2\nu$, $x_3 = 2\nu$, and $x_4 = E + 2\nu$. To match the complete elliptic integral of the first kind,

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

The branch points in x need to map onto the branch points

$$t_1 = -1/k, \qquad t_2 = -1, \qquad t_3 = 1, \qquad t_4 = 1/k.$$

In particular, the upper limit 2ν should correspond to $t = 1$. I use the cross ratio equality,

$$\frac{(x_1 - x_2)(x_3 - x_4)}{(x_1 - x_4)(x_3 - x_2)} = \frac{(-1/k + 1)(1 - 1/k)}{(-1/k - 1/k)(1 + 1)},$$

$$\frac{E^2}{(4\nu - E)(4\nu + E)} = \frac{(k-1)^2}{4k}.$$

Thus

$$k = \frac{4\nu - E}{4\nu + E}, \quad \text{or} \quad k = \frac{4\nu + E}{4\nu - E}.$$

When $0 < E < 4\nu$, the first choice is in the range $0 < k < 1$, so I select that. Then

$$\frac{(x - x_2)(x_3 - x_4)}{(x - x_4)(x_3 - x_2)} = \frac{(x - E + 2\nu)(2\nu - E - 2\nu)}{(x - E - 2\nu)(2\nu - E + 2\nu)} = \frac{(t+1)(1 - 1/k)}{(t - 1/k)(1 + 1)},$$

so

$$t = \frac{E - 2x}{E - 4\nu} \quad \text{or} \quad x = \frac{1}{2}(E(1-t) + 4t\nu).$$

The limits $x = E - 2\nu$ and $x = 2\nu$ map onto $t = -1$ and $t = 1$ respectively, so when $0 < E < 4\nu$,

$$\begin{aligned}
\tilde{D}(E) &= \frac{1}{\pi^2}\int_{E-2\nu}^{2\nu} \frac{1}{\sqrt{4\nu^2 - (E-x)^2}}\frac{1}{\sqrt{4\nu^2 - x^2}}\,\mathrm{d}x \\
&= \frac{1+k}{4\pi^2}\int_{-1}^{1}\frac{1}{\sqrt{(1-t^2)(1-k^2t^2)}}\,\mathrm{d}t \\
\tilde{D}(E) &= \frac{4}{\pi^2}\frac{1}{E+4\nu}K\left(\frac{4\nu - E}{4\nu + E}\right).
\end{aligned}$$

Thus using symmetry

$$\tilde{D}(E) = \frac{4}{\pi^2}\frac{1}{|E| + 4\nu}K\left(\frac{4\nu - |E|}{4\nu + |E|}\right).$$

So up to this point, the material developed is really a connection just between one elliptic function, namely the Jacobi function $\mathrm{sn}(x, k)$, and one type of elliptic integral, namely the Legendre elliptic integral of the first kind. Now let's return to the more general elliptic integrals.

4.5 Reduction of the most general elliptic integral in terms of the three Legendre forms

Portions of the analysis procedure below follows loosely that outlined in the book by Baker [1], although he has used a different method for reducing the surd expressions to standard form. At the start of the discussion on elliptic integrals was the general case,

$$I = \int R\left(y, \sqrt{u}\right)\mathrm{d}y,$$

where

$$u = u(y) = c_4y^4 + c_3y^3 + \cdots c_0$$

with not both c_4 and c_3 equal zero, and R some rational function. After a line or so of algebra, it turned out that I could be reduced to

$$I = \int R_1(y)\,\mathrm{d}y + \frac{R_2(y)}{\sqrt{u(y)}}\,\mathrm{d}y,$$

where $R_1(y)$ and $R_2(y)$ are rational functions of y, i.e. they are ratios of polynomials in y. The first integral is elementary, and one can deal with it using the partial fractions decomposition. The second can be reduced to a combination of three specific forms that are tabulated functions. These are the Legendre elliptic integrals of the first, second and third kind. Here they are respectively.

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

$$E(\sin^{-1}y,\, k) = \int_0^y \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\,\mathrm{d}t, \tag{II}$$

$$\Pi(\sin^{-1}y,\, q,\, k) = \int_0^y \frac{\mathrm{d}t}{(1+qt^2)\sqrt{(1-t^2)(1-k^2t^2)}}. \tag{III}$$

In each of these, the k corresponds to the elliptic modulus, and so it is typically a real number $0 \leqslant k \leqslant 1$. In the third formula, q is any real number. These integrals are seen as functions of ϕ, where $y = \sin\phi$. Discussion above concerned mostly the first kind and its relation to the *sn* function, although the comments that were made on branch cuts are generally relevant.

