Algebraic aspects of integrability and reversibility in maps

Author:
Jogia, Danesh Michael

Publication Date:
2008

DOI:
https://doi.org/10.26190/unsworks/17873

Algebraic Aspects of Integrability and Reversibility in Maps

Danesh Jogia, 2008

Submitted in accordance with the requirements for the award of
Doctor of Philosophy, UNSW.

School of Mathematics and Statistics,
University of New South Wales

Contents

Acknowledgments

My parents - for instilling in me the self-belief and the means to pursue my dreams.

My supervisor, John Roberts - for taking me on for this even *after* Honours and all the discussion, red marks and help with the academic style.

My co-supervisor, Bruce Henry - who initially spurred my interest in the broad area of dynamics.

Franco Vivaldi - for being a partner in crime and those long holiday discussions.

J.J. Duistermaat and Daniel Chan - for enlightening discussions about algebraic geometry.

Jim - for being a good officemate, when I was actually there.

This work was sponsored by an Australian Postgraduate Award and the School of Mathematics' Mathematics Research Award.

Chapter 1

Introduction

The study of time-discrete maps using algebraic techniques has been a popular approach to the subject since the 1990s (see [60, 30, 61] and references therein as well as section 3.3.1 below). However the work in this thesis came about less as an extension of these algebraic-geometric attacks on integrable maps and more as an extension of the specific test for integrability in two dimensions first proposed by Roberts and Vivaldi in [60]. This paper was the first to detect signatures of properties of maps by considering them over finite phase spaces and as a test it was particularly effective at separating nearly integrable maps from integrable maps. The heart of this test lay in considering the maps concerned over various finite fields. This poses several questions: firstly what does a map look like when considered over a finite field? Since in the continuum integrable maps have phase space portraits that look different to non-integrable maps, can we see integrability by just considering phase space portraits for maps over finite fields? In figure (1.1) we see two columns. On the left are figures obtained from the non-integrable map from [23]

$$(x, y) \to (y, -x + y + \frac{1}{y^2}) \tag{1.1}$$

while on the right we have figures from the integrable QRT map

$$(x, y) \to (\frac{y+2}{x}, \frac{y+2x+2}{xy}). \tag{1.2}$$

From top to bottom we show: three orbits of each map in the real plane in particular the orbits of $(1, 2), (1, 3)$ and $(1, 4)$, the same three orbits when the maps are considered over the finite field $\mathbb{Z}_{11}$ and finally the normalised lengths of the orbit of the point $(1,3)$ when considered over finite fields $\mathbb{Z}_p$ for primes $1000 \leq p \leq 10000$. The kind of figures the test produces show how it splits nearly integrable from integrable as can be seen in chapter two in particular figure (3.1). The other property whose signature is sought by this general method of finite phase space analysis by the authors is time-reversal symmetry in [61]. That this thesis sprang from this programme of signature detection can be clearly seen by the fact that chapter three sprang largely from the test for integrability and chapter five sprang largely from the test for time-reversal symmetry. As iconic flagships for these two themes of the thesis, we introduce here two figures and briefly explain how they relate to integrability and reversibility detection.

Figure (1.2) arises from plotting the cumulative frequency distribution of the orbit lengths from an integrable map called the Screensaver Map (due to how it arises from considering the paths traced about by polygons as they are folded across themselves)

$$x' = \frac{y+\alpha}{x} \quad y' = \frac{y+\alpha x+\alpha}{xy}. \tag{1.3}$$

when considered over a finite phase space (with $\alpha = 2$). The existence of flat regions in the graph from figure (1.2), which indicate regions such that no orbits have lengths within that region, is striking given that lengths for non-integrable maps are distributed differently. The observation that integrable maps had this feature goes back as far as [57], whereas before this the prime concern was the test for integrability. As mentioned above, the reason for originally considering these graphs had to do with a test for integrability detection. In fact, the test itself considers the lengths of orbits while such graphs as figure (1.2) come from considering the cumulative frequency distribution of the orbit lengths. The test has its roots firmly in algebraic geometry particularly with regards to elliptic curves. Chapter one gives the necessary background in algebraic geometry to prove the original results in chapters three and four that use theory from algebraic geometry. It will be explained how the theory behind this figure gives an extension of the test for integrability of [60]. As background for testing for integrability in maps, chapter two discusses previous tests of a similar nature though section 3.3.1 contains work of an original nature. Chapter two also introduces the notion of reversibility to facilitate discussion of the extended test for reversibility (that relates to figure (1.3)) of [61] in chapter five.

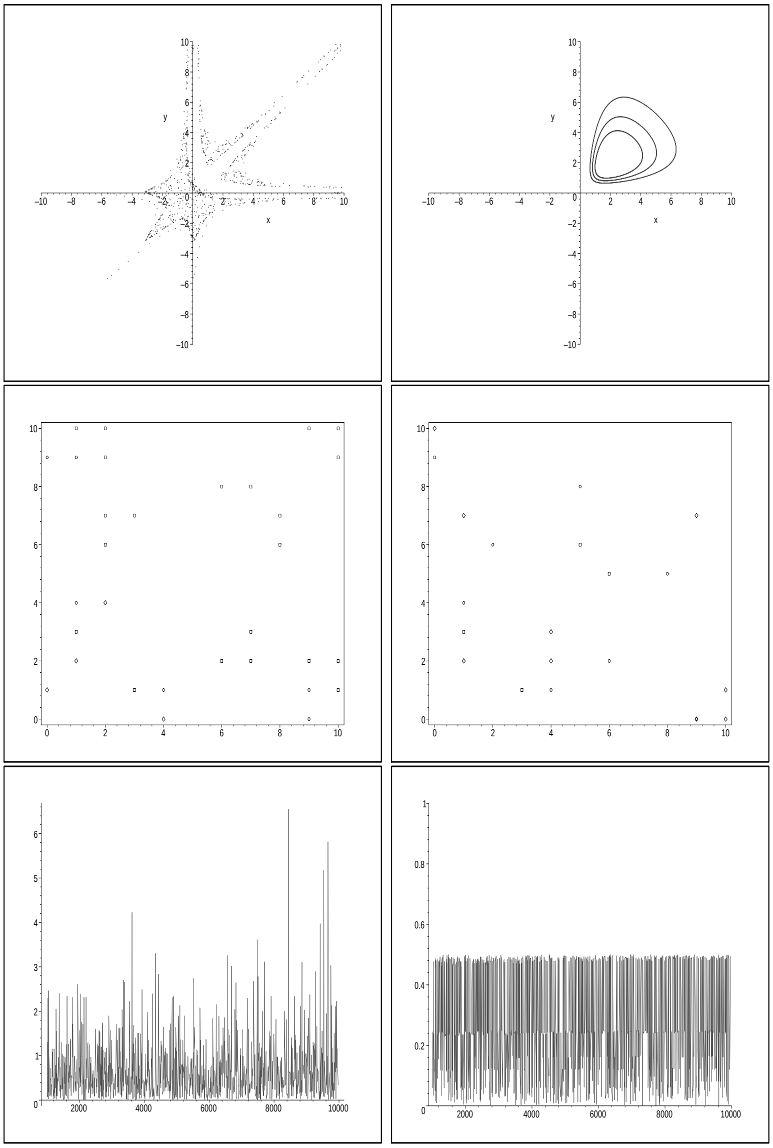

Figure 1.1: Top: The real orbits of the points $(1,2),(1,3)$ and $(1,4)$ under the maps (1.1) (left) and (1.2) (right). Middle: The same orbits when the maps are considered over the finite field $\mathbb{Z}_{11}$. Only the affine points of the orbits have been included. Bottom: The normalised lengths of the orbit of the point $(1,3)$ when considered over finite fields $\mathbb{Z}_p$ for primes $1000 \leq p \leq 10000$; the normalising factor is $p+2\sqrt{p}+1$ and we expect integrable maps to have normalised orbit lengths less than or equal to one. The orbit of the point is considered to close when either it ends in the projectively non-existent point $[0,0,0]$ or when it closes in the usual periodic manner.

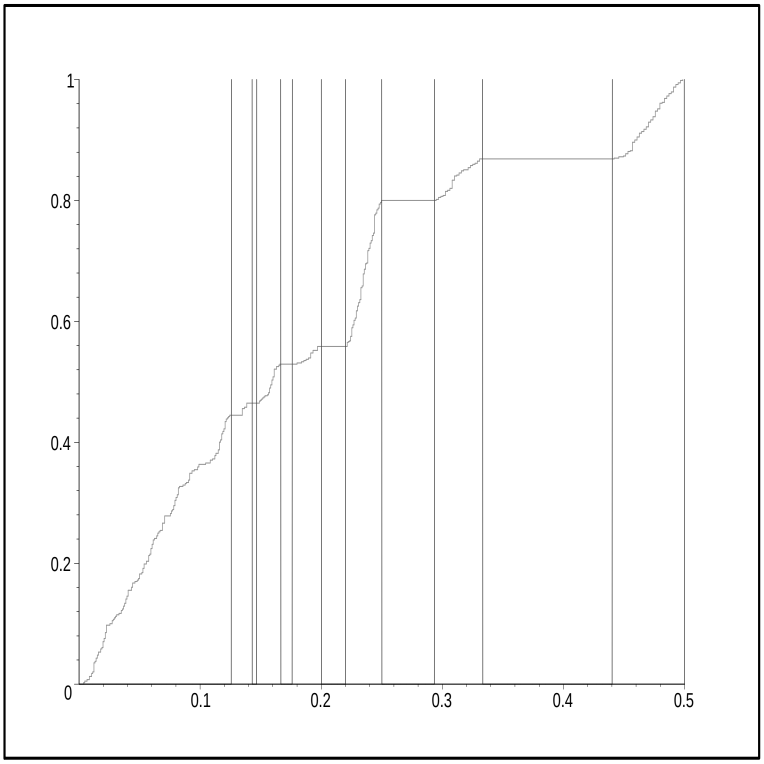

Figure 1.2: Cumulative frequency distribution of normalised orbit lengths gathered from the Screensaver map over the finite phase space $\mathbb{Z}_{997}^2$. This map is integrable, and the characteristic plateaus are explained in chapter three.

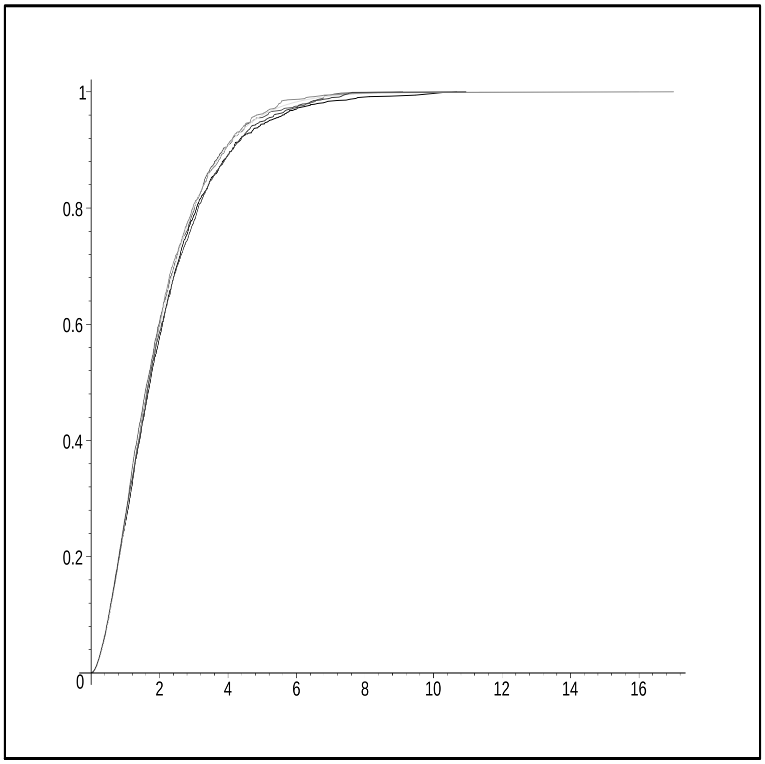

Figure 1.3: Cumulative frequency distributions obtained from orbit lengths for various type II-II reversible maps of three dimensions over the finite phase space $\mathbb{Z}_{103}^3$. The conformity of these distributions that come from differing maps to one and the same universal curve, $y = 1 - e^{-x}(1 + x)$ is what we document in chapter five.

However going beyond the programme of detecting properties of maps but remaining quite close to algebraic geometric roots, chapter three also considers the problem of creating new integrable maps from existing integrable maps.

Figure (1.3) arises from plotting the cumulative frequency distribution of the orbit lengths from several three dimensional reversible maps when considered over a finite phase space. The conformity to a universal distribution is what is of interest here. In chapter five we examine the various types of reversibility possible in three dimensions and unify their orbit length distributions into a conjecture (conjecture (6.19)) that also satisfies the observations made in two dimensions in [61]. While the emphasis here is on three dimensions, we also present some evidence that the conjecture holds in four dimensions. We lastly discuss some potential future directions for this reversibility work to go in.

Part of the work in this thesis has been published. Section 3.3.1 reprises ideas and results from [57], though applied to a different example. Sections 4.1-4.3 are essentially [30], although 4.2.2 is unpublished. Papers related to the concept of mixing in section 5.3 and to the extension to higher dimensions of the reversibility distribution in chapter 6 are presently in preparation.

Chapter 2

Foundations 1: Algebraic Geometry

This chapter goes through some of the rudiments of algebraic geometry, from explaining the difference between the notations and concepts of projective and affine spaces to some of the theory of elliptic curves that will be useful in later chapters. In the first section we will be fairly loose with the notation and definition of curves so as to focus on the differences between projective and affine space. More technicality will be introduced in section 2.3 when we start to define objects such as curves more precisely. Being introductory in nature, the chief sources for this chapter are textbooks, in particular [63] and [22].

2.1 Definition of Projective Spaces

Here we will define a projective space as a set and give some intuitions on how to think about them.

Definition 2.1. *Let K be a field. Then the projective n-space over K is the set $P(K^n) = \{(x_1, x_2, \dots, x_{n+1}) \in \bar{K}^{n+1} - \{\boldsymbol{0}\} : \frac{x_j}{x_i} \in K \text{ for some } x_i \neq 0\}$ modulo the equivalence relation given by identifying parallel vectors i.e. $\boldsymbol{x} \equiv \boldsymbol{y}$ if $\exists\, \lambda \in K$ with $\boldsymbol{x} = \lambda \boldsymbol{y}$. The point $(x_1, x_2, \dots, x_{n+1})$ is said to be written in homogeneous coordinates.*

The K in the notation will often be suppressed since the ground field is obvious from context and as a further matter of notation when a point is in homogeneous coordinates we will use square parentheses and upper case characters so that, e.g., $[X_1, X_2, X_3]$ is a point in two dimensional projective space whereas (x_1, x_2, x_3) would be a point in three dimensional affine space. For a large part of this thesis we shall be concentrating on when $n = 2$ so we shall use this as an example. Putting $n = 2$ into definition (2.1) tells us that we should be thinking of a three-space. Now the equivalence classes under the equivalence relations are lines through the origin (so one succinct way of describing a two dimensional projective space is to say it is "the set of lines through the origin in a three dimensional space") and what is left is to choose a sensible member from each class to work with. This is achieved in example (2.2).

Example 2.2. Let us construct a nice partitioning of and way of thinking about the complex projective plane, $P(\mathbb{C}^2)$. Consider the set of all triples $[X, Y, Z] \in \mathbb{C}^3$. We can cut a plane through the copy of $\mathbb{C}^3$; each line through the origin will certainly only intersect this plane at most once. So we ensure that identified points are never double counted using the plane $Z = 1$. So, supposing $Z \neq 0$, we can divide our points $[X, Y, Z]$ by Z to be left with all points $[\frac{X}{Z}, \frac{Y}{Z}, 1]$. This set will make up the bulk of our new projective space by picking out one candidate from most of the equivalence classes. Since each ordinate comes from a field, we need not write it this way and can just write it as $[X, Y, 1]$. So considering all points of this form covers all homogeneous points such that the third ordinate is non-zero. Now we need to consider all points with $Z = 0$. Here we can divide, where possible, by Y to get points of the form $[X, 1, 0]$. The final point we need to consider is when both Z and Y are 0. This gives all points of the form $[X, 0, 0]$ which under the equivalence relation is the single point [1,0,0]. By considering these three classes of points separately, we see that the projective plane can be written as a union of the affine plane (all the points of the form $[X, Y, 1]$), the affine line (all the points of the form $[X, 1, 0]$) and finally a single affine point. So to create the projective plane from an affine plane it is necessary to add an extra affine line and an extra affine point. These two combined actually form a projective line (if you go through the reasoning throughout this example, one easily sees that points of the form $[X, Y]$ after identifying points from the same equivalence class can be written as a union of an affine line and a single affine point).

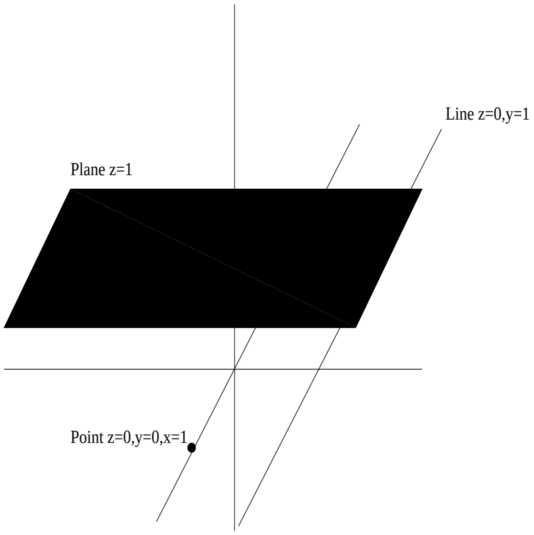

Figure 2.1: Visualisation of the real projective plane.

This extra (projective) line is often called the "line at infinity". Figure (2.1) shows this method of visualising the real projective plane as the complex projective plane is understandably difficult to graphically represent like this.

Now that we have this new definition of projective points, our typical notion of curves may also need to be modified so that they continue to work the way we might hope. We do this by considering curves (although curve will have to wait for a more precise definition) defined by homogeneous polynomials.

Definition 2.3. *A degree d polynomial is homogeneous if in each individual term $X_1^{d_1} X_2^{d_2} \dots X_n^{d_n}$ the sum of degrees $\sum d_i$ is exactly d.*

Proposition 2.4. *In homogeneous coordinates, homogeneous polynomials give rise to well defined sets of zeros.*

Proof. Let $F(X_1, X_2, \dots, X_{n+1}) = 0$ be a degree d homogeneous polynomial equation. Then $F(tX_1, tX_2, \dots, tX_{n+1}) = t^d F(X_1, X_2, \dots, X_{n+1})$ which is 0 if and only if the original equation is satisfied (since $t \neq 0$ else the original point would not exist in projective space regardless). Thus if one member of an equivalence class satisfies the polynomial equation, each other member does also. □

Many of the results used in this thesis assume that we are working in a relevant projective space. However, as the following example will illustrate, it is often not necessary (or desirable, due to the confusion and complexity that introducing another variable will add to any involved algebra) to write everything in homogeneous form.

Example 2.5. Consider the affine curve in $\mathbb{C}^2$ defined by $y^2 - x^3 - ax - b = 0$. We make this polynomial homogeneous by setting $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ and then clearing the denominator to get $Y^2Z - X^3 - aXZ^2 - bZ^3 = 0$. Note that upon substituting $Z = 1$ we are left with the original affine curve with different variable names, showing how the data of an affine curve is contained within the data of its homogenisation. To find all the points at infinity on such a curve we ensure we move away from the affine plane by putting $Z = 0$ which leaves just $X^3 = 0$, regardless of the values of a and b. This tells us that there is a triple point of contact between the curve and the line at infinity at the point $[0, 1, 0]$. This will often turn out to be the case; we can write a curve in projective space as the dehomogenised version in affine space with the proviso that we consider the several other points lying at infinity separately.

A second example that we shall be seeing a lot is the biquadratic.

Example 2.6. Let

$$B(x, y) = \alpha x^2y^2 + \beta x^2y + \delta xy^2 + \gamma x^2 + \kappa y^2 + \epsilon xy + \xi x + \lambda y + \mu. \tag{2.1}$$

To homogenise this we again put $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ and multiply throughout by Z^4 to get

$$\begin{aligned} B(X, Y, Z) =& \alpha X^2Y^2 + \beta X^2YZ + \delta XY^2Z + \gamma X^2Z^2 + \kappa Y^2Z^2 + \\ & \epsilon XYZ^2 + \xi XZ^3 + \lambda YZ^3 + \mu Z^4. \end{aligned} \tag{2.2}$$

To find where such biquadratics intersect the line at infinity we put $Z = 0$ in to get $B(X, Y, 0) = \alpha X^2Y^2$ so there are double points of contact at [1,0,0] and [0,1,0].

2.2 Projective Finite Spaces

A projective finite space arises when the ground field in definition (2.1) is finite. In this thesis we will need to consider the spaces with finite ground fields of the form $\mathbb{Z}_p$ i.e. those whose ground field is prime order. The situations in which it is desirable to work with such objects will often require a conversion from infinite fields (usually $\mathbb{Q}$) to finite fields; this is what we shall be talking about in this section. We will write the finite projective spaces spaces as $P(\mathbb{Z}_p^n)$. Some care must be taken when working with finite projective spaces as homogenising coordinates will not always lead to what is expected, as the following example shows.

Example 2.7. Consider the affine point $(\frac{1}{7}, 1) \in \mathbb{A}^2_{\mathbb{Q}}$ that we wish to represent in the finite projective space $P(\mathbb{Z}_7^2)$. Then first we homogenise the original point to get $[1, 7, 7]$ which reduces to $[1, 0, 0]$ modulo 7. This example shows that when reducing to a finite field, affine rational points can become points at infinity. This is of course because the point we were considering had a "zero" in the denominator, which we might expect to be sent to the line at infinity.

The process we followed in example (2.7) is called "normalisation" of a point written in homogeneous coordinates. Normalised coordinates are easier to work with theoretically (because they are less ambiguous) but also are very useful computationally because they allow for easy comparison between points.

Definition 2.8. *A projective point $[X_1, X_2, \ldots, X_{n+1}]$ with coordinates in a field of quotients is said to be normalised if the largest denominator across all of the X_i's is 1 and* $\gcd(X_1, X_2, \ldots, X_{n+1}) = 1$.

It is easy to see that normalised coordinates are unique up to a sign change. This is because one algorithmically normalises points by multiplying each ordinate by a number that will clear all denominators, then dividing each ordinate by their greatest common divisor. Having normalised a point, it is then easy to reduce the point's coordinates modulo p. A similar normalisation procedure applies to homogeneous polynomial curves defined by

$$F(X_1, X_2, \ldots, X_{n+1}) = 0.$$

Instead of multiplying and dividing ordinates, one just does the same operations to the coefficients of F.

2.3 Elliptic Curves

This section contains much of the basics that any undergraduate course on elliptic curves would cover. Some more mathematically complicated matters are left until the later section 2.4. The meaning of elliptic curves, their group structure (both in general and in particular over the rational numbers), the number of points in finite fields contained on them and a way of parameterising them will all be covered as these are all required background knowledge to begin exploiting elliptic curves in the context of discrete dynamical systems.

2.3.1 Defining Elliptic Curves

Elliptic curves are a particular class of algebraic variety, thus a definition of varieties will first be necessary. Before we can begin such a definition we should give the meaning of an occasional assumption that will be made on the ground field in which we are working.

Definition 2.9. *A field K is said to be perfect if every algebraic extension of it is separable (recall a particular extension of a field is separable if the minimal polynomial of every element in the extension has no multiple roots).*

Perfect fields include all fields in characteristic 0 and all finite fields but as an example of an imperfect field consider:

Example 2.10. Consider the field $\mathbb{F}_2(t)$ i.e. the smallest field extension of $\mathbb{F}_2$ containing the symbol t with no relations. Now this is an infinite field in characteristic 2 and will be imperfect. Consider the (finite) extension where we adjoin $\sqrt{t}$. The minimal polynomial for the new element is $x^2 - t$ which factorises as $(x - \sqrt{t})(x + \sqrt{t})$ which, when coefficients are being taken from $\mathbb{F}_2$, is exactly the same as $(x-\sqrt{t})(x-\sqrt{t})$ and this multiple root in the minimal polynomial is precisely what it means for a finite field extension to not be separable.

The assumption that K is perfect is often required in many standard references and as such it pertains to the background of this thesis, most notably for the material in section 2.4. However we are able to use slightly weaker results that do not require perfectness. That K has characteristic not 2 or 3 will always be assumed, though with some care most of the results in this thesis still hold for fields of those characteristics.

Now to define the basic objects of study

Definition 2.11. *An algebraic set defined over a field K is the set of simultaneous zeros to a set of polynomials in $K[x_1, x_2, \ldots, x_n]$ i.e. $V = \{x \in K^n : f_1(x) = f_2(x) = \ldots = f_m(x) = 0\}$.*

Definition 2.12. *An ideal I of a ring R is a set such that for all $x \in I$ and $r \in R$, $rx \in I$.*

The type of ideal we shall be dealing with most commonly is that generated by a polynomial.

Example 2.13. Let R be a ring of polynomials, say $K[x]$. Then for any polynomial $p(x) \in R$, the set $I = \{p(x)g(x) : g(x) \in R\ \}$ forms an ideal.

To each algebraic set V we can associate an ideal of $K[x_1, x_2, ..., x_n]$ by

$$I(V) = \{f \in K[x_1, x_2, \ldots, x_n] : f(x) = 0\ \forall\, x \in V\}.$$

Similarly we can associate algebraic sets to any given ideal of $K[x_1, x_2, \ldots, x_n]$ by considering the algebraic set defined by all the polynomials inside the ideal.

Definition 2.14. *A variety (or irreducible algebraic set) is an algebraic subset V whose ideal $I(V)$ is a prime ideal (recall an ideal I is prime if $ab \in I$ implies that either $a \in I$ or $b \in I$) in $\bar{K}[x_1, x_2, \ldots, x_n]$.*

Varieties are the central objects of study in algebraic geometry and throughout this thesis all varieties are assumed to be irreducible (indeed, varieties that are not irreducible are more commonly just called algebraic sets) unless specifically noted. Reducibility of a variety certainly does occur when one of the defining polynomials is reducible over K.

Example 2.15. Some irreducible varieties defined over $\mathbb{Q}$ are:

$$\begin{aligned} V_1 &= \{(x, y, z) : x^2 + y^2 - 1 = 0\} \\ V_2 &= \{(x, y, z) : x^2 + y^2 + z^2 - 1 = 0\} \\ V_3 &= \{(x, y) : ax + by + c = 0, a, b, c \in \mathbb{Q}\} \end{aligned}$$

while a variety that is not irreducible is:

$$\begin{aligned} V_4 &= \{(x, y) : y^2 - x^2 = 0 = (y - x)(y + x)\} \\ &= \{(x, y) : y - x = 0\} \cup \{(x, y) : x + y = 0\}. \end{aligned}$$

For V_1, V_2, V_3, the defining polynomials are irreducible over the algebraic closure of the field of definition, so the ideals generated by the polynomials are prime. For V_4 however, we see that $(y-x)(y+x) \in I(V_4)$ but $y - x \notin I(V_4)$ and $y + x \notin I(V_4)$ so that $I(V_4)$ is not prime.

Maps between varieties become important later, so here we shall give the definition of the natural kind of map between varieties, and an example thereof. We will be using projective varieties because this allows us to consider maps that would otherwise cause troubles due to vanishing denominators.

Definition 2.16. *Let $V \in P(K^n)$ be a variety. A rational function $f : V \to K$ is regular at a point P if $f(P)$ is defined.*

The rational functions that are defined at P make up the local ring of V at P. The localisation of a ring (in this case the ring of polynomials $\bar{K}[V]$) at a particular maximal ideal (in this case $M_P = \{f \in \bar{K}[V] : f(P) = 0\}$) is a common construct in commutative algebra.

Definition 2.17. *Let V_1 and V_2 be two varieties in $P(K^n)$. A morphism $f : V_1 \to V_2$ is an $(n+1)$-tuple of rational functions $f = (f_1, \ldots, f_{n+1})$ such that for each $P \in V_1$ there is a $g \in K(V_1)$ (see proposition (2.21)) such that:*

- *gf_i is regular at P for each i and*
- *$gf_i(P) \neq 0$ for at least one i.*

The first requirement is that each ordinate is regular; we are allowed to introduce the seemingly extraneous g because in projective space points along the same line through the origin are identified anyway. The second requirement is there because the 0-vector does not exist in projective space.

Definition 2.18. *Two varieties V_1 and V_2 are isomorphic if there exist morphisms $\phi : V_1 \to V_2$ and $\psi : V_2 \to V_1$ such that $\phi \circ \psi$ and $\psi \circ \phi$ are the identity maps on V_2 and V_1 respectively.*

To ground these definitions somewhat, we shall give a few examples of morphisms and isomorphisms and how to identify them.

Example 2.19. Let V be the projective variety defined by $X^2 - Y^2 = Z^2$. We aim to show that this variety is in fact isomorphic to the projective line $\mathbb{P}^1$. To do this, we have to find two morphisms $\phi : V \to \mathbb{P}^1$ and $\psi : \mathbb{P}^1 \to V$ such that $\phi \circ \psi = id$ on $\mathbb{P}^1$ and that $\psi \circ \phi = id$ on V. This will also require checking that both ϕ and ψ are regular on their domains. We begin constructing these two functions by finding a function that would map the (complex) line onto the *affine* variety $V_A : x^2 - y^2 = 1$ and then homogenise this transformation and find its inverse. It is fairly obvious that the transformation

$$\psi_A : t \to (\frac{1+t^2}{1-t^2}, \frac{2t}{1-t^2})$$

suffices to map a single complex parameter onto the entirety of V. Now to invert this we solve $x = \frac{1+t^2}{1-t^2}$ for t^2 to get $t^2 = \frac{x-1}{x+1}$ and then substitute this into $y = \frac{2t}{1-t^2}$. This then allows us to solve for t resulting in $t = \frac{y}{1+x}$. Define

$$\phi_A : (x, y) \to \frac{y}{1+x}.$$

First we check that ψ_A and ϕ_A are indeed inverses.

$$\begin{aligned}
\psi_A(\phi_A(x,y)) &= \psi_A(\frac{y}{1+x}) \\
&= (\frac{1 + \frac{y^2}{(1+x)^2}}{1 - \frac{y^2}{(1+x)^2}}, \frac{\frac{2y}{1+x}}{1 - \frac{y^2}{(1+x)^2}}) \\
&= (\frac{\frac{(1+x)^2+y^2}{(1+x)^2}}{\frac{(1+x)^2-y^2}{(1+x)^2}}, \frac{\frac{2y(1+x)}{(1+x)^2}}{\frac{(1+x)^2-y^2}{(1+x)^2}}) \\
&= (\frac{1+x^2+y^2+2x}{1+2x+x^2-y^2}, \frac{2y+2xy}{1+2x+x^2-y^2}) \\
&= (\frac{2x^2+2x}{2+2x}, \frac{y(2+2x)}{2+2x}) \\
&= (x, y),
\end{aligned}$$

$$\begin{aligned}\phi_A(\psi_A(t)) &= \phi_A(\frac{1+t^2}{1-t^2}, \frac{2t}{1-t^2}) \\ &= (\frac{2t}{1-t^2})\frac{1}{1+\frac{1+t^2}{1-t^2}} \\ &= (\frac{2t}{1-t^2})\frac{1}{\frac{1-t^2}{1-t^2}+\frac{1+t^2}{1-t^2}} \\ &= (\frac{2t}{1-t^2})\frac{1}{\frac{2}{1-t^2}} \\ &= (\frac{2t}{1-t^2})\frac{1-t^2}{2} \\ &= t.\end{aligned}$$

Having checked that these functions are inverse to each other, we now homogenise them and check that they are regular across their domain. First homogenise ϕ_A by substituting $x = \frac{X}{Z}, y = \frac{Y}{Z}$ and moving all denominators into a second coordinate (see section 3.1 for more information on homogenising maps) to get

$$\begin{aligned}\phi : (X, Y, Z) &\to \frac{\frac{Y}{Z}}{1+\frac{X}{Z}} \\ &= \frac{\frac{Y}{Z}}{\frac{Z+X}{Z}} \\ &= \frac{Y}{Z+X} \\ &= (Y, Z+X).\end{aligned}$$

Now homogenise ψ_A by substituting $t = \frac{T}{S}$ to get

$$\begin{aligned}\psi : (T, S) &\to (\frac{1+\frac{T^2}{S^2}}{1-\frac{T^2}{S^2}}, \frac{2\frac{T}{S}}{1-\frac{T^2}{S^2}}) \\ &= (\frac{S^2+T^2}{S^2-T^2}, \frac{2ST}{S^2-T^2}) \\ &= (S^2+T^2, 2ST, S^2-T^2).\end{aligned}$$

Now ψ is obviously a morphism (we can take $g = 1$ in definition (2.17)), and the only point where ϕ might not be regular is at [-1,0,1]. However in this case we can multiply ϕ by $(Z-X, Z-X)$ to get $\phi(X, Y, Z) = (Y(Z-X), Z^2-X^2) = (Y(Z-X), -Y^2) = (Z-X, -Y)$ which evaluates to [1,0] at the potentially problematic point [-1,0,1]. So at this single point we must take $g = Z - X$ in definition (2.17) but elsewhere we can take $g = 1$ for ϕ to be a morphism. Thus both ϕ and ψ are morphisms and both are the inverse of one another hence $V \cong \mathbb{P}^1$.

While we will not be using these morphisms until later when we need consider isomorphisms, this was a natural place to introduce them. Now we turn back to defining elliptic curves.

Definition 2.20. *Let V be an irreducible variety defined over a field K. The coordinate ring of V is the quotient ring $K[V] = \frac{K[x_1,x_2,\ldots,x_n]}{I(V)}$.*

Proposition 2.21. *If V is an irreducible variety, the set $K(V)$ consisting of all quotients from the ring $K[V]$ is a field and is called the function field of V.*

Proof. The polynomial ring $K[x_1, x_2, \ldots, x_n]$ is certainly an integral domain and so we show that factoring out by the prime ideal $I(V)$ does not introduce any zero divisors. Suppose $fg = 0 \mod I(V)$. Then $fg \in I(V)$ whence, since $I(V)$ is prime, either $f \in I(V)$ or $g \in I(V)$ i.e. $f = 0 \mod I(V)$ or $g = 0 \mod I(V)$ respectively. So the coordinate ring $K[V]$ is an integral domain. Now consider the set $\{\frac{f}{g} : f, g \in K[V], g \neq 0\}$. All field operations work correctly by definition and since there are no zero divisors ($K[V]$ is an integral domain), the need for a "$\frac{1}{0}$" doesn't arise. □

The function field of a curve is an important structure in and of itself, but the following example will illustrate a function field that will be of great use in this thesis.

Example 2.22. Consider the complex affine line, $\mathbb{C}$. This is an algebraic set, being all the points that satisfy the polynomial equation $f(x) = 0$ with f being the zero polynomial. The ideal generated by this algebraic set is simply the zero ideal, which is certainly prime, so $\mathbb{C}$ is an irreducible variety. Factoring the ring of one variable complex coefficient polynomials by this ideal does not change it, so $\mathbb{C}[\mathbb{C}]$ is isomorphic to $\mathbb{C}[t]$ where t is a transcendental. Another way of saying this is that by quotienting out the zero ideal we are not identifying any other elements with zero, and so it remains unchanged. Regardless, this means that the function field of the complex line is just the field consisting of all rational functions with complex coefficients in one variable. This function field will arise naturally in the context of dynamics as the field to consider certain maps (and the curves that they preserve) to be defined over.

The function field of a variety can be used to define a dimension of that variety.

Definition 2.23. *The K-dimension of an irreducible variety V, written dim(V), is the degree of the function field $\bar{K}(V)$ over the closure of the ground field, $\bar{K}$.*

Finally with a notion of dimension, we can define what a curve is.

Definition 2.24. *A curve is a variety of dimension 1.*

Elliptic curves are a specific class of curves, in one sense the simplest curves after conics. The rigourous definition requires the full definition and understanding of the notion of genus of a curve. To fully engage this idea would be to move too far astray from the crux of the thesis. The true definition of the genus of a curve is related to the dimension of a certain vector space associated with that curve. For our purposes we will go into more detail about the geometry of curves and use this to give a simpler account of genus.

Definition 2.25. *Let $C : F(X_1, X_2, \ldots, X_{n+1}) = 0$ be a curve in n-dimensional projective space. Then a point $P \in C$ is singular if*

$$(\frac{\partial F}{\partial X_1}(P), \frac{\partial F}{\partial X_2}(P), \ldots, \frac{\partial F}{\partial X_{n+1}}(P)) = (0, 0, \ldots, 0) \tag{2.3}$$

i.e. if the tangent at P is not defined.

Singular points can be further classified depending on the behaviour of the vector of double derivatives at that point and so forth. Two types of singular points which will occur frequently when studying low degree curves are ordinary double points and cusps.

Example 2.26. Consider the homogeneous curve

$$C_1 : Y^2Z - X^3 = 0.$$

Computing the vector of partial derivatives gives $(-3X^2, 2YZ, Y^2)$ so [0,0,1] is a singular point. Furthermore it is the only singular point. The behaviour around this singular point is that of two branches meeting in such a way that the tangent lines coincide and it is known as a cusp. However, the curve

$$C_2 : Y^2Z - X^3 - X^2Z = 0,$$

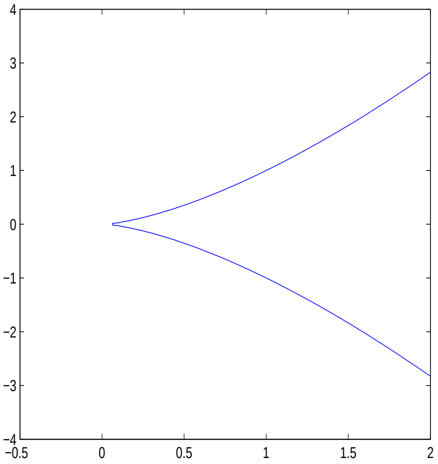

Figure 2.2: A Weierstrass cubic with a cusp

which has partial derivatives $(-3X^2-2XZ, 2YZ, Y^2-X^2)$ also has a unique singular point at [0,0,1] but it is of a different type, called an ordinary double point. In this case there are two tangent lines that intersect transversally. Figures (2.2) and (2.3) show, respectively, the scenario of C_1 and C_2 around their singular points.

An equation that gives the genus of an irreducible curve with only ordinary double points and cusps in terms of the degree of the defining polynomial and its singular points is

$$p = \frac{1}{2}(n-1)(n-2) - (\delta + \kappa) \tag{2.4}$$

where n is the degree of the defining polynomial, δ is the number of ordinary double points and κ is the number of cusps.

Definition 2.27. *An elliptic curve is an algebraic curve of genus 1, with a specified point $\mathcal{O}$ on that curve.*

An elliptic curve E is said to be defined over a field K if the equation for E has coefficients in the field K and E contains a non-singular point with coordinates in K. We shall now give an example, in detail, of an elliptic curve that will recur throughout the thesis - a genus one biquadratic.

Example 2.28. Consider the curve

$$B(x,y) = x^2y^2 - t^2(x^2+y^2) - 2xy + 1, \tag{2.5}$$

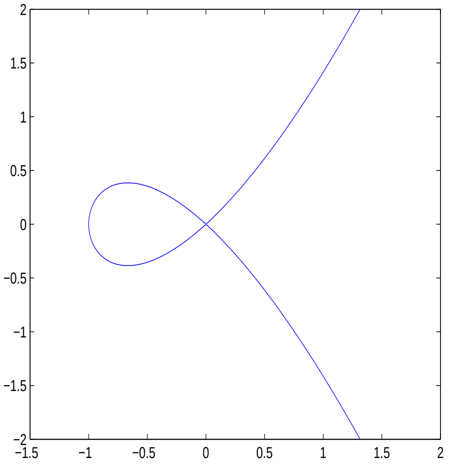

Figure 2.3: A Weierstrass cubic with a double point

which is an example of the form from equation (2.1). We first homogenise this equation to get

$$B(X,Y,Z) = X^2Y^2 - t^2Z^2(X^2+Y^2) - 2XYZ^2 + Z^4 \tag{2.6}$$

and find the singular points on a typical B to check that it is indeed elliptic. To this end, we find the partial derivatives.

$$\begin{aligned}
\frac{\partial B}{\partial X} &= 2XY^2 - 2XZ^2t^2 - 2YZ^2\\
\frac{\partial B}{\partial Y} &= 2YX^2 - 2YZ^2t^2 - 2XZ^2\\
\frac{\partial B}{\partial Z} &= -2Zt^2(X^2+Y^2) - 4ZXY + 4Z^3
\end{aligned}$$

First we check for singular points at infinity by putting $Z = 0$. This yields the two non-trivial equations $2XY^2 = 0, 2YX^2 = 0$ so there are two singular points at infinity, [1,0,0] and [0,1,0]. Incidentally these are the only points at infinity on B. Next we check for any affine singular points. In general we expect there to be none of these since a third singular point would reduce the genus of B, according to equation (2.4), to 0. The three equations we arrive at with $Z = 1$ are

$$\begin{aligned}
2XY^2 - 2Xt^2 - 2Y &= 0\\
2YX^2 - 2Yt^2 - 2X &= 0\\
-2t^2(X^2+Y^2) - 4XY + 4 &= 0.
\end{aligned}$$

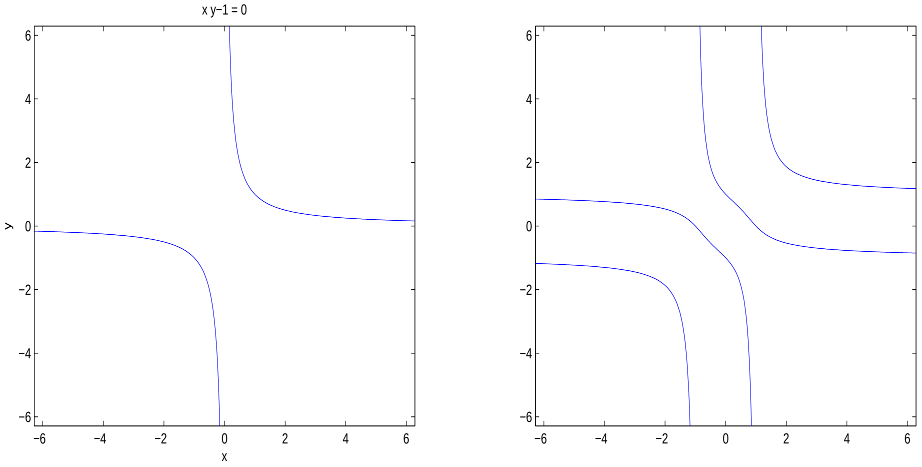

Figure 2.4: The reducible level set $t = 0$ and the level set $t = 1$ of equation (2.5).

Being over-specified as three equations in two variables, they generally have no solution but for certain values of t they do. From here we make the note that the partial derivative with respect to X and with respect to Y are the same polynomial if we swap X and Y while the partial derivative with respect to Z is symmetric in X and Y. This tells us that if we find some singular point $[X_s, Y_s, 1]$ then $[Y_s, X_s, 1]$ must also be a singular point. However, from the earlier equation (2.4) we can see that a curve of degree four can have at most 3 singular points. So we check for affine singular points when $X = Y$. Searching for such singular points results in somewhat fictitious singular points. Consider $t = 0$. Then the equations tell us that there are two singular points at [1,1,1] and [-1,-1,1]. The contradictory existence of two singular points for $t = 0$ is explained by the the fact that B factorises into $(xy - 1)^2$. Similarly for $t^2 = 2$ and $t^2 = -2$ two "singular points" show up through the equations; but they are both cases where B factorises and these singular points are really keeping track of where these two distinct factors intersect. Figures (2.4) and (2.5) each show small parts of the real, affine curves defined by putting a particular value of t into the equation for B. Figure (2.6) overlays these four figures; note that they do not intersect (except for when $t^2 = 2$, which is explained above). This is expected since t^2 can be solved for uniquely in the equation for B.