It was noted above also that if I make the trig substitution, $t = \sin\theta$, $y = \sin\phi$ then

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

Similarly for the elliptic integral of the second kind,

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

By the elliptic 'trig substitution', $t = \mathrm{sn}(x, k)$, $\mathrm{d}t = \mathrm{cn}(x, k)\ \mathrm{dn}(x, k)\mathrm{d}x$, the Legendre integral type 2 becomes

$$E(\sin^{-1}y,\, k) = \int_0^y \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\,\mathrm{d}t = \int_0^{\sin^{-1}(y,k)} \frac{\sqrt{1-k^2sn^2}}{\sqrt{1-sn^2}}\ cn\ dn\ \mathrm{d}x,$$

$$E(\sin^{-1}y,\, k) = \int_0^{\sin^{-1}(y,k)} \mathrm{dn}^2(x,\, k)\,\mathrm{d}x.$$

So, it is an integral of *dn* squared.

For the elliptic integral of the third kind,

$$\Pi(\sin^{-1}y,\, q,\, k) = \int_0^y \frac{\mathrm{d}t}{(1+qt^2)\sqrt{(1-t^2)(1-k^2t^2)}}$$
$$= \int_0^{\sin^{-1}y} \frac{\mathrm{d}\theta}{(1+q\sin^2\theta)\sqrt{(1-k^2\sin^2\theta)}}.$$

Otherwise, making the elliptic function substitution,

$$t = \mathrm{sn}(x,\, k), \qquad \mathrm{d}t = \mathrm{cn}(x,\, k)\,\mathrm{dn}(x,\, k)\,\mathrm{d}x,$$

$$\Pi(\sin^{-1}y,\, q,\, k) = \int_0^{\sin^{-1}y} \frac{\mathrm{d}x}{1+q\,\mathrm{sn}^2(x,\, k)}.$$

A method for transforming the general quartic polynomial

$$u = c_4 t^4 + c_3 t^3 + \cdots + c_0$$

into the Legendre form $(1 - t^2)(1 - k^2 t^2)$ was the topic of chapter 3.

Let's assume for the moment that $u = \alpha y^4 + \beta y^3 + \gamma$. The general elliptic integrand reduced as

$$R(y, \sqrt{u}) = R_1(y) + \frac{R_2(y)}{\sqrt{u(y)}},$$

and I have noted that the first integrand $R_1(y)$ is rational and so the integral can be done by elementary methods. Let's return to the problem of reducing the general elliptic integral in terms of Legendre's three standard forms. So to that end the next task in view is to reduce the second integral to some combination of Legendre forms. The integrand is

$$\frac{R_2}{\sqrt{u}} = \frac{1}{\sqrt{u}} \frac{G_0 + G_1 y + G_2 y^2 + G_3 y^3 + \cdots}{H_0 + H_1 y + H_2 y^2 + H_3 y^3 + \cdots}.$$

Divide powers in the denominator into odd and even.

$$\frac{R_2}{\sqrt{u}} = \frac{1}{\sqrt{u}} \frac{G_0 + G_1 y + G_2 y^2 + G_3 y^3 + \cdots}{H_0 + H_2 y^2 + \cdots + y(H_1 + H_3 y^2 + \cdots)}.$$

and multiply top and bottom by $H_0 + H_2 y^2 + \cdots - y(H_1 + H_3 y^2 + \cdots)$, so that

$$\frac{R_2}{\sqrt{u}} = \frac{1}{\sqrt{u}} \frac{M_0 + M_2 y^2 + \cdots + y(M_1 + M_3 y^2 \cdots)}{N_0 + N_2 y^2 + N_4 y^4 + \cdots}.$$

So this shows

$$\frac{R_2(y)}{\sqrt{u}} = \frac{R_3(y^2)}{\sqrt{u}} + \frac{R_4(y^2) y}{\sqrt{u}},$$

where again R_3 and R_4 are rational functions, but this time they are rational functions of y^2. Since u is also a function of y^2, substituting z for y^2 in the second integrand makes the corresponding integral elementary,

$$\int \frac{R_4(y^2) y \,\mathrm{d}y}{\sqrt{u}} = \frac{1}{2} \int \frac{R_4(z) \,\mathrm{d}z}{\sqrt{\alpha z^2 + \beta z + \gamma}}.$$

What's left is to reduce the other integral to a combination of Legendre forms. The rational function can be written by partial fractions as a finite sum of terms, and so the general form is

$$\frac{R_3(y^2)}{\sqrt{u}} = \left(a_0 + a_1 y^2 + \cdots + a_{2n} y^{2n} + \sum_{k=1}^{N} \sum_{m=1}^{n_k} \frac{a_{k,-m}}{(y^2 - b_k)^m} \right) \frac{1}{\sqrt{u}}.$$

The latter decomposition shows the elliptic integrals depend on a set of more elementary integrals with integrands of the forms

$$\frac{1}{\sqrt{u}},\ \frac{y^2}{\sqrt{u}},\ \frac{y^4}{\sqrt{u}}\ \cdots \quad \text{and} \quad \frac{1}{(y^2-b)\sqrt{u}},\quad \frac{1}{(y^2-b)^2\sqrt{u}},\ \cdots$$

Integrals of the first set are related by a recursion. That this is so follows from integrating by parts. Consider

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y}\left(y^{2m-3}\sqrt{u}\right) &= (2m-3)y^{2m-4}\sqrt{u} + y^{2m-3}\frac{1}{2\sqrt{u}}\frac{\mathrm{d}}{\mathrm{d}y}\left(\alpha y^4 + \beta y^2 + \gamma\right)\\ &= (2m-3)y^{2m-4}\frac{1}{\sqrt{u}}(\alpha y^4 + \beta y^2 + \gamma) + y^{2m-3}\frac{1}{\sqrt{u}}(2\alpha y^3 + \beta y).\end{aligned}$$

Collecting like powers of y on the right side,

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(y^{2m-3}\sqrt{u}\right) = (2m-3)\gamma\,\frac{y^{2m-4}}{\sqrt{u}} + (2m-2)\beta\,\frac{y^{2m-2}}{\sqrt{u}} + (2m-1)\alpha\,\frac{y^{2m}}{\sqrt{u}}$$

So, when $m \geqslant 1$,

$$(2m-1)\alpha\int\frac{y^{2m}}{\sqrt{u}}\,\mathrm{d}y = y^{2m-3}\sqrt{u} - (2m-2)\beta\int\frac{y^{2m-2}}{\sqrt{u}}\,\mathrm{d}y - (2m-3)\gamma\int\frac{y^{2m-4}}{\sqrt{u}}\,\mathrm{d}y,$$

which shows that all of the elliptic integrals of the form

$$I_m = \int\frac{y^{2m}}{\sqrt{u}}\,\mathrm{d}y$$

reduce to expressions involving only

$$I_0 = \int\frac{1}{\sqrt{u}}\,\mathrm{d}y \qquad \text{and} \qquad I_1 = \int\frac{y^2}{\sqrt{u}}\,\mathrm{d}y.$$

In the standard case when

$$\begin{aligned}&u = (1-y^2)(1-k^2y^2), \quad \text{or} \quad \alpha = k^2, \quad \beta = -(1+k^2), \quad \gamma = 1,\\ &I_0 = \int\frac{1}{\sqrt{u}}\,\mathrm{d}y = F(\sin^{-1}y, k),\\ &I_1 = \frac{1}{k^2}\int\frac{k^2y^2 - 1 + 1}{\sqrt{u}}\,\mathrm{d}y = \frac{1}{k^2}((F(\sin^{-1}y, k) - E(\sin^{-1}y, k)).\end{aligned}$$

Finally, the other sequence of integrals needs to be reduced, namely the ones with integrands

$$\frac{1}{(y^2-b)\sqrt{u}},\quad \frac{1}{(y^2-b)^2\sqrt{u}},\quad \frac{1}{(y^2-b)^3\sqrt{u}},\ \cdots.$$

First, it is easy to see that

$$\frac{\partial}{\partial b}\int \frac{1}{(y^2-b)^m\sqrt{u}}\,\mathrm{d}y = -m\int \frac{1}{(y^2-b)^{m+1}\sqrt{u}}\,\mathrm{d}y.$$

However, this is not of much utility, since one does not have information about how to differentiate the relevant integrals with respect to b. So again integration by parts is useful. Consider

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{y\sqrt{u}}{(y^2-b)^{n-1}}\right) &= \frac{1}{(y^2-b)^{n-1}}\left(\sqrt{u}+\frac{1}{2\sqrt{u}}\,y\frac{\mathrm{d}}{\mathrm{d}y}(\alpha y^4+\beta y^2+\gamma)\right)\\ &\quad - y\sqrt{u}(n-1)\frac{1}{(y^2-b)^n}2y\\ &= \frac{3\alpha y^4+2\beta y^2+\gamma}{(y^2-b)^{n-1}\sqrt{u}} - 2(n-1)\frac{y^2(\alpha y^4+\beta y^2+\gamma)}{(y^2-b)^n\sqrt{u}}.\end{aligned}$$