The definition of elliptic curves given in definition (2.27) requires that a point is specified on the curve in question. In the next section we shall consider the

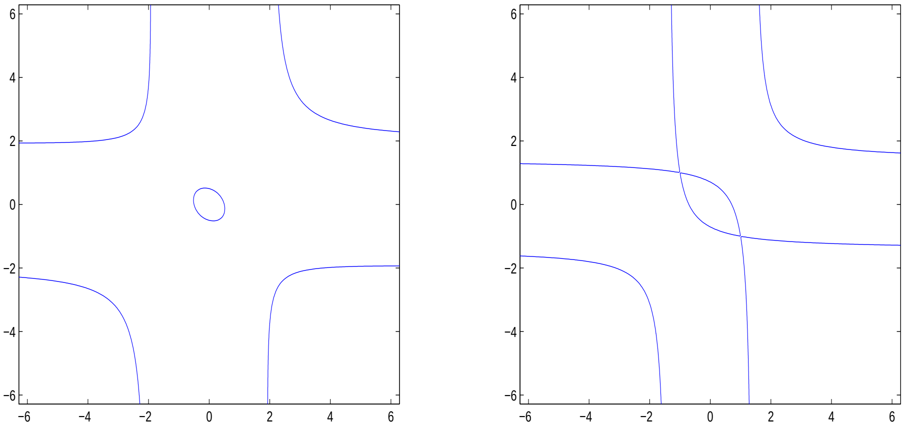

Figure 2.5: The level set $t = 2$ and the reducible level set $t = \sqrt{2}$ of equation (2.5).

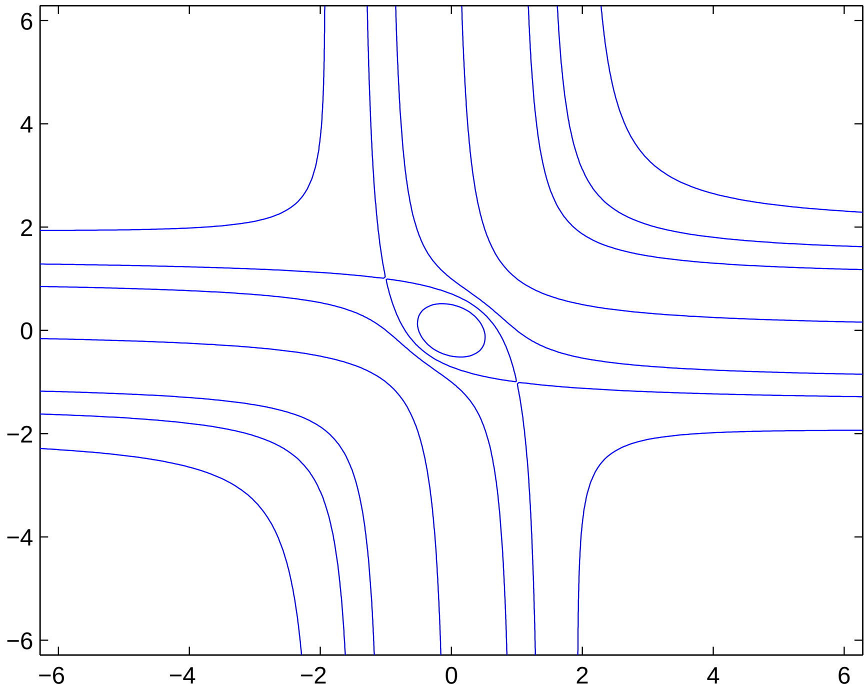

Figure 2.6: The four level sets from figures (2.4) and (2.5) overlaid.

difference between two elliptic curves when the curve's equation remains the same, but the specified point $\mathcal{O}$ is changed.

2.3.2 Normal Forms for Elliptic Curves

One of the most important tools for studying elliptic curves is knowing that any genus one curve containing a point with coordinates in a field K can have its equation brought through birational transformations defined over K to a simple form. The theory that says this is the case is a little far afield to prove in this thesis but a compact treatment of the material is given in proposition 3.1 (and its preceding material) of [63]. We reproduce the relevant part of proposition 3.1 of [63] here to make this notion precise.

Proposition 2.29. *Let E be an elliptic curve defined over K, with a point $\mathcal{O} \in E$. There exist functions $x, y \in K(E)$ such that the map*

$$\phi : E \to P(K^2)$$

$$\phi = [x, y, 1]$$

gives an isomorphism of E/K (that is, the curve E defined over K) onto a curve given by a Weierstrass equation

$$C : Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6 \tag{2.7}$$

with coefficients $a_1, \ldots, a_6 \in K$ and such that $\phi(\mathcal{O})$ =[0,1,0] . The functions x and y are called Weierstrass coordinate functions on E.

The heart of the proof lies in showing that the functions $1, x, x^2, x^3, y, xy, y^2$ cannot be linearly independent, and hence a relationship of the type that defines the curve's equation must exist. The definition of isomorphism here is that used in definition (2.18)

Now we can make the substitution $X \to x, Y \to \frac{1}{2}(y - a_1x - a_3)$ [1] to reduce the equation (2.7) to

$$C' : y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6$$

[1] A note on notation. The transformation denoted by $z_i \to f(z_1, z_2, \ldots)$ means to replace each instance of z_i with the expression f. It is a convenient way of writing coordinate changes without having to change the variable names each time

where $b_2 = a_1^2 + 4a_2$, $b_4 = 2a_4 + a_1a_3$ and $b_6 = a_3^2 + 4a_6$. A second substitution of $x \to \frac{x-3b_2}{36}, y \to \frac{y}{216}$ eliminates the x^2 term so that, with the exception of fields with characteristic 2 or 3, we may assume that an elliptic curve can be written in the form

$$C'' : y^2 = x^3 + Ax + B. \tag{2.8}$$

This is the form that we will usually refer to as Weierstrass form. We will also say the curve has Weierstrass equation $y^2 = x^3 + Ax + B$ in this case.

Finding the Weierstrass equation of a curve can be done by following the process given in the proof of proposition (2.29) in [63]. As it transpires, this is the method followed by the algorithm implemented in the MAPLE computing package. The exact algorithm comes care of [66], which creates the functions x and y in the prescribed way (for example just above we saw that they were really a culmination of substitutions) then finds the relation that exists due to linear dependence.

We now turn to examining some properties of elliptic curves in Weierstrass form. The first question regards the uniqueness of the Weierstrass equation for a particular elliptic curve. This question is in fact answered by the second part of proposition 3.1 of [63], the first part of which was reproduced in proposition (2.29). We reproduce this second part now.

Proposition 2.30. *Any two Weierstrass equations for E as in proposition (2.29) (note that this means the point $\mathcal{O}$ must be mapped to $[0, 1, 0]$ in each case) are related by a linear change of variables of the form*

$$X \to u^2X + r$$

$$Y \to u^3Y + su^2X + t$$

with $u, r, s, t \in K$ and $u \neq 0$.

The proof of this proposition uses dimension/basis arguments on any two pairs of Weierstrass coordinate functions (x, y) and (x', y') which give different Weierstrass equations for a single curve E to show that such relations between them must exist. Note that this proposition refers to the “extended” Weierstrass form as given in proposition (2.29), not the shorter form from equation (2.8) which we typically use. To illustrate the kind of calculations that occur when working with differing short

Weierstrass equations, we look more in depth into how unique short Weierstrass forms can be, using proposition (2.30) as a starting point.

Example 2.31. Let $C : y^2 = x^3 + Ax + B$ define an elliptic curve over some field K. As this is just a special case of the long Weierstrass form, we know that if C' is any other Weierstrass equation for C then they must be related by a change of variables $x = u^2x' + r, y = u^3y' + su^2x' + t$. Putting such a change of variables into the equation for C gives

$$\begin{aligned} C' : u^6y'^2 + 2u^5sx'y' + 2u^3ty' =& u^6x'^3 - (s^2u^4 - 3u^4r)x'^2 - \\ & (2u^2st - 3u^2r^2 - Au^2)x' - \\ & (t^2 - r^3 - Ar - B). \end{aligned}$$

For this to be in (short) Weierstrass form, we need $s = t = r = 0$ (and then to divide throughout by u^6), so the only transformations that map from Weierstrass form to Weierstrass form are those of the form $x \to u^{-2}x, y \to u^{-3}y$.

Furthermore, we can see that any other Weierstrass form for $C : y^2 = x^3 + Ax + B$ is just $C' : y^2 = x^3 + Au^{-4}x + Bu^{-6}$. By similar arguments we can check when such a transformation would in fact map C to itself, the calculations for which we shall perform later.

A second and most obvious fact about curves in Weierstrass form is the simplicity of their intersection with the line at infinity. Let us homogenise a Weierstrass equation and check this intersection with the line at infinity.

Example 2.32. Let C have equation $y^2 = x^3 + Ax + B$. Then homogenising this gives

$$Y^2Z = X^3 + AXZ^2 + BZ^3. \tag{2.9}$$

Substituting $Z = 0$ leaves $0 = X^3$ so we see that there is a triple point of contact at the point $[0, 1, 0]$. The fact that this is a triple point of contact common to all curves in Weierstrass form, even long Weierstrass form, proves to be important in the next section.

The last property of curves in Weierstrass form we shall comment on is precisely when they are NOT in fact elliptic curves. That is to say, what characteristic should we look for in a Weierstrass equation to tell us if that curve is elliptic or not? The fact that the Weierstrass equation is degree three coupled with the formula (2.4) tells us that we should be looking for when the equation has a singular point. The following example looks into this situation.

Example 2.33. Let C/K be a curve with equation $F : y^2 - x^3 - Ax - B = 0$. Computing the three partial derivatives of the homogeneous version of this equation (2.9) gives

$$\begin{aligned}\frac{\partial F}{\partial X} &= -3X^2 - AZ^2\\ \frac{\partial F}{\partial Y} &= 2YZ\\ \frac{\partial F}{\partial Z} &= Y^2 - 2AXZ - 3BZ^2\end{aligned}$$

To find any singular points we need to find when these three equations simultaneously vanish. This tells us, from the second, that either $Y = 0$ or $Z = 0$. Supposing $Z = 0$, the other two equations then tell us that $X = Y = 0$ which is not a point in projective space so instead we try $Y = 0$ and $Z = 1$. This simplifies the equations to finding when

$$\begin{aligned}\frac{\partial F}{\partial X} &= -3X^2 - A\\ \frac{\partial F}{\partial Y} &= 0\\ \frac{\partial F}{\partial Z} &= -2AX - 3B\end{aligned} \tag{2.10}$$

simultaneously vanish. This gives $X^2 = \frac{A}{3} = \frac{9B^2}{4A^2}$. Note that this is a condition solely on A and B. Consider when the polynomial $f(X) = X^3 + AX + B$ has a root of multiplicity greater than one. This occurs if and only if there exists an α such that

$$\begin{aligned}\alpha^3 + A\alpha + B &= 0\\ 3\alpha^2 + A &= 0\end{aligned}$$

are both satisfied (i.e. there is some value of X that makes both the equation and its derivative vanish). Multiplying the first condition by 3 and the second by α and subtracting gives a third condition

$$2A\alpha + 3B = 0$$

Now this condition and one of the the original conditions $3\alpha^2+A=0$ are exactly the two previous conditions (2.10) needed for a singular point to exist. Thus a singular point exists on the curve $y^2=f(x)=x^3+Ax+B$ precisely when $f(x)$ has a root of multiplicity two or more. Graphs of such curves were shown earlier in example (2.26).

As stated in definition (2.27), an elliptic curve is a curve together with a specified point on it. We will now look at taking an equation for a curve, finding its Weierstrass form (and appropriate conversion functions) using MAPLE, then finding a potentially different Weierstrass form by using a different specified point.

Example 2.34. Let $B(x,y)=x^2y^2-t^2(x^2+y^2)-2xy+1$. For most values of t, $B(x,y)=0$ is a curve of genus one. It is not hard to check that $(0,\frac{1}{t})\in B$. We can use the MAPLE function `Weierstrassform` to find a Weierstrass form for B, inputting the point $(0,\frac{1}{t})$. The function will use the input point and choose Weierstrass coordinate functions such that the image of the input point is [0,1,0]. Performing this calculation, the result is:

$$\begin{aligned}
W(u,v) &= u^3+(-\frac{1}{3}t^8-4t^4)u-\frac{8}{3}t^8+\frac{2}{27}t^{12}+v^2\\
s_1: u &= \frac{t(t^3x^2-6t-6x-6t^2y+6x^2y)}{3x^2}\\
v &= \frac{2t^2(t^3x^2-2t-2x-2t^2y+2x^2y)}{x^3}\\
x &= \frac{-18tv}{t^8-36t^4-6ut^4+9u^2}\\
y &= \frac{-5t^8+12ut^4+36t^4+9u^2+18v}{t(9u^2-6t^4u+36u+t^8-12t^4)}.
\end{aligned}$$

Notice that W is written not in the form $y^2=x^3+Ax+B$ but rather the form y^2+x^3+Ax+B. The two are birationally equivalent via the simple transformation $x\leftrightarrow -x$. Unraveling this data, it is saying that should we take a point $(p_x,p_y)\in B$ then by substituting $x=p_x, y=p_y$ into the equations for u and v we would find a point that satisfies the equation of W. Furthermore, should we put the resulting (p_u,p_v) pair into the equation for x and y, we would end up with the same (p_x,p_y) with which we started. Notice that putting the input point $(0,\frac{1}{t})$ into the formula for u and v forces us to the line at infinity, where there is only one point on any curve in the form of W. Thus, as stated, the image of $(0,\frac{1}{t})$ is indeed [0,1,0]. Another way to show this is to homogenise each of the equations and transformations listed above. Doing this gives the following (see definition (3.2) for how one homogenises transformations):

$$\begin{aligned}
W(U,V,T) = &\quad U^3 + (-\tfrac{1}{3}t^8 - 4t^4)UT^2 - \tfrac{8}{3}t^8T^3 + \tfrac{2}{27}t^{12}T^3 + V^2T \\
S_1 : U = &\quad tX(t^3X^2Z - 6tZ^3 - 6XZ^2 - 6t^2YZ^2 + 6X^2Y) \\
V = &\quad 6t^3Z(t^3X^2Z - 2tZ^3 - 2XZ^2 - 2t^2YZ^2 + 2X^2Y) \\
T = &\quad 3ZX^3 \\
X = &\ -18VtT(t(9U^2T^2 - 6tU^4 + 36UT^3 + t^8T^4 - 12t^4T^4)) \\
Y = &\quad -(5t^8T^2 - 12Ut^4T - 36t^4T^2 - 9U^2 - 18VT)T^2 \\
&\quad (t^8T^2 - 36t^4T^2 - 6Ut^4T + 9U^2) \\
Z = &\quad (t^8T^2 - 36t^4T^2 - 6Ut^4T + 9U^2)t \\
&\quad (9U^2T^2 - 6tU^4 + 36UT^3 + t^8T^4 - 12t^4T^4).
\end{aligned}$$

Now with these equations it is quite obvious that putting $X = 0, Y = 1, Z = t$ into $[U,V,T]$ gives $[0,1,0]$. Now let us find a second point on B. Since B is symmetric under $x \leftrightarrow y$, all we need do is switch the coordinates of our point and take $\mathcal{O}'$ $= (\frac{1}{t}, 0)$. The data for the Weierstrass form with this chosen point is:

$$\begin{aligned}
W(u,v) = &\quad u^3 + (-\tfrac{1}{3}t^8 - 4t^4)u - \tfrac{8}{3}t^8 + \tfrac{2}{27}t^{12} + v^2 \\
s_2 : u = &\quad \tfrac{t(t^5x^2+4t^4x-5t^3-6tx^2-6x-6t^2y+6x^2y)}{3(t^2x^2-2tx+1)} \\
v = &\ 2t^2\tfrac{(2t^4x^2-2t^2-2tx^3-2x^2+yt^6x-t^5y-2yt^2x-yt^4x^3+}{t^3x^3-3t^2x^2+3tx-1} \\
&\quad \tfrac{yt^3x^2+2yx^3)}{t^3x^3-3t^2x^2+3tx-1} \\
x = &\quad \tfrac{-5t^8+36t^4+12ut^4+9u^2-18v}{9u^2t+t(-6t^4+36)u+t(t^8-12t^4)} \\
y = &\quad \tfrac{18tv}{t^8-36t^4-6ut^4+9u^2}.
\end{aligned}$$

The curve W is the same in both cases but the Weierstrass coordinate functions, and hence their inverses, all differ. In the next section, when we explore the group structure of elliptic curves in Weierstrass form, we shall see that they don't differ in an arbitrary fashion. Also, while it may not currently be clear why the point [0,1,0], which is usually written $\mathcal{O}$ for Weierstrass curves, is so important, it will become clear when exploring the group structure of elliptic curves. Later when we

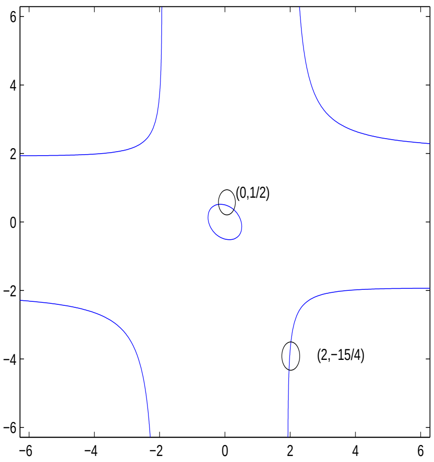

Figure 2.7: Some rational points on B from example (2.34).

have introduced some more algebraic sophistication we will discuss the question of the number of ways that one can move from one Weierstrass form to another using regular birational functions. In figure (2.7) we show the level set of B defined by $t = 2$ with two points on it; $(0, \frac{1}{2})$ (note this is one of the points that gets used as the identity in our transformation to Weierstrass form) and an arbitrarily chosen rational point. In figure (2.8) we see these two points on the curve W following the two different transformations used to map B to W - one making the point $(0, \frac{1}{2})$ the identity (s_1) and the other making the point $(\frac{1}{2}, 0)$ the identity (s_2). There is nothing profound in these points; it is just to show how such a thing looks in the real plane.

2.3.3 Elliptic Curves as Abelian Groups

It transpires that elliptic curves have not only the structure of an algebraic variety, but also that of an Abelian group. There are two ways to approach the construction of the group law on an elliptic curve. The first is to follow the theory of divisors and construct a group law from these. The second is to assume the curve is in Weierstrass form and construct a geometric group law. The two turn out to be equivalent but only the geometric construction will be given here. A treatment of divisors and how they can be used to give, for elliptic curves in Weierstrass form, the same group

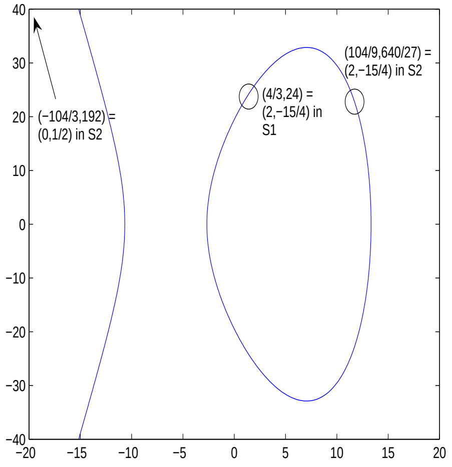

Figure 2.8: The images of the rational points (shown in figure (2.7)) on the Weierstrass cubic W in example (2.34) under the two different functions that convert to Weierstrass form.

structure as the geometric construction can be found in [63] or with more of an eye to complex elliptic curves in [69].

Let $C : y^2 = x^3 + Ax + B$ be an elliptic curve defined over K in Weierstrass form. Let $P = (p_x, p_y)$ and $Q = (q_x, q_y)$ be two points on C with coordinates in k, some subfield of K. Define an operation on two such points as follows:

Definition 2.35. *Let C be an elliptic curve in Weierstrass form and $P, Q \in C$ as above. Let L be the line that joins P and Q. Define $R = P * Q$ to be the third point of intersection between L and C.*

We shall have to prove a few facts about this star operation before using it to define a group law on the curve.

Proposition 2.36. *Let C be an elliptic curve in Weierstrass form and $P, Q \in C$ with coordinates in some field K. Let $R = P * Q$. Then*

(a) R exists and has coordinates in K.

(b) The $*$ operation does not turn the points of C into a group.

Proof. (a) Let $P = (p_x, p_y)$ and $Q = (q_x, q_y)$. Then the equation of L, the line that joins P and Q is $y - p_y = \lambda(x - p_x)$ where $\lambda = \frac{q_y - p_y}{q_x - p_x}$. To find the third point of intersection between L and C we substitute $y = \lambda(x - p_x) + p_y$ into the equation for C and solve for x. Following this step-by-step gives

$$\begin{aligned} (\lambda(x - p_x) + p_y)^2 &= x^3 + Ax + B \\ \lambda^2(x - p_x)^2 + 2p_y\lambda(x - p_x) + p_y^2 &= x^3 + Ax + B \\ \lambda^2(x^2 - 2xp_x + p_x^2) + 2p_y\lambda(x - p_x) + p_y^2 &= x^3 + Ax + B \\ x^2(\lambda^2) + x(-2\lambda^2 p_x + 2\lambda p_y) + (\lambda^2 p_x^2 - 2\lambda p_y p_x + p_y^2) &= x^3 + Ax + B \\ x^3 + x^2(-\lambda^2) + x(2\lambda^2 p_x - 2\lambda p_y + A) + (B + 2\lambda p_x p_y - \lambda^2 p_x^2 - p_y^2) &= 0 \end{aligned}$$

Since two roots to this cubic are already known, namely p_x and q_x we can use the sum of roots formula to find the third root, which we call r_x. The sum of roots formula applied to this cubic tells us that

$$p_x + q_x + r_x = \lambda^2$$

and thus that

$$r_x = (\frac{q_y - p_y}{q_x - p_x})^2 - p_x - q_x.$$

Note that since $p_x, p_y, q_x, q_y \in K$, r_x is also in K. To find the y-coordinate we simply put $x = r_x$ into $y = \lambda(x - p_x) + p_y$ to find r_y, which will also be in K. Thus $R = (r_x, r_y)$ exists so long as λ does, and always has its coordinates in K. If λ does not exist (i.e. $p_x = q_x$) then an alteration is required. If $p_y = q_y$, then $P = Q$ and the line L (which is meant to be the line that connects P and Q) is the tangent line to C at the point P and one follows the procedure as above. If $p_y \neq q_y$ then the line L is a vertical line $x = p_x$. In homogeneous form, this has the equation $X - p_x Z = 0$. Putting $Z = 0$ to check where this intersects the line at infinity tells us that $X = 0$ and hence Y cannot be 0 which is to say that L passes through the point [0,1,0]. As discussed in example (2.32), every curve in Weierstrass form possesses this point; thus the third point of intersection between L and C in this case is the point $R = [0, 1, 0]$.

(b) It is enough to note that no single point acts as an identity to show that the star operation does not turn the points on the curve into a group. To have $P * Q = P$ we require that $Q = P * P$. Now if $P = [0, 1, 0]$, then $P * P = [0, 1, 0]$ but it is clear that for general R, $R * [0, 1, 0] \neq R$, since the star of a point by $[0, 1, 0]$ reflects that

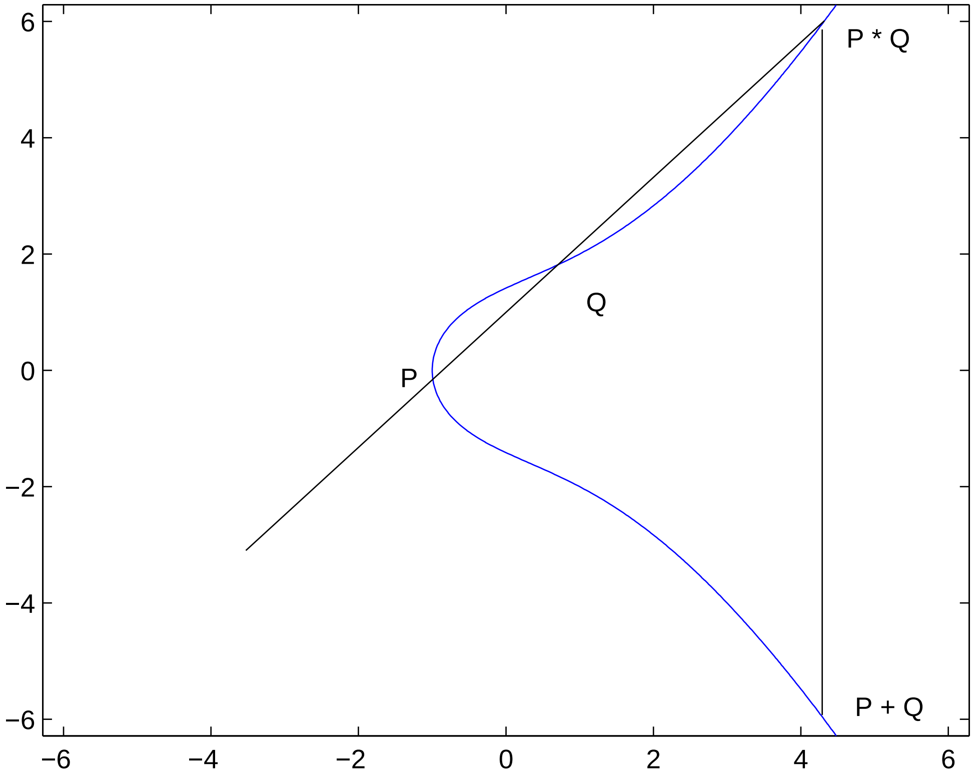

Figure 2.9: The star operation of definition (2.35) as well as the group law made up of two consecutive star operations on a Weierstrass cubic. The equation of the cubic is of a fairly generic type used for such illustrations; $y^2 = f(x)$ with $f(x) = 0$ having exactly one real solution.

point in the x-axis (this is easy to see because of the vertical lines mentioned just above). □

The question then is how we can modify the star operation to turn it into a group law. Two consecutive star operations turns out to be the answer. First recall that [0,1,0] lies on every curve in Weierstrass form, thus this ubiquitous point is used as the identity in the construction of the group law on elliptic curves in Weierstrass form and is of great import. Figure (2.9) shows the star operation and the group law on a possible graph of a Weierstrass equation.

Proposition 2.37. *Let $C : y^2 = x^3 + Ax + B$ be an elliptic curve in Weierstrass* *form and $\mathcal{O}$ = [0,1,0]. Then defining $P + Q = (P * Q) * \mathcal{O}$ for any points $P, Q \in C$ defines an Abelian group law on the points of C.*

Proof. The most comprehensive way to prove that this double star satisfies the properties for a group operation is to construct the formulae for the action and show that it satisfies those properties. So let $P = (p_x, p_y), Q = (q_x, q_y)$ and $R = (r_x, r_y)$ all be points with coordinates in some field K. First we show that $\mathcal{O}$ acts as the identity element. Adding $\mathcal{O}$ to any point P consists of drawing the vertical line through P, and jumping to the other point of intersection with the curve to give $P * \mathcal{O}$, followed by drawing the vertical line through this point to return to P. Thus $\mathcal{O}$ acts as identity point. Assume for now that $p_x \neq q_x$. Noting that the only effect starring a point by $\mathcal{O}$ has is to change the sign of the y-coordinate, we can use the calculations in proposition (2.36) to say that $P + Q = (\lambda - p_x - q_x, -(\lambda(\lambda^2 - 2p_x - q_x) + p_y))$. While it is clear from the geometric construction of the star law that $P+Q = Q+P$, associativity is not obvious. To prove this we calculate $(P + Q) + R$ and compare it to $P + (Q + R)$. To do this, assume our curve in Weierstrass form is given by $y^2 + x^3 + Ax + B$. We assume this non-standard form when performing long calculations since it is the form which MAPLE's Weierstrass conversion procedures work with, and hence the form all personal programs were written to work with. Now let $P = (p, \pm\sqrt{-p^3 - Ap - B}), Q = (q, \pm\sqrt{-q^3 - Aq - B}), R = (r, \pm\sqrt{-r^3 - Ar - B})$; the choice of whether we take the positive or negative square root in each individual point does not matter so long as we are consistent throughout the whole calculation. Now putting these points into $(P+Q)+R$ and $P+(Q+R)$ (see appendix for MAPLE code that can do this) eventually gives two equal expressions. The symbolic nature of this calculation means a separate set of calculations must be performed assuming that some of the points have the same x-coordinate, but the method and result are the same. Lastly, for the existence of inverses we again check what this means geometrically. Given a point P, we want to find a point P^{-1} such that when we join the two with a line the third point of intersection of this line and the curve is a point that is the reflection of $\mathcal{O}$ in the x-axis. Of course, the reflection of $\mathcal{O}$ in the x-axis is just $\mathcal{O}$ again, so we're looking for the vertical line through P. Thus for $P = (p_x, p_y)$, $P^{-1} = (p_x, -p_y)$.

One further thing to note about this group construction for elliptic curves is that all operations preserve the field in which the points lie. So supposing k is a subfield of K, the group of points on C with coordinates in k form a subgroup of the group of points on C with coordinates in K. □

Example 2.38. Let's return to example (2.34) and make a comment on what the conversion functions produced by MAPLE are doing. In the first case we used the point $(0, \frac{1}{t})$ as the identity, and called the transformation and, abusing notation a little, the whole situation s_1. In the second we used $(\frac{1}{t}, 0)$, and called the transformation and situation s_2. Now

$$s_1((\frac{1}{t}, 0)) = P_2 = (\frac{-t^2(5t^2+6)}{3}, -2t^4(t^2+2))$$

and

$$s_2((0, \frac{1}{t})) = P_1 = (\frac{-t^2(5t^2+6)}{3}, 2t^4(t^2+2)).$$

Notice that $P_2 = \mathcal{O} - P_1$ (this is easy to tell at a glance since only the second coordinate differs by being the negative of one another), and conversely that $P_1 = \mathcal{O} - P_2$. This suggests checking if the difference between s_1 and s_2 is just a shift by the relevant point. To check this we find a third point on the original biquadratic, find its image under both s_1 and s_2 and see if they differ by a shift of P_2. We use the point $Q = (-\frac{2t}{t^4-1}, -\frac{t^8-6t^4+1}{t(t^8-2t^4-3)})$ as our test. Then

$$s_1(Q) = Q_2 = (\frac{-2t^4}{3} - 1, t^4 - 1)$$

while

$$\begin{aligned} s_2(Q) = Q_1 =&(\frac{-t^2(5t^{10}+2t^8-14t^6+16t^4+29t^2+6)}{6t^4-12t^2+12t^6+3+3t^8}, \\ &\frac{-2(t^{14}-7t^{10}-2t^8+7t^6+4t^4+15t^2+6)t^4}{-4t^6-9t^4+9t^8+6t^2+6t^{10}-1+t^{12}}). \end{aligned}$$

Now it is easy to check that $Q_1 = Q_2 - P_2$. One can carry out this same exercise with a general point (x, y) performing the role of Q above to find the more general statement that to change identity points on this Weierstrass we just have to subtract the point to be used as the new identity.

Example 2.39. Consider again the Weierstrass curve

$$W : u^3 + (-\frac{1}{3}t^8 - 4t^4)u + \frac{2}{27}t^{12} - \frac{8}{3}t^8 + v^2$$

Now this contains the point $P = (-\frac{1}{3}t^2(5t^2+6), -2t^4(t^2+2))$ and if we attempt to use `Weierstrassform` to convert W using the point P as the new identity we actually come to a curve with a different equation. Of course it should be possible to maintain the same equation for the curve but the algorithm works in a set way and the way it works dictates that the equation change. As it transpires, the new equation is

$$W_2 : U^3 + (-177147t^8 - 2125764t^4)U + 28697814t^{12} - 1033121304t^8 + V^2$$

and we know that we should be able to move between W and W_2 without changing the identity point by a much simpler transformation of the form $(U, V) = (\mu^2 u, \mu^3 v)$. We can find μ^6 by taking the term $28697814t^{12}$ from W_2 and dividing by the corresponding term from W, $\frac{2}{27}t^{12}$. This yields $\mu = 27$ and applying this transformation does indeed move from W to W_2 as it should.

Before moving on, we note here that while we have assumed our elliptic curves to be in Weierstrass form for the geometric construction of the group law, any elliptic curve can be considered to have a group structure simply by inheriting that of the group structure on the Weierstrass form that it is conjugate to. Geometrically the group law for an elliptic curve in Weierstrass form is simple enough to work with even if it is a little unmotivated. Divisors are somewhat the opposite; more difficult (in their abstraction) to work with but better motivated.

With the construction of the group law dealt with, we now turn to what kind of structure this group law gives to the points on an elliptic curve. No matter which field we are working over the group is always Abelian but the finer structure of the group depends upon which field one is working over.

Elliptic curves over $\mathbb{Q}$

Let $C : y^2 = x^3 + Ax + B$ be an elliptic curve defined over $\mathbb{Q}$. Let us consider the group formed by the points $C(\mathbb{Q})$ under the group law as defined in section (2.3.3). One of the first major results regarding elliptic curves at all, proven first for this case of rational elliptic curves by Mordell in 1922, is the Mordell-Weil Theorem:

Theorem 2.40. *(Mordell) Let C be an elliptic curve defined over $\mathbb{Q}$. Then the group $C(\mathbb{Q})$ is finitely generated.*

For a relatively elementary proof of this theorem, see [64]. Knowing that this group is both finitely generated and Abelian allows us to use the Fundamental Theorem of Finitely Generated Abelian Groups to give the general structure of $C(\mathbb{Q})$ as

$$C(\mathbb{Q}) \cong \mathbb{Z}_{p_1^{a_1}} \oplus \mathbb{Z}_{p_2^{a_2}} \oplus \ldots \oplus \mathbb{Z}^r$$

The quantity r here, which gives how many copies of $\mathbb{Z}$ are inside $C(\mathbb{Q})$ is called the *rank* of C. As usual, the subgroup of finite order elements are called the *torsion* of $C(\mathbb{Q})$. More can be said about the rank and torsion of a rational elliptic curve. A theorem on the torsion subgroup of $C(\mathbb{Q})$ comes due to Mazur [44].

Theorem 2.41. *(Mazur) Let C be an elliptic curve defined over $\mathbb{Q}$. Then the torsion subgroup of $C(\mathbb{Q})$ is isomorphic to one of the following groups:*

$$\begin{aligned} \mathbb{Z}/n\mathbb{Z}, &\quad 1 \leq n \leq 10 \text{ or } n = 12 \\ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}, &\quad n = 2, 4, 6, 8 \end{aligned}$$

Mazur's theorem tells us the possible finite orders for points with rational coordinates on a rational elliptic curve. Actually finding such points of finite order is another question and efficient ways to find them have been made based on the Nagell-Lutz theorem, proven independently by Nagell and Lutz in the mid-1930s. Before giving this theorem we shall prove a little result on Weierstrass equations.

Lemma 2.42. *Let W be an elliptic curve defined over $\mathbb{Q}$ with equation $y^2 = x^3 + \frac{A_1}{A_2}x + \frac{B_1}{B_2}$. Then there is an equivalent Weierstrass form with integer coefficients.*

Proof. Testing the transformations $x \to \mu^2 x$ and $y \to \mu^3 y$ gives the new equation

$$\mu^6 y^2 = \mu^6 x^3 + \mu^2 \frac{A_1}{A_2} x + \frac{B_1}{B_2}$$

which simplifies to

$$y^2 = x^3 + \mu^{-4} \frac{A_1}{A_2} x + \mu^{-6} \frac{B_1}{B_2}$$

Now to make all coefficients integers, it suffices to take μ^{-1} to be the lowest common multiple of A_2 and B_2. □

Theorem 2.43. *(Nagell, Lutz) Let W be an elliptic curve defined over $\mathbb{Q}$ with equation $y^2 = x^3 + Ax + B$ where A, B are both integers (note that Lemma (2.42)* *tells us this is possible without loss of generality). Then if $(x_p, y_p) \in W$ is a point of finite order, x_p and y_p are both integers. Furthermore either $y_p = 0$ (and the point is of order two) or y_p^2 divides $4A^3 + 27B^2$.*

While the torsion subgroups of elliptic curves defined over the rational numbers are well understood, the rank is more mysterious. Knowing that there is only a finite number of possibilities for the torsion part of a group $E(\mathbb{Q})$, one can pose a question regarding rank by asking what the maximal rank is (if one exists) for elliptic curves with a given torsion subgroup. A. Dujella maintains an up-to-date website at [15] that lists the current answer to this question though it is conjectured that there is no bound for any of the allowable torsion subgroups. The current maximal rank across all allowable torsion subgroups is 24.

Elliptic Curves over $\mathbb{R}$ and $\mathbb{C}$

While elliptic curves defined over the real and complex numbers are never required in any detail throughout this thesis, some of the structural theory of them is presented here for completeness sake. The major difference between working with the rational numbers and the real or complex is that the latter two form a continuum. This shows itself while working with elliptic curves defined over them when the group law becomes continuous and the group of points on the curve becomes a Lie group. In the case of elliptic curves defined over the real numbers there is not much to say - the only one dimensional compact (and elliptic curves in Weierstrass form are compact; the point at infinity sees to that) connected Lie group is the group of rotations of the unit circle. Now real elliptic curves can be classified into two classes by writing them as $y^2 = (x - e_1)(x - e_2)(x - e_3)$ and seeing if all three of the e_i's are real, or just one. If only one is real, then the entire curve is connected and so the group of points on it is isomorphic to the multiplicative group $S^1 = \{e^{i\theta} | \theta \in [0, 2\pi)\}$. Alternatively, if all three of the e_i's are real, the curve is made up of two connected components; one loop and one arc. In this case the group of real points on the curve is isomorphic to $C_2 \times S^1$ with the S^1 component again coming from the arc part of the curve containing the identity point.

The structure for complex elliptic curves allows us to mention the historical context behind the use of "Weierstrass equation", "Weierstrass curve" and "Weierstrass form" as well as the very name "elliptic curve". Knowing that an elliptic curve can be written with equation $y^2 = x^3 + Ax + B$, a simple rescaling results in the form $y^2 = 4x^3 - g_2 x - g_3$ which is an equation frequently seen in the theory of complex numbers. A fundamental mathematical object that we will need in the following discussion is a (point) lattice.

Definition 2.44. *Let ω_1 and ω_2 be two complex numbers that are linearly independent over $\mathbb{R}$. Then the lattice spanned by ω_1 and ω_2 is*

$$L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \{n_1\omega_1 + n_2\omega_2 | n_1, n_2 \ \in \ \mathbb{Z}\}$$

Figure (2.10) shows part of the lattice generated when $\omega_1 = 1+i$ and $\omega_2 = 1+2i$. The fundamental parallelogram (see below) is also drawn in.

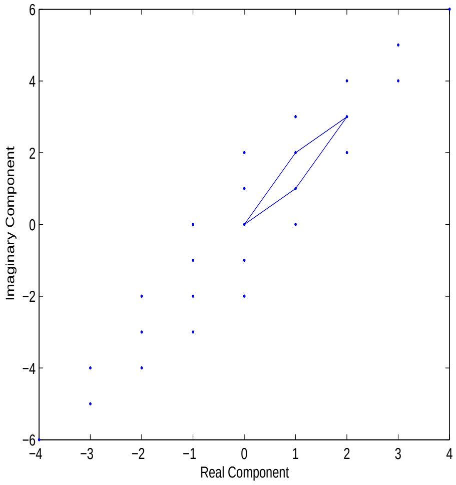

Figure 2.10: A lattice defined by two complex numbers with the fundamental parallelogram drawn included.

The differential equation

$$(\frac{df}{dz})^2 = 4(f(z))^3 - g_2 f(z) - g_3 \tag{2.11}$$

has a meromorphic solution called the Weierstrass p-function defined by

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in L} (\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2})$$

where L is the lattice spanned by two complex numbers ω_1 and ω_2 which can be found from g_2 and g_3 via some rather complicated integral calculations. The integral calculations required to find the ω_i's from the g_i's are called elliptic integrals. Such integrals arise, in one instance, in finding the arc length of ellipses; this begins to explain the name of elliptic curves. Now given a Weierstrass p-function $\wp(z)$ we can make the identification $x = \wp(z), y = \wp'(z)$ and substitute these into the differential equation (2.11) to get the usual equation for an elliptic curve. A legitimate question to ask is whether these functions parameterise the elliptic curve totally (i.e. is the map $z \to (\wp(z), \wp'(z))$ a surjective map onto the elliptic curve)? The answer is yes though the proof is understandably difficult. More interestingly, this map is not injective. Indeed it is the case that $\wp(z + \omega) = \wp(z) \; \forall \omega \in L$. This relation is why the Weierstrass $\wp$-functions are called "doubly periodic" as they are periodic in two linearly independent directions. This tells us that we can fill the entire elliptic curve using complex numbers just in the interior (and some of the boundary of) of one of the lattice's parallelograms. The natural parallelogram (called the *fundamental parallelogram*) to choose is of course that with the vertices $0, w_1, w_2, w_1+w_2$. Various properties of Weierstrass p-functions can be used to show that

$$\wp(z_1 + z_2) = \wp(z_1) + \wp(z_2).$$

This turns out to give another equivalent formulation of the group law for complex points on elliptic curves; we can use the addition identities for Weierstrass p-functions which can be proven to give the same group structure as the geometric group law and the group law defined through divisors on the same elliptic curve. Writing the geometric group law addition as $\oplus$ and writing complex addition (let p and q be two complex numbers) modulo the lattice L as $+$, we can express this group isomorphism as

$$\wp(p) \oplus \wp(q) = \wp(p + q). \tag{2.12}$$

It is by considering the group of complex points on an elliptic curve as the interior of the fundamental parallelogram that we can get a simple idea of the structure - it is simply isomorphic to the group $\mathbb{C} \backslash L$.

2.4 Advanced Theory of Elliptic Curves

In this section we build upon the basic theory of elliptic curves given in section (2.3) and give results that we will be directly invoking in the original work in this thesis. Most of the results in this section will be given without proof, but the lemmas that can be shown using elementary methods will be proven. This section is largely dedicated to maps between elliptic curves, which will become important as we will be considering maps of an elliptic curve to itself. Also discussed is the existence of actual points on elliptic curves, which will become doubly important; firstly as a way of constructing new integrable maps and secondly as an expansion of the numerical method of detecting integrability given in [57].

2.4.1 Maps Between Elliptic Curves

With the formal definition of morphism previously given in definition (2.17) we can turn to how such maps link with the algebraic structure of elliptic curves. The theory given without proof will be important in proving the original results in this thesis. The proof of these can be found in [63], although it will occasionally refer back to [22].

Definition 2.45. *An isogeny $\iota : E_1 \to E_2$ is a morphism such that $\iota(O_1) = O_2$, i.e. an isogeny sends the identity point to the identity point.*

Theorem 2.46. *All isogenies preserve the group law on the elliptic curves they map between.*

Proof. See [63], Chapter III Theorem 4.8 □

Theorem (2.46) tells us that isogenies are homomorphisms and conversely homomorphisms are certainly isogenies since we require that the identity point is always mapped to the identity point. So with theorem (2.46) taken into account, an isogeny is just another name for a homomorphism of the group of points on an elliptic curve. We now give the example which will be most useful throughout the thesis and also a class of examples of particular isogenies.