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{y\sqrt{u}}{(y^2-b)^{n-1}}\right) = \frac{(3\alpha y^4+2\beta y^2+\gamma)(y^2-b)-2(n-1)y^2(\alpha y^4+\beta y^2+\gamma)}{(y^2-b)^n\sqrt{u}},$$

so

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{y\sqrt{u}}{(y^2-b)^{n-1}}\right) = \frac{\alpha(5-2n)y^6+(2\beta(2-n)-3\alpha b)y^4+(\gamma(3-2n)-2\beta b)y^2-\gamma b}{(y^2-b)^n\sqrt{u}}.$$

Now the strategy is to replace the different powers of y^2 in terms of powers of $z = y^2 - b$ which appears in the denominator. Then by cancellation there will be a recursion involving terms with constants in the numerator and various powers of z in the denominator. Thus a recursion relation will result.

$$\begin{aligned}&y^2-b=z, \quad \text{so}\\ &\quad y^2 = z+b, \quad y^4 = z^2+2bz+b^2 \quad y^3 = z^3+3bz^2+3b^2z+b^3.\end{aligned}$$

So, expanding the numerator in terms of z, each term cancels a certain number of factors in the denominator $z^n\sqrt{u}$. Then, substituting back in terms of y and integrating, the general recursion turns out to be

$$\begin{aligned}\frac{y\sqrt{u}}{(y^2-b)^{n-1}} &= -(2n-5)\alpha\int\frac{\mathrm{d}y}{(y^2-b)^{n-3}\sqrt{u}} - (2n-4)(\beta+3\alpha b)\int\frac{\mathrm{d}y}{(y^2-b)^{n-2}\sqrt{u}}\\ &\quad - (2n-3)(3\alpha b^2+2\beta b+\gamma)\int\frac{\mathrm{d}y}{(y^2-b)^{n-1}\sqrt{u}}\\ &\quad - (2n-2)(\alpha b^3+\beta b^2+\gamma b)\int\frac{\mathrm{d}y}{(y^2-b)^n\sqrt{u}}.\end{aligned} \tag{4.4}$$

For the special case, $n = 2$,

$$\frac{y\sqrt{u}}{(y^2 - b)} = \alpha \int \frac{(y^2 - b)\, \mathrm{d}y}{\sqrt{u}} - (3\alpha b^2 + 2\beta b + \gamma) \int \frac{\mathrm{d}y}{(y^2 - b)\sqrt{u}} - 2(\alpha b^3 + \beta b^2 + \gamma b) \int \frac{\mathrm{d}y}{(y^2 - b)^2 \sqrt{u}}.$$

From the latter, it is clear that

$$\int \frac{\mathrm{d}y}{(y^2 - b)^2 \sqrt{u}}$$

can be evaluated via

$$\int \frac{\mathrm{d}y}{(y^2 - b)\sqrt{u}}, \qquad \int \frac{y^2 \mathrm{d}y}{\sqrt{u}} \quad \text{and} \quad \int \frac{\mathrm{d}y}{\sqrt{u}},$$

and hence in view of the previous discussion, all the elliptic integrals can be evaluated in terms of the three basic elliptic integrals of Legendre.

So in this section the reduction for a general elliptic integral of the form (4.1) in the case where $\sqrt{u}$ is of the standard form $\sqrt{(1 - y^2)(1 - k^2 y^2)}$ has led to an expression, possibly quite complicated, involving the Legendre integrals. Because the general quartic case has been mapped onto this standard form by means of a linear fractional transformation, the general elliptic integral is thus so reduced. A similar reduction is carried out for the cubic case, and can be found for example in the first chapter of the book by Baker [1]. These general elliptic integrals result from a somewhat larger range of physical problems than is represented by the elliptic functions themselves. Along the way, by looking at the elliptic integrals in the complex plane, it was seen that the elliptic functions have two independent periodicities. In fact, I mentioned that an elliptic function can be defined as a function of a complex variable z that has two independent periods in the z plane, and no singularities except poles in the finite plane. This is the alternative approach to the elliptic functions that was pursued by Weierstrass. In a later volume as part of a study of applied complex variables, the Weierstrass theory of elliptic functions will be taken up. In this volume, the intention has been to provide first an elementary, trigonometric introduction to the elliptic functions of Jacobi. Then after some brief review of complex variables, the aim was to give enough of the details of reducing differential equations to a form solvable by elliptic functions or elliptic integrals so that the reader can apply the theory to actual physical situations.

Reference

[1] Baker A L 1890 *Elliptic Functions: An Elementary Text-Book for Students of Mathematics* (New York: Wiley) pp 4–14