Example 2.47. Suppose we have an isogeny $\iota : E \to E$, then ι is called an endomorphism of E. These form a ring under addition ($(\iota + \kappa)(A) = \iota(A) + \kappa(A)$) and composition ($\iota \circ \kappa(A) = \iota(\kappa(A))$). The units (i.e. the invertible elements) of $End(E)$ form a group called the automorphism group of E, $Aut(E)$. The structure of this automorphism group will become important later in this section.

The most obvious specific examples of isogenies are those endomorphisms of any elliptic curve that send a point to its double, or triple, etc. according to the group law. Let us denote the map that sends a point P to mP as $[m]$. Then these certainly send O to itself, meaning that they preserve the group law by theorem (2.46). They clearly add and compose with each other as $[m]+[n]=[m+n]$ and $[m]\circ[n]=[mn]$ giving a copy of $\mathbb{Z}$ inside the ring $End(E)$. Indeed for most elliptic curves this is the entire ring.

As mentioned in the above example, the automorphisms of an elliptic curve are the invertible isogenies from that elliptic curve to itself. We can use the Weierstrass normal form for elliptic curves to completely describe this group. Suppose an elliptic curve E has Weierstrass equation $y^2=x^3+Ax+B$. Now consider any map that transforms E to another elliptic curve E' with Weierstrass equation $y^2=x^3+A'x+B'$. Referring back to example (2.31) it is clear that any transformation that maintains this Weierstrass form is necessarily of the form $(x,y)\to(u^2x,u^3y)$ with $u\in K^*$. Using this kind of substitution, we see that this implies relations between A, A', B, B' of

$$A' = u^4 A$$
$$B' = u^6 B.$$

Now for such a transformation to be an automorphism of E we require that E and E' are the same so $A'=A, B'=B$. If $AB\neq 0$, we see that the only way these conditions can be satisfied is if $u=\pm 1$. However, if $B=0$ the additional possibilities of $u=\pm i$ are introduced while if $A=0$, the additional possibilities of u being any sixth root of unity are introduced. This discussion suggests the following theorem which is proven formally in ([63], Chapter III Section 10):

Theorem 2.48. *The group $Aut(E)$ of automorphisms of an elliptic curve is isomorphic to either C_2, C_4 or C_6 i.e. the cyclic group of order 2, 4 or 6.*

Which specific structure it has is determined by the values of A and B. If $A=0$ we have $Aut(E)\cong C_6$ while if $B=0$ we have $Aut(E)\cong C_4$ and finally if $AB\neq 0$ (which is the typical case, and thus the case which we will assume when it simplifies calculations) we have $Aut(E)\cong C_2$.

One very important fact (or, at least, one that we shall be using a lot) about the morphisms from an elliptic curve to itself is that there are very few of them, in the following sense:

Theorem 2.49. *Let f be a morphism of an elliptic curve E. Then f is the composition of an isogeny of E and a translation on E.*

Proof. Let $f(\mathcal{O}) = Q$. Let q be the (invertible) map $q : P \to P - Q$. Then the map $\tau : q \circ f$ is an isogeny since it maps $\mathcal{O}$ to itself. Thus $f = q^{-1} \circ \tau$ with τ an isogeny and q^{-1} a translation as desired. □

This theorem allows us to introduce some more sophistication into the earlier example (2.34) regarding how many "different" ways we can convert two isomorphic elliptic curves to one another.

Corollary 2.50. *Let E_1 and E_2 be two elliptic curves. Suppose $\phi_1 : E_1 \to E_2$ and $\phi_2 : E_1 \to E_2$ are (one direction of) isomorphisms between E_1 and E_2. Then $\phi_1(P) = \phi_2(\tau(P) + Q_T)$ where $\tau \in Aut(E_1)$ and $Q_T = \phi_2^{-1} \circ \phi_1(\mathcal{O}) \in E_1$.*

Proof. The composition $\phi_2^{-1} \circ \phi_1$ is a morphism of E_1 hence $\phi_2^{-1} \circ \phi_1 = T \circ \tau$ where T is a translation (by the point Q_T say) and τ is an isogeny. However, since the left hand side of this equality is invertible, the right hand side must also be invertible so τ is an invertible isogeny i.e. an automorphism of E_1. Thus $\phi_1(P) = \phi_2(\tau(P) + Q_T)$ with $\tau \in Aut(E_1)$. □

In words, corollary (2.50) tells us that two invertible maps between the same Weierstrass curves can only differ by a morphism in the argument.

A final theorem that seems unmotivated for now but very important in later chapters is Hurwitz' Theorem.

Theorem 2.51. *Let C be a non-singular curve with genus $g \geq 2$. Then the automorphism group[2] of C has order at most $84(g-1)$.*

While this theorem does not say anything directly about elliptic curves, it does tell us that if we know that a curve has a large group of automorphisms (i.e. of

[2] Here the automorphism group refers to all the birational maps of the curve of genus g. It is unrelated to the internal group structure of an elliptic curve.

infinite size) then the curve we are examining must be singular or of genus zero or one. One way to confirm that any automorphism group is infinite order is by discovering an automorphism of infinite order.

2.4.2 Maps Preserving Conics

Theorem (2.51) ensures that in the context of interesting (that is, infinite order) curve-preserving maps we may restrict our search to curves of genus zero and one. As is clear, the majority of this thesis deals with the genus one case. In this section, we briefly examine the possibilities in the genus zero case. Let C be any genus zero curve defined over a field K, and L a map from C to itself also defined over K. The question we wish to ask is whether C can be given a group structure, and if L then necessarily acts in some particular way with relation to that group structure.

From a geometric approach, the fact that elliptic curves could be birationally reduced to cubics played an important role in constructing a group law on them. Conics are simpler in that they can be rationally parameterised. That is to say, any point on a conic C can be given by $(x(r), y(r))$ where x, y are rational functions in r. Now taking r to be all values in a field K will give all the K-rational points on C. Thus a natural imposition of a group upon C comes by considering the group $(K, +)$ acting on the parameter r.

Example 2.52. Let C be the curve given by the equation $y - x^2 - t = 0$. Now a general point on C is given by $R(r) = (r, r^2 + t)$. Now any change in r gives a corresponding change to the point R. Adding points on C, considering C to be an additive group as inherited from the parameter r, is given by

$$R(r) + Q(q) = (r + q, (r + q)^2 + t)$$

To examine how such a map looks in the Cartesian plane (as opposed to in the parameter space) we fix one point to be adding, say $Q = (1, 1 + t)$ and try to interpret this in the (x, y)-plane. Since C and its parameterisation is particularly simple, we can immediately see that $x = r$ (since any point (x, y) on C can be written as $(r, r^2 + t)$). Now referring back to equation (2.13) we see that for this point Q, the map on x is simple - $x' = x + 1$. To determine the map on y, we just use the fact that if $(x, y) \in C$ then also $(x', y') \in C$. That is, $y' - x'^2 - t = 0$ whence

$y' = x'^2 + t = (x+1)^2 + t$. Thus addition of the point $(1, 1+t)$ on the conic C is represented by the map $(x, y) \to (x+1, (x+1)^2+t)$. Clearly one can construct maps of r that are more complicated than this kind of translation. There are all the maps of the form $(x, y) \to (ax, (ax)^2 + t)$ which play a similar role to the endomorphisms $P \to kP$ on elliptic curves, but these maps are invertible whereas the "multiplication endomorphisms" on elliptic curves are not.

2.4.3 The j-invariant

The j-invariant is a quantity that can be attached to any elliptic curve which can be used to tell, at a glance, if two elliptic curves are isomorphic to each other.

Definition 2.53. *Let $C : y^2 = x^3 + Ax + B$ be an elliptic curve defined over a field K. Then the j-invariant of C is*

$$j(C) = 1728 \frac{4A^3}{4A^3 + 27B^2}. \tag{2.13}$$

We shall prove here that any two elliptic curves in Weierstrass form with the same j-invariant are isomorphic, and conversely that any two isomorphic curves in Weierstrass form have the same j-invariant.

Proposition 2.54. *Let C and C' be two elliptic curves in Weierstrass form defined over K. Then C is isomorphic to C' if and only if $j(C) = j(C')$. However, while the isomorphism will be defined over $\overline{K}$, it may not be defined over K.*

Proof. Let $C : y^2 = x^3 + Ax + B$ and $C' : (y')^2 = (x')^3 + A'x' + B'$. Suppose $j(C) = j(C')$ so that

$$\begin{aligned}
1728 \frac{4A^3}{4A^3 + 27B^2} &= 1728 \frac{4A'^3}{4A'^3 + 27B'^2} \\
4A^3(4A'^3 + 27B'^2) &= 4A'^3(4A^3 + 27B^2) \\
A^3B'^2 &= A'^3B^2
\end{aligned}$$

We now look for an isomorphism of the form $(x', y') = (u^2x, u^3y)$ for some u. Suppose first that $A = 0$. Then B is non-zero since otherwise C is not genus one. Also, since $A' = 0$ (from the equality of j-invariants), B' is non-zero or else C' would not be genus one. Let $u = (\frac{B'}{B})^{1/6}$. Substituting this into the proposed isomorphism and then into the equation for C' gives

$$\begin{aligned} \frac{B'}{B}y^2 &= \frac{B'}{B}x^3 + B' \\ y^2 &= x^3 + B \end{aligned}$$

i.e. as long as $(x, y) \in C$, then $(u^2x, u^3y) \in C'$ for the above u. Similarly, if $B = 0$ then we take $u = (\frac{A'}{A})^{1/4}$ and find the same. Lastly, suppose $AB \neq 0$. From above, we know that $\frac{A^3}{A'^3} = \frac{B^2}{B'^2}$ whence taking $u = (\frac{A'}{A})^{1/4} = (\frac{B'}{B})^{1/6}$ gives the correct isomorphism.

Conversely, if two elliptic curves in Weierstrass form are isomorphic they are related by a change of variables $(x', y') = (u^2x, u^3y)$. Putting this into the equation for C' gives $u^6y^2 = u^6x^3 + A'u^2x + B'$ which we rewrite as $C : y^2 = x^3 + (A'u^{-4})x + B'u^{-6}$. Putting this new curve's coefficients into the formula for the j-invariant gives

$$j(C) = 1728\frac{4(A'u^{-4})^3}{4(A'u^{-4})^3 + 27(B'u^{-6})^2}.$$

Multiplying numerator and denominator by u^{12} gives $j(C')$, so $j(C) = j(C')$ when the two curves differ only by the allowed transformations. □

A nice example illustrating the proviso that the isomorphism of proposition (2.54) need only be defined over $\overline{K}$ comes care of [69].

Example 2.55. Consider the two curves $W_1 : y^2 = x^3 - 25x$ and $W_2 : y^2 = x^3 - 4x$. These two curves are defined over $\mathbb{Q}$ and have the same j-invariant yet it can be shown that there is a point of order infinity $(-4, 6) \in W_1$ (that this point has infinite order can be seen by checking that nP is not the identity for $1 \leq n \leq 12$ and invoking theorem (2.41)) while the only rational points W_2 possesses are those of order 2; $(2, 0), (-2, 0)$ and $(0, 0)$. Thus there can be no transformation with rational coefficients that maps the two curves to one another. However we know that a transformation of the form $(x, y) \to (\mu^2x, \mu^3y)$ can be used to map the two to each other. To calculate μ here we divide the term $-25x$ by the corresponding term $-4x$ and notice that this quotient must be μ^4. So, $\mu^2 = \frac{5}{2}$ and $\mu = \frac{\sqrt{10}}{2}$.

As a final example, we will show how an elliptic curve with any given j-invariant can be created.

Example 2.56. Consider a curve $y^2 = x^3 + Ax + A$. We compute the j-invariant of such an elliptic curve as $j(C) = 1728\frac{4A^3}{4A^3+27A^2} = 1728\frac{4A}{27+4A}$. Now we solve $j(C) = J$ for A and find that $A = -\frac{27J}{4(J-1728)}$. Thus an elliptic curve of the form $y^2 = x^3 - \frac{27J}{4(J-1728)}x - \frac{27J}{4(J-1728)}$ has a j-invariant of J. Clearly this will not work for $J = 0$ (the curve is not elliptic) or $J = 1728$ (the coefficients do not exist). However, $y^2 = x + Ax$ has a j-invariant of 1728, and $y^2 = x + B$ has a j-invariant of 0 for all choices of A and B.

2.5 The Hasse-Weil Bound and Similar Theorems

The focus in this section is the theory that governs the size of curves over finite fields. After running through the few problems that may arise when considering curves defined over finite fields, we shall give the theorem (the Hasse-Weil Bound) that sparked interest in this area and then explain why and how. Finally a generalisation of the Hasse-Weil Bound to higher dimensions will be given as a possible future direction for extending the results of the work here that pertains to [60].

Curves "work" in most finite fields just as they work in more standard fields such as the rational numbers and the complex numbers as has been described up to this point. The two exceptions are when the fields are of characteristic 2 or 3, in which cases certain formulae break down and one must re-approach that particular situation from the start to fix the problem. However we will be working exclusively with finite fields of prime order. Small primes will be used for illustration of principles (small meaning around 11) and larger primes will be used for demonstrating numerical evidence. Depending on the calculations involved, large can mean anything between 100 (for particularly labourious calculations) up to 10000.

Aside from problems in characteristics two and three, which are dealt with throughout by simply avoiding such fields, one other problem can arise when working with curves over finite fields. Generally we are led to consider a particular curve C over a finite field by first considering it over the rational numbers and reducing it to a different curve $\widetilde{C}$ that is defined over the finite field instead of the rational numbers. For fields of prime order p this is done in the obvious way; by reducing each coefficient in the curve's equation modulo p. Similarly, rational points on the curve

t^2	Equation of W	Number of points on B	Number of points on W
0	u^3+v^2	10	11
1	$u^3+3u+8+v^2$	4	7
2	$u^3+8u+2+v^2$	2	12
3	$u^3+3u+3+v^2$	12	15
4	$u^3+u+9+v^2$	12	15
5	u^3+7u+v^2	8	11
6	u^3+7u+v^2	12	11
7	$u^3+u+9+v^2$	16	15
8	$u^3+3u+3+v^2$	16	15
9	$u^3+8u+2+v^2$	20	12
10	$u^3+3u+8+v^2$	8	7

Table 2.1: Number of affine points lying on level sets of the family of curves of example (2.57).

are reduced and become points with coordinates from the field $\mathbb{Z}_p$. Problems can arise when two unique points that have some significance to the curve are reduced to a single point. Such a situation will be well demonstrated by an example that arose unexpectedly in the previous work [30].

Example 2.57. Again take the curve

$$B(x,y) = x^2y^2 - t^2(x^2+y^2) - 2xy + 1$$

which has Weierstrass equation

$$W(u,v) = u^3 + (-\frac{1}{3}t^8 - 4t^4)u + \frac{2}{27}t^{12} - \frac{8}{3}t^8 + v^2$$

We wish to consider these two curves modulo 11, i.e. check how the curves look as curves defined over $\mathbb{F}_{11}$ rather than over $\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$. Table (2.1) contains data regarding the number of (affine) points on the curves B and W for values of t^2. The graphs of these curves look nothing like their counterparts defined over $\mathbb{R}$. Figures (2.11) and (2.12) show, for B and W respectively, plots of the level sets $t^2 = 0$ (denoted by circles), $t^2 = 2$ (denoted by squares) and $t^2 = 5$ (denoted by diamonds) over the finite affine plane $\mathbb{F}_{11}^2$. Features worthy of note include the

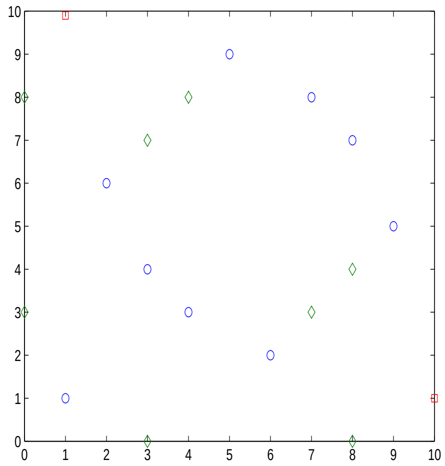

Figure 2.11: Some biquadratics from example (2.57) defined over $\mathbb{F}_{11}$.

symmetry displayed by the first plot that is not present in the second. The symmetry displayed is that given by the symmetry in the equation for B of switching x and y. After giving the Hasse-Weil Bound, we shall see that something is amiss with some of the entries in table (2.1) and resolve them.

The theorem governing how many points can lie on a curve with coordinates from some finite field is named the Hasse-Weil Bound after Hasse, who proved the theorem for the case of elliptic curves in 1933 and Weil who proved the theorem for arbitrary genus in 1948.

Theorem 2.58. *(Hasse, Weil) If C is a non-singular, irreducible curve of genus g defined over the finite field $\mathbb{F}_p$ then the number of points on C with coordinates in $\mathbb{F}_p$ is $p+1+\epsilon$ where ϵ is an error term satisfying $|\epsilon| \leq 2g\sqrt{p}$. Alternatively we can write both sides of this inequality as*

$$p+1-2g\sqrt{p} \leq |C| \leq p+1+2g\sqrt{p}$$

Note that since we only ever work in finite fields of prime order, we use the symbol p here, and usually throughout the thesis , rather than the more traditional q.

For $p = 11$ and $g = 1$, the lower bound is 6 and the upper bound is 18. Now we can see that the entries in table (2.1) with $t^2 = 2$ and $t^2 = 9$ are both awry as the

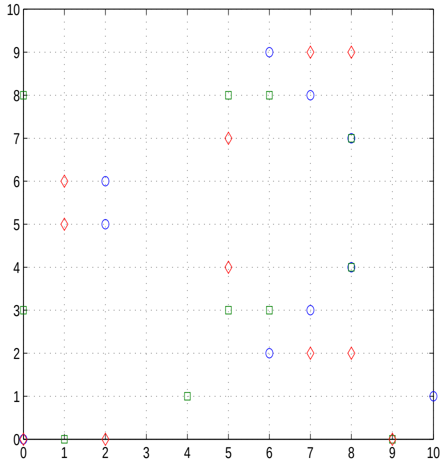

Figure 2.12: Some Weierstrass cubics from example (2.57) defined over $\mathbb{F}_{11}$.

corresponding curves B contain 2 and 20 points respectively. The problem here is that these two curves are reducible. The curve with $t^2 = 9$ is reducible over $\mathbb{F}_{11}$ (but not over $\mathbb{Q}$ or $\mathbb{C}$; it factorises into $((y+8)x+3y+1)((y+3)x+8y+1))$ while the curve with $t^2 = 2$ is reducible over an extension of $\mathbb{F}_{11}$, the extension being by $\sqrt{2}$ (it factorises into $((x+10\alpha)y+1+10\alpha x)((x+\alpha)y+1+\alpha x)$, where $\alpha = \sqrt{2}$). It is easy to see how this could confound experimental results especially if the number of "exceptional" cases (where reducibility of a curve comes from the particular choice of finite field) is relatively large compared to the entire set of curves. A perceptive reader would have noted that according to table (2.1), the entry for $t^2 = 1$ also fails to satisfy the Hasse-Weil bound. However the table does not take into account the two points at infinity which lie on each of the curves whereas the Hasse-Weil bound assumes this projectivity in curves. Beyond the Hasse-Weil concerns, one might also question why the numbers are so different between the two tables. This is due to the fact that the conversion functions between the two curves can sometimes fail to be 1-1 for some points when considered over finite fields as well as the fact that points with coordinates in an extension of $\mathbb{F}_{11}$ can be mapped to points in the field proper by the conversion functions.

It was the existence of the Hasse-Weil Bound that served as inspiration for the basis of this work. A common interest in dynamics is with maps whose orbits all lie on a particular curve. Thus by checking the length of an orbit of any map considered over a finite field and comparing this to the upper Hasse-Weil Bound, we can rule out such an orbit being restrained to a curve of a particular genus. More of this discussion and the arguments driving it will be seen shortly in chapter 3. While the Hasse-Weil bound only deals with the number of points on curves (recall these are varieties of dimension one), a generalisation by Lang and Weil published in 1954 [39] gives a similar theorem for varieties of higher dimension.

Theorem 2.59. *Let V be a variety in n-dimensional projective space with dimension r and degree d defined over a finite field $\mathbb{F}_p$. Then there exists a constant $A(n,d,r)$ such that*

$$|N - p^r| \leq (d-1)(d-2)p^{r-\frac{1}{2}} + A(n,d,r)p^{r-1}$$

where N is the number of points on V.

This asymptotic result has great potential to extend the results of [60] on planar integrable systems to higher dimensions. However, the algebraic geometric technology to exploit this theorem lies beyond the scope of this thesis.

Chapter 3

Foundations 2: Dynamics and Maps

This chapter is a brief overview of the history and theory of the type of dynamical systems with which we are chiefly concerned in this thesis. As such this chapter will mostly pertain to integrable and reversible dynamical systems that are discrete in time.

3.1 Maps

This section consists largely of a historical overview of integrable maps (being the focus and motivation for the original parts of this thesis) and the reversing symmetry group of (not necessarily integrable) maps. We start by defining the basics of time-discrete maps.

Definition 3.1. *A general map of a space K^n is a function $L : K^n \to K^n$. If $v = Lu$, v is called the iterate or image point of u. The orbit of a point u is the (ordered) set $\{u, Lu, L^2u \ldots\}$.*

In terms of notation, we shall take the following convention of writing maps

$$\begin{aligned} x_1' &= f_1(x_1, x_2, \ldots, x_n) \\ x_2' &= f_2(x_1, x_2, \ldots, x_n) \\ \vdots &= \vdots \\ x_n' &= f_n(x_1, x_2, \ldots, x_n) \end{aligned}$$

where x_i' denotes the iterate of x_i. Any number of these coordinates may never change, the equation for their iterate being given by $x_i' = x_i$. We will often adopt the Greek alphabet to label such coordinates and call them *parameters*. This reduces the technical difficulty of studying the map. Note that rational maps (those where each f_i is a rational function of the n coordinates) can be easily extended to projective maps by the following procedure of homogenising maps. The purpose of homogenising a map is to remove any singularities. To do this, we introduce an $(n+1)$th coordinate and ensure that all singularity information is contained within that extra coordinate.

Definition 3.2. *Let L be a map of K^n given by*

$$\begin{aligned} x_1' &= f_1(x_1, x_2, \ldots, x_n) \\ x_2' &= f_2(x_1, x_2, \ldots, x_n) \\ \vdots &= \vdots \\ x_n' &= f_n(x_1, x_2, \ldots x_n) \end{aligned}$$

with each f_i rational. Then we make a series of substitutions $x_i = \frac{X_i}{X_{n+1}}$ and multiply each of the ordinates of the map by X_{n+1}' to reduce each ordinate down to a simple fraction F_i of polynomials in $X_1, X_2, \ldots, X_{n+1}$ multiplied by a factor of X_{n+1}'. At this stage we have the map L given by

$$\begin{aligned} X_1' &= X_{n+1}' \, F_1(X_1, X_2, \ldots, X_{n+1}) \\ X_2' &= X_{n+1}' \, F_2(X_1, X_2, \ldots, X_{n+1}) \\ \vdots &= \vdots \\ X_n' &= X_{n+1}' \, F_n(X_1, X_2, \ldots, X_{n+1}) \end{aligned}$$

From here, we define X_{n+1}' to be the lowest common multiple of the denominator of each F_i, as this is the smallest factor needed to clear each denominator. This process is called homogenising the map L.

We shall illustrate this process thoroughly on a simple example.

Example 3.3. Consider the map of $\mathbb{Q}^2$ given by

$$\begin{aligned} L : x' &= y \\ y' &= \frac{2y^3 - x(y^4 - 1)}{y^4 - 1 + 2xy} \end{aligned}$$

We make the substitutions $x = \frac{X}{Z}, y = \frac{Y}{Z}$ to get

$$
\begin{aligned}
X' &= Z'\frac{Y}{Z} \\
Y' &= Z'\frac{2\frac{Y^3}{Z^3} - \frac{X}{Z}(\frac{Y^4}{Z^4} - 1)}{\frac{Y^4}{Z^4} - 1 + 2\frac{X}{Z}\frac{Y}{Z}}
\end{aligned}
$$

We first simplify the second ordinate to get

$$
Y' = Z'\frac{2Y^3Z^2 - XY^4 - XZ^4}{Z(Y^4 - Z^4 + 2XYZ^2)}
$$

Now we define $Z' = Z(Y^4 - Z^4 + 2XYZ^2)$ as this is the lowest common multiple of the two denominators in each ordinate in the map L. This finally leaves us with a map we shall denote L_H

$$
\begin{aligned}
X' &= (Y^4 - Z^4 + 2XYZ^2)Y \\
Y' &= (2Y^3Z^2 - XY^4 - XZ^4) \\
Z' &= Z(Y^4 - Z^4 + 2XYZ^2)
\end{aligned}
$$

Note that the singularities of the map are now all contained within the zeros of the added ordinate Z'. The affine singularities are in the factor $Y^4 - Z^4 + 2XYZ^2$, which all come from the original y'. The first component of the map, x', has no affine singularities but does have a projective singularity at the line at infinity. This gives the factor of Z in Z'. We check now, numerically, that such a procedure has done what we hope to L.

- $L(2, 3) = (3, \frac{-53}{46})$
- $L_H([2, 3, 1]) = [276, -106, 92] = [3, \frac{-53}{46}, 1]$
- $L(0, 1) = (1, \infty)$
- $L_H([0, 1, 1]) = [0, 2, 0] = [0, 1, 0]$
- $L_H([X, Y, 0]) = [Y^5, -XY^4, 0] = [Y, -X, 0]$

We can see that the projective version of the map behaves the same as the affine version of the map on points that remain confined to the affine part of the plane, while those points that satisfy $Y^4 - Z^4 + 2XYZ^2 = 0$ or $Z = 0$ are mapped to the line at infinity.

3.2 A History of Integrable maps

The notion of integrability for planar, time-discrete maps is of prime importance for the subject of this thesis. As such it is appropriate to describe what one could call the evolution of the theory of discrete integrable systems, which, in recent years has had somewhat of a renaissance. The idea of integrability is inherited from continuous dynamical systems (i.e. systems whose behaviour is governed by differential equations). In continuous systems one looks at the properties that a solution has, by analogy we study the properties that orbits have for discrete systems. We will look at the history of the discovery of maps that we can consider as integrable under our definition followed by more recent work done in detecting integrability by looking at various features of orbits of maps.

The definition of integrability we will use here will be one strong enough that any map satisfying it will be considered universally to be integrable. It has been used previously and with a discussion of greater depth in [53]. The major feature of an integrable map will be that it leaves many curves fixed. Many curves in this case means a foliation as defined (in the simple two dimensional cases with which we are concerned) by:

Definition 3.4. *Let $\mathcal{C} = \{C_\alpha\}$ where each $C_\alpha : C(x, y, \alpha) = 0$ is an equation for an algebraic curve defined over $\mathbb{C}$. Then $\mathcal{C}$ is a foliation of the projective plane $P(K^2)$ if all but a finite number of points $[X, Y, Z]$ lie on exactly one C_α. Furthermore, any point that lies on more than one C_α must lie on each C_α; they are called base-points.*

For the complete technical definition of foliation as it relates to studying manifolds see, for example, the review of the topic [40]. A map is said to preserve a foliation if the image of every curve in the foliation is also a curve in the foliation. Of more immediate interest however is the stricter condition that a map leaves fixed every curve in a foliation. Since there is no shorthand for this property, we shall just say that the map leaves fixed every curve in the foliation. For terminology, we can usually solve the generic equation for the curves in the foliation C_α for the parameter α, so we are left with each curve written in the form $\alpha = I(x, y)$. We call $I(x, y)$ the integral of the map L and we call each individual curve $I(x, y) = \alpha$ a level set of the integral. In such a case, α is called the height of that particular level set. Curves are one of a few examples of invariant sets for maps. Other examples include attractors and even the full orbit of any point.

A secondary requirement, considering the algebraic geometric approach of this thesis, for a map to be integrable is that it is measure preserving. The simplest form of measure preservation is area preservation.

Definition 3.5. *A map L defined by $(x', y') = (f(x,y), g(x,y))$ is said to be area preserving if the determinant of the Jacobian matrix of L*

$$J(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} \tag{3.1}$$

has absolute value equal to 1 for all x and y.

More intuitively, a map is area preserving if, when the image of a portion of the plane is taken under the map, the resulting set has the same area. The generalisation of area preservation is measure preservation.

Definition 3.6. *A map L defined by $(x', y') = (f(x,y), g(x,y))$ is said to be measure preserving if the determinant of the Jacobian matrix of L satisfies*

$$J = \pm \frac{m(x,y)}{m(x',y')}$$

for some positive and smooth function $m(x,y)$. In such a case, $m(x,y)$ is called the density.

The definition of measure preservation extends in the natural way to higher dimension. With measure preservation and foliation defined, we can give a definition of integrability.

Definition 3.7. *An planar, invertible map L with meromorphic ordinate functions is integrable if L leaves fixed every curve in a foliation and is measure-preserving for some density m.*

Using these two properties as the indicator of integrability directs us first to a paper of E. McMillan, a Nobel laureate in physics. Before we review the maps from this paper however, we consider a precursor of sorts. While planar maps written in the style of this thesis were not so frequently considered before McMillan, recurrence relations had been. Coupled with the fact that nth order recurrence relations can be rewritten as (particularly simple in all but one variable) maps of K^n, this means that certain recurrence relations considered by Lyness and others can be considered integrable.

In [41] we can see Lyness arriving at an integrable dynamical system starting from a completely different question. Beginning with a number theoretic problem of finding three integers such that the sum or difference of any two of them gives a square number, Lyness found what he called a "5-cycle" corresponding to the equation

$$u_{n+1}u_{n-1} = a(u_n + a). \tag{3.2}$$

Using the substitutions $u_{n-1} = x, u_n = y$ this gives the following planar map

$$\begin{aligned} L : x' &= y \\ y' &= \frac{a(y+a)}{x}. \end{aligned} \tag{3.3}$$

This map has the property that $L^5(x, y) = (x, y)$ for every point (x, y), thus the original recurrence relation is called a 5-cycle. Specialising to $a = 1$, it is noted that a consequence of the pattern of iterates given by the map L (or rather the corresponding recurrence relation; for our purposes we shall use the equivalent planar map), that pattern being

$$\begin{aligned} &(x,y) \xrightarrow{L} (y, \frac{y+1}{x}) \xrightarrow{L} (\frac{y+1}{x}, \frac{x+y+1}{xy}) \xrightarrow{L} (\frac{x+y+1}{xy}, \frac{x+1}{y}) \xrightarrow{L} \\ &(\frac{x+1}{y}, x) \xrightarrow{L} (x,y), \end{aligned} \tag{3.4}$$

is that if any point (x_0, y_0) lies on the curve with equation

$$(x+1)(y+1)(x+y+1) = kxy$$

then the point $L(x_0, y_0)$ will also lie on that curve. This is the case because upon solving the above expression for k we see that the expression is actually just the product $xx'x''x'''x''''$. This product, upon substituting $x = x_0, y = y_0$ is clearly invariant under the iteration denoted by $'$. The value of k is determined by the particular (x_0, y_0) pair. This, of course, is precisely saying that the quantity given by

$$k = I(x,y) = \frac{(x+1)(y+1)(x+y+1)}{xy}$$

is an invariant (or an integral) for the map L. This map is, in modern terms, perhaps less interesting than most because it is finite order throughout the whole plane. Nevertheless, the way at which it was arrived and the fact that it is one of the first maps noted (if indirectly) for its possession of an integral makes it noteworthy.

While the work of Lyness and his contemporaries who also remarked on these so called cycles did not receive much attention historically (but, see [33] and [34]), another work with roots in another area did. The physicist McMillan published a paper designed to answer a question in the stability of dynamical systems [45]. The question itself was spawned (for McMillan at least) from works of people considering the stability of the solar system, and the stability, or lack of, in systems coming from storage rings. The exact question McMillan sought to answer was “Is there a non-linear area preserving transformation with a finite region of guaranteed stability?”. Guaranteed stability in this context is a term used to describe regions of the plane whose orbits remain confined to some finite area. To relate this to a map possessing an invariant curve we must consider when a map has a closed invariant curve. In such a case, either the interior of this curve must remain within the interior of the curve or the interior of the curve must be mapped to the exterior of the curve. This is so due to the assumed continuity of the map. However if it is known that there is also a fixed point of the map in the interior of the curve, then we are left immediately with a finite region of guaranteed stability - the interior of the closed invariant curve. With this planned method for answering his question, McMillan considered the form of planar map

$$\begin{aligned} x' &= y \\ y' &= -x + f(y) \end{aligned} \tag{3.5}$$

which has inverse

$$\begin{aligned} x' &= -y + f(x) \\ y' &= x. \end{aligned} \tag{3.6}$$

A map of the form (3.5) is automatically area preserving so the task remaining is to ensure it has a closed invariant curve. McMillan does this by supposing that a curve with equation $x = \phi(y)$ is invariant, meaning that $x' = \phi(y')$ and $\phi^{-1}(x') = y'$, assuming an inverse exists. Substituting this into the second coordinate of equation (3.5) gives

$$\phi^{-1}(x') = -\phi(y) + f(y). \tag{3.7}$$

The first coordinate of the map lets us replace x' by y resulting in

$$f(y) = \phi(y) + \phi^{-1}(y). \tag{3.8}$$

Before constructing examples however an approach that will become important later in this thesis is taken; the map (3.5) is split as $L = G \circ H$ into two involutions where

$$\begin{aligned} H : x' &= y \\ y' &= x \\ G : x' &= x \\ y' &= -y + f(x). \end{aligned} \tag{3.9}$$

Now constructing a curve which is invariant under both of these (and the first is easy; just ensure the curve remains the same when we change x and y) will give a curve invariant under the original map which is the composition $G \circ H$. It is harder to sort out what kind of curve is invariant under G but McMillan notes in his paper that choosing curves which are quadratic in y give the simplest cases that had not already been studied. This leaves the map

$$\begin{aligned} x' &= y \\ y' &= -x + (-\frac{By^2 + Dy}{Ay^2 + By + C}) \end{aligned} \tag{3.10}$$

which has invariant curve

$$Ax^2y^2 + B(x^2y + xy^2) + C(x^2 + y^2) + Dxy = constant. \tag{3.11}$$

Note that the constant in equation (3.11) does not turn up in the map (3.10); this is because the map has each such curve as an invariant.

While the McMillan family of curve preserving maps was a large advance in the theory of discrete dynamical systems, it was expanded upon by Quispel, Roberts and Thompson when they constructed their QRT family of maps in [50] and [51].

These maps expanded upon the McMillan family quite directly. Consider a general biquadratic

$$\begin{aligned} B(x,y,t) =&\alpha(t)x^2y^2 + \beta(t)x^2y + \delta(t)xy^2 + \gamma(t)x^2 + \kappa(t)y^2 + \epsilon(t)xy + \\ &\xi(t)x + \lambda(t)y + \mu(t). \end{aligned} \tag{3.12}$$

To make the family of curves defined by $B(x,y,t) = 0$ foliate the plane, we must ensure any particular pair (x_0, y_0) satisfies $B(x_0, y_0, t_0) = 0$ for exactly one value of t_0. One way to be certain of this is by having each Greek letter at most linear in t. This was the path that Quispel, Roberts and Thompson took, though later work by Iatrou and Roberts ([27], [28] and [29]) looks at ways of relaxing this linearity requirement while still maintaining the property that each point in the plane lies on exactly one level set. The method of creating the QRT maps that will preserve each of the biquadratics is quite ingenious in its simplicity. Given any point (x,y) one simple operation which will map the point to another point on the same biquadratic is to "move horizontally until we hit the same curve again". A second operation is to "move vertically until we hit the same curve again". Since the curves in question are biquadratics, these operations are well-defined - consider the number of intersections between a (vertical) line and a biquadratic. This amounts to fixing $x = x_0$ and then solving the resulting biquadratic for y, giving at most two distinct values. One of these values corresponds to the input point, the other to the image point under our operation of moving vertically. Of course Bezout's Theorem tells us there are four projective points of intersection; the other two are on the line at infinity and thus they are easily distinguished and ignored for the purposes of the operation. Since there are only two such affine points of intersection between a line and a biquadratic, it is clear that composing these operations with themselves yields the identity map. However it is when we compose one with the other that we will most likely come up with an interesting map. This construction is the essence of the QRT family of maps. With this reasoning behind the construction, we can go through the algebra to find the form of the QRT maps. Consider the general biquadratic given by equation (3.12). Suppose we begin at a point (x_0, y_0) which must lie on $B(x,y,t)$ for some value of t; we fix $x = x_0$ so that we consider this purely as a quadratic in y (thus we are moving vertically here). The result is

$$\begin{aligned}B(x_0, y, t) =&y^2(\alpha(t)x_0^2 + \delta(t)x_0 + \kappa(t)) + y(\beta(t)x_0^2 + \epsilon(t)x_0 + \lambda(t))+\\ &(\gamma(t)x_0^2 + \xi(t)x_0 + \mu(t)) = 0.\end{aligned}$$

Now we know that $y = y_0$ is one solution to this quadratic and using the sum of roots formula the other is $y' = -y_0 - \frac{\beta(t)x_0^2+\epsilon(t)x_0+\lambda(t)}{\alpha(t)x_0^2+\delta(t)x_0+\kappa}$. Thus the map to move vertically on a biquadratic is

$$\begin{aligned}G : x' &= x\\ y' &= -y - \frac{\beta(t)x^2 + \epsilon(t)x + \lambda(t)}{\alpha(t)x^2 + \delta(t)x + \kappa(t)}\end{aligned}$$

and through similar reasoning the map to move horizontally is

$$\begin{aligned}H : x' &= -x - \frac{\delta(t)y^2 + \epsilon(t)y + \xi(t)}{\alpha(t)y^2 + \beta(t)y + \gamma(t)}\\ y' &= y.\end{aligned}$$

While composing these two maps in either order gives a (generally) infinite order map, the QRT maps were defined by $L = G \circ H$ and reversing the order gives L^{-1}. Written as is, this map L is only a true map of the plane as long as you pre-calculate the correct value of t, according to the biquadratic $B(x, y, t)$, for the point (x, y) with which one is working. To avoid this, one can substitute for t in terms of x and y using equation (3.12), which works so long as each of the Greek letters in that equation are at most linear in t. Assuming that this is the case, and that $\alpha = \alpha_0 + \alpha_1 t$, and so on up to $\mu = \mu_0 + \mu_1 t$ then solving for t gives

$$t = -\frac{\alpha_0x^2y^2 + \beta_0x^2y + \delta_0xy^2 + \gamma_0x^2 + \kappa_0y^2 + \epsilon_0xy + \xi_0x + \lambda_0y + \mu_0}{\alpha_1x^2y^2 + \beta_1x^2y + \delta_1xy^2 + \gamma_1x^2 + \kappa_1y^2 + \epsilon_1xy + \xi_1x + \lambda_1y + \mu_1}.$$

Substituting this back into $L = G \circ H$ gives a rather complicated map, which is most succinctly written in the following "matrix form".

$$\begin{aligned}L : x' &= \frac{f_1(y) - xf_2(y)}{f_2(y) - xf_3(y)}\\ y' &= \frac{g_1(x') - yg_2(x')}{g_2(x') - yg_3(x')}\end{aligned} \tag{3.13}$$

where the polynomials f_i and g_i are each quartics defined by

$$\begin{pmatrix} f_1(z)\\ f_2(z)\\ f_3(z)\end{pmatrix} = \begin{pmatrix}\alpha_0z^2 + \beta_0z + \gamma_0\\ \delta_0z^2 + \epsilon_0z + \xi_0\\ \kappa_0z^2 + \lambda_0z + \mu_0\end{pmatrix} \times \begin{pmatrix}\alpha_1z^2 + \beta_1z + \gamma_1\\ \delta_1z^2 + \epsilon_1z + \xi_1\\ \kappa_1z^2 + \lambda_1z + \mu_1\end{pmatrix} \tag{3.14}$$

and

$$\begin{pmatrix} g_1(z) \\ g_2(z) \\ g_3(z) \end{pmatrix} = \begin{pmatrix} \alpha_0 z^2 + \delta_0 z + \kappa_0 \\ \beta_0 z^2 + \epsilon_0 z + \lambda_0 \\ \gamma_0 z^2 + \xi_0 z + \mu_0 \end{pmatrix} \times \begin{pmatrix} \alpha_1 z^2 + \delta_1 z + \kappa_1 \\ \beta_1 z^2 + \epsilon_1 z + \lambda_1 \\ \gamma_1 z^2 + \xi_1 z + \mu_1 \end{pmatrix}. \tag{3.15}$$

However, it is computationally advantageous to leave L in its "t-dependent form" when generating phase spaces. At the beginning of each orbit, one calculates the value of t for that orbit using the initial condition, then continues to use that value of t for the rest of the orbit. More shall be said on the advantages in leaving maps such as the QRT maps in a t-dependent form later when we use them frequently. In addition to leaving fixed every curve in a foliation, the QRT maps are also measure-preserving. The density that they preserve is

$$m(x,y) = (\alpha_1 x^2 y^2 + \beta_1 x^2 y + \delta_1 x y^2 + \gamma_1 x^2 + \kappa_1 y^2 + \epsilon_1 xy + \xi_1 x + \lambda_1 y + \mu_1)^{-1}$$

This means that they fit into the definition of planar integrability of definition (3.7). The QRT family of maps has been important as a source to draw upon for examples when testing various conjectures about integrability. Furthermore it is still not known whether they are (up to birational conjugacy) the only integrable planar maps though examples have lately been showing up in the literature which are, to current knowledge, unrelated to any QRT map[1].

One planar map that arose relatively recently in literature which is integrable but which may not be a QRT map in disguise (i.e. birationally equivalent to a QRT map) was first created as a recurrence relation in [32] and studied further in [68]. This map is defined by

$$\begin{aligned} x' &= y \\ y' &= -\frac{-y^2xap^2 + xa + ap^2y - y^3a + y^2 - 1 + y^2a^2 - a^2}{a(p^2y^2-1)(yx-1)} \end{aligned} \tag{3.16}$$

with a, p free parameters and it preserves a family of biquartic curves defined by

$$((x-y)^2 - p^2(xy-1)^2)((x+y-b)^2 - p^2(xy-1)^2) - K(xy-1)^2 = 0 \tag{3.17}$$

where $b = a + \frac{1}{a}$ and K is the foliating parameter.

[1] However, a preprint of a text by Duistermaat [14] uses algebraic geometric techniques beyond the scope of this thesis to suggest that all maps of "rational elliptic surfaces" are, by a series of blow ups at base points, related to QRT maps.

A second example comes from [31] where the authors manipulate integrable lattice equations into integrable maps. Of the three examples given in that paper which are denoted as possibly being non-QRT in nature, two examples preserve curves of genus zero (not noted in the paper, but easy to calculate) while only one preserves curves of genus one. The algebraic form of the map preserving elliptic curves is impractical to reproduce on paper and its formula is given in terms of the group law on elliptic curves. The context in which this example is embedded (we consider the genus zero examples to be of considerably less interest, as crafting non-QRT maps in this case is a little easier; see the section on maps of conics in chapter 2) is interesting as it ties in quite tightly with the later topic of mixing in chapter 5.

3.3 Testing for Integrability

With the idea that integrability of a map implies that the map is stable or predictable (consider McMillan's original problem of a finite region of guaranteed stability, and how this resulted in an integrable map), the ability to discern whether a map is integrable or not may become of some import in certain situations. The devising of integrability tests did not begin until after the construction of the QRT family of maps with QRT maps first surfacing in 1988 and the first test for integrability surfacing in 1991. In addition to any other concerns this was at least because to construct a test for integrability, some range of integrable maps is required to apply one's test to. We shall here review four different methods that can be used to test for integrability. In chronological order of their appearance they are checking for "singularity confinement", calculating the "algebraic entropy" of the map, reducing the orbits of the map over a finite field and lastly checking for "Diophantine integrability"[2].

The notion of singularity confinement is one directly inherited from the study of differential equations. For partial differential equations, the so-called Painlevé Criterion allows integrability to be quite reliably tested for by checking the singularity structure of the equation [20]. The construction of such a check for difference

[2] These four methods lend themselves well to algorithmic and computational testing for integrability. Some different approaches include those described in [52] and [1].

equations was the goal in [20]. In this paper, a map is said to satisfy singularity confinement if (verbatim):

The movable singularities of integrable maps are confined, i.e., they are canceled out after a finite number of steps. Moreover, the memory of the initial condition is not lost whenever a singularity is crossed.

After this definition attention is given over to examples. We shall follow the same approach here, applying the singularity confinement method to two maps; one QRT map that is integrable and one perturbed QRT map that is not.

Example 3.8. The QRT map we consider is one of the so-called symmetric QRT maps. These are thus named because rather than having the two usual involutions G and H as above, H is instead the simple involution that interchanges x and y. The other involution G remains the same, yielding the map

$$\begin{aligned} L : x' &= y \\ y' &= -x - \frac{\beta(t)y^2 + \epsilon(t)y + \lambda(t)}{\alpha(t)y^2 + \delta(t)y + \kappa(t)}. \end{aligned} \tag{3.18}$$

The fact that H is the interchange of x and y means that the biquadratic that L preserves, being invariant under this interchange, is symmetric in x and y. The parameters we choose to illustrate singularity confinement are $\kappa = 0, \alpha = 1, \delta = 1, \beta = 1, \epsilon = 3, \lambda = 1$ which leaves

$$\begin{aligned} L : x' &= y \\ y' &= -x - \frac{y^2 + 3y + 1}{y(y+1)}. \end{aligned} \tag{3.19}$$

Clearly the map (3.19) has singularities when $y = 0$ and $y = -1$. To check if these singularities are confined, we iterate the point (x, η) and check at each successive iterate if substituting $\eta = 0$ and $\eta = -1$ gives an affine answer which also depends on x. The affine answer tells us that the singularity is confined, but the dependence on x tells us that the initial condition is not being lost as we cross the singularity. In this case, the singularity is confined by the third power; $L^3(x, 0) = (-1, -1 - x)$ and $L^3(x, -1) = (0, -1 - x)$.

However, now we change the map slightly while retaining area preservation to

$$\begin{aligned} L : x' &= y \\ y' &= -x - \frac{y^3 + y^2 + 3y + 1}{y(y+1)}. \end{aligned} \tag{3.20}$$

In this case we find out that even going up to the seventh power, neither singularity is ever confined i.e. we never return to an affine point. While this does not tell us that the singularity is never confined, the method remains a simple first test for integrability driven more by the fact that it gets results than rigour. Indeed, in [23] Hietarinta and Viallet exhibit a map that satisfies singularity confinement but is known not to be integrable.

The roots of the notion of algebraic entropy started in 1990 when V.I. Arnold introduced the dynamical complexity of a homeomorphism in [3]. From then on, throughout the 1990s, a link was noticed between this dynamical complexity of the map and the degree of the nth power of the map. These ideas were made rigourous with the definition of algebraic entropy in 1999 by M.P. Bellon and C.-M. Viallet [10]. Defining algebraic entropy first of all requires working over projective space. Recall that rational maps in n variables can be converted to polynomial maps in $n+1$ variables by the process of homogenisation outlined in definition (3.2) and the example following. Furthermore this process leaves us with a polynomial map with a single degree d common in each coordinate. Thus it is natural that when composed with itself, the nth iterate of such a map has degree d^n. The observation that the introduction of algebraic entropy sought to explain was that for maps that had reason to be thought integrable the degree grew not exponentially but polynomially. To define algebraic entropy properly, we introduce the notation that d_n is the degree of the nth iterate of some map ϕ after it has been reduced. Reduced in this context means factors common in each coordinate of ϕ^n are removed (in projective space, removal of common factors is allowed). Then the definition of algebraic entropy is as follows.

Definition 3.9. *Let ϕ be a rational map whose nth power has degree d_n. Then the algebraic entropy of ϕ is $\lim_{n\to\infty} \frac{\log d_n}{n}$.*

The authors note in [10] that the proof of the existence of this limit is a straightforward consequence of the inequality $d_{n_1+n_2} \leq d_{n_1} d_{n_2}$ and go on to mention that in their research the sequence of degrees itself always satisfied a finite linear recurrence relation with integer coefficients, though a proof of this fact is as yet unknown. Nevertheless this observation allows the (probable) algebraic entropy of a map to be calculated without checking symbolically how the degree of the map behaves at high iterates. With this definition, it is clear that any map whose sequence of degrees grows polynomially will possess an algebraic entropy of zero while any map whose sequence of degrees grows exponentially will possess a non-zero algebraic entropy. The final piece that would allow us to separate integrable maps from non-integrable would be to prove that maps considered integrable through other criteria do in fact have sequences of degrees that grow only polynomially. This is conjectured in the conclusion of [10]. In the same year, Bellon proved that for maps that foliate the plane with invariant curves this is indeed the case and hence that for maps that leave fixed every curve in a foliation the algebraic entropy is zero [9].

The next development in the detection of integrability came with the advent of a purely numerical test care of J.A.G. Roberts and F. Vivaldi in 2003 [60]. This test was part of a larger scheme by the authors whose purpose was to find signatures of properties (one of them happening to be integrability) of maps by considering the actions of these maps over finite fields. The other paper in this series looks at detecting time-reversal symmetry [61]. Due to the importance of this work in developing the original work documented in this thesis we shall reserve this method of integrability detection for a more thorough examination in the next section.

The most recent, and perhaps simplest, proposed method of integrability detection comes care of Halburd in [21]. This test hinges on noticing that the heights of rational orbits of integrable systems grow more slowly than heights of rational orbits of non-integrable systems. The height here is the usual quantity when dealing with rational numbers:

Definition 3.10. *Let $x = \frac{p}{q}$ be a rational number with p and q having no common factors. Then the height of x is*

$$H(x) = max\{|p|, |q|\} \tag{3.21}$$

and $H(0) = 1$.

We can extend this definition to define the height of an orbit in the obvious way.

Definition 3.11. *Let $O = \{(x_1, y_1), (x_2, y_2), \ldots\}$ be an orbit composed of rational numbers. Then the height of O at the nth iterate is*

$$H(O_n) = max\{H(x_n), H(y_n)\}. \tag{3.22}$$

If needs be, one can also extend these definitions to projective spaces by including the height of the third homogenising coordinate also. With these definitions in mind, Halburd defines a map to be Diophantine integrable if the sequence $\{h(O_n)\} = \{\log(H(O_n))\}$ grows no faster than a polynomial in n for each orbit O. The author notes that this is related to, and possibly equivalent to, the notion of algebraic entropy described earlier. However, it is easier to numerically check for Diophantine integrability than it is to check for zero algebraic entropy. Let us refer back to the same examples used in our discussion of singularity confinement (example (3.8)) and check how they fit into the Diophantine integrability scheme.

Example 3.12. Let L be the map from equation (3.19). Let our initial condition be $P = (\frac{2}{3}, \frac{-1}{2})$. Then $H(O_1) = 3$ and the sequence $h(O_n)$ under L begins as (to one decimal place)

$$\{1.1, 1.6, 2.1, 5.7, 8.6, 13.4, 18.7 \ldots\}$$

However if we use the same initial condition and this time use the map L_2 from equation (3.20) the same sequence under L_2 begins as (to one decimal place)

$$\{1.1, 2.6, 7.2, 19.6, 53.0, 144.0, 392.2, \ldots\}$$

The difference here is quite remarkable, and very easily calculated.

3.3.1 Maps over Finite Fields

Here we shall be giving an in-depth (compared to the overviews of other forms of integrability detection already looked at) description of the test first conceived by Roberts and Vivaldi in [60] and further developed by Roberts, Jogia, and Vivaldi in [57], the latter paper being original work included in this thesis. The test requires us to reduce a map to an action on a finite phase space. This has the immediate consequence of restricting us to rational maps with rational coefficients since we cannot guarantee the existence of, for example, $\sqrt{2}$ in a finite field. It remains true that any algebraic number can be expressed over infinitely many finite fields but for the sake of simplicity and generality we will restrict ourselves to rational maps with rational coefficients. Furthermore we require that a map be "representable" over any particular finite field in which we may work; this is taken to mean that there are no denominators in the map divisible by the order of the finite field.

To describe the integrability test, first we shall follow a straightforward argument originally used by Veselov in [67]. Suppose L is an infinite order birational map of the plane which leaves fixed some curve C. Then by definition, L is an automorphism of C and $\{L^n | n \in \mathbb{Z}\}$ is a subgroup of $\mathrm{Aut}(C)$. Since L is infinite order, the group generated by L is infinite order and so $\mathrm{Aut}(C)$ must also be infinite order. Since C has an infinite automorphism group, by Hurwitz' theorem (theorem (2.51)) C must be genus 0 or 1 or singular. This argument explains why the theory of elliptic curves is so important for integrable planar maps. Keeping in mind the fact that the largest genus curves we can encounter as being preserved by infinite order maps is genus one, recall the Hasse-Weil bound from theorem (2.58). The important inequality is that for any (irreducible) curve C with genus g defined over a finite field of order p

$$p + 1 - 2g\sqrt{p} \leq |C| \leq p + 1 + 2g\sqrt{p}$$

where $|C|$ denotes the number of projective points on the curve with coordinates in F_p. In particular for $g = 0$ and $g = 1$ we get the special cases, respectively

$$|C| = p + 1$$

$$p + 1 - 2\sqrt{p} \leq |C| \leq p + 1 + 2\sqrt{p}$$

and taking the least sharp bound of the two scenarios leaves the second equation. To incorporate a map into this theory, suppose that L is an infinite order map that leaves the curve C fixed. Then taking any point $P \in C$ and generating the orbit of P gives a set of points each of which lie on C. Reducing the map L, the curve C and the orbit of P down to some finite field $\mathbb{Z}_p$ for a prime p for which each of those three objects is still defined will not change this situation[3]; the reduced orbit will still lie on the reduced curve, and the reduced map will still generate the reduced

[3] This is due to the fact that reduction modulo p is a field homomorphism and curves are defined by field operations.

orbit. From the Hasse-Weil theorem we have an upper bound on the size of this orbit - its length is at most $p + 1 + 2\sqrt{p}$. To frame this in an alternative way, if we have an orbit O generated by an infinite order map M with reduced versions $\widetilde{O}$ and $\widetilde{M}$ modulo p with the property that $|\widetilde{O}| > p + 1 + 2\sqrt{p}$ then we know that such an orbit must not lie on a curve of genus 0 or 1 and consequently that O does not lie on such a curve either or if it does then it must be as a component of a reducible curve. Coupling this with the fact that for a map to be integrable we require a foliation of curves to be left fixed, we are left with a negative test for integrability. Note that the lower Hasse-Weil bound does not help this test since a reduced map can partition any curve in a finite field into several small orbits. This test is essentially summed up in theorem 2 of [57], which we reproduce here as theorem (3.13).

Theorem 3.13. *Let L be an infinite order birational map which is representable over the finite field $\mathbb{F}_p$. Let $\mathcal{O}_p(X)$ denote the maximal orbit (see remarks below) of its projective version $\widetilde{L}$ containing a given point $X \in P(\mathbb{Z}_p^2)$ (using L^{-1} when necessary to find the pre-images of X to generate the maximal orbit). If L has a rational integral that is representable over $\mathbb{F}_p$ and the level set C containing X is irreducible then*

$$|\mathcal{O}_p(X)| \leq p + 1 + 2\sqrt{p} + \#C_s \tag{3.23}$$

where $\#C_s$ is the number of singular points on C. Furthermore, if $\mathcal{O}_p(X)$ is a cycle then all its points are either singular points or all non-singular points on the level set C. In the former case, $|\mathcal{O}_p(X)| \leq \#C_s$ and in the latter case $|\mathcal{O}_p(X)| \leq p+1+2\sqrt{p}$.

The proof of theorem (3.13) is mostly a straightforward application of the Hasse-Weil bound to orbits combined with the argument given previously that requires the genus of a level set to be either 0 or 1. The second part stating that orbits consist solely of singular points or solely of non-singular points is a consequence of the chain rule applied to the derivative of X and $L(X)$.

While the heart of the test is quite simple, there are some complications which we will now go through. Firstly we shall discuss how to turn this into a numerical test for integrability. For any given map we need to, in one manner or another, generate sufficient data to draw conclusions. For a negative conclusion we should find a reasonable number of orbits that are too large to obey the Hasse-Weil bound.

One case may not suffice; suppose $I(x, y)$ is an integral for a map L. Thus, the preserved curves are $I(x, y) = k$, and the generic level set is at most genus 1. For some values of k the curve may be reducible. In principle, at least, this could lead to $I(x, y) - k$ factorising into many genus 0 components, which in turn could lead to allowable orbit lengths of any integer multiple of $p + 1$, as the action of the map bounces between each genus 0 factor of $I(x, y) - k$. So, while one abnormally large orbit length does not automatically mean the map fails to preserve a family of curves, each such orbit one finds makes it exceedingly unlikely. Conversely, no matter how many orbit lengths do lie under the Hasse-Weil bound one finds it technically is never enough to conclude that a foliation of curves is left fixed. In practice, however, non-integrable maps rarely have the short orbit property. This becomes especially true as the size of the prime p increases; the differences between integrable and non-integrable become more and more marked. The effectiveness of this method for testing integrability will be demonstrated by performing it and displaying some of the numerics. Other results can be seen in [60] and [57].

The second complication meriting a mention here is that of singularities in finite fields. It is quite common for orbits in a finite field not to close properly. By this we mean that under successive iteration by a map L, the orbit of a point will often end at [0,0,0] which does not exist in projective space. Furthermore, iterating the same point under the map L^{-1} will lead to $[0, 0, 0]$ also. This problem forces us to take some care with how we measure the length of orbits. It is a common situation when working with maps over finite fields that a loss of invertibility occurs at a couple of points; many orbits that start out separate lead into one common point, which in turn trails off to [0,0,0]. Because of this, if we were to consider the finite phase space $P(\mathbb{Z}_p^n)$ as a discrete graph where the vertices are the points in the phase space and there are (directed) edges between points that are the image of one another under L or its inverse, we would wind up with an erroneous picture. Counting the size of the connected components of this graph and comparing this to the Hasse-Weil bound would risk counting multiple orbits as one and flagging a possibly integrable map as non-integrable. Therefore it is wiser, and faster, to instead just pick a point in the finite phase space, iterate this forward under L until either it returns to the same point (in which case the orbit is periodic and its length is obvious) or until it reaches the point [0,0,0]. In this case, one then also iterates the same point backwards under L^{-1} until it too reaches [0,0,0] and then concatenates the two orbits together. After removing the two end points (both [0,0,0]) and any repeated points (see example (3.14) for how this can occur) one can then count the length of the remainder in the obvious way. This is what is meant by “maximal orbit” in theorem (3.13). While all this sounds complicated it is in fact rather simple and can be easily demonstrated by almost any map.

Example 3.14. We return again to the maps we have been using as primary examples in discussing tests for integrability i.e. equations (3.19) and (3.20). Let us start with the first map. Homogenising this map gives

$$\begin{aligned} X' &= Y^2(Y+Z) \\ Y' &= -(XY^2+XYZ+ZY^2+3YZ^2+Z^3) \\ Z' &= ZY(Y+Z) \end{aligned} \tag{3.24}$$

which has inverse

$$\begin{aligned} X' &= -(XYZ+X^2Y+ZX^2+3XZ^2+Z^3) \\ Y' &= X^2(X+Z) \\ Z' &= ZX(X+Z). \end{aligned} \tag{3.25}$$

Taking the randomly chosen point [1,2,3] we find that under forward iteration the following semi-orbit is generated:

$$[1,2,3] \rightarrow [7,1,4] \rightarrow [5,5,7] \rightarrow [1,7,4] \rightarrow [6,3,9] \rightarrow [4,0,12] \rightarrow [0,1,0] \rightarrow$$
$$[1,0,0] \rightarrow [0,0,0]$$

and under backward iteration the following semi-orbit is generated

$$[0,0,0] \leftarrow [0,1,0] \leftarrow [1,0,0] \leftarrow [0,4,12] \leftarrow [1,2,3]$$

By comparing the two different halves of the single orbit, one can see that there is a loss of invertibility when going from [4,0,12] to [0,1,0] as applying the inverse map to [0,1,0] results in [0,0,0]. In terms of counting orbit lengths, there are two ways to deal with this problem. The first is, as already mentioned, to combine the "forwards" orbit and the "backwards" orbit and count that length (correctly). The second is to ignore any orbit which is not periodic. For many maps this will render a large part of the phase space useless and therefore a larger phase space must be taken. However the results are much sharper when considering only periodic orbits.

One thing that all tests for integrability should do is distinguish clearly between "near integrable" and actually integrable. A consequence of the famous KAM theorem (see [2] and [43] among many others for further details) is that small perturbations of real integrable maps will generate phase space portraits that look very similar to the phase space portrait of the initial integrable map. In particular, it will appear that some of the preserved curves have remained untouched and are still preserved by the perturbation. Mathematically, the equations of these curves have altered in such a way that while they may look the same to the naked eye they are, from an algebraic point of view, potentially very different objects. The "growth type" tests mentioned earlier (algebraic entropy and Diophantine integrability) both split these near integrable maps from their integrable counterparts and so does this orbit growth test. Why this occurs is easy to explain conceptually - there is no concept of nearness in finite fields. Therefore one cannot have a small perturbation of an integrable map. Mathematically however the situation is a little more complicated. Certainly the altered curves that remain in the real phase space portrait for the perturbed map are curves in some sense, so why do they not remain seen in the finite field setting? The answer is most likely that they are rarely, if ever, algebraic curves though this seems impossible to check. We close this discussion with some plots demonstrating the test's usefulness in separating integrability and near-integrability. Figure (3.1) shows six images. The left column shows three orbits (for the initial conditions $(1,0),(1,1)$ and $(1,2)$) for three maps each from the same family, that being equation (3.26).

$$\begin{aligned} x' &= -x - \frac{\delta y^2 + \epsilon y + \xi}{y^2+1} \\ y' &= -y - \frac{\epsilon x' + \lambda}{x'^2+1}. \end{aligned} \tag{3.26}$$

The top phase space portrait has $\delta = 0$, the middle has $\delta = 10^{-4}$ and the bottom has $\delta = 1$. All three have $\epsilon = \frac{2}{3}, \xi = \frac{1}{5}, \lambda = -4$. For this family of maps, δ is the parameter which controls integrability. When $\delta = 0$, the map is a QRT map and hence integrable, moving δ away from 0 takes the map further from integrability. The right column shows the results of applying the simple orbit length test for integrability. The plots here are the normalised (by dividing by the upper Hasse-Weil bound) orbit lengths of the point $[1, 1, 1]$ in $P(\mathbb{Z}_p^2)$ as they vary with prime p up to $p = 5000$. Here we have taken the orbit to mean both the forward orbit and the backward orbit concatenated together and, as such, both periodic orbits and aperiodic orbits are included.

As the test for integrability is asymptotic in nature, we wish to consider the maps and corresponding orbits over $\mathbb{Z}_p$ for large values of p. At this point we have two options to gather large amounts of data. Either we can fix a large value of p and collect data pertaining to the entire phase space. This entails finding the orbits for $p^2 + p + 1$ projective points (though of course not all of these are used as an initial condition; each orbit will generally consume more than one point in the phase space) which is quite laborious. The alternative is to fix one initial condition and generate the orbits for this initial condition for many values of p. Naturally this allows us to take much larger values of p and, perhaps surprisingly, generates similar statistical distributions when the appropriate data is plotted. We shall show some plots of data gathered using both techniques in chapter 4 but for computational simplicity, this latter approach is the one we shall usually follow in numerical undertakings. Both the acknowledgement that this test generally distinguishes between integrable and near-integrable and the use of the Monte-Carlo method with fixed initial condition and varied prime were original ideas first seen in [57]. Further discussion on this Monte-Carlo method can be found in section 4.2.3.

3.4 Some Interesting Relations Between maps

In this part we describe several ways of historical import that maps can be related to one another. In later chapters we will discuss how the original material of this thesis can be used to give us information about the relations defined in this part.

The first relation between maps that we wish to introduce almost needs no introduction. It is almost the simplest relation possible, that of conjugacy. Since later on we will require that our maps be birational, we impose that restriction here also.

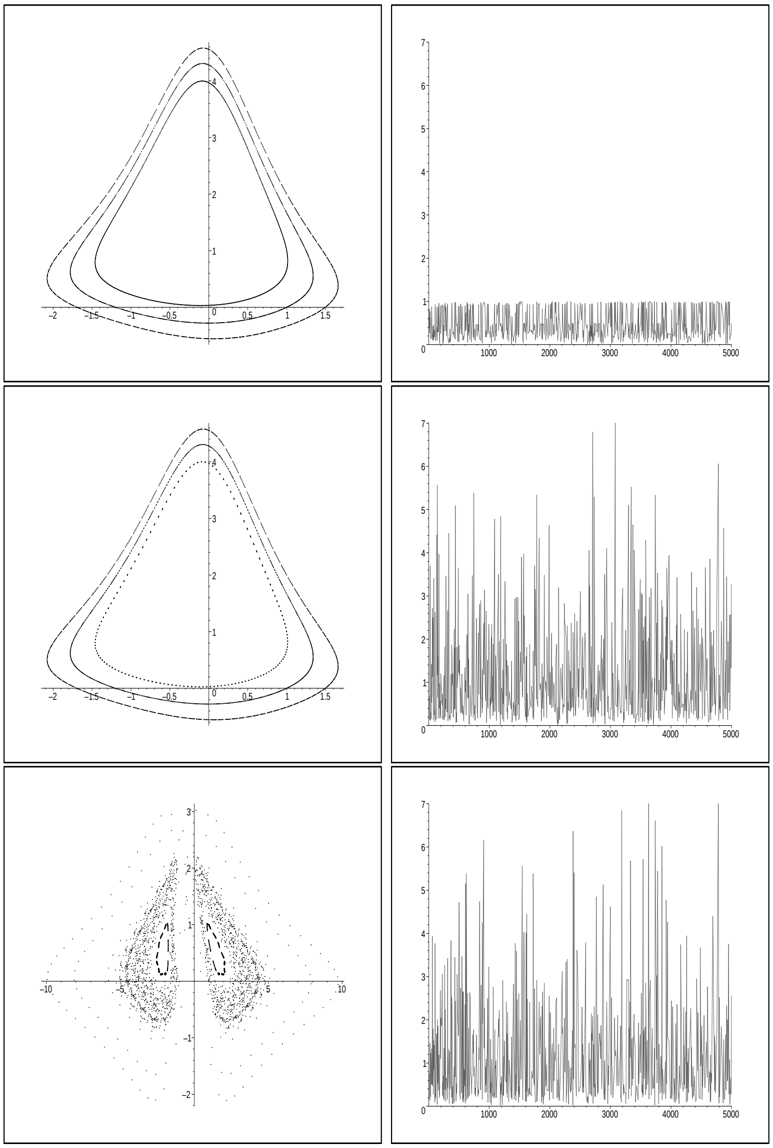

Figure 3.1: Evidence of the efficacy of the Hasse-Weil Bound based test from [60] and [57] to distinguishing near-integrability and integrability. The phase space portraits are for the map (3.26) with parameter values $(\epsilon, \xi, \lambda) = (\frac{2}{3}, \frac{1}{5}, -4)$ with $\delta = 0, 10^{-4}, 1$ from top to bottom. The parameter δ controls the integrability of the map with the top being integrable, the middle being near-integrable and the bottom being far from integrable. Notice that the phase space portraits of the integrable and near-integrable are similar while the orbit length plots (shown on the right column with normalised orbit length being plotted against prime) for the same are very different.

Definition 3.15. *Let L and M be two birational maps of the same space. They are said to be conjugate to one another if there exists a birational map ϕ such that*

$$L = \phi M \phi^{-1}$$

Interest in conjugate maps comes from two places; firstly the relation is treatable with the algebraic-geometric approach taken in this thesis and secondly it is a common construct in basic dynamics. A second relation which also fits these criteria is that of maps being power-related.

Definition 3.16. *Let L and M be two birational maps of the same space. Such maps will be called power-related if there exists integers l and m such that*

$$L^l = M^m.$$

Both conjugacy and power-relation are easily and quickly dealt with in later chapters, the last two relations we deal with are of greater substance. We define now symmetries and reversors of a map L and briefly give some of their basic properties which will be of use later. We define the space on which L acts fairly generally, as the relations are able to be defined in this generality but one can think of L as being a map of the plane.

Definition 3.17. *Let L be an invertible map of a set. Then define the set*

$$\mathcal{S}(L) = \{S : S \textit{ is an invertible map of the same set and } S \circ L = L \circ S\}.$$

We call the elements of $\mathcal{S}(L)$ symmetries of L.

Definition 3.18. *Let L be an invertible map of a set. The define the set*

$$\mathcal{R}(L) = \{R : R \textit{ is an invertible map of the same set and } R \circ L = L^{\pm 1} \circ R\}.$$

The elements of this set that are not symmetries are called reversors of L.

Proposition 3.19. *Let L be an invertible map of some set. Then $\mathcal{R}(L)$ is a group (called the reversing symmetry group of L) and $\mathcal{S}(L)$ is a subgroup (called the symmetry group of L) of $\mathcal{R}(L)$.*

Proof. To prove that $\mathcal{R}(L)$ is a group, we shall defer the calculations until the proof of proposition (3.20). To prove that $\mathcal{S}(L)$ is a subgroup, let S_1 and S_2 be symmetries of L. Then $S_1S_2L = S_1LS_2 = LS_1S_2$, so the product of two symmetries is also a symmetry. For inverses, $S_1L = LS_1 \Leftrightarrow LS_1^{-1} = S_1^{-1}L$. □

It is often desirable to make a distinction between a symmetry of a map and the other kind of maps inside the reversing symmetry group (i.e. those with $LR = RL^{-1}$). We will call these latter maps reversors. With this terminology, the following proposition becomes a lot easier to dictate.

Proposition 3.20. *Let L be any map with reversors R_1 and R_2 and a symmetry S. Then the product $R_1 \circ R_2$ is a symmetry of L and the product $R_1 \circ S$ is a reversor of L.*

Proof. For the former, consider the composition $LR_1R_2 = R_1L^{-1}R_2 = R_1R_2L$. For the latter similarly consider $L^{-1}R_1S = R_1LS = R_1SL$. Note that these calculations also prove closure of $\mathcal{R}(L)$ under composition; the existence of inverses is a similar one line proof. □

The structure of the symmetry group $\mathcal{S}(L)$ and the reversing symmetry group $\mathcal{R}(L)$ of a planar map have a large body of literature behind them. Two reviews of reversing symmetries for dynamical systems may be found in [49] and [37] the latter containing many references for further reading. Lamb in [35] and Goodson in [19] give some group theoretic results regarding reversing symmetry groups. In [5, 6] an algebraic approach to studying these groups was taken for maps that were toral automorphisms and maps that were related to toral automorphisms. In [7, 56], a similarly algebraic approach was taken to study these groups for when L is a polynomial automorphism (i.e. a polynomial map with polynomial inverse). The most recent paper by these authors, [8], discusses some general aspects of the group structure of the reversing symmetry group. Since we make extensive use of some of the theory that can be found in this series, we will reproduce some of the preliminary results from [8]. First we must make some group theoretic definitions which can be found in most introductory algebra texts for example [38].

Definition 3.21. *A non-identity element f of a group G is an involution if f^2 is the identity element .*

Definition 3.22. *Let G be a group. Then N is a normal subgroup of G if for every $g \in G$ and $n \in N$ we have that $gng^{-1} \in N$.*

Definition 3.23. *Let N and H be groups, $\phi : H \to Aut(N)$ a homomorphism and $G = \{(x, h) : x \in N, h \in H\}$. Then with the composition law*

$$(x_1, h_1)(x_2, h_2) = (x_1\phi(h_1)(x_2), h_1h_2)$$

G becomes a group called the semidirect product of N and H with respect to ϕ, written $G = N \rtimes_\phi H$. With this definition, N (or rather, all the pairs (n, id) with $n \in N$) is normal in G.

With these definitions we give Fact 1 and Lemma 1 from [8] as two lemmas, the proof of both these can be found in that cited paper.

Lemma 3.24. *Let f be an element of a group G. Then the symmetry group $\mathcal{S}(f)$ is a normal subgroup of the reversing symmetry group $\mathcal{R}(f)$ and the factor group $\mathcal{R}(f)/\mathcal{S}(f)$ is either the trivial group or C_2, the cyclic group of order 2.*

Lemma 3.25. *Let f be an element of a group G with $f^2 \neq 1$ and symmetry group $\mathcal{S}(f)$. If f has an involutory reversor r (that is, a reversor which is also an involution), the reversing symmetry group within G is $\mathcal{R}(f) = \mathcal{S}(f) \rtimes C_2$ with $C_2 = \{r, id\}$.*

To interpret these lemmas in the case of planar maps we should identify $f \in G$ as an element of the group of (say) rational, invertible maps of the plane. A last lemma that we shall implicitly use a fair bit later on is the following.

Lemma 3.26. *Let f be a member of a group G. Suppose that f has an involutory reversor, that is, there exists an $r \in G$ such that $rfr = f^{-1}$ and $r^2 = 1$. Then f can be written as the composition of two involutions.*

Proof. Since r is an involution, $f = (fr)(r)$. Now r is trivially an involution and $(fr)(fr) = f(rfr) = f(f^{-1}) = id$ so fr is also an involution. □

As for how the reversing symmetry group relates to the thesis, in chapter 6, a test for detecting reversibility (that is, the existence of a reversor) in a map will be given. Originally the test was devised for two dimensional maps [61]. In chapter 6 we extend its viability to the more complicated situation of reversibility in three dimensions and improve on some of the empirical conjectures regarding the test. In addition to this use of reversors in chapter 6, the structure of the reversing symmetry group of integrable maps will be discussed in chapter 5 as an extended application of the theory in chapter 4, this being where we will use the above lemmas.

Chapter 4

A Theorem on Curve Preserving maps

In this chapter we fully detail the main theorem of this thesis. A good deal of the work in this chapter was published in 2006 [30]. This theorem uses elliptic curve theory to give an alternative way of considering time-discrete maps that preserve a curve. After the theorem is given and proven, which, with the setup of the foundational chapters does not take too much work, some applications are given. The way in which it arose has already been largely introduced in chapter 3.

The drive to explain some universal aspects of the orbit length distribution for all integrable maps led to the hypothesis that a group action underlay such integrable maps. Such a group action that somehow varied throughout the plane would, when reduced to a group action over a finite field, lead exactly to the kind of equidistribution where the variation was between level sets of the integral in question. Pointing us in the right direction was much recent work done in proving that various specific cases of integrable maps were in fact acting as simple group addition on an elliptic curve. This suggested the group action for which we should be aiming; algebraic geometry would give the correct framework with which to prove it.

The first paper to give a direct and calculated link between an integrable map and the group addition on an associated Weierstrass cubic was in 2001 [16]. In this paper, the authors seek to describe the dynamics traced out by the corners of planar quadrilaterals when folded across themselves repeatedly. This leads them to consider a map with one parameter whose square is a simple QRT map. They note that the curves this map preserves are each conjugate to a Weierstrass cubic via explicit calculation of the functions needed to shift between the two sets of coordinates. They go on to show that the particular QRT map in the original setting is conjugate to the usual group addition in the Weierstrass setting. In [65] in 2004, a similar result was proved that applied to all maps in QRT form. In this paper, a general biquadratic (which covers the totality of QRT integrals) was reduced to a Weierstrass cubic containing the point $(0, 0)$ and it was shown that the QRT map acts on this cubic as addition by that common point. The method of proof was a mixture of explicit and algebraic geometric in nature. The coordinate transformations were explicitly calculated but the proof that the QRT map acted as addition by $(0, 0)$ followed a geometric approach. While this covered a lot of the maps used as examples in the literature of integrability, another paper in 2004 [68] mentioned a map that was non-QRT in its form but which also followed the same addition on associated Weierstrass cubic pattern. Even before these results were known, certain integrable maps had been solved (in the sense of their future orbits begin constructed as functions $(x(n), y(n)))$, with the solutions being elliptic functions (see, for example, [25, 28]). Furthermore one can see a general feeling in some work that integrability should be associated to Weierstrass addition. For example, in the paper that introduced algebraic entropy [10] the authors note that they consider integrability to mean that the map is essentially translation on a torus. In the two dimensional case this is equivalent to addition on an elliptic curve (see section 2.3.3).

We should also mention that a major contribution towards an algebraic geometric analysis of planar integrable maps is the forthcoming book of Duistermaat [14].

The theorem we present now is a generalisation of the specific and explicit results referred to above. Therefore, it is also a vindication of the belief in the community that integrability and rotation on a torus are indeed intimately related.

4.1 Statement and Proof of the Theorem

With the theoretical setup provided by previous chapters, the proof of the theorem is quite simple, and the statement easy to understand.

Theorem 4.1. *Let L be a birational map defined over a field K that leaves fixed an elliptic curve E/K with a corresponding Weierstrass curve W/K. Then L is conjugate to a birational map $\widetilde{L}$ which fixes W/K and can be expressed in terms of the group law $+$ on W as either:*

1. $\widetilde{L} : P \mapsto P + \Omega$
2. $\widetilde{L} : P \mapsto \iota(P) + \Omega$

where $\Omega = \widetilde{L}([0, 1, 0]) \in W(K)$ and ι is an automorphism of W/K of order 2, or possibly orders 4 (if $j(E) = 1728$), or 3 or 6 (if $j(E) = 0$). In the second case, $\widetilde{L}$ (and hence L) is of finite order with the same order as ι.

Proof. Let $\phi : E \to W$ be the birational map defined over K that takes E to its Weierstrass form W. Then define $\widetilde{L} = \phi \circ L \circ \phi^{-1}$. As a composition of birational maps, $\widetilde{L}$ is also birational and furthermore it clearly leaves W fixed. The inverse of $\widetilde{L}$ is easy to calculate; it is $\widetilde{L}^{-1} = \phi \circ L^{-1} \circ \phi^{-1}$. Let $\Omega = \widetilde{L}([0, 1, 0]) \in W(K)$. From theorem (2.49) we know that $\widetilde{L}$, as a morphism of the elliptic curve W, can be written as the composition of an isogeny and a translation. Let T_A denote the map of W that translates by the point $A \in W$. From the details of the proof of this theorem we furthermore know that if we define $\iota = T_{-\Omega} \circ \widetilde{L}$ then ι is an isogeny of W defined over K. Similarly the inverse of ι exists and can be written as $\iota^{-1} = \widetilde{L}^{-1} \circ T_{\Omega}$. However ι^{-1} is demonstrably an isogeny; $\iota^{-1}([0, 1, 0]) = \widetilde{L}^{-1}(T_{\Omega}([0, 1, 0])) = \widetilde{L}^{-1}(\Omega) = [0, 1, 0]$. Thus ι is an automorphism defined over K of W with

$$\widetilde{L} = T_{\Omega} \circ \iota : P \mapsto \iota(P) + \Omega$$

The possible automorphism groups of W which are defined over K are given in theorem (2.48). If ι is the identity automorphism then we are left with the first case above, a plain translation on W. Otherwise we have the second case where ι is a non-trivial finite order automorphism. To work through the consequences of this situation we tacitly identify the respective automorphism groups with their isomorphic images consisting of roots of unity of the appropriate orders. So, for example, in the case the $Aut(W) \cong C_4$ and ι is of order 4, we shall write ι as simply i. Now we work through the various cases systematically.

First let us suppose that ι is order 2 so we write $\iota = -1$. In this case the automorphism group can be isomorphic to either of the three allowable possibilities (they are C_2, C_4 and C_6) and we can write $\widetilde{L} : P \mapsto -P + \Omega$. Composing this map with itself gives $\widetilde{L}^2 : P \mapsto -(-P+\Omega)+\Omega = P - \Omega + \Omega = P$ so $\widetilde{L}$ is order 2 as required.

Now suppose that ι is order 3 so we write $\iota = \omega$ where ω is a primitive third root of unity. In this case the automorphism group may only be isomorphic to C_6. Now

$$\begin{aligned}\widetilde{L}^3 : P &\mapsto \omega(\omega(\omega P + \Omega) + \Omega) + \Omega \\ &= \omega^3 P + \omega^2\Omega + \omega\Omega + \Omega \\ &= P + (\omega^2 + \omega + 1)\Omega \\ &= P + 0\Omega \\ &= P.\end{aligned}$$

Here the map 0 is the constant map that sends each point to the identity on W. In this case $\widetilde{L}$ is order 3 again as required.

Thirdly suppose that ι is order 4 and we may write $\iota = i$ where $i^2 = -1$. Calculating the second power of this map gives

$$\begin{aligned}\widetilde{L}^2 : P &\mapsto i(iP + \Omega) + \Omega \\ &= -P + (i\Omega + \Omega).\end{aligned}$$

This square map is of the same form as the case with $\iota = -1$ which we know has order 2, thus this case gives an order 4 map as desired.

Finally suppose that ι is order 6 and write $\iota = -\omega$ where again ω is a primitive third root of unity. Calculating the second power of this map yields

$$\begin{aligned}\widetilde{L}^2 : P &\mapsto -\omega(-\omega P + \Omega) + \Omega \\ &= \omega^2 P + (-\omega\Omega + \Omega).\end{aligned}$$

Since ω is a primitive cube root of unity, ω^2 must also be and thus we are left with a situation similar to above when ι was order 3. So, with $\widetilde{L}^2$ being order 3, $\widetilde{L}$ is order 6 as required. □

An alternative way of describing this theorem is by using it in the context of the group of birational maps that preserve an elliptic curve E/K. Let $\mathcal{L}$ be the group of birational maps defined over K that preserve the elliptic curve E/K. For a Weierstrass curve W/K that corresponds to E consider the set

$$\widetilde{\mathcal{L}} = \{P \mapsto \iota(P) + \omega : \iota \in Aut(W), \omega \in W(K)\}. \tag{4.1}$$

The set $\widetilde{\mathcal{L}}$ consists of all possible compositions of the automorphisms and translations on W and in fact can be written as a group:

$$\widetilde{\mathcal{L}} = \mathcal{T} \rtimes Aut(W) \tag{4.2}$$

where $\rtimes$ denotes semi-direct product and $\mathcal{T}$ is the Abelian group of translations on W. The group $\mathcal{T}$ is a normal subgroup of $\widetilde{\mathcal{L}}$ and intersects $Aut(W)$ only in the identity automorphism. With these definitions, theorem (4.1) sets up a group isomorphism between $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ via the isomorphism

$$\Phi : \mathcal{L} \to \widetilde{\mathcal{L}} \quad L \mapsto \widetilde{L} = \phi \circ L \circ \phi^{-1} \tag{4.3}$$

where $\phi : E \to W$ is a fixed conversion function. That Φ is indeed an isomorphism is easily checked by first noting that it has an inverse given by

$$\Phi^{-1} : \widetilde{\mathcal{L}} \to \mathcal{L} \quad \widetilde{L} \mapsto L = \phi^{-1} \circ \widetilde{L} \circ \phi \tag{4.4}$$

and secondly by noting that

$$\begin{aligned} \Phi(L \circ M) &= \phi \circ L \circ M \circ \phi^{-1} \\ &= (\phi \circ L \circ \phi^{-1}) \circ (\phi \circ M \circ \phi^{-1}) \\ &= \Phi(L)\Phi(M). \end{aligned}$$

This isomorphism of groups can be summarized by the commuting diagram in figure (4.1)

Since the group of translations $\mathcal{T}$ is isomorphic to the group of points $W(K)$, with the isomorphism mapping each point to the translation that translates by that input point, we can further refine (4.2) to

$$\widetilde{\mathcal{L}} \cong W(K) \rtimes Aut(W). \tag{4.5}$$

$$\begin{tikzcd} E \arrow[r, "L"] \arrow[d, "\phi"] & E \arrow[d, "\phi"] \\ W \arrow[r, "\widetilde{L}: P \mapsto \iota(P)+\Omega"] & W \end{tikzcd}$$

Figure 4.1: Commuting diagram implied by theorem (4.1).

Typically, with the cases being separated according to the j-invariant of W, $Aut(W) \cong C_2$. So in this situation a birational map leaving invariant an elliptic curve corresponds to, on W, either:

1. a translation $P \mapsto P + \Omega$ (which, as an element of $\mathcal{T}$, commutes with each other element of $\mathcal{T}$); or
2. an involution $P \mapsto -P + \Omega$.

It is possible to draw a relation between these involutions and translations in a way that is of dynamical interest by the following proposition:

Proposition 4.2. *A birational map L on an elliptic curve E which corresponds to a translation $\widetilde{L} : P \mapsto P + \Omega$ on a Weierstrass curve W is reversible, i.e., can be written as a composition $G \circ H$ of birational involutions with H corresponding to $\widetilde{H} : P \mapsto -P + S$ and G corresponding to $\widetilde{G} : P \mapsto -P + \Omega + S$ where $S \in W$ is arbitrary.*

Proof. That $\widetilde{G}$ and $\widetilde{H}$ are involutions is clear; their composition gives

$$\begin{aligned} \widetilde{G} \circ \widetilde{H} &= -(-P + S) + \Omega + S \\ &= P - S + \Omega + S \\ &= \widetilde{L} \end{aligned}$$

as desired. □

So it is clear that if it is known that L corresponds to a translation on W (note that a sufficient, but not necessary, condition is that L is infinite order), it can be decomposed into rational involutions in many ways by varying the choice of the point S.

With theorem (4.1) providing a lexicon that relates (infinite order) maps on curves to points on the same, we get some elucidation into the matter of comparing birational maps preserving elliptic curves. Recall that one of the type of relations between maps L, M from section (3.4) was having $M^m = L^l$ for some integers m, l. With our map/elliptic curve lexicon we have the follow proposition.

Proposition 4.3. *Let L_1 and L_2 be birational maps on an elliptic curve that correspond, respectively, to translations $P \mapsto P + \Omega_1$ and $P \mapsto P + \Omega_2$ on an associated Weierstrass curve. For $m, n \in \mathbb{Z}$ we have*

$$L_1^m = L_2^n \Leftrightarrow m\Omega_1 - n\Omega_2 = \mathcal{O}_W = [0,1,0], \tag{4.6}$$

where

$$j\Omega_i = \underbrace{\Omega_i + \Omega_i + \ldots + \Omega_i}_{j \ times}.$$

Proof. This is an automatic consequence of the isomorphism Φ, noting that L_i^j is conjugate to $P \mapsto P + j\Omega_i$. □

Proposition (4.3) shows that two maps are power-related if and only if their corresponding points on W are linearly-dependent over $\mathbb{Z}$. As a particular case, taking L_2 as the identity map shows that L_1 is of finite order m on the curve E/K if and only if

$$m\Omega_1 = \mathcal{O}_W = [0,1,0], \tag{4.7}$$

meaning that Ω_1 is a point of order m on W. Dynamically speaking, Ω_1 being of finite order m on W is equivalent to saying that all points on the curve E have one and the same period m under L.

A second relation regarding which we can obtain a result is that of conjugacy.

Proposition 4.4. *Let L_1, L_2 be infinite order birational maps on, respectively, elliptic curves E_1 and E_2 with L_1 corresponding to the infinite order translation $P \mapsto P + \Omega_1$ on the associated Weierstrass curve W_1. If L_2 is birationally conjugate to L_1, i.e., if there exists G birational such that*

$$L_2 = G\, L_1\, G^{-1}, \tag{4.8}$$

then L_2 corresponds to an infinite order translation $P \mapsto P + \Omega_2$ on W_1 with

$$\Omega_2 = \iota(\Omega_1) \tag{4.9}$$

and $\iota \in Aut(W_1)$.

$$\begin{tikzcd}
E_1 \arrow[r, "L_1"] \arrow[d, "\Phi_1"] & E_1 \arrow[d, "\Phi_1"] \arrow[dd, bend left, "G"] \\
W_1 \arrow[r, "P+\Omega_1"] & W_1 \\
E_2 \arrow[u, "\Phi_2"] \arrow[r, "L_2"'] \arrow[uu, bend left, "G^{-1}"] & E_2 \arrow[u, "\Phi_2"']
\end{tikzcd}$$

Figure 4.2: Commuting diagram for proposition (4.4). In the figure, L_1 is known to be conjugate to $P \to P + \Omega_1$ on the Weierstrass cubic W_1. The proposition shows a relationship between Ω_1 and the corresponding translative point Ω_2 for L_2.

Proof. Firstly note that G birationally maps E_1 to E_2 so that in particular they share the same j-invariant and they share some common Weierstrass form W_1.

Reviewing the commuting diagram in figure (4.2) we see that

$$L_2 = \phi_2^{-1}(\iota_2 P + \Omega_2)\phi_2$$

where $P \mapsto \iota_2 P + \Omega_2$ is the action of L_2 on W_1 implied by theorem (4.1). However also we have that

$$L_2 = GL_1G^{-1}.$$

Equating these two yields

$$\begin{aligned}
\phi_2^{-1}(\iota_2 P + \Omega_2)\phi_2 &= GL_1G^{-1} \\
(\iota_2 P + \Omega_2) &= \phi_2 G L_1 G^{-1}\phi_2^{-1} \\
&= (\phi_2 G \phi_1^{-1})(P + \Omega_1)(\phi_1 G^{-1}\phi_2^{-1}).
\end{aligned}$$

Now $(\phi_1 G^{-1}\phi_2^{-1})$ is a map from W_1 to itself so we may write it as $P \mapsto \iota^{-1}P + \Theta$

which has inverse $P \mapsto \iota P - \iota\Theta$. Substituting this into our last equation gives

$$\begin{aligned}(\iota_2 P + \Omega_2) &= (\iota P - \iota\Theta)(P + \Omega_1)(\iota^{-1}P + \Theta) \\ &= (\iota P - \iota\Theta)(\iota^{-1}P + \Theta + \Omega_1) \\ &= P + \iota\Theta + \iota\Omega_1 - \iota\Theta \\ &= P + \iota\Omega_1.\end{aligned}$$

Now putting $P = \mathcal{O}$ tells us that $\Omega_2 = \iota\Omega_1$ after which we see immediately that ι_2 must be the identity automorphism. □

Note that since generically the only isogenies that exist on an elliptic curve are the identity map and the negation map, proposition (4.4) tells us that maps of elliptic curves that are conjugate to one another are, when reduced to acting on the same Weierstrass curve, the same map or inverses of one another.

The Mordell-Weil theorem (2.40) tells us that the set of rational points $E(\mathbb{Q})$, or $W(\mathbb{Q})$ for the associated Weierstrass, is a finitely-generated abelian group. This means that any point $\omega \in W(\mathbb{Q})$ can be written as a linear combination

$$\omega = k_1\omega_1 + k_2\omega_2 + \cdots + k_r\omega_r + T. \tag{4.10}$$

In equation (4.10), the ω_i's have infinite order and are linearly independent, T is a member of the torsion subgroup, k_i are integers uniquely determined by ω and r is the rank of the elliptic curve. Via the isomorphism Φ, a birational map L over $\mathbb{Q}$ of an elliptic curve $E/\mathbb{Q}$ that corresponds to a translation can be written uniquely as the composition

$$L = L_1^{k_1}\, L_2^{k_2} \ldots L_r^{k_r}\, L_T \tag{4.11}$$

where L_i is an infinite-order birational map corresponding to $P \mapsto P + \omega_i$, L_T is a finite order birational map and elements of $\{L_i, L_T\}$ pairwise commute. Conversely, we can, in principle use the isomorphism to construct birational maps over $\mathbb{Q}$ that preserve a given rational elliptic curve $E/\mathbb{Q}$ and correspond to translations. We do this by finding in equation (4.10) appropriate ω_i's and torsion points on the corresponding Weierstrass and then use Φ^{-1}.

4.2 Applications of Theorem to Finite Fields

The beauty of theorem (4.1), apart from its simplicity, lies in the fact that the field K is allowed to be quite general. The two kinds of fields that are of immediately obvious use are finite fields and function fields, the latter being the case of the field of rational functions of one variable.

4.2.1 Equidistribution and Plateaus

An observation that was made as early as during the production of [60] is that there appears to be more of a pattern than a simple cap on orbit length driving the lengths of orbits of integrable maps when considered over finite fields. Indeed, when the normalised orbit lengths (when normalising orbit lengths of planar maps considered over $\mathbb{Z}_p$ we divide the orbit lengths by the Hasse-Weil upper bound so that anything exceeding 1 we can flag as suspicious) are plotted as a function of the prime p, one notices that in integrable cases the normalised orbit length $\frac{|O|}{HW(p)}$ always lies close to $\frac{1}{n}$ for some natural number n. To display this orbit length data, we consider the cumulative distribution function generated by the normalised orbit lengths. With the notation (originally from [60]) that C_p is the set of points with periodic orbits and $T(y)$ is the period of a point y the distribution can be defined as (figure (1.2) of the introduction has already illustrated $D_p(x)$ for a particular example):

$$D_p(x) = \frac{|\{y \in C_p : T(y) \leq rx\}|}{|C_p|}$$

where r is the normalising factor being used. Originally, the feeling around the time of the work that generated [60] was that this distribution formed a “Devil's staircase” with fractal behaviour as larger primes were considered. These forbidden regions were noted quite universally in integrable maps, and plots of the cumulative frequency of normalised orbit length were particularly stark. It was this observation that led to the proof of the main original theorem in this thesis. In an attempt to explain these “forbidden regions”, so called since orbits with a normalised length could not lie in these regions for integrable maps, a conjecture regarding the “equidistribution” of orbit lengths was made.

Definition 4.5. *Let L be a map of the plane that leaves fixed a foliation of curves $\mathcal{C}$. Suppose that both L and $\mathcal{C}$ are representable modulo p (recall this means that there are no denominators divisible by p). Then we shall say that the orbits of L are equidistributed if each particular level set of $\mathcal{C}$ is partitioned by L into orbits of equal length.*

The conjecture was that every integrable planar map had this equidistribution property. Assuming this was the case, the following reasoning could be followed. Supposing each level set is generically genus 1, the Hasse-Weil bound coupled with equidistribution implied

$$p - 2\sqrt{p} + 1 \leq n|O| \leq p + 2\sqrt{p} + 1$$

where $|O|$ is the fundamental orbit length for that particular level set, which, by the equidistribution assumption, exists uniquely for each level set and n is the number of orbits of this length lying on that level set. From here it is simple to divide both sides by the upper Hasse-Weil Bound to get

$$1 - \frac{4\sqrt{p}}{p + 2\sqrt{p} + 1} \leq n\frac{|O|}{p + 2\sqrt{p} + 1} \leq 1$$

and dividing throughout by n (which, while unknown, is certainly an integer) to get

$$\frac{1}{n} - \frac{4\sqrt{p}}{n(p + 2\sqrt{p} + 1)} \leq \frac{|O|}{p + 2\sqrt{p} + 1} \leq \frac{1}{n}. \tag{4.12}$$

This equation is saying precisely that the normalised orbit lengths which we observe must lie within (small) windows to the left of the reciprocals of the natural numbers. The assumption of equidistribution allows us to refine our test of integrability. Rather than having a coarse statement along the lines of "If any normalised orbit length is greater than 1, then the map is (probably) not integrable" we instead can say "If any normalised orbit length does not lie in an allowable window, then the map is (probably) not integrable". Now these allowable windows eventually overlap for sufficiently large values of n, meaning that for each particular prime p there is a largest value of n for which the allowable windows remain distinguished. This n can be calculated as a function of p in the following way.

Proposition 4.6. *For any given p, the maximal value of n that p can distinguish is the floor (recall that the floor of a number is the greatest integer less than that number) of the quantity $\frac{1}{4}\sqrt{p} - \frac{1}{2} + \frac{1}{4\sqrt{p}}$.*

Proof. We wish to find the first n such that $\frac{1}{n} - \frac{4\sqrt{p}}{n(p+2\sqrt{p}+1)} \leq \frac{1}{n+1}$ as this is the condition for two consecutive windows to overlap. Equality in this relation gives

$$n = \frac{1}{4}\sqrt{p} - \frac{1}{2} + \frac{1}{4\sqrt{p}}$$

□

Proposition (4.6) allows us to set some threshold window number n that we wish our test to be accurate up to and choose our value of p accordingly. The crucial assumption of equidistribution is discussed just below. The need to give some reason for the apparent existence of equidistribution was the drive behind the original work of this chapter.

Knowing that reduction from the rational numbers to a finite field is a field homomorphism gives us great scope for applying theorem (4.1) over finite fields. We begin by justifying the assumption of equidistribution. Consider an infinite order birational map L that leaves fixed a foliation of generically elliptic curves with level sets defined by $t = E(x, y)$. Now each choice of $t = t_0$ such that the curve defined by $t_0 - E(x, y) = 0$ is rational can be reduced over some finite field $\mathbb{F}_p$ and the homomorphic nature of reduction ensures that this new reduced curve is preserved by the reduction of L over the same finite field. Over the finite field $\mathbb{F}_p$ there are only $p + 1$ curves in this new reduced foliation - one for each choice of the foliating parameter t (we include $t = \infty$ for projectivity) - and each is preserved by the reduction of L. Now we may apply theorem (4.1) to L as it preserves each of these $p + 1$ curves independently. Doing so informs us that associated to each curve there is a point Ω_t such that the action of the reduced map on that curve is simply addition by Ω_t. Therefore, the length of each orbit on these reduced curves is equal to the order of Ω_t for that particular level set. At this stage we have no way of relating the different points associated with the different level sets, but our purpose of equidistribution does not require this. The only assumption we needed to justify is that on each level set there is a fundamental period length that each orbit on that level set possesses and this “curve at a time” approach does that. This proof of equidistribution, coupled with the argument following definition (4.5) tells us that we should expect normalised orbit lengths (recall that we normalise orbit lengths for this integrability detection work by dividing the orbit length by the upper Hasse-Weil bound) to lie in small windows to the left of the reciprocals of the integers.

In practice there are a number of things that can go wrong. The biggest of these is the problem of aperiodic orbits as discussed in section (1.5) which we also discuss further in the next section. Other issues can all be put down to “bad reduction”. Two such issues that have been identified include curves that are irreducible over the rational numbers being reducible over certain finite fields and equations with no multiple roots having multiple roots over finite fields. The second two are, from experience, rare and there is nothing to be done about them in any case. To minimise their impact, one just has to be sure to gather enough numerical data that cases of bad reduction make up a small fraction of that data.

Now we move on to showing how equidistribution manifests itself in the data we collect from studying maps. We have a few properties we can turn on and off here. Firstly to show the difference between an integrable map and a non-integrable map we will turn integrability on and off. Secondly, to show the difference between considering only periodic orbits, where equidistribution occurs automatically, and considering all orbits, where equidistribution only appears after the complicated gluing process described in the next section, we turn on and off the presence of singularities in the affine plane. As usual we will be creating the cumulative frequency distributions generated by calculating normalised orbit lengths. It is in these plots that we expect equidistribution to show itself by orbits having normalised lengths in a window to the left of an inverse of a natural number. So the cumulative frequency distributions should climb to the left of the inverses of the natural numbers and be flat elsewhere. The exact formula for where normalised orbit lengths for integrable maps can lie was given above in equation (4.12). So we have several layers of complexity when it comes to equidistribution which we summarise.

- If the map is a permutation on a particular finite phase space then the orbits are all periodic and on any given level set for an integrable map we will see only one orbit length.

- If the map is not a permutation then we can consider only the periodic orbits in which case again for an integrable map we will see only one orbit length on any given level set. However, data will be a little sparser as we are excluding some of the phase space.

- We can include aperiodic orbits in our distribution in which case plateaus in the distribution $D_p(x)$ are not ensured although the ad hoc process of "gluing" (as given in the next section) can help remedy this.

To illustrate equidistribution's signature, we shall use the QRT map known as the Screensaver map of [16] with equations

$$x' = \frac{y+\alpha}{x} \quad y' = \frac{y+\alpha x+\alpha}{xy}. \tag{4.13}$$

For comparison to a non-integrable map, we perturb this Screensaver map slightly to

$$\begin{aligned} x' &= \frac{y+\alpha}{x} \\ y' &= \frac{y+\alpha x+\alpha}{xy} + \epsilon. \end{aligned} \tag{4.14}$$

We shall decompose the projective finite phase spaces $P(\mathbb{Z}_p^3)$ for $p = 101$ and $p = 997$ under the two different maps and consider separately the periodic orbits and all orbits. The parameters are set at $\alpha = 2$ and $\epsilon = \frac{1}{10}$. Figure (4.3) shows the cumulative frequency diagrams for these two maps with prime $p = 101$. All orbits are included in this figure so we expect that the signature of equidistribution will not necessarily be very strong for the integrable map. In figure (4.4), where we only plot the data for periodic orbits, we still don't see any real indication of equidistribution (and thus integrability in the map) through flat plateaus away from the allowable windows to the left of the inverses of the natural numbers. This is partially because the biggest window resolved by $p = 101$ is only $n = 2$, meaning that for $n = 3$ and beyond, the regions where data is allowed to fall overlap each other and partially because the prime $p = 101$ is quite small. So while the plateaus exist, they are difficult to detect on a macroscopic scale.

Figures (4.5) and (4.6) show the same data for $p = 997$. The number of resolved windows for this prime is $n = 7$, and six of these seven windows are shown in the second figure (the first window has no data points in it). For both primes 101 and 997, there are not enough periodic orbits for the perturbed Screensaver map to get

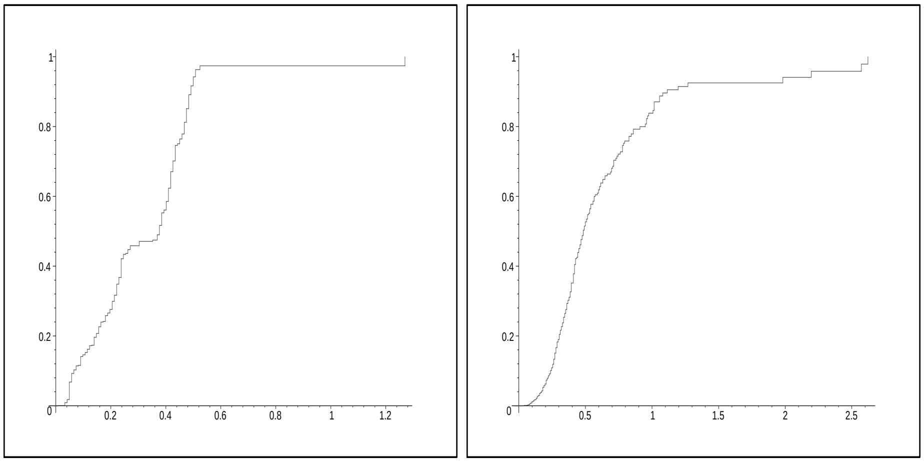

Figure 4.3: Cumulative frequency distribution for (left) the Screensaver map of equation (4.13) and (right) a perturbed Screensaver map, equation (4.14). Orbit lengths have been normalised by the upper Hasse-Weil bound and all orbits including aperiodic ones are included. The prime for this data set is $p = 101$.

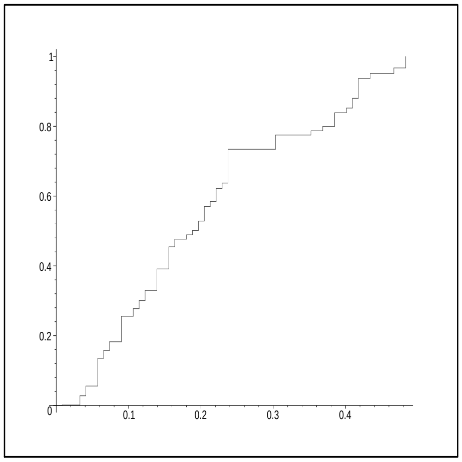

Figure 4.4: Cumulative frequency distribution for the Screensaver map. Only periodic orbits are shown here so we expect very clean plateaus from the equidistribution property of integrable maps over finite fields. The prime for this data set is $p = 101$. Unfortunately, the small amount of periodic data makes for an unsmooth looking distribution.

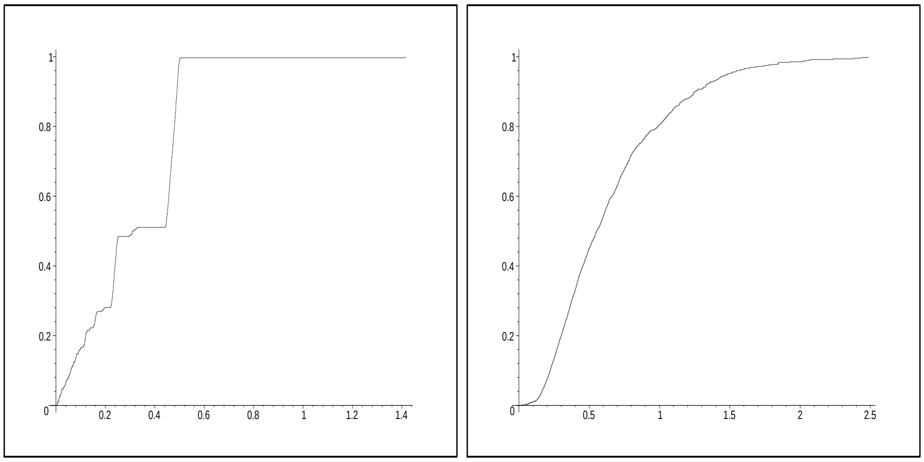

Figure 4.5: Cumulative frequency distribution for (left) the Screensaver map of equation (4.13) and (right) a perturbed Screensaver map, equation (4.14). Orbit lengths have been normalised by the upper Hasse-Weil bound and all orbits including aperiodic ones are included. The prime for this data set is $p = 997$. Notice that the plateaus become quite ill-defined as we move away from the larger windows due to aperiodic orbits being able to lie in the forbidden regions. Contrast this to figure (4.6) where the plateaus are very noticeable for all windows distinguishable by $p = 997$.

a reasonable spread of data, hence those two cumulative frequency distributions are not shown.

4.2.2 Aperiodic Orbits and "Gluing"

The problem of aperiodic orbits will occur any time it is possible for points in the affine plane to be mapped to the projective line which in turn will occur any time there is a denominator in the map that has a root in the finite field one is working over. As mentioned earlier, the easiest way of avoiding this problem in regards to testing for integrability is to only consider the periodic orbits. But in light of theorem (4.1) telling us that these aperiodic orbits that leak off to the point [0,0,0] are in fact conjugate to periodic orbits on some Weierstrass curve, we can see that there must be more than meets the eye to this situation. What follows is a discussion of an ad hoc method of "gluing together" aperiodic orbits into periodic orbits in the cases where theorem (4.1) applies. Following this we present some pictorial representations of

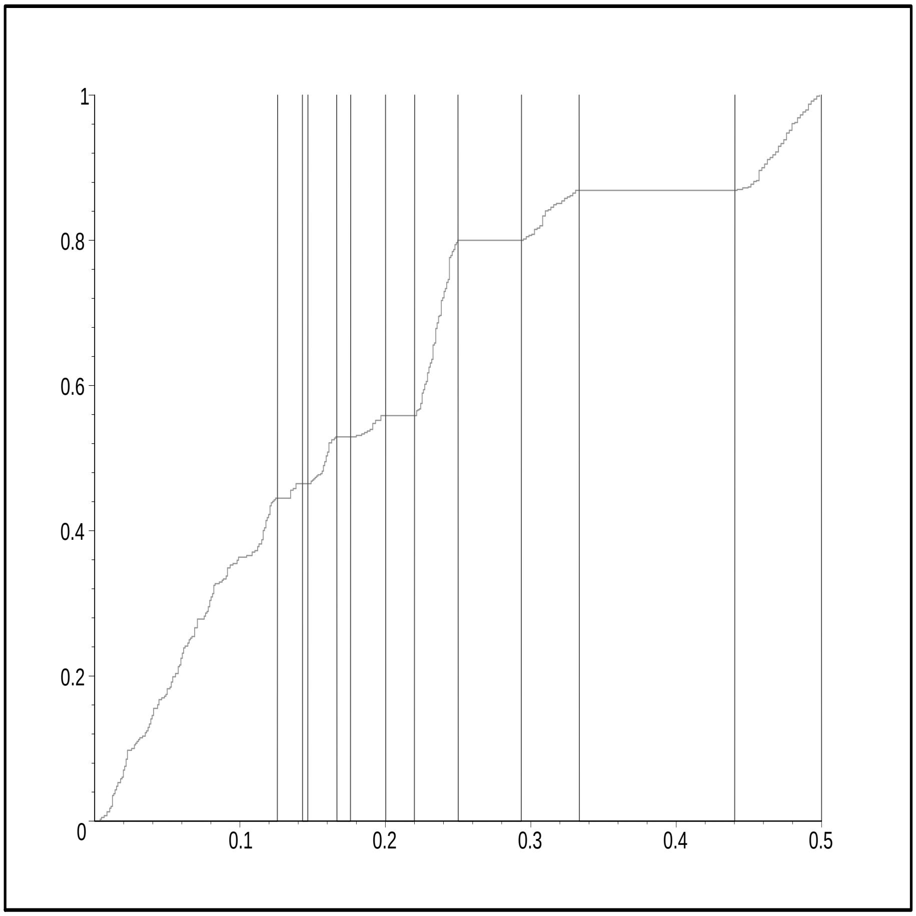

Figure 4.6: Cumulative frequency distribution for the Screensaver map. The plateaus here signifying forbidden regions for orbit lengths are quite stark. The prime for this data set is $p = 997$. Also drawn are six of the seven allowable windows that are resolved by this prime - the first is excluded because it contains no data points. Only periodic orbits are shown resulting in beautiful and cleanly defined plateaus.

equidistribution in integrable maps when considered over finite fields.

One thing to be noted about the gluing together of aperiodic orbits is that it is of no practical use in terms of refining the test of integrability based on finite fields. As we will see, throughout the gluing process we will need to know that the map is integrable. What it does is explain what the aperiodic orbits are doing by using the fact that they are essentially periodic orbits broken up by the coordinate change between the Weierstrass setting and the original setting.

Due to the hands-on nature of the gluing process, we illustrate it via an example.

Example 4.7. The QRT map we use in this example comes from [16], where it is examined in the context of folding plane quadrilaterals. We begin with the family of maps given by equation (4.13), which preserves the integral with level sets

$$I(x,y) = \frac{(x+1)(y+1)(x+y+\alpha)}{xy}. \tag{4.15}$$

It is important to note that the genus of most of the level sets of this family of curves is 1. The level set $I(x,y)=0$ is always reducible into $(x+1)(y+1)(x+y+\alpha)$ and $I(x,y)=1$ is reducible when $\alpha=2$ (we'll be taking this value later) into $(x+y+1)(xy+x+y+2)$ but for the other level sets it is possible to find a birational transformation between each particular level set and a cubic in Weierstrass form. Indeed, the explicit transformations are given in [16] in two stages; one called η^{-1} to move from this QRT setting to the setting most obviously related to folding quadrilaterals and a second labeled ϕ to move from this quadrilateral setting to the standard Weierstrass cubic setting. Composing these in the appropriate order gives a combined transformation that maps points on any given level set of $I(x,y)$ to points on a particular Weierstrass cubic. The same transformations take the map (4.13) to a different map that preserves the Weierstrass cubic, which is itself the transformed equivalent of the integral (4.15). By theorem (4.1), this conjugate map is simply translation by a fixed point on that Weierstrass cubic.

Numerically, we show the conjugacy by applying the map (4.13) to an entire finite phase space ($P(\mathbb{Z}_{17}^2)$ in this case) and showing that for every orbit created by the QRT map on each level set of its integral, there is a conjugate orbit (or sometimes two orbits) on a Weierstrass cubic. The conjugacy transformations contain three parameters, C, L and Q, pertaining to the quadrilateral folding interpretation. Since we are not interested in this interpretation, we can select one of these quite arbitrarily (we take $C = 1$ throughout this example) with L and Q then partially determined by what particular level set of $I(x, y)$ we are working on. To fully determine all but one parameter, we set $\alpha = 2$, which imposes restrictions on Q in terms of L (see below). Let us denote, for any point (x, y), the value of $I(x, y)$ to be β. Then we have the following equations and transformations:

$$
\begin{aligned}
\alpha &= 2 - \frac{L}{C}(Q + \frac{L^2}{C}) \\
\beta &= -\frac{L^3}{C^2} \\
Q &= -L^2 \\
C &= 1 \\
\eta^{-1} &: [X, Y, Z] \mapsto [\frac{X+Z}{L}, \frac{Y+Z}{L}, \frac{Z}{C}] \\
\phi &: [X, Y, Z] \mapsto [\frac{1}{24}(X+Y)(Q^2 - 4L) + \frac{1}{24}Z(12C - 8LQ - Q^3), \\
&\quad \frac{1}{2}(X - Y)(C - LQ), \frac{1}{2}(X + Y - QZ)]
\end{aligned}
$$

Table (4.1) shows the orbits made by the QRT map in $P(\mathbb{Z}_{17})$ along with the level set of $I(x, y)$ they lie on (that is, the value of β). Those values of β with no entry have no *closed* orbits lying upon them. The singular orbits have been found by taking an affine part "midway" along the orbit and following the map forward and the inverse map backwards to generate the entire singular orbit. The subscripts a_1, a_2, b and c mean the various "chains to singularity". These are the different chains of points that move an orbit from the affine part of the plane to the point [0,0,0]. Note that a_1 moves to [0,0,0] only in the inverse direction, a_2 moves to [0,0,0] only in the forward direction and b and c can move to [0,0,0] in both directions. Which end is actually used is obvious from context since [0,0,0] must occur at an end of an open orbit. The four chains to singularity in this case are:

- $a_1 = [0, 0, 0] \hookleftarrow [0, 1, 0] \hookleftarrow [1, 0, 0] \hookleftarrow [0, 15, 1]$
- $a_2 = [15, 0, 1] \hookrightarrow [0, 1, 0] \hookrightarrow [1, 0, 0] \hookrightarrow [0, 0, 0]$
- $b = [0, 0, 0] \hookleftarrow [0, 16, 1] \hookrightarrow [16, 1, 0] \hookrightarrow [16, 0, 1] \hookrightarrow [0, 0, 0]$
- $c = [0, 0, 0] \hookleftarrow [0, 1, 0] \hookrightarrow [1, 0, 0] \hookrightarrow [0, 0, 0]$.

All of the points in these chains to singularity are also base points of the projective integral. We can see this by projectivising the integral to get $\frac{(X+Z)(Y+Z)(X+Y+2Z)}{XYZ}$

and then solving simultaneously (over $\mathbb{Z}_{17}$) for both denominator and numerator equal to zero. This gives exactly the seven distinct points in the above chains to singularity.

Finding L and Q for each relevant value of β is a simple task, and using these to find η^{-1}, ϕ and the Weierstrass cubic is best accomplished through formulae in [16]. It also gives a simple expression for the point whose addition is conjugate to the QRT map, given as A below:

$$\begin{aligned}
v^2 &- 4u^3 + g_2 u + g_3 = 0 \\
g_2 &= \frac{1}{12}Q^4 + \frac{4}{3}Q^4 L + \frac{4}{3}L^2 - 2QC \\
g_3 &= -\frac{1}{216}Q^6 - \frac{1}{9}Q^4 L + \frac{1}{6}Q^3 C - \frac{5}{9}L^2 Q^2 + \frac{8}{27}L^3 + \frac{4}{3}CLQ - C^2 \\
A &= [\frac{Q^2 + 8L}{12}, -C, 1].
\end{aligned}$$

Armed with this simple way of finding our Weierstrass cubic and the important point on it, we can either numerically calculate orbits or we can find the algebraic expression for adding a generic point (ρ_1, ρ_2) to a free point (u, v), as in equation (4.16).

$$\begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \frac{m_1^2}{4} - u - \rho_1 \\ -(m_1(\frac{m_1^2}{4} - u - \rho_1) + \xi_1) \end{pmatrix} \tag{4.16}$$

$$\begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \frac{m_2^2}{4} - u - \rho_1 \\ -(m_2(\frac{m_2^2}{4} - u - \rho_1) + \xi_2) \end{pmatrix} \tag{4.17}$$

where $m_1 = \frac{v - \rho_2}{u - \rho_1}$, $\xi_1 = \rho_2 - \frac{\rho_1(v - \rho_2)}{u - \rho_1}$, $m_2 = \frac{12\rho_1^2 - g_2}{2\rho_2}$ and $\xi_2 = \rho_2 - \frac{\rho_1(12\rho_1^2 - g_2)}{2\rho_2}$. Equation (4.17) is the duplication formula, that is only applicable when $u = \rho_1$ and $v = \rho_2$. With these equations we would just have to substitute $(\rho_1, \rho_2) = A$ to find the formula that performs the maps action on the Weierstrass curve. This symbolic algebraic approach would be more instructive if we could expect to find the relation $QRT = \eta \circ \phi^{-1} \circ \text{group law} \circ \phi \circ \eta^{-1}$ to hold but unfortunately we cannot since the left hand side (QRT) acts on the entire family of curves $I(x, y) = \beta$ whereas the group law acts on the Weierstrass curves individually. So to get a symbolic equality between the two maps it would be necessary to somehow incorporate the difference

β	Periodic orbit lengths	Aperiodic orbit lengths
0	1	${}_{b}22_{a_2}$, ${}_{a_1}22_{b}$
1	9,9	${}_{a_1}5_{a_2}$, ${}_{b}6_{b}$
2		${}_{b}4_{b}$, ${}_{a_1}3_{a_2}$
3	5,5	${}_{b}2_{b}$, ${}_{a_1}1_{a_2}$
4		${}_{a_1}8_{b}$, ${}_{b}8_{a_2}$
5	1	${}_{a_1}4_{a_2}$, ${}_{b}5_{b}$
6		${}_{a_1}6_{b}$, ${}_{b}6_{a_2}$
7		${}_{a_1}5_{b}$, ${}_{b}5_{a_2}$
8	4,4	${}_{b}1_{b}$
9	7,7	
10	9	${}_{a_1}1_{b}$, ${}_{b}1_{a_2}$
11	11	${}_{b}2_{a_2}$, ${}_{a_1}2_{b}$
12		${}_{b}7_{b}$, ${}_{a_1}6_{a_2}$
13		${}_{b}3_{a_2}$, ${}_{a_1}3_{b}$
14		${}_{a_1}6_{b}$, ${}_{b}6_{a_2}$
15		${}_{a_1}9_{a_2}$, ${}_{b}10_{b}$
16		${}_{a_1}7_{a_2}$, ${}_{b}8_{b}$
∞		${}_{c}1_{c}$, ${}_{c}2_{a_2}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$,
		${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{a_1}2_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$, ${}_{c}3_{c}$
$\frac{0}{0}$		b

Table 4.1: How $P(\mathbb{Z}_{17}^2)$ decomposes under the Screensaver map. The notation for the singular orbits follows the convention that the subscript characters are the "chains to singularity" that always iterate to the point $[0, 0, 0]$ either under the map (if the subscript character is after the number) or the inverse map (for subscript characters preceding the number). The single number between the two subscript characters is the number of mundane (that is, not involved in a chain to singularity) affine points in the orbit. Note that β denotes the value of the integral $I(x, y)$ for each level set.

$$\begin{tikzcd}
(13,13) \arrow[r, mapsto] \arrow[d, "\phi\eta^{-1}"] & (9,10) \arrow[r, mapsto] \arrow[d, "\phi\eta^{-1}"] & (7,6) \arrow[r, mapsto] \arrow[d, "\phi\eta^{-1}"] & (6,7) \arrow[r, mapsto] \arrow[d, "\phi\eta^{-1}"] & (10,9) \arrow[r, mapsto] \arrow[d, "\phi\eta^{-1}"] & (13,13) \arrow[d, "\phi\eta^{-1}"] \\
(3,0) \arrow[r, mapsto] & (11,2) \arrow[r, mapsto] & (1,14) \arrow[r, mapsto] & (1,3) \arrow[r, mapsto] & (11,15) \arrow[r, mapsto] & (3,0)
\end{tikzcd}$$

Table 4.2: A QRT orbit of period five from the level set $\beta = 3$ of table (4.1) and its group law equivalent. Projective notation is suppressed for brevity since the third ordinate in each point is $Z = 1$.

in the curves they act upon. Thus it is just as useful and easier to numerically calculate the orbits under the group law. The procedure used is:

- Find a point whose orbit lies on $I(x, y) = \beta$.

- Find what this point becomes under the conjugacy transformation $\phi \circ \eta^{-1}$.

- Add the point A to this following the group law in equation (4.16), or (4.17) if the duplication law is necessary.

- See if and when this orbit closes.

- Check that the points in this orbit are the images of the orbit on $I(x, y) = \beta$ under the conjugacy transformation.

We begin then, with the level set $I(x, y) = 3$. The two level sets $I(x, y) = 0$ and $I(x, y) = 1$ are skipped because the curves themselves are not genus one and hence there is no Weierstrass cubic and none of the theory behind the method carries through. The level set $I(x, y) = 2$ is skipped because there are no periodic points on it. It is easy to find that $(13, 13)$ lies on $I(x, y) = 3$ and then to construct table (4.2). The only special feature here is that under the QRT map, the orbit had an $x-y$ symmetry which was lost (or at least transformed into a less obvious symmetry) once we turned to the Weierstrass setting. This continues for all other orbits listed in table (4.1); either each orbit under the QRT map is symmetric, or two orbits on the same level set form an asymmetric pair. We now look at the three other types of orbits in table (4.1): the ${}_{a_1}X_{a_2}$ kind, the ${}_{b}X_{b}$ kind and the ${}_{a_1}X_{b}/{}_{b}X_{a_2}$ pair kind. With these orbits it becomes necessary to denote whether the inverse map or the map itself is performing the iteration at each arrow. For periodic orbits it is irrelevant, but when the possibility of the "black hole" $[0, 0, 0]$ arises, one must be more careful with notation.

Consider first the level set $\beta = 5$. This level set has a fixed point ($[2, 2, 1]$ which happens to be a singular point on the curve when considered modulo 17) and both an ${}_{a_1}4_{a_2}$ non-periodic orbit and a ${}_{b}5_{b}$ non-periodic orbit. Since this level set has an extra singular point, the curve and thus its Weierstrass form are in fact genus 0 and we might expect any group theoretic attacks to fail. However so long as one avoids the singular point, a notion of group structure survives on the Weierstrass curve and this is why the approach works even in this case (see [64] where the group structure of singular Weiertsrass curves is considered). The point that the group law should be adding each time is $A = [10, 16, 1]$. We construct table (4.3) to show the ${}_{b}5_{b}$ orbit. In the case of table (4.3), the conjugate orbit on the Weierstrass cubic is exactly the subgroup generated by the point A (that is, A and all its powers). It also shows that while the orbit may be non-periodic in the QRT setting due to the "black hole", in the group law setting, the orbit is closed. From a numerics point of view, these orbits are a little more complicated simply because the formula for the group law changes. When adding the point A to itself, the duplication formula must be used and when adding the point A to either the identity $[0, 1, 0]$ or A^{-1}, the projective formula must be used. Despite the added difficulties with the calculations, the geometric concept remains exactly the same.

In table (4.4), we see a non-periodic QRT orbit becoming a completely mundane (that is, not even the duplication formula or projective version must be used) orbit under the group law. Furthermore, despite the apparent difference in length between an ${}_{a_1}4_{a_2}$ orbit and a ${}_{b}5_{b}$ orbit, one can count that each orbit contains exactly 8 unique points (in both QRT and group law settings).

One feature of the preceding three orbits discussed is that they have been symmetric; if a point $[x, y, z]$ is in an orbit, then the point $[y, x, z]$ is also. The presence of $[16, 1, 0]$ without a partner seems to contradict this, but in $P(\mathbb{Z}_{17})$, $[16, 1, 0] = [1, 16, 0]$ so it is its own partner. The fact that each point must be partnered comes from the symmetry inherent in the integral (4.15). Now we turn to an example of an asymmetric orbit, or more accurately a partnered pair of asymmetric orbits. In this case, the presence of a point $[x, y, z]$ on the first orbit implies the presence of

$$\begin{tikzcd}
\text{QRT} & & \text{Group Law} \\
{[0,0,0]} & & \\
{[0,16,1]} \arrow[u] \arrow[d, leftrightarrow] & & {[10,1,1]} \arrow[d, leftrightarrow] \\
{[16,1,0]} \arrow[d, leftrightarrow] & & {[0,1,0]} \arrow[d, leftrightarrow] \\
{[16,0,1]} & & {[10,16,1]} \arrow[d, leftrightarrow] \\
{[15,9,1]} \arrow[u] \arrow[d, leftrightarrow] & & {[15,11,1]} \arrow[d, leftrightarrow] \\
{[3,10,1]} \arrow[d, leftrightarrow] & & {[5,13,1]} \arrow[d, leftrightarrow] \\
{[4,4,1]} \arrow[d, leftrightarrow] \arrow[rr, "\phi\eta^{-1}"] & & {[6,0,1]} \arrow[d, leftrightarrow] \\
{[10,3,1]} \arrow[d, leftrightarrow] & & {[5,4,1]} \arrow[d, leftrightarrow] \\
{[9,15,1]} \arrow[d] & & {[15,6,1]} \arrow[d, leftrightarrow] \\
{[0,16,1]} \arrow[d, leftrightarrow] & & {[10,1,1]} \arrow[d, leftrightarrow] \\
{[16,1,0]} \arrow[d, leftrightarrow] & & {[0,1,0]} \arrow[d, leftrightarrow] \\
{[16,0,1]} \arrow[d] & & {[10,16,1]} \\
{[0,0,0]} & &
\end{tikzcd}$$

Table 4.3: A ${}_b5_b$ non-periodic orbit from the level set $\beta = 5$ of table (4.1). We see it becoming a periodic orbit of length 8 on the associated Weierstrass curve.

$$\begin{tikzcd}
{[0,0,0]} & \\
{[0,1,0]} \arrow[u] \arrow[d, leftrightarrow] & {[4,4,1]} \arrow[d, leftrightarrow] \\
{[1,0,0]} & {[4,13,1]} \arrow[d, leftrightarrow] \\
{[0,15,1]} \arrow[u] & {[2,5,1]} \arrow[d, leftrightarrow] \\
{[9,3,1]} \arrow[u] \arrow[d, leftrightarrow] & {[7,3,1]} \arrow[d, leftrightarrow] \\
{[10,4,1]} \arrow[r, "\phi\eta^{-1}"] \arrow[d, leftrightarrow] & {[8,4,1]} \arrow[d, leftrightarrow] \\
{[4,10,1]} \arrow[d, leftrightarrow] & {[8,13,1]} \arrow[d, leftrightarrow] \\
{[3,9,1]} \arrow[d] & {[7,14,1]} \arrow[d, leftrightarrow] \\
{[15,0,1]} \arrow[d] & {[2,12,1]} \arrow[d, leftrightarrow] \\
{[0,1,0]} \arrow[d, leftrightarrow] & {[4,4,1]} \arrow[d, leftrightarrow] \\
{[1,0,0]} \arrow[d] & {[4,13,1]} \\
{[0,0,0]} &
\end{tikzcd}$$

Table 4.4: An ${}_{a_1}4_{a_2}$ non-periodic orbit from $\beta = 5$ of table (4.1). We see it becoming a periodic orbit of length 8 on the associated Weierstrass curve.

$[y, x, z]$ on the second orbit and vice-versa.

$a_1 2b$ orbit			$b 2 a_2$ orbit		
QRT		Group Law	QRT		Group Law
$[0,0,0]$			$[0,0,0]$		
$\uparrow$			$\uparrow$		
$[0,1,0]$		$[15,10,1]$	$[0,16,1]$		$[3,1,1]$
$\updownarrow$		$\updownarrow$	$\updownarrow$		$\updownarrow$
$[1,0,0]$		$[15,7,1]$	$[16,1,0]$		$[0,1,0]$
$\uparrow$		$\updownarrow$	$\updownarrow$		$\updownarrow$
$[0,15,1]$		$[1,8,1]$	$[16,0,1]$		$[3,16,1]$
$\uparrow$		$\updownarrow$	$\uparrow$		$\updownarrow$
$[4,14,1]$		$[0,13,1]$	$[15,4,1]$		$[10,6,1]$
$\updownarrow$		$\updownarrow$	$\updownarrow$		$\updownarrow$
$[4,15,1]$	$\phi\eta^{-1} \rightarrow$	$[10,11,1]$	$[14,4,1]$	$\phi\eta^{-1} \rightarrow$	$[0,4,1]$
$\downarrow$		$\updownarrow$	$\downarrow$		$\updownarrow$
$[0,16,1]$		$[3,1,1]$	$[15,0,1]$		$[1,9,1]$
$\updownarrow$		$\updownarrow$	$\downarrow$		$\updownarrow$
$[16,1,0]$		$[0,1,0]$	$[0,1,0]$		$[15,10,1]$
$\updownarrow$		$\updownarrow$	$\updownarrow$		$\updownarrow$
$[16,0,1]$		$[3,16,1]$	$[1,0,0]$		$[15,7,1]$
$\downarrow$			$\downarrow$		
$[0,0,0]$			$[0,0,0]$		

Table 4.5: Paired asymmetric non-periodic orbits from the level set $\beta = 11$ of table (4.1).

Table (4.5) comes from the level set $\beta = 11$, which, as we can see, also contains a periodic orbit of length 11. Looking purely at the table (4.1), the untrained eye would find it hard to see the link between the periodic orbit of length 11 and the two non-periodic orbits of length ${}_{a_1}2_b$ and ${}_b2_{a_2}$. However, with the expanded data presented in table (4.5) the task becomes much easier. What has happened in this case is the “black hole” has split up what should be considered as a single orbit into two pieces. To weld the two back together, we overlap the two b components and

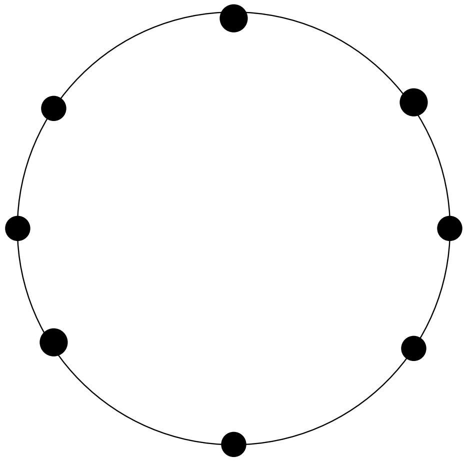

Figure 4.7: Pictorial of a periodic orbit.

overlap the two a components, thus turning the two pieces into a single closed loop. It is true that the QRT map cannot follow this closed loop around in any sense. However, the group law does not see two separate, aperiodic orbits, it sees only the one unified orbit. This idea suggests an alternative way of representing the data in table (4.1), which is given below in table (4.6). The data in this table has been interpreted by "gluing together" certain of the aperiodic orbits. In an a_1Xa_2 orbit, we glue the repeated points $[1,0,0]$ and $[0,1,0]$ to their repetitions, thus turning the disconnected orbit into a connected cycle. Similarly with bXb orbits. However, for a_1Xb orbits, we need their pairs under the symmetry. Note that for every a_1Xb orbit there is a bXa_2 orbit of the same length on the same level set of the integral. This is no accident; the doppelganger comes from the time-reversal symmetry of the map. We glue together the pairs by overlapping first the b chain. It is then an a_1XbXa_2 orbit, which is of the form $a_1X'a_2$, the method of gluing of which has already been described. It is important to note that this gluing is not just a convenient artefact; when following the conjugate orbits under the group law, one notices that exactly this method of gluing has occurred. To help in understanding the process of gluing, we can draw figures that represent the four different types of orbits. These pictures can be seen in figures (4.7), (4.9), (4.8) and (4.10).

In terms of the points and orbits involved, the gluing counts each distinct point on the new glued orbit once, removing any repetitions.

Table (4.6) allows us to count the number of points more easily, as well as high-

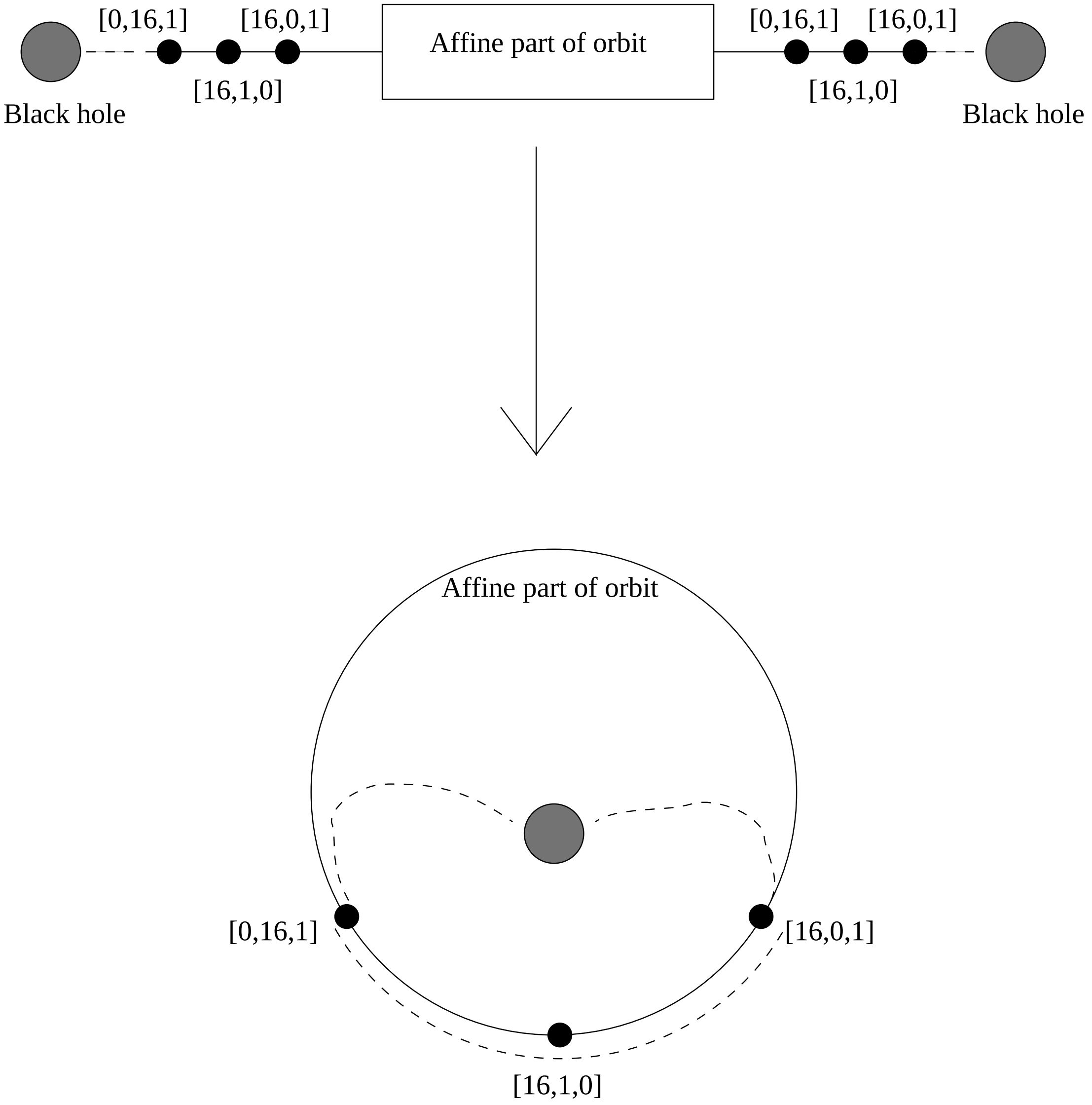

Figure 4.8: Pictorial of a bXb orbit having its ends overlayed to make it more recognisable as a periodic orbit that is seen on the Weierstrass cubic under the image of $\phi\eta^{-1}$ e.g table (4.3). Note that the points labelled on the "stitched together" part of the diagram are being abused for illustrative purposes; rather than being (for example) [16,0,1] it should strictly be labelled as the image of [16,0,1] under $\phi\eta^{-1}$.

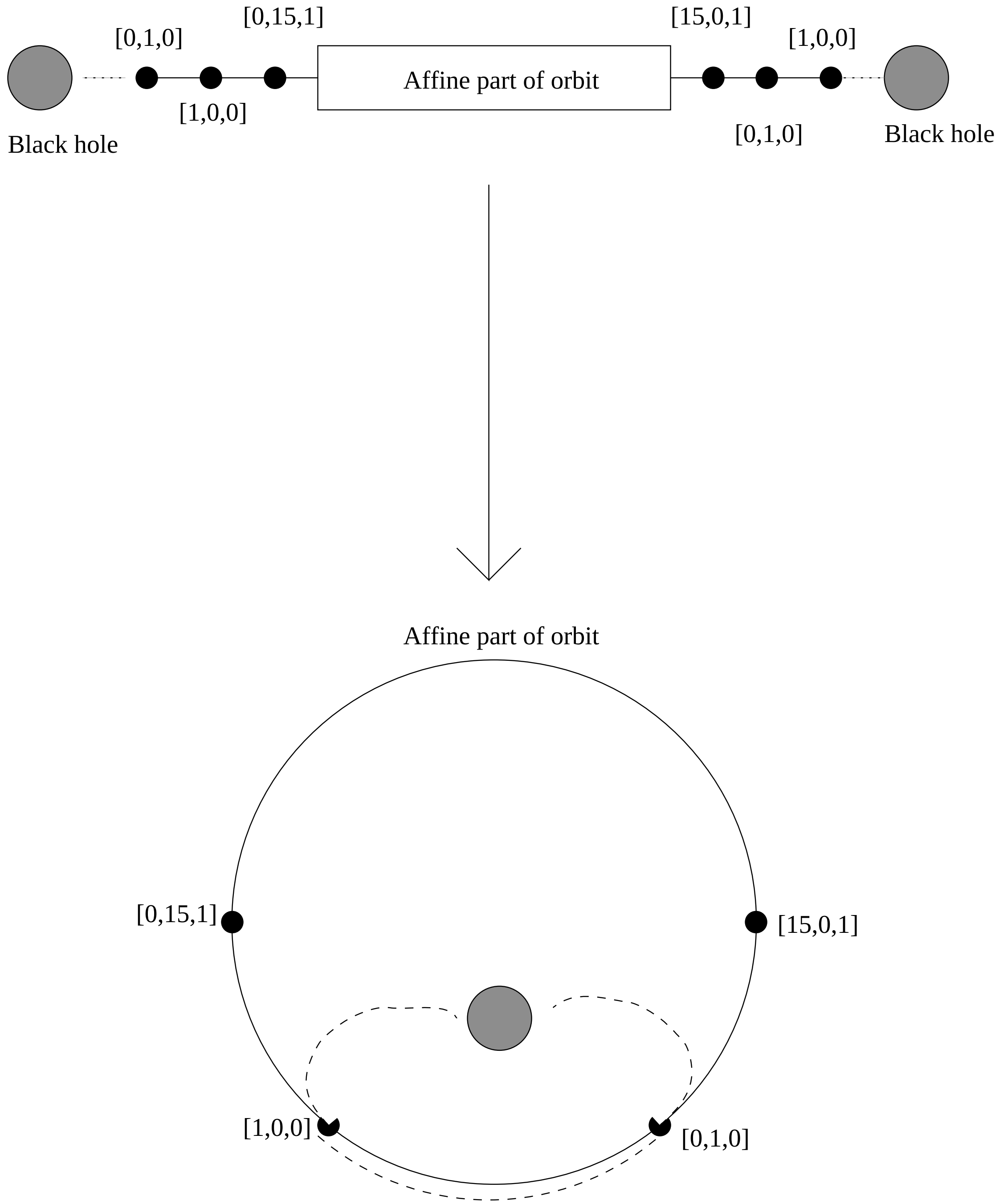

Figure 4.9: Pictorial of an aXa orbit having its ends overlayed to make it more recognisable as a periodic orbit that is seen on the Weierstrass cubic under the image of $\phi\eta^{-1}$ e.g table (4.4). Note that the points labelled on the "stitched together" part of the diagram are being abused for illustrative purposes; rather than being (for example) [15,0,1] it should strictly be labelled as the image of [15,0,1] under $\phi\eta^{-1}$.

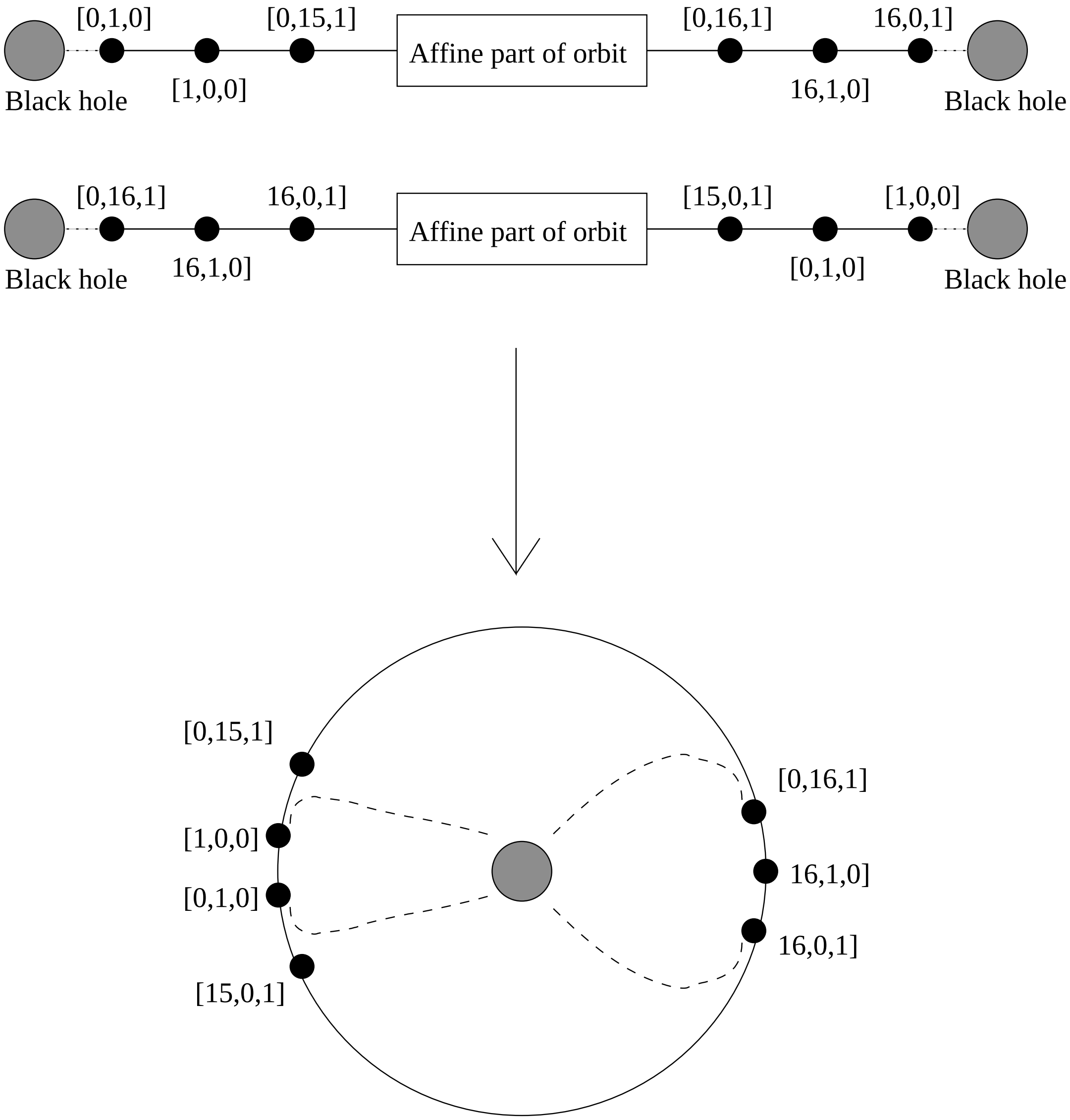

Figure 4.10: Pictorial of an aXb pair of orbits being glued together to make them more recognisable as a single periodic orbit as seen on the Weierstrass cubic under the image of $\phi\eta^{-1}$ e.g. table (4.5). Note that the points labelled on the "stitched together" part of the diagram are being abused for illustrative purposes; rather than being (for example) [16,0,1] it should strictly be labelled as the image of [16,0,1] under $\phi\eta^{-1}$.

β	Periodic orbit lengths	Aperiodic orbit lengths
0	1	51
1	9,9	9,9
2		7,7
3	5,5	5,5
4		23
5	1	8,8
6		19
7		17
8	4,4	4
9	7,7	
10	9	9
11	11	11
12		10,10
13		13
14		19
15		13,13
16		11,11
∞		${}_c1_c$, ${}_c3_c$, 8,8,8,8,8,8,8
$\frac{0}{0}$		b

Table 4.6: How $P(\mathbb{Z}_{17}^2)$ decomposes under the Screensaver map (equation (4.13)) after applying the gluing procedure. Contrast the uniformity in periods across any given level set with the (lack of) uniformity shown in table (4.1).

light the equality of orbit lengths on each level set. The level set $\beta = 0$ appears to contain 52 points. The equation defining this level set factors into three genus 0 curves (really just lines), meaning that there should be exactly 54 points. The discrepancy is explained by going into the details of what points lie on each of the three lines. There is one point that really lies on all 3 of the lines; this point needs to be counted an extra two times, leaving exactly 54 points. The level set $\beta = 1$ contains exactly 36 points and the equation defining it factors into two genus 0 curves. Finally, the level sets $\beta = \infty$ and $\beta = \frac{0}{0}$ remain mysterious. The former contains, from the table, 48 points. However each of the affine level sets now follow exactly the structure we expect from the group law - a single period length for each level set.

One can see that this gluing process is very context specific. To go about constructing a gluing procedure, we must already know the integral, the Weierstrass form of the integral's level sets and the conversion functions between them. In each case, the number and form of the chains to singularity vary so how they can be glued together must be figured out anew. For these reasons, the gluing procedure necessarily remains a heuristic method to apply when completely deconstructing particular examples rather than an abstract theory.

What the gluing procedure ultimately does is show us how level sets containing aperiodic orbits of different lengths can be viewed as containing orbits of a single length as required by theorem (4.1). That is, it shows how equidistribution is really present when it appears not to be.

4.2.3 The Monte-Carlo Alternative

Fixing a prime number and decomposing $P(\mathbb{F}_p^2)$ under a map is one way of attaining a cumulative frequency distribution which will flag potentially integrable maps. Another way is to vary the prime and from each phase space choose a point (either fix a point to be used in all phase spaces, or pick a random point each phase space) and find its orbit. Then we can store the normalised length of that orbit and generate a cumulative frequency distribution from these data points. This distribution must also have plateaus in the case of integrable maps since it is the normalised lengths we are counting. For any prime, periodic orbit lengths must still come in quantities of approximately Hasse-Weil, one half Hasse-Weil, a third Hasse-Weil and so forth. One advantage to this Monte-Carlo type approach is that it allows us to consider larger primes than would otherwise be possible. Another is that since we are only taking one orbit for each prime, any “bad prime” we encounter where, say, something unusual occurs to the form of the integral, making it generically a conic, plays a small role in the numerics. Figure (4.11) shows the cumulative distribution frequency distributions for the Screensaver and perturbed Screensaver maps using this method for the fixed initial condition $[\frac{1}{7}, 4, 1]$. The “double orbit” for this point is calculated (i.e. the forward orbit concatenated with the orbit under the inverse map) and its length stored. The primes range from 997 to 10007. While there is not a great deal of analysis presented here to suggest that the Monte-Carlo approach yields similar results to single prime decompositions, a good deal of numerical evidence does exist to support the hypothesis that the two approaches do give the same orbit length distributions. Prior to [57] only the single prime decomposition method was used; this paper was the first to show the use of the Monte-Carlo approach and the work done for this paper led to the belief that the two were equivalent. From then on the Monte-Carlo approach was almost exclusively used due to its beneficial properties.

4.3 Application of Theorem to Function Fields

Recall that $\mathbb{C}(t)$ is the field of rational functions with coefficients in $\mathbb{C}$ with t being the free variable. This field forms the function field of the complex line and by applying theorem (4.1) with $K = \mathbb{C}(t)$ we obtain a result that directly applies to many traditional discrete planar integrable systems. The crucial step lies in thinking of a one-parameter family of elliptic curves, such as equation (2.5) as a single elliptic curve over the field $\mathbb{C}(t)$. This process is valid so long as the dependence on the parameter (in this case t) is given by rational functions.

To make things precise, let $C(x, y, t) = 0$ be a family of curves with complex coefficients, each coefficient being parameterised by the parameter t. The equation $C(x, y, t) = 0$ defines a foliation of the $x-y$ plane if there exists a function $\tau : \mathbb{C}^2 \to \mathbb{C}$, $(x, y) \mapsto \tau(x, y)$ which is defined apart from, possibly, finitely many points and such

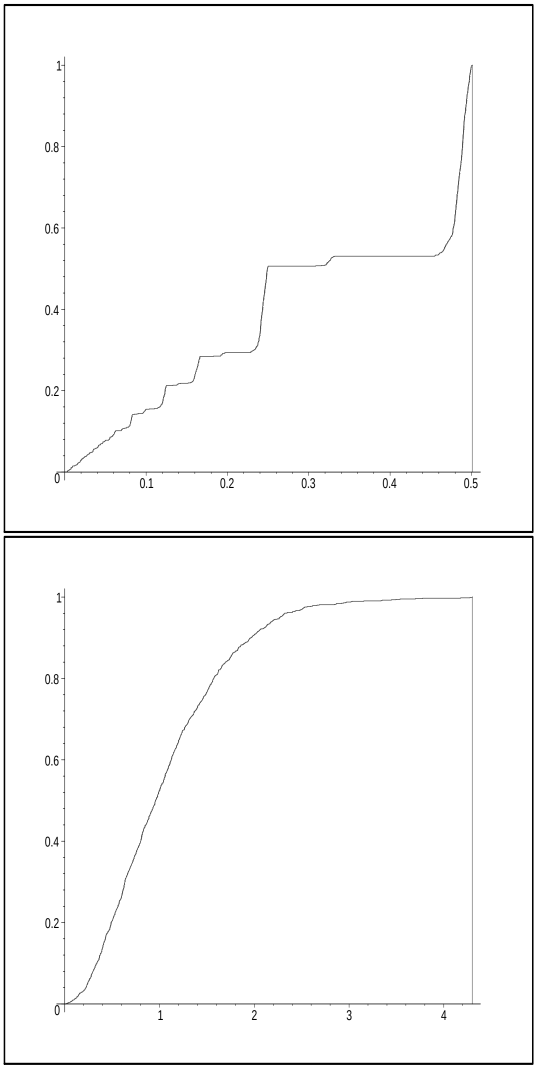

Figure 4.11: Cumulative frequency distributions from the Screensaver and perturbed Screensaver map by using the Monte Carlo approach to data collection. Note that the two are remarkably similar to the counterpart distributions in figure (4.5) attained by fixing $p = 997$ and decomposing the whole phase space - the largest clump of orbit lengths is around one half Hasse-Weil with the next at one quarter (note that the scales on the two are different which make them appear different). Such shape related features have been retained through the Monte-Carlo switch in the maps we have performed this analysis on.

that $C(x, y, \tau(x, y)) = 0$. The finitely many exceptional points are the base-points of the foliation. A map $L : (x, y) \mapsto (x', y')$ that preserves each curve in the foliation will satisfy the condition

$$C(x, y, \tau(x, y)) = 0 \Rightarrow C(x', y', \tau(x', y')) = 0$$

highlighting that $\tau(x, y)$ is an integral of motion under L.

To allow an algebraic-geometric approach, we specialise to the case in which $C(x, y, t)$ is algebraic (in which case we talk of an algebraic foliation), and L is birational . We can allow L to have an explicit algebraic dependence on t. In such a case where there is an explicit dependence on t, we may - in the language of [27], [28] - call L a curve-dependent map. However the line of thinking being supported here requires us to think of such a map as a single map defined over the function field $\mathbb{C}(t)$.

Combining theorem (4.1), theorem (2.51) and proposition (4.2) gives

Theorem 4.8. *Let L be an infinite order birational map defined over $\mathbb{C}(t)$ that leaves fixed each curve in an algebraic foliation $C(x, y, t) = 0$ where $C = E/\mathbb{C}(t)$ is an elliptic curve. Then L is conjugate to a map $\widetilde{L} : P \mapsto P + \Omega(t)$ on the associated Weierstrass curve $W/\mathbb{C}(t)$, where $\Omega(t) = (\omega_1(t), \omega_2(t))$ with $\omega_i(t) \in \mathbb{C}(t)$. Furthermore, L is reversible, i.e. can be written as the composition of two rational involutions over $\mathbb{C}(t)$ and the dynamics of L on each curve can be parameterised in terms of Weierstrass elliptic functions (see equation (2.12) and the preceding text).*

Proof. The proof is a straightforward application of the theorems quoted above. □

Some comments to be made about theorem (4.8) are:

1. to use the theorem as stated, one must find a point $C(x, y, t) = 0$ with $x, y \in \mathbb{C}(t)$. However if $C = E/K$ with K some subfield of $\mathbb{C}(t)$ (e.g. $\mathbb{R}(t)$ or $\mathbb{Q}(t)$) or K some extension of $\mathbb{C}(t)$, and if L is also defined over K then the theorem stands with $\mathbb{C}(t)$ replaced by K;

2. the inference in the statement of the theorem that the algebraic foliation $C(x, y, t) = 0$ is actually an elliptic curve defined over $\mathbb{C}(t)$ is an application of Hurwitz' theorem, given previously as theorem (2.51);

3. theorem (4.8) can be viewed as an analogue of the Arnol'd-Liouville theorem in the sense that it tells us that a birational map leaving fixed each curve in an algebraic foliation is conjugate to a translation by a point that depends on the particular curve within the foliation. For rational measure-preserving maps of the plane with a rational integral of motion, a discrete Liouville theorem due to Veselov [67] gives that the dynamics on a compact, non-singular level set of the integral is conjugate to the rotation $\theta \mapsto \theta + \omega(t)$. As a result of the algebraic nature of all objects and arguments involved here, theorem (4.8) has no measure-preservation requirement;

4. the decomposition of integrable planar maps, such as the QRT maps, as the composition of two involutions has been much exploited to elucidate their properties (in fact, reversibility was presumed from the outset in order to create the QRT maps in [50] and [51])

5. returning for a moment to the finite field case we can see that an interpretation of the distribution of orbit lengths is to instead consider the order of the point $\Omega(t)$. Since this decides the period of orbits on given level sets by considering particular values $t = t_0$ we can use this as an alternate way to generate the orbit length distribution.

A further consequence of theorem (4.8) is that it suggests how to determine all the possible finite orders of maps L satisfying the assumptions of the theorem (this now refers to the global finite order of L, not just its action on any particular curve). From equation (4.7), the issue is to calculate the possible finite orders of the translative point $\Omega = (\omega_1(t), \omega_2(t)) \in W/\mathbb{C}(t)$. This has been resolved in [13] and [47] for elliptic curves E defined over $\mathbb{C}(t)$ with $j(E)$ not belonging to $\mathbb{C}$ and having a long Weierstrass equation $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ with coefficients a_i being polynomials in t of degree at most i. Elliptic curves defined over $\mathbb{C}(t)$ that satisfy this condition on the coefficients are called rational elliptic surfaces. Under these conditions, it is known from [47] that the group of $\mathbb{C}(t)$ points on the Weierstrass equation of the curve is one of a very small number of possibilities. These possibilities are given in that paper and are as follows:

$$\begin{gathered}\mathbb{Z}^r (1 \leq r \leq 8), \mathbb{Z}^r \oplus \mathbb{Z}/2\mathbb{Z}(1 \leq r \leq 4), \mathbb{Z}^r \oplus \mathbb{Z}/3\mathbb{Z}(1 \leq r \leq 2), \qquad (4.18)\\ \mathbb{Z}^r \oplus (\mathbb{Z}/2\mathbb{Z})^2 (1 \leq r \leq 2), \mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}, \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}, (\mathbb{Z}/3\mathbb{Z})^2,\\ (\mathbb{Z}/2\mathbb{Z})^2, \mathbb{Z}/6\mathbb{Z}, \mathbb{Z}/5\mathbb{Z}, \mathbb{Z}/4\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}, 0.\end{gathered}$$

In this list, r is the rank of the rational elliptic surface and is equal to the number of linearly independent infinite order points in the group.

Proposition 4.9. *Let L be a birational map preserving a rational elliptic surface whose j-invariant is not in $\mathbb{C}$. Then, if L has finite order its order does not exceed 6.*

Proof. Firstly, we recall that any map of the form $P \mapsto \iota P + \Omega$ when ι is not the identity automorphism automatically has order equal to the order of ι. Since the maximal order of such automorphisms is 6, we may assume that $\iota = id$ and thus any maps of order greater than six have the form of translation. From here the proof is a direct application of the above list to the translative point $\Omega = (\omega_1(t), \omega_2(t))$ as given in theorem (4.8). □

In [65], it has been shown, constructively, that the QRT maps preserving a general biquadratic with coefficients that depend affinely on t are equivalent to translations on a rational elliptic surface. The finite order possibilities for such QRT maps are found and examples of each given.

Example 4.10. Our first example illustrates theorem (4.8).

Consider the one-parameter family of curves

$$B(x, y, t) = x^2y^2 - t^2(x^2 + y^2) - 2xy + 1 = 0 \qquad (4.19)$$

which constitutes a single algebraic curve defined over $\mathbb{Q}(t)$ (a subfield of $\mathbb{C}(t)$), since it contains the smooth point $(0, \frac{1}{t})$.[1] This means that the conversion-to-Weierstrass functions ϕ and ϕ^{-1} are defined over $\mathbb{Q}(t)$. The associated Weierstrass equation is [2]

$$W(u, v, t^2) = v^2 + u^3 + \big(- \frac{1}{3}(t^2)^4 - 4(t^2)^2\big)u + \frac{2}{27}(t^2)^6 - \frac{8}{3}(t^2)^4 = 0. \qquad (4.20)$$

[1] An alternative starting point is to note that equation (4.19) is an elliptic curve over $\mathbb{Q}(\sqrt{2}, i = \sqrt{-1}, t^2)$ since it contains the point $(\frac{(1-i)}{\sqrt{2}}, \frac{-(1+i)(t^2-i)}{\sqrt{2}(t^2+i)})$.

[2] Our Weierstrass here, and in example (4.11) below, is in the form outputted by MAPLE and is related to the usual form from equation (2.8) by $u = -x$, $v = y$.

Theorem 4.8 applies for any infinite order birational maps that preserve B. One such map is

$$L: \; x' = y\,, \; y' = -x + \frac{2y}{y^2 - t^2}. \qquad (4.21)$$

Note that L is defined over $\mathbb{Q}(t)$; it is the curve-dependent McMillan map leaving fixed each curve in the algebraic foliation $B(x, y, t) = 0$ [27, 28]. Since the function $\tau(x, y)$ obtained from solving (4.19) for t^2 is rational in x and y, replacing t^2 in (4.21) by τ produces an alternative form of (4.21) which is still birational

$$L: \; x' = y \quad y' = \frac{2y^3 - x(y^4 - 1)}{y^4 - 1 + 2xy}. \qquad (4.22)$$

This is a symmetric QRT map. (It is shown in [27, 28] that all symmetric and asymmetric QRT maps admit a curve-dependent McMillan description.)

Using theorem (4.8) we can find the additive point $\Omega(t) = (\omega_1(t), \omega_2(t))$ utilising a transformation ϕ taking B to W (an explicit form for ϕ can be found using the algorithm given in [66] and implemented in the MAPLE computing package). We find that

$$\Omega(t) = \big(-\frac{1}{3}(5t^2 - 6)t^2, \, -2(t^2 - 2)t^4 \big). \qquad (4.23)$$

As expected, the coordinates of Ω are in $\mathbb{Q}(t)$.

In table (4.7), we list various birational maps that leave fixed each curve in the foliation $B = 0$ and their corresponding actions on W, including L and L^2. The other entries relate to finite order maps. In particular, we show the standard involutions G and H such that $L = G \circ H$. Another decomposition into orientation-reversing involutions is $L = N \circ R$.

If we take $z = t^2$, we observe that $W(u, v, z)$ of (4.20) is a rational elliptic surface with j-invariant equal to $16(z^2 + 12)/(z^2 - 4)^2$. Furthermore, all the points involved in the maps of table (4.7), including (4.23), are in $\mathbb{Q}(z)$. Since in our table we have one point of infinite order (corresponding to the map given by (4.21) and (4.22)) and also three points of order two (corresponding to the simple involutions I_1, I_2 and I_3, which form a group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$) we are immediately left with only one possible structure from equation (4.18) for the group of $\mathbb{C}(z)$ points on W, namely

$$\mathbb{Z}^r \oplus (\mathbb{Z}/2\mathbb{Z})^2$$

Map on W	Map on B
$P \mapsto -P$	$N(x,y) = (-x, y - \frac{2x}{x^2-t^2})$
$P \mapsto -P + (-\frac{t^2}{3}(5t^2-6), 2(t^2-2)t^4)$	$R(x,y) = (-y,-x)$
$P \mapsto -P + (-\frac{t^2}{3}(5t^2+6), -2(t^2+2)t^4)$	$H(x,y) = (y,x)$
$P \mapsto -P + (-\frac{2}{3}t^4, 0)$	$G(x,y) = (x, -y + \frac{2x}{x^2-t^2})$
$P \mapsto P + (-\frac{2t^4}{3}, 0)$	$I_1(x,y) = (-x,-y)$
$P \mapsto P + (\frac{1}{3}t^4 + 2t^2, 0)$	$I_2(x,y) = (\frac{1}{x}, \frac{1}{y})$
$P \mapsto P + (\frac{1}{3}t^4 - 2t^2, 0)$	$I_3(x,y) = (-\frac{1}{x}, -\frac{1}{y})$
$P \mapsto P + (-\frac{t^2}{3}(5t^2-6), -2(t^2-2)t^4)$	$L(x,y) = (y, -x + \frac{2y}{y^2-t^2})$
$P \mapsto P + (-\frac{2}{3}t^4 - 1, -1 + t^4)$	L^2

Table 4.7: Examples of birational maps preserving (4.19) and their corresponding description on (4.20).

with $1 \leq r \leq 2$ [47]. In particular, this structure restriction tells us that the three simple involutions I_1, I_2 and I_3 are the *only* finite order rational maps corresponding to points in $\mathbb{C}(z)$ that commute with the QRT map. Of course there are other possibilities for finite order maps leaving fixed each curve in the foliation $B = 0$ that do not commute with the QRT map. However these must all be involutions. By theorem (4.1), they all act on W of equation (4.20) as $P \mapsto \iota(P) + \omega$ and since W has j-invariant not equal to 0 or 1728, the only possibility for ι is ± 1. When $\iota = -1$, the resulting transformation is necessarily of order two, but does not commute with infinite order maps of translative type.

Example 4.11. Our second example uses propositions (4.3) and (4.4) and the concept of the rank of an elliptic curve to find two independent birational maps leaving fixed each curve the same foliation.

The one-parameter family of curves

$$B(x,y,t) = (1-t)x^2y^2 - t(x^2+y^2) - 3txy + x - 3y = 0 \qquad (4.24)$$

is an elliptic curve defined over $\mathbb{C}(t)$, in fact over its subfield $\mathbb{Q}(t)$, since it contains e.g. the point $(\frac{1}{t}, 0)$. It has associated Weierstrass equation

$$W(u,v,t) = v^2 + u^3 + (-\frac{25}{48}t^4 + \frac{29}{2}t(t-1))\,u - \frac{125}{864}t^6 + \frac{601}{24}t^3(t-1) - \frac{9}{4}(t-1)^2 = 0. \tag{4.25}$$

Since (4.24) is an asymmetric biquadratic, it is preserved by an asymmetric QRT map L_1, whose component polynomials are (see equation (3.14))

$$\begin{aligned}(f_1, f_2, f_3)(y) &= (-10y^2, y(y^3 + 3y^2 + 3), -3y^3 + y^2 + 1)\\ (g_1, g_2, g_3)(x) &= (6x^2, x(x^3 - x^2 - 1), -3(x^3 + x^2 + 1).\end{aligned}$$

These give the equation

$$\begin{aligned} L_1 : x' &= \frac{-10y^2 - x(y^4 + 3y^3 + 3y)}{y^4 + 3y^3 + 3y - x(-3y^3 + y^2 + 1)}\\ y' &= \frac{6x'^2 - y(x'^4 - x'^3 - x')}{x'^4 - x'^3 - x' + y(3x'^3 + 3x'^2 + 3)}. \end{aligned} \tag{4.26}$$

We find that L_1 and the involutions G and H such that $L_1 = G \circ H$ correspond to the following elements of (4.1) acting on (4.25):

$$\begin{aligned} \widetilde{L}_1 : \quad P &\mapsto P + \Omega(t), \quad \Omega(t) = (-\frac{17}{12}t^2\,,\ \frac{3}{2}\,(t^3 - t + 1))\\ \widetilde{G} : \ P &\mapsto -P + \tau_1(t), \ \ \tau_1(t) = (-\frac{91}{36}t^2\,,\ \frac{209}{54}t^3 + \frac{3}{2}(t-1))\\ \widetilde{H} : \ P &\mapsto -P + \tau_2(t), \ \ \tau_2(t) = (-\frac{233}{12}t^2\,,\ -\frac{171}{2}t^3 - \frac{3}{2}(t-1)) \end{aligned}$$

The points $\Omega(t)$, $\tau_1(t)$ and $\tau_2(t)$ belong to $W(\mathbb{Q}(t))$ with $\Omega(t) = \tau_1(t) - \tau_2(t)$ as expected. Consider the new translation on W based upon the point $\tau_2(t)$:

$$\widetilde{L}_2 : P \mapsto P + \tau_2(t). \tag{4.27}$$

Evidently, $\widetilde{L}_2 = \widetilde{H} \circ \widetilde{N}$, where $\widetilde{N} : P \mapsto -P$. Via the isomorphism of figure (4.1), $\widetilde{L}_2$ generates another birational map $L_2 = H \circ N$ leaving fixed each curve in the foliation (4.24) that commutes with the QRT map L_1. The explicit form of the involution N is found to be

$$\begin{aligned} x' &= \frac{-(110x^2y^2 + 27y^3x^3 - 9y^4 + 27yx^3 - 27x^2y^3 - 252xy^2 - 27y^3 + 19y^4x^3 - 27x^2y - 152x - 627y - 114y^3x - 9x^2y^4)}{(-110xy^2 + 21x^2y^3 + 114x^2y - 9x^2y^2 - 81y^2 + 9x^3y^2 + 33y^3x^3 + 8x^2y^4 - 60xy - 361 - 27y^3 - 9x^2 + 9x^3)}\\ y' &= \frac{-y^3x^3 + 3x^4y^2 + 152y - 28x^2y - 3xy^2 - 3x^3 + 3x^4 - 171x + 19x^4y^3 - y^3x - 114x^2y^2 + 46yx^3 - 3x^3y^2}{8x^4y^2 - 9y^3x^3 - 17x^3y^2 - x^3 + 3x^2y^3 + 9x^2y^2 - 114x^2y + x^2 + 46xy^2 + 108xy + 3y^3 + 9y^2 + 361}. \end{aligned}$$

From proposition 4.4, L_2 and L_1 are not birationally conjugate since $\Omega(t) \neq \pm\tau_2(t)$ (here $j(W) \neq 0, 1728$, so that $Aut(W) = \{P \mapsto \pm P\}$. We also claim that L_2 is not

t-value	QRT Point $\Omega(t)$	Point $\tau_2(t)$	Decomposition of $\tau_2(t)$
2	$(\frac{-17}{3},\frac{21}{2})$	$(\frac{-233}{3},\frac{-1371}{2})$	$-(\frac{-17}{3},\frac{21}{2})-(\frac{-35}{3},\frac{81}{2})-(\frac{-59}{12},\frac{45}{8})$
$-\frac{31}{10}$	$(\frac{-16337}{1200},\frac{-77073}{2000})$	$(\frac{-223913}{1200},\frac{5106561}{2000})$	$0(\frac{-16337}{1200},\frac{-77073}{2000})+(\frac{-223913}{1200},\frac{5106561}{2000})$
11	$(\frac{-2057}{12},\frac{3963}{2})$	$(\frac{-28193}{12},\frac{-227631}{2})$	$(\frac{-1337}{12},\frac{1593}{2})+0(\frac{-2057}{12},\frac{3963}{2})-(\frac{-1073}{12},\frac{629}{2})$

Table 4.8: Calculations showing the linear independence of $\Omega(t)$ and $\tau_2(t)$ for various $t \in \mathbb{Q}$ on the elliptic curve $W(u,v,t)$ of (4.25). The last column expresses $\tau_2(t)$ in terms of infinite order and linearly independent elements of $W(\mathbb{Q})$, one of which is always $\Omega(t)$.

power-related to the QRT map L_1 as described in proposition 4.3. If this were true, it would mean there exist integers m, n satisfying

$$m\,\Omega(t) - n\,\tau_2(t) = [0,1,0], \tag{4.28}$$

an identity in t. Using the elliptic curve computational package Apecs (Arithmetic of Plane Elliptic Curves)[3], we can specialise $W(u,v,t)$, $\Omega(t)$ and $\tau_2(t)$ to various $t \in \mathbb{Q}$. Table (4.8) indicates some of the results showing that (4.28) cannot be satisfied, so $\tau_2(t)$ is linearly independent of $\Omega(t)$.[4] In line with the Mordell-Weil theorem over $\mathbb{C}(t)$ [47] and (4.10), this indicates that the rank of $W(\mathbb{C}(t))$ is at least 2. Furthermore, the torsion group in $W(\mathbb{C}(t))$ is also found to be trivial here, so from the list (4.18) the structure of $W(\mathbb{C}(t))$ appears to be $\mathbb{Z}^r$ with $2 \leq r \leq 8$.

4.4 The Rotation Number

A second type of analysis that can be used to highlight differences between two maps leaving fixed each curve in the same foliation is rotation number analysis. This analysis works only when the map has a fixed point in the affine plane around which orbits consist of closed curves surrounding this fixed point.

[3] http://www.math.mcgill.ca/connell/

[4] In contrast, we remark that the corresponding $L_2 = H \circ N$ that could be created in example (4.10) from H and N of table (4.7) *is* power-related to the QRT map L of (4.21), satisfying $L_2^2 = L^{-2}$. This follows since the point of $\mathbb{C}(t^2)$ corresponding to G in table (4.7) has order two so that when we add two of these points it vanishes from the map's composition entirely leaving just the point inherited from H.

4.4.1 Basic Rotation Number Theory

Consider a discrete map from the unit circle to itself. No matter how the map is acting, by considering each orbit singly we can view the map as rotating the circle at each iteration by some angle. This angle by which the map rotates will, on the face of it, vary depending upon the input point. In this section we seek to make precise this notion of the "rotation number" of an integrable map, and see how this number can change with various factors. As usual we will be interested in the case of maps which are globally integrable, in the sense that the invariant quantity is a globally defined one. To define rotation number however we need simply to start with a homeomorphism (continuous transformation) of the circle and the excellent review [17]. Let $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$ be the circle, $\pi : \mathbb{R} \to \mathbb{T}^1$ the natural projection such that $\pi(x) = x + \mathbb{Z}$. Then we have the following definitions:

Definition 4.12. *A map of the circle $\mathbb{T}^1$ is orientation preserving if the orientation of any directed arc segment is the same as the orientation of the image of that arc segment.*

Definition 4.13. *Let $f : \mathbb{T}^1 \to \mathbb{T}^1$ be an orientation preserving homeomorphism. Then a homeomorphism $F : \mathbb{R} \to \mathbb{R}$ is a lift of f if, $\forall x \in \mathbb{R}$, we have*

$$f(\pi(x)) = \pi(F(x)).$$

We can use the lift of a function to give a rigourous definition of the rotation number in the following way.

Definition 4.14. *Let f be an orientation preserving homeomorphism of the circle and F a lift of f. Then the translation number of a point $x \in \mathbb{R}$ under F is given by*

$$\tau(x, F) = \lim_{n\to\infty} \frac{F^n(x) - x}{n}.$$

i.e. the translation number is the average amount that F translates x by.

The translation number has a number of nice properties; two useful ones are summarised in proposition 4.15, the proof of which can be found in the survey [17] and references therein.

Proposition 4.15. *Let f be a homeomorphism of the circle and F a lift of it. Then $\tau(x, F)$ exists $\forall x \in \mathbb{R}$ and secondly the quantity $\tau(x, F)$ is independent of x allowing us to define a single translation number $\tau(F)$.*

The existence of this unique translation number for a homeomorphism of a circle gives us an obvious definition for the rotation number.

Definition 4.16. *Let f be an orientation preserving homeomorphism of the circle and F a lift of it. Then the rotation number of f is*

$$\phi(f) = \pi(\tau(F)).$$

With the definition of rotation number now given, we turn to how it can be used in the context of integrable maps where in general we will have elliptic curves as our object of interest rather than the circle.

4.4.2 Computational Aspects of the Rotation Number

Above we used the simple example of a map acting on the unit circle. In general an integrable map will leave fixed invariant elliptic curves, so the first question to answer is how we recover some notion of rotation in this context. The way to do this is to consider maps where, in some region of the plane, the invariant elliptic curves are topologically equivalent to a circle - they are just closed curves. Then on each such closed curve, which will correspond to (some part of) a level set of an elliptic foliation of the plane, say $E(x, y, t^*) = 0$ for some value $t^* \in \mathbb{C}$, we will be able to consider our idea of rotation inherited from the unit circle case. Indeed, since the definition of rotation number hinges on the map being a homeomorphism, we can automatically include rational maps acting on elliptic curves that are topologically equivalent (i.e. homeomorphic) to the circle since these rational maps on the closed elliptic curves would be homeomorphic maps of the circle. It is quite typical of integrals that yield such invariant closed curves that these curves are concentric (indeed, it is difficult to conceive of a simple alternative when the invariant curves $E(x, y, t) = 0$ must foliate the plane) and so we first decide on what should be the "centre" of these curves for the purpose of angle measurement. The feature we are looking for, then, is a single isolated point (to act as the centre) surrounded by concentric curves, all of which are (parts of) level sets of some invariant quantity $t = t(x, y)$.

One way of ensuring such a feature exists in the foliation $E(x, y, t) = 0$ is to ensure that $t = t(x, y)$ has a local minimum or maximum in the real affine plane. This we can do using the standard second partial derivative test for functions of two variables - if $\frac{\partial^2 t}{\partial x^2}\frac{\partial^2 t}{\partial y^2} - (\frac{\partial^2 t}{\partial x \partial y})^2$ is positive when evaluated at a critical point (x^*, y^*) then that critical point is either a minimum or a maximum.

Proposition 4.17. *Let $E(x, y, t) = 0$ define a foliation of the real plane by elliptic curves. Suppose that E is solvable globally and uniquely for t, so that $t = t(x, y)$. Then each point (x, y) which gives a local minimum or local maximum of t is a point in the plane surrounded by closed concentric curves.*

Proof. Suppose $(x^*, y^*) \in \mathbb{R}^2$ is a local maximum for $t = t(x, y)$. Let $t^* = t(x^*, y^*)$. Now in the neighbourhood of the point (x^*, y^*, t^*) the graph of the function $t(x, y)$ is hemispherical in shape, with lines of latitude being lines of constant t value. Projecting each curve of constant t, $E(x, y, T) = 0$ with T in some neighbourhood of t^* onto one $x - y$ plane gives a single point, that being the value $t = t^*$, surrounded by closed concentric curves, those being the values of t in a neighbourhood of t^*. A similar argument applies for local minima. □

Assume now that $t = t(x, y)$ gives rise to an elliptic foliation of the plane. Assume also that it has a local maximum (or minimum), so we have some elliptic curves in the plane for which some region is a single point surrounded by closed concentric curves. Now suppose also that there is a continuous map that preserves not only each curve in the foliation but also preserves these particular closed contours (we make the distinction between curve and contour here because we wish to introduce the possibility that the closed contour around the fixed point is not the entire curve for that value of t) surrounding the single isolated point. Then the isolated point is necessarily a fixed point of the map. This is a simple consequence of the continuity of the map. Within a small neighbourhood of the isolated point, each small closed contour is being mapped to itself, resulting in only one nearby place for the single point to be mapped to - itself. A fixed point in the phase space of a planar discrete map surrounded by closed curves which are also trajectories of the map is called an elliptic fixed point. Thus we now have two characterisations of the type of phenomena in phase space we are looking for - either an elliptic fixed point of a map, or a local minimum/maximum of the integral of motion of the map. In the former case, the seemingly innocuous requirement that each closed curve surrounding the isolated point be mapped to itself is actually quite crucial. Without this requirement the possibility of an orbit leaving the closed curve exists which would ruin the measurement of the angle between such points. Fortunately for integrals that arise from QRT maps this is always the case, so from here we restrict ourselves to these biquadratics and study them in more detail. Another discussion on the presence of fixed points in maps possessing an integral can be found in the first appendix of [28].

Let $B(x, y, t) = 0$ be a biquadratic in x and y where each coefficient is at most linear in t (thus it defines a typical QRT integral). Suppose that the integral $t = t(x, y)$ possesses a local maximum at (x^*, y^*, t^*). We wish to prove the above claim: that each closed contour surrounding the isolated point is mapped to itself. Recall that the QRT map acts in two stages. First it alters the x-ordinate of the point while leaving y fixed then does the opposite. For the first stage, since B is a quadratic in x, leaving y fixed gives two solutions for x, the point we are mapping to and hence the point we must map to. This image point is clearly on the same closed contour as the preimage. Similarly for the second stage that leaves x fixed. Thus the final result is that the QRT map maps any point on one closed contour to another point on the same closed contour. However the case for non-QRT maps that preserve QRT invariant curves are not subject to the same argument. Nevertheless it turns out that the nonQRT maps we construct that preserve the QRT invariant curves also preserve these individual contours. In fact the individual contours surrounding the fixed point are the only pieces of each level set $t(x, y) =$ constant that are in the real affine plane, so there is no other option but to preserve them as the maps themselves are real.

With the convention that a map always rotates the plane in an anti-clockwise direction, it is fairly clear how to measure the angle between two points relative to a fixed reference point. The lift of a map, while necessary for the theoretical definition of the rotation number, is not so useful for its calculation. In terms of a numerical implementation it is convenient to make use of any function to calculate the argument of a complex number that most mathematical packages contain. To do this, one constructs $z = (x - x*) + i(y - y*)$ and $z' = (x' - x*) + i(y' - y*)$.

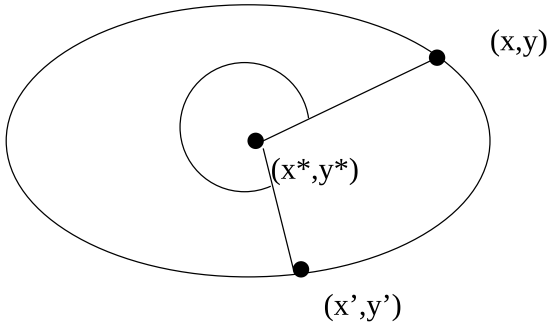

Figure 4.12: Measuring the rotation number.

Then subject to slight adjustments based on which half of the complex plane the two numbers lie in, the rotation from (x, y) to (x', y') relative to $(x*, y*)$ is just $\theta = arg(z') - arg(z)$. Now to assign each curve a rotation number we must take a limit of such quantities. Consider an orbit of points that all lie on some closed curve $\{p_1, p_2, p_3, \ldots\}$. This orbit gives a sequence of rotational angles $\{\theta_1, \theta_2, \ldots\}$. Let $\Theta_k = \frac{1}{k}\Sigma_{i=1}^{k}\theta_i$, so that Θ_k is measuring a cumulative average rotation. If the limit $\lim_{k\to\infty}\Theta_k$ exists, then this, normalised by 2π, is the rotation number we assign to that curve.

The final hurdle in performing experiments to do with rotation numbers is producing examples of maps to which the analysis can be applied. Theoretically, we know which features we are looking for, but constructing examples which contain them is not so simple. A first attempt at creating a QRT map in a way such that all algebraic conditions for its integral containing a level set which is an isolated point went as follows. Let $B(x, y) = \alpha x^2y^2 + \beta x^2y + \delta xy^2 + \gamma x^2 + \kappa y^2 + \epsilon xy + \xi x + \lambda y + \mu$, where each lower case Greek letter is (at most) linear in a symbol t. Solving $B(x, y)$ for $t = t(x, y)$ would give the integral of motion for the QRT map, which we can easily create for any such choices of $\alpha, \beta, \ldots$. Since our eventual goal is to compare the rotation numbers of QRT maps with the rotation numbers of non-QRT maps, we first set $\mu = 0$, since in this case a lot of points on $B(x, y)$ are easily found which in turn leads to more easily created non-QRT maps. Now we just have to ensure that B contains an isolated point (as described above, this is when the function $z = t(x, y)$ contains a local minimum or maximum). We continue now with some actual examples of integrable maps where we can find the rotation number.

4.4.3 Examples and Usage of the Rotation Number

Ultimately our hope for the rotation number is that it can be used to distinguish between different maps by considering their rotation profile. The way we display this data is by plotting the rotation number (normalised by 2π) for a level set as a function of its distance from the fixed point. This distance is not unique; we must choose one direction in which we stray from the fixed point and define the distance to be the length traveled along this direction from the fixed point to the point of intersection with the level set in question. In our calculations, the direction we use to define distance is simply the horizontal distance to the right of the fixed point. Carrying on from example (4.11), there is a fixed point of both the QRT map L_1 and the new map L_2 at $P_f = (-.6317112699, .5844144553)$. The fact that it is a fixed point of both these maps is not surprising; they both preserve the same set of curves and this fixed point is a topological limit of these preserved curves in the sense that near to the fixed point the preserved curves manifest as small closed loops surrounding the fixed point, which is just the situation needed to measure the rotation number. The phase space portrait under L_1 of the area surrounding the fixed point P_f is given in figure (4.13), as well as the line along which we shall be measuring distance from the fixed point. The phase space under L_2 is fairly identical to the naked eye since both maps fill their preserved curves densely. However, the rotation number profiles for the two maps reveal a difference. Figure (4.14) shows the rotation number profile for both maps. As it turns out, by looking at the phase space portraits of the two maps in more detail one can see the rotation number of particular level sets manifest itself. By connecting successive iterates with a line (we will call this a time-phase space portrait), one can see how the map is moving around the curve with each iterate. Note that a high rotation number close to 1 represents a small clockwise rotation while a small rotation number close to 0 represents a small anti-clockwise rotation, so the two are in fact closely related. Rotation numbers away from the two extremes generate a distinctive time-phase space portrait. Figure (4.15) shows the time-phase space portrait for the map L_1 for the orbit closest to the fixed point P_f along our line of constant y (i.e. the orbit of the point $(-.6217112699, .5844144553)$). Figure (4.16) shows the time-phase space portrait for the map L_2 for the orbit of the same

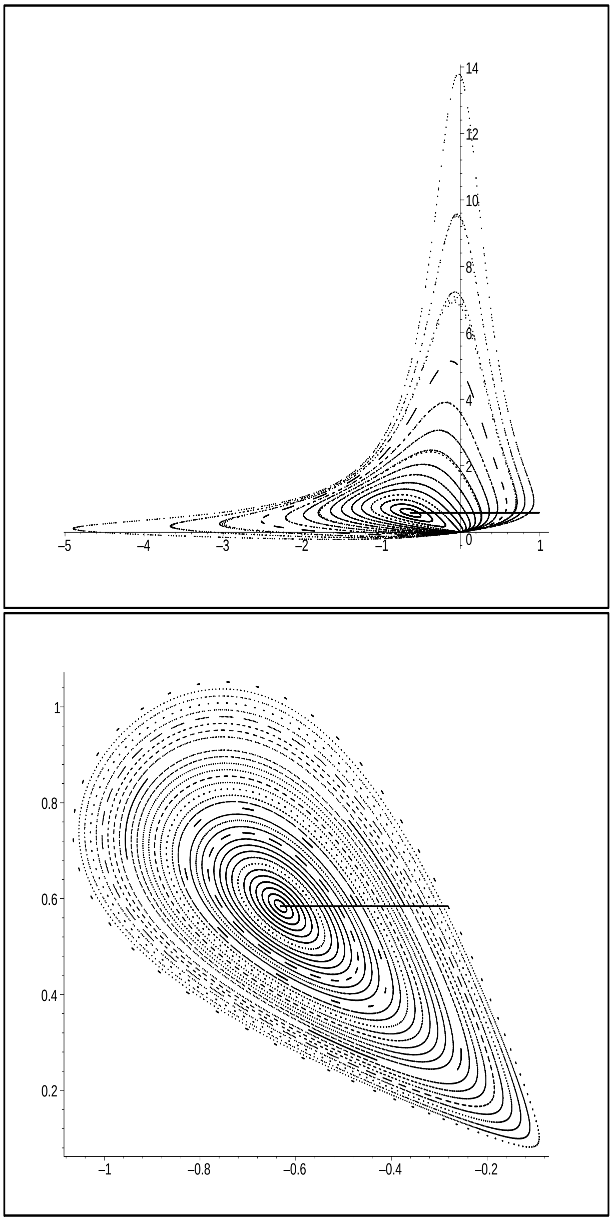

Figure 4.13: Top: Phase space of the QRT map (4.26) around a fixed point, see example (4.11) and the succeeding discussion. Bottom: The part of this phase space where we calculate rotation numbers; this is the part between the fixed point P_f and the base point of the integral at $(0, 0)$.

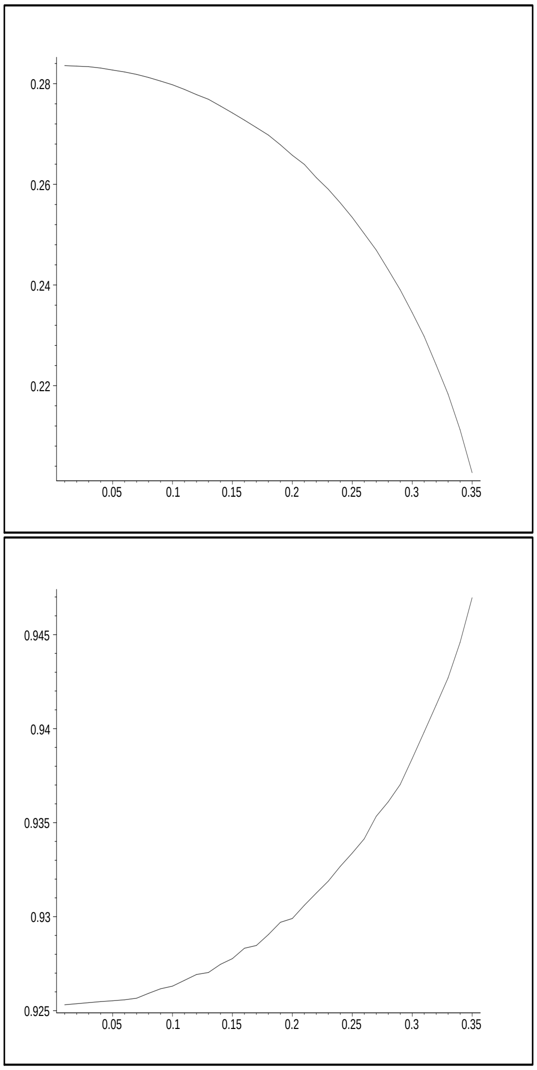

Figure 4.14: Top: Rotation number against distance from fixed point for the QRT map L_1 of example (4.11). Bottom: Rotation number against distance from fixed point for the non-QRT map L_2 of example (4.11). In both cases we measure the rotation number at 0.01 intervals in the x direction.

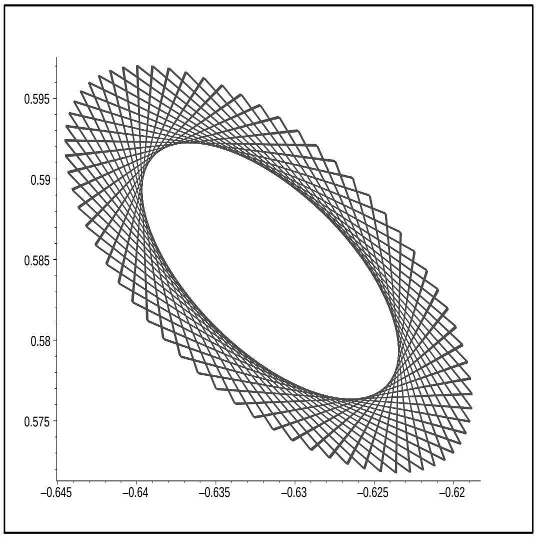

Figure 4.15: Time-phase space plot of the orbit of the point $(-.6217112699, .5844144553)$ for the QRT map L_1. It has a large (that is, far from both 0 and 1) rotation number.

point. Note that the former "criss-crosses" the middle which is always the case for large (that is, far from both 0 and 1) rotation numbers while the latter follows the preserved curve fairly closely.

Despite the apparent differences in the two rotation number profiles in figure (4.14) we note that there seems to be a fractional linear transformation linking the two profiles on the interval we consider. The link was calculated by solving three equations

$$\frac{\alpha\omega_{L_2} + \beta}{\gamma\omega_{L_2} + \delta} = \omega_{L_1} \tag{4.29}$$

where ω_{L_i} is the rotation number for the map L_i. The three equations come by sampling the rotation number at three different points; the "start" (i.e. the curve that is 0.1 units away from the fixed point in the positive x direction), close to the "end" (the curve that is 0.33 units away from the fixed point in the positive x direction) and one curve in the middle of the two (the curve that is 0.17 away from the fixed point in the positive x direction). Solving these three equations with the additional restriction that $\alpha\delta - \beta\gamma = -1$ yields the solution

$$(\alpha, \beta) = (1.573874330, -1.599156584)$$

$$(\gamma, \delta) = (-1.451767190, 0.8397132062).$$

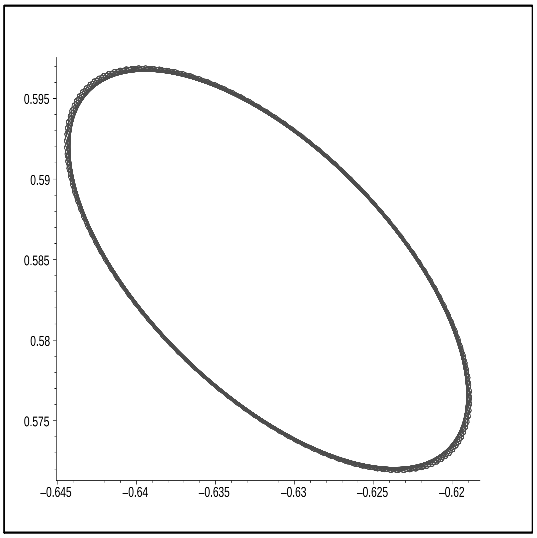

Figure 4.16: Time-phase space plot of the orbit of the point $(-.6217112699, .5844144553)$ for the non-QRT map L_2. It has a small (that is, close to either 0 or 1) rotation number.

So, we assume a fractional linear relation between the two profiles at three different points and we find that away from these points the relation is still followed closely - see figure (4.17). From an elliptic curves perspective, however, it is puzzling that we must resort to fractional linear transformations to find a link between the two. Given that there is no concept of division in the group of points of elliptic curves (and hence in the maps that correspond to the points) we might expect that the relation between the two rotation number profiles should, if one exists at all, be affine in nature. A relation of the type $\omega_{L_1} = \alpha\omega_{L_2} + \beta$ is explainable in terms of the group structure on an elliptic curve while a fractional linear one is not. Trying to fit such an affine relation to the two rotation number profiles in a similar manner fails. The interesting point being shown here is that the maps L_1 and L_2 of example (4.11) appear to be distinct from an algebraic point of view. But instead of being quite different, the rotation number profiles they induce on invariant curves appear to be related. Perhaps the maps are related in a non-algebraic way?

In this final example we perform similar calculations as in the previous. Here we go into a little more detail, examining a QRT map, and both the involutions that make it up. The end result is that we see the rotation number acting in a way that respects map composition. Another important thing to note about this example is

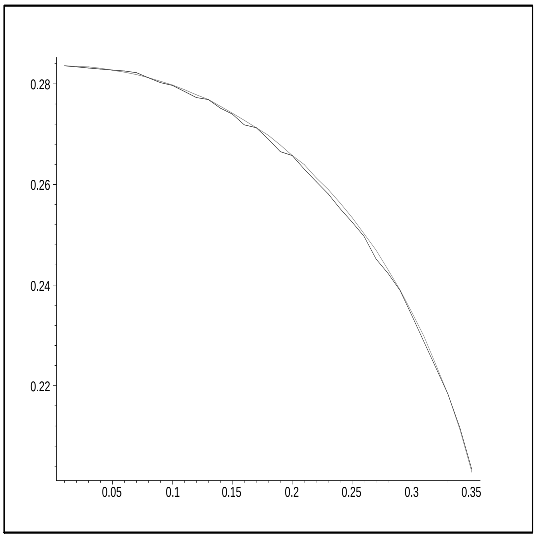

Figure 4.17: Plotted here is the rotation number profile for the QRT map L_1 of example (4.11) as well as the transformed rotation number profile of the non-QRT map L_2 of example (4.11). The transformation done is fractional linear (see equation (4.29)); the exact parameters can be seen in the text. The similarity between the two is suggestive that the two rotation number profiles are exactly related by a fractional linear transformation on the interval we consider.

that the (biquadratic) curve involved was created with the aim of being able to find two independent maps that preserve it in mind. It was not difficult to do, and leads to the question of whether it is difficult to find a curve that has three, or even more, independent maps preserving it.

Example 4.18. We again seek to create two maps that leave fixed each curve in the same foliation and find their rotation number profiles. Let

$$t = -\frac{x^2y^2 - x^2y + y^2 - xy - x + y}{xy^2 - x^2 + xy - x - y}. \tag{4.30}$$

The biquadratics in the ratio on the right hand side of this equation came from a random biquadratic generator which randomly selects a value of $-1, 0$ or 1 for each coefficient. This equality corresponds to a curve defined over $\mathbb{Q}(t)$ given by

$$C(x, y, t) = x^2y^2 - x^2y + txy^2 - tx^2 + y^2 + (-1+t)xy + (-1-t)x + (1-t)y = 0 \tag{4.31}$$

which itself can be thought of as a foliation of $\mathbb{C}^2$ by defining a family of curves by realising t as an actual complex number as opposed to a free variable.

The first map we shall be considering that preserves this curve is the usual QRT map from equation (3.13). This map is, as usual, defined by $L = G \circ H$ where G is the unique involution that preserves C whilst leaving x fixed and H is the unique involution that preserves C whilst leaving y fixed. It is easy to check that C contains the $\mathbb{C}(t)$ point $(0, t-1)$ and so we may use this point to convert C to a Weierstrass form. The result, in standard MAPLE form, is

$$\begin{aligned}W(u,v) =& u^3 + (-\frac{11}{12}t^3 + \frac{11}{8}t^2 - \frac{49}{48}t^4 + \frac{23}{48} + \frac{37}{12}t)u - \frac{581}{216}t^3 + \frac{83}{288}t^2 + \\ &\frac{503}{288}t^4 + \frac{77}{144}t^5 + \frac{65}{144}t - \frac{629}{864} - \frac{521}{864}t^6 + v^2.\end{aligned} \tag{4.32}$$

A second point on B is $(\frac{-1}{t} + 1, 0)$ and we use this point along with theorem (4.1) to calculate the translative points present in L, H and G. Let

$$\begin{aligned}\widetilde{L} : P &\mapsto P + \Omega_L \\ \widetilde{G} : P &\mapsto -P + \Omega_G \\ \widetilde{H} : P &\mapsto -P + \Omega_H.\end{aligned}$$

Then the three points are

$$\begin{aligned}\Omega_L &= (-\frac{5}{12}t^2 + \frac{5}{6}t - \frac{5}{12}, \frac{1}{2}t^3 - \frac{1}{2}t^2 + 1) \\ \Omega_G &= (\frac{7t^4 + 20t^3 - 66t^2 - 4t - 5}{12(-1+t)^2}, -\frac{t^6 - 3t^5 - 4t^4 + 9t^3 + 6t^2 - 2t + 1}{(-1+t)^3})\end{aligned}$$

$$\begin{aligned}\Omega_H = (&-\frac{5t^8 + 28t^7 - 170t^6 + 112t^5 + 529t^4 - 1076t^3 + 764t^2 - 240t + 48}{12(-2+t)(t^3 - 3t - 2)t^2}, \\ &\frac{(t^{12} - 8t^{11} + 23t^{10} + 32t^9 - 313t^8 + 490t^7 + 319t^6 - 1762t^5 + 2174t^4 - 1368t^3 + 532t^2 - 120t + 16)}{2(t^2 - t - 2)(-2+t)(t^3 - 3t - 2)t^3}).\end{aligned}$$

One question to pose immediately is to ask what the rank of W is. If the rank is 1, then searching for any real difference in the rotation number profile of maps generated by points on W is likely futile. However, as it turns out, evidence suggests that W is at least rank 2, as seen in table (4.9). The collapse of the two apparently independent, infinite order points that occurs at $t = -1$ is shown in the points Ω_L, Ω_H and Ω_G at $t = -1$. For this t value, the points are, respectively, $(-\frac{5}{3}, 0), (-\frac{5}{3}, 0)$ and $[0, 1, 0]$ all three of which are clearly finite order. A similar situation occurs for $t = 0$ and $t = 1$; where one of Ω_G or Ω_H is the identity point resulting in Ω_L being rather simply related to the third point. This is of course consistent with the knowledge that $\Omega_L = \Omega_G - \Omega_H$.

t_0	Rank of $C(x,y)_{t=t_0}$
-3	3
-2	2
-1	0
0	1
1	1
2	2
3	2
$-\frac{4}{7}$	5
$\frac{4}{7}$	2

Table 4.9: The minimal rank, from APECS, of W from equation (4.32) for varying values of the parameter t. We specialise the value of t to get an idea of the rank of W over the field $\mathbb{C}(t)$.

With three points, two of which seem to be linearly independent, we can construct a second map that preserves the original curve C by transforming the maps $\widetilde{G'}$: $P \mapsto P + \Omega_G$ and $\widetilde{H'}$: $P \mapsto P + \Omega_H$ so that they act on C instead of W, with these transformed versions being denoted by G' and H'. The algebraic form of these maps, being the composition of three separate birational maps (the conversion function, its inverse, and the addition map on W), is rather cumbersome but they remain easy to operate with numerically. In figures (4.19), (4.20) and (4.21) we see the rotation profiles for, respectively, the maps which add Ω_L, Ω_G and Ω_H on the biquadratic C. The fixed point around which these calculations are performed is $P_f = (1.899129855, .7927931762)$. Distance is again being measured horizontally to the right (the phase space can be seen in figure (4.18)). The profiles for Ω_L and Ω_G look very similar while that of Ω_H is vastly different. Note the difference in scales; the profile for Ω_H is within a very narrow band below 1, meaning that the amount its map rotates by is very small (indicated by how close it is to 1) and that the rotation is small in the clockwise direction (whereas if the rotation number is close to 0, the rotation is small in the anticlockwise direction). Indeed, looking at the numbers that make up these rotation number profiles it is the case that $\omega_L = \omega_{G'} - \omega_{H'}$ where ω_L is the rotation number for the map L and so forth. Since $\widetilde{L} = \widetilde{G} \circ \widetilde{H} : P \to P + \Omega_G - \Omega_H$,

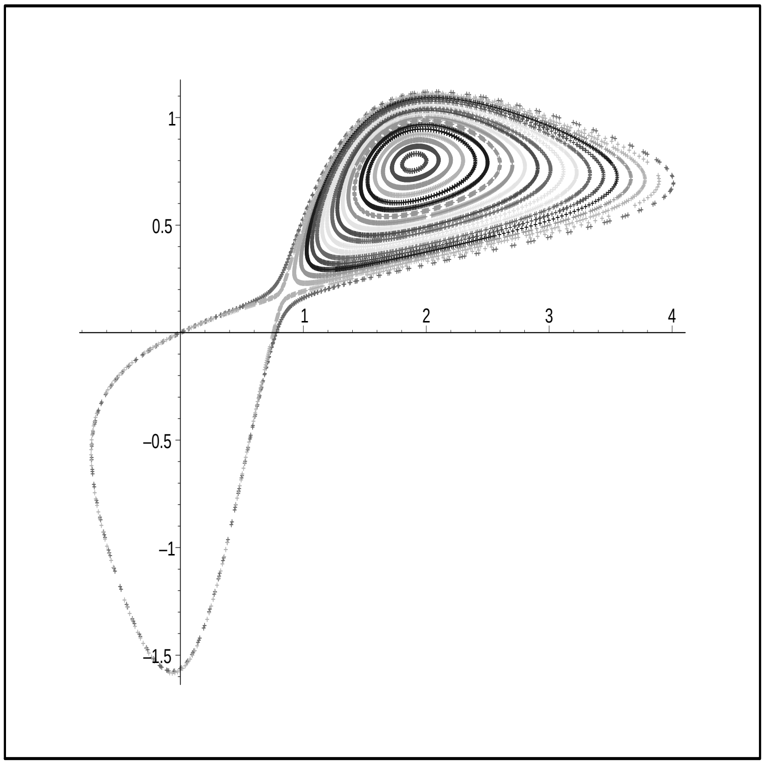

Figure 4.18: The phase space of the QRT map L of example (4.18) around the fixed point in the first quadrant of the real plane. Note that at some distance from the fixed point the level sets of the integral cease being small closed loops around the fixed point.

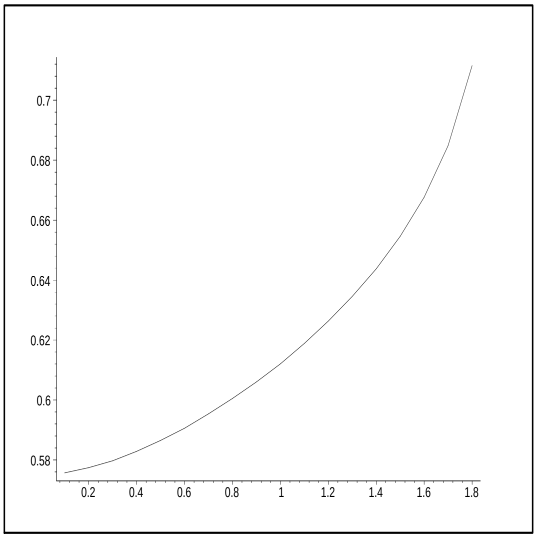

Figure 4.19: Rotation number against distance from fixed point for the QRT map L that preserves the ratio of biquadratics from equation (4.30).

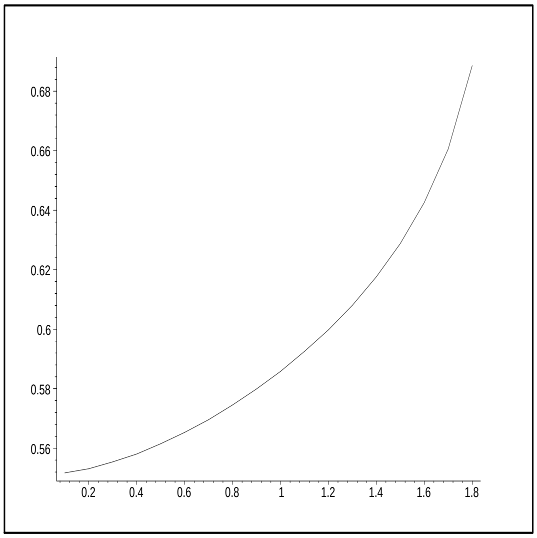

Figure 4.20: Rotation number against distance from fixed point for the non-QRT map corresponding to addition by Ω_G, called G' in the text.

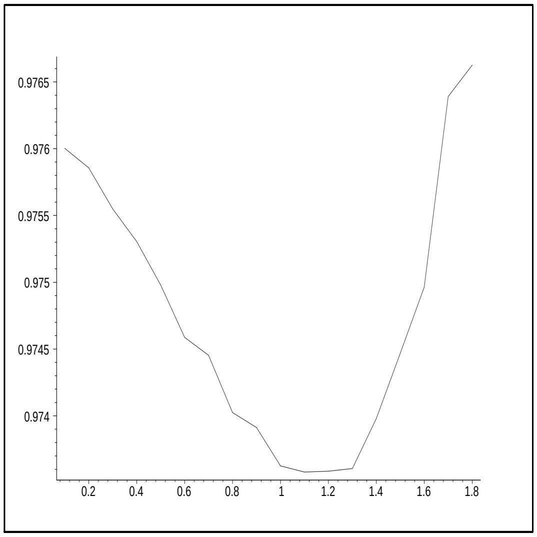

Figure 4.21: Rotation number against distance from fixed point for the non-QRT map corresponding to addition by Ω_H, called H' in the text.

we can alternatively write this as $\widetilde{L} = \widetilde{G'} \circ \widetilde{H'}^{-1}$ so we see that the rotation number adds in the same manner as the translative points, with both of these phenomenon being direct implications of the relation $L = G \circ H$.

Up to this point we have been considering changes in the rotation number of a map as we vary the distance from the fixed point around which closed curves exist. Now we consider instead how it changes as we vary the height of the level set instead. This amounts to a relabeling of the x-axis in previous diagrams. The reason for this is a cue from a note made in an unpublished book of Duistermaat [14] where it is wondered whether a conjecture of E.C. Zeeman, proved in [11], holds true for the rotation number of all maps rather than just the Lyness map for which Zeeman originally conjectured. This conjecture is that the rotation number always varies in a monotonic fashion with the level set of the integral upon which the rotation number was calculated so long as no singular level sets were crossed. So, in addition to the graphs of the rotation number plotted against distance from the fixed point, we also show graphs of the rotation number plotted against the height of the level set. These different graphs are shown in figures (4.22,4.23,4.24). The range of heights we examine takes us from close to the fixed point situated at $(1.899129855, .7927931762)$ which has $t = -0.7156590115$ out to the next singular level set which occurs at $t = -0.5705412594$ (this is the level set of the separatrix seen in figure (4.18)). The nature of these two singular level sets which we look between is realised when one calculates the j-invariant of the curve from equation (4.31). This j-invariant is a rational function of t (i.e. a member of $\mathbb{C}(t)$) and has poles at the above t values, among others - the j-invariant is

$$j(C(t)) = \frac{-(-66t^2 + 44t^3 - 148t + 49t^4 - 23)^3}{(236 + 4420t^3 + 3270t^9 + 89t^{12} - 462t^{11} - 676t^{10} - 5486t^7 + 980t^8 - 1690t^5 + 778t^6 + 7t^4 + 842t^2 - 148t)}. \tag{4.33}$$

A pole in the j-invariant means that in the Weierstrass form $y^2 = f(x)$ of the curve the cubic f has a multiple root, meaning in turn that the curve is not elliptic at all. Note that it is also possible for $t = \infty$ to be a pole of the j-invariant and for this value of t to be approached within the affine plane. Indeed, this is the situation of the previous example, example (4.11). In that example the two singular curves we calculate rotation numbers between are a fixed point, which has $t = 9.749253062$ (this t value being a pole of the j-invariant) and the second singular curve occurs as we approach the base point $(0, 0)$. As we move closer to this base point (from the direction of the second quadrant, where our fixed point lies), the t value of the invariant curves on which nearby points lie goes to $+\infty$ which in turn sends the j-invariant to ∞. Thus in this case the singular curve has a t value of ∞ which is a pole in the j-invariant.

For the current example, in the interval in which we look the behaviour of the rotation numbers for the QRT map L and the map G' is clearly monotonic. The graph of the rotation number for H' looks contradictory to the conjecture. However the point Ω_H for the values of t shown in the graph is relatively large and therefore numerical errors creep in quite easily - as an example for $t = -0.65$ both Ω_G and Ω_L are within a circle of radius two from the origin while $\Omega_H = (-204, 2913)$ to the nearest integer. As an example of these numerical errors, the zoomed part of figure (4.24) show an apparent dip in the rotation number. But the orbit under H' that gave that dipped value of the rotation number is, upon examination, clearly plagued by a numerical error as the orbit actually leaves the level set by a significant amount; see figure (4.25). Given this disclaimer regarding the accuracy of the data for the map H', one could be forgiven for thinking that perhaps H' is simply the identity map as it appears to be rotating by a number close to 1. However this is not the case and the nearness to 1 is a manifestation of the fact that “large points” on Weierstrass elliptic curves can be thought of as “near” the identity point in so far as how the geometry of their addition acts.

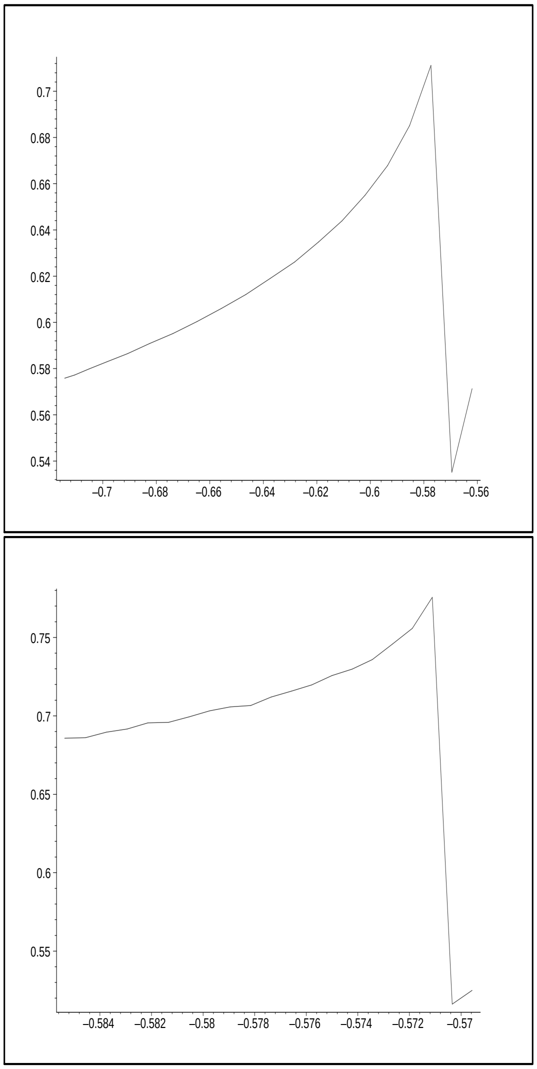

Figure 4.22: Top: Rotation number plotted against height of level set for the QRT map L of example (4.18). Bottom: A zoomed portion of the top figure closer to the singular height $t = -0.5705412594$.

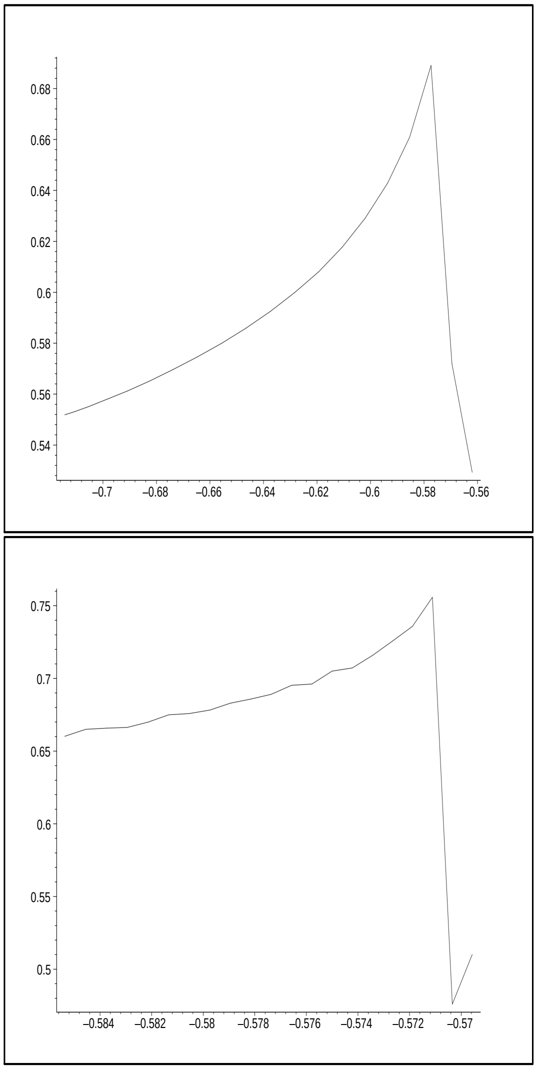

Figure 4.23: Top: Rotation number plotted against height of level set for the map G' of example (4.18). Bottom: A zoomed portion of the top figure closer to the singular height $t = -0.5705412594$.

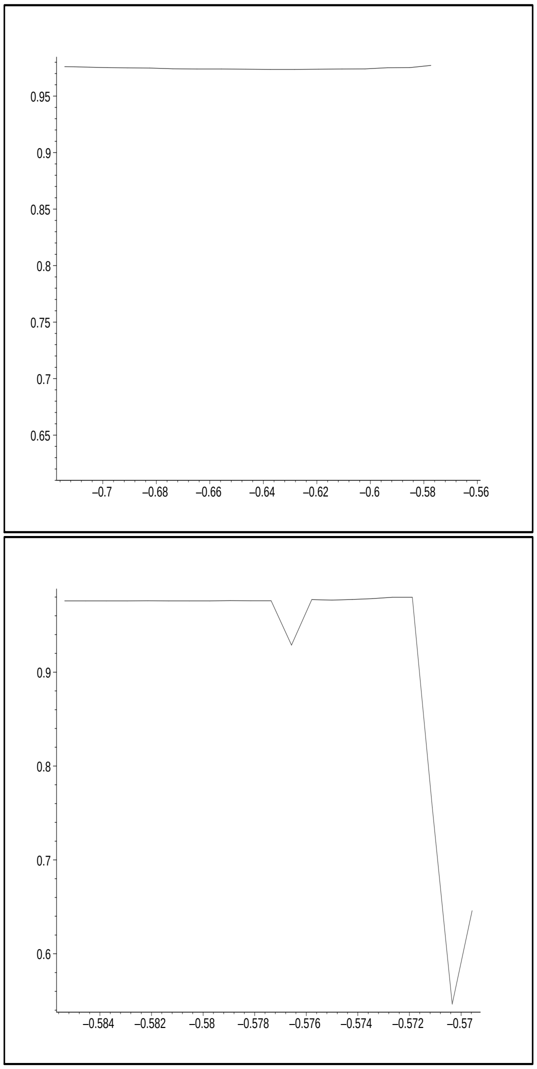

Figure 4.24: Top: Rotation number plotted against height of level set for the map H' of example (4.18). Bottom: A zoomed portion of the top figure closer to the singular height $t = -0.5705412594$.

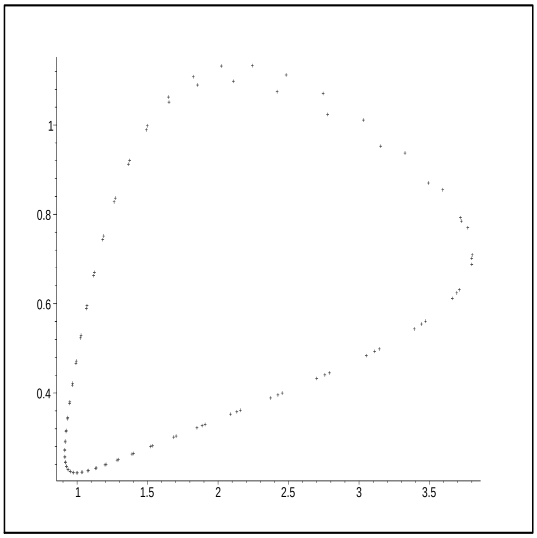

Figure 4.25: Evidence that the numerical calculations for the map H' are not wholly reliable and the reason for the dip in the zoomed part of figure (4.24). Notice that the orbit is not confined to a single curve as we know it should be.

Chapter 5

Broader Applications of the Theorem

In this chapter we turn to a broad application of the main theorem to determine the possible structures of the reversing symmetry group of a typical integrable map. In addition to this we give a generalised notion of integrability by expanding the definition in the algebraic-geometric direction and a few preliminary results about this generalisation. While this generalisation is not an application of the main theorem, it is certainly based on the same algebraic-geometric theory.

5.1 Composition of the Reversing Symmetry Group

We can break the search for the reversing symmetry group of a map into two categories: the symmetries and the actual reversors, see definitions (3.17) and (3.18). Throughout we shall use this division because the defining property of each give different conclusions. The concept of a reversing symmetry group can be applied in the obvious way to any group as the most important property, that of $RL = L^{\pm 1}R$, is stated plainly in the language of groups.

In chapter 4, theorem (4.1) is used to prove that the group of birational maps that preserve an elliptic curve (with both map and curve defined over a field K) is isomorphic to the semidirect product $W(K) \rtimes Aut(W)$ where $W(K)$ is the set of K-rational points on W and $Aut(W)$ is the group of automorphisms of W. This isomorphism leads us to consider the reversing symmetry group of a more general semidirect product. From there, we can specialise to the particular semidirect product $W(K) \rtimes Aut(W)$.

Let G be any group with operation denoted by $\oplus$, and let $H = G \rtimes Aut(G)$. As a matter of notation, a lower case g denotes a member of G and a lower case f denotes a member of $Aut(G)$. Then with the group law is given by

$$(g_1, f_1) \circ (g_2, f_2) = (f_1(g_2) \oplus g_1, f_1 f_2)$$

and

$$(g_1, f_1)^{-1} = (f_1^{-1}(g_1^{-1}), f_1^{-1}).$$

Now we can see what elements of the reversing symmetry group of a member of H "look like". For $S = (g_s, f_s)$ to be a symmetry of $L = (g_l, f_l)$ we require that $SL = LS$. Expanding this equality gives

$$(f_s(g_l) \oplus g_s, f_s f_l) = (f_l(g_s) \oplus g_l, f_l f_s). \tag{5.1}$$

The second component of this equality tells us that f_s must be a symmetry of f_l inside the group $Aut(G)$.

For $R = (g_r, f_r)$ to be a reversor of L, we require that $RL = L^{-1}R$. Expanding we get

$$(f_r(g_l) \oplus g_r, f_r f_l) = (f_l^{-1}(g_r) \oplus f_l^{-1}(g_l^{-1}), f_l^{-1} f_r) \tag{5.2}$$

Again, the second component of the equality tells us that f_r must be a reversor of f_l inside $Aut(G)$.

Examining these relationships in the context of $G = W(K)$ simplifies the first components of these two equalities greatly. Since the group of points on an elliptic curve is Abelian, we shall replace $\oplus$ by the more standard $+$ and write a standard element of $W(K) \rtimes Aut(W)$ as (θ, ι). Furthermore, since we are mainly interested in infinite order maps, which from theorem 4.1, corresponds to $\iota_l = id$ (i.e. pure translations), we will restrict to this case when necessary.

Recall from chapter 2 that the group of automorphisms of an elliptic curve is typically isomorphic to C_2 with possible exceptions C_4 and C_6. Since the group of automorphisms is Abelian, the condition for symmetries that $\iota_s \iota_l = \iota_l \iota_s$ gives no additional information as all ι_s satisfy this. For the condition arising from the first component of equation (5.1) however, we restrict to pure translative maps (i.e. $\iota_l = id$) and have that $\iota_s(\theta_l) + \theta_s = \theta_s + \theta_l$. Of course $\iota_s(\theta_l) = \theta_l$ is true if $\iota_s = id$. If θ_l has order two (this means that L is of order two), then we see that $\iota_s = -1$ is also allowable. This discussion proves some of the following:

Proposition 5.1. *Let L be an infinite order birational map of an elliptic curve W in Weierstrass form. Then the only symmetries of L that also act on this curve are translations. Conversely, all translations on W are symmetries of L.*

Proof. We have proven this proposition except for the cases when $Aut(W)$ is isomorphic to either C_4 or C_6. Let $\theta_l = (x_l, y_l)$ be the translative point for L. Suppose in the preceding discussion that ι_s is a generator of C_4. Then as mentioned in the discussion around theorem (2.48), any Weierstrass equation $y^2 = x^3 + Ax + B$ for the curve has $B = 0$. This corresponds to $\mu = i$, whence the condition $\iota_s(\theta_l) = \theta_l$ becomes $(-x_l, -iy_l) = (x_l, y_l)$ and hence $x_l = y_l = 0$. This restriction on the point θ_l forces it (and so the map L) to be order two. Lastly, suppose that $Aut(W) \cong C_6$. Now the possible orders (that have not already been dealt with) of ι_s and hence of μ are 3 and 6. The second again forces θ_l to be of order two. The first forces $x_l = 0$, and, coupled with $B = 0$ this forces θ_l to be order three. In all cases the map L can not be infinite order as required. The converse is a simple consequence of the fact that the group of points on W is Abelian. □

A similar result holds for the reversors of an infinite order map:

Proposition 5.2. *Let L be an infinite order birational map of an elliptic curve W in Weierstrass form. Then the only reversors of L that also act on this curve are involutions of the form $P \to -P + \theta_r$. Conversely, any map R of W that has the form $R : P \to -P + \theta_r$ is a reversor of L.*

Proof. Let $\widetilde{L} \cong (\theta_l, id) = ((x_l, y_l), id)$. Suppose (θ_r, ι_r) is a reversor of $\widetilde{L}$. Then referring to the above calculation (5.2) of the equality $R\widetilde{L} = \widetilde{L}^{-1}R$, we see that since $f_l = id$ the only non-trivial condition in the case of $G = W(K)$ is that $\iota_r(\theta_l) + \theta_r = \theta_r - \theta_l$. Noting that $-\theta_l = (x_l, -y_l)$, we consider again the possibilities $ord(\iota_r) = 1, 2, 3, 4, 6$ with corresponding values $\mu = 1, -1, \sqrt[3]{1}, i, \sqrt[3]{-1}$ in the relation $(\mu^2 x, \mu^3 y) = (x, -y)$. Clearly $\mu = 1$ fails while $\mu = -1$ satisfies this no matter what x, y are. Having $\mu = \sqrt[3]{1}$ or $\mu = i$ forces $y = 0$ and hence θ_l to be a point of order two (again, this forces $\widetilde{L}$ to be of order two) while having $\mu = \sqrt[3]{-1}$ forces $x = 0$ which once again, coupled with the condition necessary for this type of automorphism to exist (that being $B = 0$), forces θ_l to be of order three. Thus, the only reversors of the map L are those of the form $(\theta_r, -1)$ where the choice of θ_r is arbitrary. Conversely, since L is infinite order it is certainly a pure translation. Then a simple calculation shows that $R\widetilde{L} = \widetilde{L}^{-1}R$. □

Propositions (5.1) and (5.2) serve to give, for a map $\widetilde{L} : P \to P + \Omega$ of an elliptic curve in Weierstrass form W, that subgroup of the reversing symmetry group of $\widetilde{L}$ which also preserves W. The symmetry group consists solely of the translations by points on W (these are maps $S : P \to P + \Psi$) and the reversing symmetry group consists solely of the symmetry group and the "reversing translations", maps of the form $R : P \to -P + \Psi$. In the next section, the finer structure of the reversing symmetry group of an integrable map is studied using these results.

5.2 Structure of the Reversing Symmetry Group

We can use the results of the preceding section, coupled with the structure theorem from [47] given earlier in chapter 4 as equation (4.18) to characterise the symmetry group of certain integrable maps. We look at these groups to find the symmetries of a map that preserves the rational elliptic surface. The requirement that the (reversing) symmetry preserve the same elliptic curve in question seems restrictive. This was explicitly assumed in the previous section, but we present some minor propositions to suggest that the assumption is reasonable.

Proposition 5.3. *Let L be a birational map that preserves an elliptic curve E : $C(x, y) = 0$. Suppose that S is a birational symmetry of L so that $SL = LS$. Then $S(E) = \{\underline{p} : C \circ S^{-1}(\underline{p}) = 0\}$ is preserved by L.*

Proof. Let $\underline{p} \in S(E)$. Then $S^{-1}(\underline{p}) \in E$, but E is preserved by L so $LS^{-1}(\underline{p}) \in E$. Using the fact that S is a symmetry gives $S^{-1}L(\underline{p}) \in E$ or $L(\underline{p}) \in S(E)$. □

It perhaps takes a moment's thought to check that the definition of $S(E)$ given above is correct. What we wish for this to mean is "the image under S of all those points lying on E". Alternatively, these are all the points such that if we map them under S^{-1}, then they lie on E. These are the points that the definition picks out.

A similar result applies for reversors of L, one just has to make the additional assumption that if L^{-1} is birational and preserves a curve then L also preserves this curve.

Proposition 5.4. *Let L be as in proposition (5.3) and suppose R is a birational reversor of L. Then $R(E) = \{\underline{p} : C \circ R^{-1}(\underline{p}) = 0\}$ is preserved by L.*

Proof. Let $\underline{p} \in R(E)$. Then $R^{-1}(\underline{p}) \in E$, but E is preserved by L so $LR^{-1}(\underline{p}) \in E$. Using the fact that R is a reversor gives $R^{-1}L^{-1}(\underline{p}) \in E$ or $L^{-1}(\underline{p}) \in R(E)$. So L^{-1} preserves $R(E)$ i.e. $L^{-1}(R(E)) = R(E)$whence $R(E) = L(R(E))$. □

Following the theme of earlier work, we look at the consequences of propositions (5.3) and (5.4) when the preserved curve is an elliptic curve defined over the field $\mathbb{C}(t)$. We shall, for now, consider only the case when we have two maps that preserve the same (elliptic) curves. In light of the above two propositions we have only two possibilities for maps that leave fixed each curve in a foliation of the plane: first that a given symmetry or reversor also preserves each curve in the foliation or that a given symmetry or reversor permutes the foliation. We leave the second possibility for later.

Theorem 5.5. *Let $E : C(x, y, t) = 0$ be an elliptic curve defined over $\mathbb{C}(t)$ with j-invariant $\neq 0, 1728$. Let W be a Weierstrass equation for E. Suppose that L is an infinite order birational map from E to itself and that birational M also maps E to itself. Then if $\widetilde{L}$ acts on W as a translation, either:*

- *M is an involutory reversor of L and M is conjugate to $\widetilde{M} : P \to -P + \Psi$ on W*
- *M is a symmetry of L and M is conjugate to $\widetilde{M} : P \to P + \Psi$ on W.*

Proof. Since M maps E to itself, we can apply the main theorem to M and hence know that on W it acts as either a translation or an involution of the form $P \to -P + \Psi$. It was shown earlier in proposition (5.1) and proposition (5.2) that the former are symmetries of other translations while the latter are reversors of translations, and interpreting this in the context of maps of E the theorem is proven. □

Theorem (5.5) is telling us that inside the group $\mathcal{L}$ of birational maps of an elliptic curve, any map $M \in \mathcal{L}$ is always related to an infinite order map $L \in \mathcal{L}$ by being either a symmetry or reversor of it. This is quite strong as compared to the generic case of, say, birational maps of the plane where it is quite rare for two arbitrarily chosen maps to be related in such a way.

There are many examples in previous literature that illustrate theorem (5.5), as any integrable map with a symmetric integral serves that purpose. Below is a family of maps to which it applies.

Example 5.6. Consider any biquadratic of the form

$$B(x,y,t) = \alpha(t)x^2y^2 + \beta(t)(x^2y + xy^2) + \gamma(t)(x^2 + y^2) + \epsilon(t)xy + \xi(t)(x+y) + \mu(t)$$

Then due to the symmetry of the curve, the order two map that interchanges x and y preserves B. The usual QRT map preserving this curve can also be created by following the usual QRT construction. This situation is realised in the following example where we set $\mu(t) = 0$ to assist with Weierstrass conversion calculations. To find a point to use as the Weierstass identity, we set $x = 0$ and are left with $B(0,y,t) = \gamma(t)y^2 + \xi(t)y$ which has solutions $y = 0, y = -\frac{\xi(t)}{\gamma(t)}$. We use the generically non-singular point $(0,0)$ to convert B to Weierstrass form and get

$$\begin{aligned}W(u,v,t) =& u^3 + (2\alpha\xi^2\gamma - \frac{1}{3}\gamma^4 - \frac{1}{3}\beta^2\xi^2 - \frac{1}{2}\epsilon\xi^2\alpha + \frac{1}{6}\epsilon^2\gamma^2 + \frac{1}{3}\epsilon^2\xi\beta + \frac{2}{3}\xi\beta\gamma^2 - \xi\epsilon\beta\gamma - \frac{1}{48}\epsilon^4)u + \\ & \frac{2}{27}\gamma^6 - \frac{1}{12}\epsilon^3\gamma\beta\xi + \frac{1}{72}\epsilon^4\gamma^2 + \frac{1}{36}\beta\xi\epsilon^4 - \frac{2}{9}\beta\xi\gamma^4 - \frac{1}{864}\epsilon^6 - \frac{1}{4}\xi^4\alpha^2 - \frac{7}{9}\gamma^2\beta^2\xi^2 - \frac{1}{24}\epsilon^3\alpha\xi^2 - \\ & \frac{2}{27}\beta^3\xi^3 + \frac{4}{3}\gamma^3\alpha\xi^2 + \frac{2}{3}\epsilon\gamma\beta^2\xi^2 + \frac{1}{6}\epsilon^2\alpha\xi^2\gamma - \frac{5}{6}\epsilon\gamma^2\alpha\xi^2 + \frac{1}{3}\beta\alpha\xi^3\epsilon - \frac{1}{3}\beta\alpha\xi^3\gamma - \frac{5}{36}\epsilon^2\beta^2\xi^2 + \\ & \frac{1}{3}\beta\xi\epsilon\gamma^3 - \frac{1}{18}\beta\xi\epsilon^2\gamma^2 - \frac{1}{18}\gamma^4\epsilon^2 + v^2.\end{aligned}$$

The symmetric QRT map that preserves B is

$$\begin{aligned} L: x' &= y \\ y' &= -\frac{\alpha x y^2 + \beta x y + \gamma x + \epsilon y + \beta y^2 + \xi}{\alpha y^2 + \beta y + \gamma} \end{aligned}$$

With $L = G \circ H$ where G and H are the following involutions

$$\begin{aligned} G: x' &= x \\ y' &= -\frac{\alpha x^2 y + \beta x y + \gamma y + \epsilon x + \beta x^2 + \xi}{\alpha x^2 + \beta x + \gamma} \\ H: x' &= y \\ y' &= x \end{aligned}$$

If we did not know *a priori* that L was made up of these two involutions, we could see by inspection that the map H preserved B. Therefore, we know from theorem (5.5) that it is in the reversing symmetry group of L and we might ask the following questions. Does H commute with L? Does it anti-commute with L? This would tell us whether H is a symmetry or a reversor of L.

The following less trivial example is a map that was first exhibited in [32] and further studied in [68], in which the action was found to be conjugate to addition on a Weierstrass cubic. Here we show that it is also reversible.

Example 5.7. Consider the map

$$\begin{aligned} L : x' &= y \\ y' &= \frac{12xy^2 - 3x - 12y + 3y^3 - 10y^2 + 10}{3(4y^2 - 1)(xy - 1)} \end{aligned} \tag{5.3}$$

This map is a birational map of the elliptic curve defined over (possibly some extension of) $\mathbb{C}(t)$ with equation

$$B(x, y) = ((x - y)^2 - 4(xy - 1)^2)((x + y - \frac{10}{3})^2 - 4(xy - 1)^2) - t(xy - 1)^2 = 0 \tag{5.4}$$

For the purpose of illustrating the reversibility of the map (5.3), we note that $M : (x, y) \to (y, x)$ leaves each level set of the integral of the map fixed. Thus immediately from theorem (5.5) we know that this map is either a symmetry or reversor of the map. By looking at the x-coordinate of MLM we can see that M must in fact be a reversor of L.

Let us turn back to the structure of the reversing symmetry group of a rational integrable map and give this structure quite precisely.

Theorem 5.8. *Let L be an infinite order birational map that preserves a rational elliptic surface E defined over $\mathbb{C}(t)$ whose j-invariant is not constant (i.e. not a member of $\mathbb{C}$) and whose Weierstrass equation is W. Then the group of birational symmetries that preserve the same surface defined over $\mathbb{C}(t)$ of L is isomorphic to one of the following groups:*

$$\mathbb{Z}^r (1 \leq r \leq 8), \mathbb{Z}^r \oplus \mathbb{Z}/2\mathbb{Z} (1 \leq r \leq 4), \mathbb{Z}^r \oplus \mathbb{Z}/3\mathbb{Z} (1 \leq r \leq 2),$$
$$\mathbb{Z}^r \oplus (\mathbb{Z}/2\mathbb{Z})^2 (1 \leq r \leq 2), \mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}, \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}, (\mathbb{Z}/3\mathbb{Z})^2,$$
$$(\mathbb{Z}/2\mathbb{Z})^2, \mathbb{Z}/6\mathbb{Z}, \mathbb{Z}/5\mathbb{Z}, \mathbb{Z}/4\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}, 0$$

Proof. From proposition (5.1), we know that any birational symmetry of an infinite order map whose integral has a non-constant j-invariant (hence necessarily it is neither 0 nor 1728) is given by a map of the form $T_\Psi : P \to P + \Psi$. Then let ϕ be the function that maps any symmetry to the point by which it translates, i.e. $\phi(T_\Psi) = \Psi$. Then ϕ is a group isomorphism from the group of symmetries (under composition) to the group of points on W under the associated addition. That ϕ is an isomorphism is easy to check:

- $\phi(T_\Psi \circ T_\Theta) = \phi(T_{\Psi+\Theta}) = \Psi + \Theta = \phi(T_\Psi) + \phi(T_\Theta)$
- $\phi(T_\Psi) = \phi(T_\Theta) \Leftrightarrow \Psi = \Theta$.

From here, the result is a direct application of the Corollary 2.1 of [47]. □

A similar statement can be made regarding the reversing symmetry group.

Theorem 5.9. *Let L be an infinite order birational map that preserves a rational elliptic surface E defined over $\mathbb{C}(t)$ whose j-invariant is not constant (i.e. not a member of $\mathbb{C}$) and whose Weierstrass equation is W. Suppose that the group of symmetries of L that preserves the same surface is $\mathcal{S}(L)$. Then the reversing symmetry group of L is isomorphic to $\mathcal{S} \rtimes C_2$.*

Proof. We know from proposition (5.1) that any symmetry of an infinite order takes the form of a translation $P \to P + \Psi$ and from proposition (5.2) that any reversor takes the form of a reversing translation $P \to -P + \Psi$. Now all that is left to check the composition law. Let us take two general members of the reversing symmetry $R_1 : P \to \iota_1 P + \Psi_1$ and $R_2 : P \to \iota_2 P + \Psi_2$. Then the composition is

$$\begin{aligned} R_1 \circ R_2 &: P \to \iota_1(\iota_2 P + \Psi_2) + \Psi_1 \\ &: P \to \iota_1 \iota_2 P + \iota_1 \Psi_2 + \Psi_1 \end{aligned}$$

This composition law is exactly the semi-direct product law, here the first group is just that made up of the identity and the involution $P \to -P$ and the second group is the group of all points on the elliptic curve, which, from theorem (5.8), is exactly the symmetry group of the curve. □

A more general version of theorem (5.9) was already known - recall lemma (3.25). It was shown that any map that possesses an involutory reversor (that is, a reversor with order two) has reversing symmetry group of the form $\mathcal{R}(L) \cong \mathcal{S}(L) \rtimes C_2$. Of course in the case of infinite order maps on elliptic curves such a reversor always exists - since on the Weierstrass equation of the elliptic curve the involution $P \to -P$ always exists.

5.3 Almost Integrability - Mixing

One concept of a map being almost integrable already exists through continuity. A map might be called almost integrable if it differs from a truly integrable map by only some “small” perturbation. However any such small perturbation will typically render an algebraic-geometric approach to studying such maps impossible. Therefore we now develop a different concept which one may consider as near integrability, one that is susceptible to attack from algebraic-geometric methods. Taking the preservation of each curve in a foliation of (algebraic) curves as the most important aspect of an integrable map, the notion that we shall call “mixing” presents itself quite readily. Rather than requiring a map to preserve each individual curve in the foliation we relax this requirement back to the original definition of foliation preservation. Thus we instead ask that the map merely maps each curve in the foliation to another curve in the foliation, not necessarily the same. Thus, the map “mixes” the foliation. A precise definition of this is:

Definition 5.10. *Let $\mathcal{F}$ be a family of curves. We shall say that a birational map M mixes $\mathcal{F}$ if $C \in \mathcal{F} \Rightarrow M(C) \in \mathcal{F}$.*

It is important to note that this use of mixing is totally distinct from the notion of mixing from ergodic theory. One might question if it is necessary to give this a term at all however it is rather cumbersome and indeed somewhat confusing to state that a map preserves a foliation when we are focusing on the fact that the curves within the foliation are being permuted by the map. This fairly abstract definition can be realised in the case of planar maps in the following way. If M is to be our mixing map and $\mathcal{F}$ our family of curves parameterised by the parameter t, we can write M as

$$\begin{aligned} M : x' &= f(x,y,t) \\ y' &= g(x,y,t) \\ t' &= \tau(t). \end{aligned} \tag{5.5}$$

Here we see that the map can depend, spatially, on t but that the actual mixing which is being performed by the function τ is independent of the spatial coordinates. This characterisation of mixing allows us to treat the mixing as being done by some function τ. It is important to be clear about what M is actually doing. Typically we look for ordinary maps of the plane that *induce* a map τ of the foliating parameter. Thus the function τ is implicit and must be found by examining each example in turn - up to now since we have been considering only integrable maps τ has been the identity function $\tau(t) = t$. Despite its implicit nature, we can see some fairly immediate consequences for this function τ.

Proposition 5.11. *Let M be a (birational) mixing map for some family of elliptic curves $\mathcal{F} = \{\{(x,y) : E(x,y,t) = 0\}, t \in \mathbb{C}\}$. Suppose τ is the function that M induces on t. Then the j-invariant of E is invariant under τ i.e. $j(E(t)) = j(E(\tau(t)))$.*

Proof. Consider any elliptic curve $C \in \mathcal{F}$ with equation $E(x,y,t) = 0$. Let the image of C be the curve $M(C) \in \mathcal{F}$. Let $t_1 \in \mathbb{C}$ correspond to C. Let the induced function on t be τ so that $\tau(t_1)$ corresponds to $M(C)$. Since C and $M(C)$ are birationally equivalent, they have the same j-invariant by proposition (2.54) i.e. $j(E(t_1)) = j(E(\tau(t_1))$. This is true for all $t_1 \in \mathbb{C}$ that make the curve $E(x,y,t_1) = 0$ elliptic thus $j(E(t)) = j(E(\tau(t)))$. □

To begin studying mixing maps, we start with the ideas that first motivated the definition of mixing maps. Consider having a map L that preserves an entire family of elliptic curves defined over $\mathbb{C}$ and a second map M that maps each of these curves to another (or, potentially, the same) curve in the same family.

A first example of such a mixing map can be easily constructed out of QRT maps and integrals; it will illustrate the existence of the function τ and we shall see that $j(E(t)) = j(E(\tau(t))$.

Example 5.12. Let $E : B(x,y,t) = x^2y^2 + t(x-y) + 1 = 0$. Then E is an elliptic curve defined over $\mathbb{C}(t)$ but can also be thought of as a family of curves each defined over $\mathbb{C}$. This is done by allowing t to vary through $\mathbb{C}$, with each different value giving a different (generically elliptic) curve. Then the usual QRT map preserves E and also preserves each curve within the family E. However the map $S : (x,y) \to (y,x)$ does not preserve E but does mix the family E by mapping from the curve with $t = t_0$ to the curve with $t = -t_0$ hence in this case the function τ induced by S is $\tau(t) = -t$. We calculate the j-invariant of the foliation $B(x,y,t) = 0$ and get

$$j(E) = -\frac{4096}{t^4(16t^4 - 27)}$$

which is clearly invariant under $t \to -t$.

This example, with S being finite order, is rather trivial. Several questions are begged. Can we construct mixing maps that are infinite order? Can we construct mixing maps that are unrelated to the QRT map? Firstly we state and prove a theorem claiming that any mixing map cannot mix a family of algebraic curves in an infinite order manner i.e. that the induced function τ must be finite order.

Theorem 5.13. *Let $E(x,y,t) = 0$ be an elliptic curve defined over $\mathbb{C}(t)$ with j-invariant $j(E) \notin \mathbb{C}$. Consider the family of (generically) elliptic curves $\mathcal{F} = \{\{(x,y) : E(x,y,t) = 0\} : t \in \mathbb{C}\}$. Then any birational map M defined over $\mathbb{C}$ that mixes $\mathcal{F}$ has some power that preserves each curve in $\mathcal{F}$.*

Proof. Since E is defined over $\mathbb{C}(t)$, $j(E) \in \mathbb{C}(t)$ i.e. $j(E) = \frac{N(t)}{D(t)}$ where N and D are both polynomials with degree d_N and d_D respectively. Consider the sets $\mathcal{F}_{J_0} = \{C \in \mathcal{F} : j(E) = J_0\}$ i.e. the curves in $\mathcal{F}$ with a particular j-invariant. These are found by finding the solutions to $j(E) = \frac{N(t)}{D(t)} = J_0$ of which there are at most $\mu = max(d_N, d_D)$ and thus can be thought of as the set $\{t_1, t_2, \ldots, t_m\}$ with $m \leq \mu$. In particular each such set is finite and μ is a global bound on their sizes.

Since M is birational, it preserves the j-invariant of elliptic curves. By hypothesis it also mixes $\mathcal{F}$, so it must map curves in a particular $\mathcal{F}_{J_0}$ to curves in the same $\mathcal{F}_{J_0}$. Now we consider how M acts on the sets $\mathcal{F}_{J_0}$. Since M^{-1} exists and is continuous almost everywhere, it cannot map one member of $\mathcal{F}_{J_0}$ to two or more members of $\mathcal{F}_{J_0}$ and hence M must be injective. Also, since M^{-1} exists and is defined globally, M

must be surjective (because clearly M maps $M^{-1}(C)$ to C for all curves $C \in \mathcal{F}_{J_0}$). Thus M is bijective on each set $\mathcal{F}_{J_0}$ and hence it must induce a permutation on the set $\{t_1, t_2, \ldots, t_m\}$.

Now take $n = lcm(\mu, \mu - 1, \mu - 2, \ldots, 1)$. Then M^n acts as the identity on each set $\mathcal{F}_{J_0}$ in that it maps each curve in that set to itself. This is ensured since no matter how $\{t_1, t_2, \ldots, t_m\}$ decomposes into cycles under M's induced permutation, M^n will act as the identity permutation on the above set of t values. This is saying exactly that M^n preserves each curve in $\mathcal{F}_{J_0}$, and, since μ was a global bound, n achieves this simultaneously for all $J_0 \in \mathbb{C}$ i.e. for all curves in $\mathcal{F}$. □

It is clear how this applies to the scenario in example (5.12) - the value of n we must take is just two, even though, as it turns out, the value of μ is 8. This goes to show that the bound given is rather crude for simple examples and indeed when the map M is clearly of finite order, such as in the example given, then taking n to be that finite order is better. However, one can construct a mixer of infinite order (to be clear, infinite order in x and y; the induced τ is necessarily finite order) by, for example, taking the composition of a finite order mixer such as $(x, y) \to (y, x)$ and a map that is known to preserve each curve such as the QRT map (on the face of it, an infinite order mixer that mixes the family trivially). This is the kind of map to which theorem (5.13) is meant to be applied.

If we assume the function τ to also be birational (which is plausible since it is being induced by a birational spatial transformation) then we are considering the finite order birational functions of one complex variable. Knowing that the function τ is finite order gives an easy proof that it is invertible also. The set of invertible, rational functions of one variable is well known to be the Möbius, or fractional linear, transformations.

Definition 5.14. *Let $f : \mathbb{C} \to \mathbb{C}$ be a function of the form*

$$f(z) = \frac{az + b}{cz + d}.$$

Then f is a Möbius transformation whenever $a, b, c, d \in \mathbb{C}$ satisfy $ad - bc \neq 0$. It is possible, by defining $f(\frac{-d}{c})$ and $f(\infty)$ appropriately, to extend a Möbius transformation to the extended complex plane (i.e. the complex plane and a single point at infinity).

This certainly has implications for the way in which τ can mix the family $\mathcal{F}$. For example, τ can have a maximum of two fixed points (if it is not the identity) and thus any mixing map can leave at most two of the mixed elliptic curves fixed. Other facts about Möbius transformations can also be interpreted in our context however the reason they are important at all is because they form the automorphism group of the extended complex plane and therefore they are the invertible functions of our parameter t.

With what could be called the main introductory theorem regarding mixing maps stated and proven, we turn our attention to several examples that have already arisen in literature and one that we construct. Following this we develop some more theory in an attempt to characterise mixing maps as best as is possible.

Taking a concept similar to mixing in a slightly different direction is an unpublished example due to Quispel and Roberts (private communication by J. A. G. Roberts). A way to characterise this difference in words is to say that mixing as defined above concerns maps that map level sets to different level sets within the same foliation. The example we consider now can be said to take level sets to the same level set but in a different foliation. Strictly speaking the map is non-autonomous i.e. it evolves with time, but despite the fact that this thesis is concerned only with autonomous maps it is still a useful case study. Consider the map of the plane

$$\begin{aligned} M_1 : x' &= y \\ y' &= \frac{b_1 + c_1 y}{xy(c_2 + b_3 y)}. \end{aligned} \tag{5.6}$$

The map M_1 maps the curve

$$B_1(x,y) = b_2b_3x^2y^2 + b_1b_4 + b_3c_1y^2x + b_1c_2x + b_2c_2x^2y + b_4c_1y - txy = 0 \tag{5.7}$$

to the curve

$$B_2(x,y) = b_3b_4x^2y^2 + b_2b_1 + b_4c_2y^2x + b_2c_1x + b_3c_1x^2y + b_1c_2 - txy = 0. \tag{5.8}$$

Three more maps M_2, M_3 and M_4 also exist, each map B_i to B_{i+1} with the subscripts i being reduced modulo 4. The other three maps M_{j+1} are defined by increasing the subscripts of b and c in M_j by 1 with the proviso that $b_{j+4} = b_j$ and $c_{j+2} = c_j$. The definitions of these maps can be simplified to just one expression so long as the simple time dependence (that of upshifting the parameters to the next state) in the maps is taken into account; this is why this series of maps can be considered a single non-autonomous map. Each curve B_i has the same Weierstrass form and so we can, by using the correct coordinate transformations, reduce all four of the maps to an action on the same curve. A common Weierstrass is given by

$$\begin{aligned}W =(&\frac{1}{72}b_1b_3c_2c_1t^4 + \frac{1}{72}t^4c_1c_2b_4b_2 - \frac{1}{864}t^6 + \frac{2}{27}b_2^3b_3^3b_1^3b_4^3 + \frac{1}{12}t^3b_2c_2b_3c_1b_1b_4 - \frac{1}{36}t^2b_2c_2^2b_3c_1^2b_1b_4-\\ &\frac{1}{3}tb_2^2c_2b_3^2c_1b_1^2b_4^2 - \frac{5}{9}b_1^2c_2^2b_2^2b_3^2b_4^2c_1^2 - \frac{1}{9}b_1^2c_2^3b_2b_3^2b_4c_1^3 - \frac{1}{9}b_1c_2^3b_2^2b_3b_4^2c_1^3 - \frac{1}{3}b_1^2c_2^2b_2b_3^2c_1^2tb_4-\\ &\frac{1}{9}b_1^3c_2b_2^2b_3^3c_1b_4^2 - \frac{1}{3}b_4^2c_1^2b_2^2c_2^2b_3tb_1 - \frac{1}{9}b_4^3c_1b_2^3c_2b_3^2b_1^2 - \frac{1}{9}b_2b_3^3b_1^3b_4c_2^2c_1^2 - \frac{1}{9}b_2^3b_3b_1b_4^3c_1^2c_2^2-\\ &\frac{1}{18}b_2^2b_3^2b_1^2b_4^2t^2 + \frac{2}{27}b_1^3c_2^3b_3^3c_1^3 - \frac{1}{36}b_1^2c_2b_2b_3^2c_1t^2b_4 + \frac{2}{27}b_4^3c_1^3b_2^3c_2^3 - \frac{1}{36}b_4^2c_1b_2^2c_2b_3t^2b_1-\\ &\frac{1}{18}t^2b_3^2b_1^2c_2^2c_1^2 - \frac{1}{18}t^2b_2^2b_4^2c_1^2c_2^2 + \frac{1}{72}t^4b_2b_3b_1b_4)+\\ &(-\frac{1}{3}b_1^2b_3^2c_2^2c_1^2 - \frac{1}{3}b_1^2b_3^2b_2^2b_4^2 + \frac{1}{3}b_1^2b_3^2c_1b_2b_4c_2 + \frac{1}{3}b_1b_3b_2c_2^2c_1^2b_4 + \frac{1}{6}b_1b_3b_2b_4t^2 + b_1b_3c_1c_2tb_4b_2+\\ &\frac{1}{3}b_1b_3b_2^2c_1b_4^2c_2 + \frac{1}{6}b_1b_3c_2c_1t^2 - \frac{1}{3}b_2^2c_1^2b_4^2c_2^2 + \frac{1}{6}b_2b_4t^2c_1c_2 - \frac{1}{48}t^4)u + u^3 + v^2.\end{aligned} \tag{5.9}$$

On this curve, each of the four maps correspond to translation by a point, these points being, respectively,

$$\begin{aligned}\Omega_1 &= (\frac{b_1b_2b_3b_4 + b_1b_3c_1c_2 - 2b_2b_4c_1c_2}{3} - \frac{t^2}{12}, \frac{b_2b_4c_1c_2(t + 2b_1b_3)}{2})\\ \Omega_2 &= (\frac{b_1b_2b_3b_4 + b_2b_4c_1c_2 - 2b_1b_3c_1c_2}{3} - \frac{t^2}{12}, \frac{b_1b_3c_1c_2(t + 2b_2b_4)}{2})\\ \Omega_3 &= (\frac{b_1b_2b_3b_4 + b_1b_3c_1c_2 - 2b_2b_4c_1c_2}{3} - \frac{t^2}{12}, \frac{b_2b_4c_1c_2(t + 2b_1b_3)}{2})\\ \Omega_4 &= (\frac{b_1b_2b_3b_4 + b_2b_4c_1c_2 - 2b_1b_3c_1c_2}{3} - \frac{t^2}{12}, \frac{b_1b_3c_1c_2(t + 2b_2b_4)}{2}).\end{aligned}$$

Notice that $\Omega_1 = \Omega_3$ and $\Omega_2 = \Omega_4$. Mathematically, this is because in each term in each point, b_i always occurs with a partner b_{i+2} (and of course c_i necessarily occurs with a c_{i+2} partner; itself) so upon upshifting twice we are left with the same expression. Knowing that this common Weierstrass exists however, we can turn any of these maps into a map that preserves any one of the B_i's by using the appropriate coordinate change functions. It is difficult to categorise this map. On one hand it certainly shares some properties in common with mixing maps and therefore warrants a mention here, but on the other hand it is ultimately a non-autonomous dynamical system which puts it into a considerably different category.

A paper in the literature where mixing has shown up without specific recognition is in [31]. Mentioned earlier as a paper where some non-QRT maps are exhibited, the way these are constructed is, from one point of view, similar to mixing maps.

While the maps ultimately come from lattice equations, it is noted that the maps can also be viewed as maps that do not quite preserve a quantity $t = t(x, y)$ but instead switch that quantity t to either $\frac{1}{t}$ or $-t$, whence the second iterate returns the quantity to t again. These two functions $t \to \frac{1}{t}$ and $t \to -t$ are very simple Möbius transformations. An implication of this is that, in the first case, $t + \frac{1}{t}$ is preserved and in the second case t^2 is a preserved quantity. Of the four examples that are detailed, one map is QRT and another two act on curves of genus zero. The fourth has a genuinely elliptic biquadratic as the family of curves it is mixing, but the map itself is lengthy enough to be excluded from the paper (though details enough are given to reconstruct it) and is noted to be a manifestation of the group law on elliptic curves. It is this fourth example that has the strongest relation to our discussion on mixing, but we are able to construct a much simpler example through elementary methods.

Earlier we noted that one way to create a potentially infinite order mixing map is to split it into two components; one spatially finite order map that performs the mixing and another (spatially infinite order) map that preserves each curve in the foliation. A logical first attempt at constructing a mixing map through such a method is to take a QRT map to be the map that fixes each curve in the foliation and a simple involution to perform the mixing. A modification of this approach will yield our mixing map through elementary methods in the following example.

Example 5.15. Creating our mixing map through the two component method means that first we must ensure our family of curves can be mixed and that we can find a mixer. Secondly, because we'll need some other map that leaves fixed each curve in the foliation we choose to remain with biquadratics. This way we always have everything built around the QRT map to work with. The first mixer one is likely to think of, given that it has arisen a lot previously, is the switch $(x, y) \to (y, x)$. Furthermore, with this choice of mixer the easiest way to actually mix the family is to send $t \to -t$. For this to happen we need to choose our ratio of biquadratics such that

$$t = \frac{N(x, y)}{D(x, y)}$$

where $N(x, y) = -N(y, x)$ and $D(x, y) = D(y, x)$. Enforcing these conditions on generic biquadratics yields the necessary ratio

$$t = \frac{\beta_0(x^2y - y^2x) + \gamma_0(x^2 - y^2) + \xi_0(x - y)}{\alpha_1 x^2y^2 + \beta_1(x^2y + xy^2) + \gamma_1(x^2 + y^2) + \epsilon_1 xy + \xi(x + y) + \mu}. \tag{5.10}$$

Now equation (5.10) is a relation with $t(x, y) = -t(y, x)$ so that the family of curves defined in the obvious way is mixed by the map $S : (x, y) \to (y, x)$. Now let $L = G \circ H$ be the QRT map with invariant quantity given by $t(x, y)$ above, with G being the involution that leaves x fixed and H the involution that leaves y fixed. Our first attempt at an infinite order mixing map is $M_1 = S \circ L$. However, M_1 turns out to be a map of order two. This is the case because of the particular symmetries and anti-symmetries in the form of $t(x, y)$ - it is not difficult to check on a computer that $S \circ G \circ S = H$ (equivalently $S \circ H \circ S = G$). With this relation in mind we see that

$$M_1^2 = (SL)(SL) = (SGH)(SGH) = (SGSGS)(SGSGS),$$

which quickly simplifies to the identity map. As a heuristic, let us consider the map M_1 as it pertains to a Weierstrass curve. It is composed of three involutions, each of which are most likely of the form $P \to -P + \Omega_i$ and when three such maps are composed together the result is a fourth map of the form $P \to -P + \Omega$ which is necessarily of order two. Note that despite S not preserving a generic biquadratic, all three involutions must be reducible to actions on some Weierstrass equation as the biquadratics being mixed must have the same j-invariant. This thinking suggests that we should instead consider $M_2 = S \circ H$ (or similarly $S \circ G$). We can easily prove that this map is infinite order. Consider

$$M_2^2 = SHSH = GH = L.$$

So we have in fact constructed a mixing map of a family of biquadratics that is the square root of the QRT map for that family of biquadratics.

Example (5.15) is a little reminiscent of a paper of Quispel, [48]. In this paper a map whose square is the QRT map is detailed. However in this case, the map is an alternating map meaning that the map is non-autonomous, but the action of the map is the same every second iterate. The similarity is superficial though since the methods of creating the maps in question are completely different and motivated by different things. Regardless, this example raises the question of which roots of the QRT maps can be realised with autonomous maps and also which QRT maps (meaning the QRT maps with which integrals) can have their roots taken.

Now that some examples have been considered we further the theory of mixing maps. We seek to characterise the kind of mixing maps that we most commonly encounter - cases where the family being mixed is made up of level sets of an elliptic curve defined over $\mathbb{C}(t)$. Let our mixing map be M and our elliptic curve C_1 defined over $\mathbb{C}(t)$ have the equation $E_1(x, y, t) = 0$. In the case we are considering M maps C_1 to another curve elliptic curve C_2 also defined over $\mathbb{C}(t)$. Let the equation for this curve be $E_2(x, y, t) = 0$. In fact E_2 must be the same as E_1 with a substitution $t \to \tau(t)$ where τ is the function that M induces on t (recall proposition (5.11)). Since C_1 is mapped to C_2 by the birational map M, they have the same j-invariant and we can take a Weierstrass curve W (let us abuse notation and give $W(u, v, t) = 0$ for its equation also) for C_1 and know that it is also a Weierstrass form for C_2. Let $\phi_1 : C_1 \to W$ be a conversion function from C_1 to W and $\phi_2 : C_2 \to W$ be a conversion function from C_2 to W. Consider the map on W obtained from the composition $\phi_2 \circ M \circ \phi_1^{-1}$; call this map $\widehat{M}$. Being a birational map of W, $\widehat{M}$ follows the usual rule for morphisms (theorem (2.49)) of an elliptic curve in Weierstrass form - it is the composition of an isogeny and a translation. Thus we can say that mixing maps are “almost conjugate” to a map of the form $P \to \pm P + \Omega$ on a Weierstrass curve by the relation $M = \phi_2^{-1} \circ \widehat{M} \circ \phi_1$.

Interestingly, we can follow this argument conversely. Let us now suppose that we have an elliptic curve C_1 defined over $\mathbb{C}(t)$ with the equation $E_1(x, y, t) = 0$. Consider the j-invariant of this curve and write it as $j(C_1) = f(t)$. Suppose that for some birational $\tau : \mathbb{C} \to \mathbb{C}$ we have the invariance $f(\tau(t)) = f(t)$. This was a necessary condition for a mixer to exist in the above paragraph; we now show that it is also sufficient. We use τ to create a second foliation C_2 with equation $E_1(x, y, \tau(t)) = 0$. Now we know that the j-invariant of this second curve has $j(C_2) = f(\tau(t)) = f(t)$ and so we can find a single Weierstrass curve W (again, with equation $W(u, v, t) = 0$) which is conjugate to both C_1 and C_2. Let $\phi_1 : C_1 \to W$ and $\phi_2 : C_2 \to W$ be two conversion functions. Now we create our mixer. As suggested by the above paragraph, we form the composition $\phi_2^{-1} \circ \widehat{M} \circ \phi_1$ where $\widehat{M}$ can be any morphism of W. The first obvious idea is to take the identity map but any composition of an isogeny and translation on W will also give us a mixer from C_1 to C_2.

Example 5.16. Let us revisit the earlier example (5.12) in light of this new theory. Here we have the curve

$$C_1 : B(x,y,t) = x^2y^2 + t(x-y) + 1 = 0$$

which has j-invariant

$$j(E) = -\frac{4096}{t^4(16t^4 - 27)}.$$

Now we can see that this is invariant under $t \to \iota t$ where $\iota^4 = 1$ but to avoid having to use complex coefficients we let $\iota = -1$. This provides our second curve C_2 with equation $B(x,y,-t) = x^2y^2 - t(x-y) + 1 = 0$. Using MAPLE to find the Weierstrass forms of C_1 and C_2 is particularly helpful as it gives us the same Weierstrass for both C_1 and C_2 automatically. From here we note the conversion functions, also from MAPLE, $\phi_1 : C_1 \to W$, $\phi_1^{-1} : W \to C_1$, $\phi_2 : C_2 \to W$ and $\phi_2^{-1} : W \to C_2$. Constructing a mixer by taking the composition $M = \phi_2^{-1} \circ \phi_1$ (so here we have $\widehat{M}$ as the identity function) and then substituting in $t = \frac{-x^2y^2-1}{x-y}$ and simplifying ends up yielding the surprising result $M : (x,y) \to (-x,-y)$. Upon reflection, this is obviously a mixer of the foliation defined by C_1 but it was not noticed until this approach showed its existence. Before this, we had only noticed the mixer $S : (x,y) \to (y,x)$. Regarding this second mixer, it is possible to form the composition $\widehat{S} = \phi_2 \circ S \circ \phi_1^{-1}$ which, according to our theory, must be a composition of an isogeny and a translation on W. In practice, it is difficult to discern just what the isogeny and translation must be. On one hand, it is impossible to determine them by the kind of observe-and-guess approach since the algebraic form of $\widehat{L}$ is incredibly complicated, but on the other hand it is difficult to find enough points on W to determine what the isogeny and translative point must be by using substitution.

The above constructions and discussion is proof of the following theorem.

Theorem 5.17. *Let C be an elliptic curve defined over $\mathbb{C}(t)$ with equation $E(x,y,t) = 0$. Let the j-invariant of C be $f(t)$. Then there is a mixer for the foliation defined by varying the parameter t through $\mathbb{C}$ if and only if $f(t) = f(\tau(t))$ for some finite order fractional linear τ.*

Proof. Suppose that $f(t) = f(\tau(t))$ for τ fractional linear. Then we construct a second curve C_2 with equation $E(x, y, \tau(t)) = 0$ which necessarily has j-invariant $f(\tau(t)) = f(t)$. Let $\phi_1 : C \to W$ and $\phi_2 : C_2 \to W$ be two conversion functions. Then the composition $M = \phi_2^{-1} \circ \phi_1$ maps C to C_2. Since τ is a Mobius transformation it is invertible and thus the foliation defined by varying the parameter t through $\mathbb{C}$ is the same for both C and C_2 and so M mixes this foliation. Conversely, suppose that we have a mixer M of the foliation defined from C. Then M induces a transformation of t which we denote τ. Construct C_2 defined over $\mathbb{C}(t)$ with equation $E(x, y, \tau(t)) = 0$. Then by definition, M maps C to C_2 and thus the j-invariants of these two curves are equal whence $f(t) = f(\tau(t))$. That τ must be fractional linear here comes since it must be invertible (by virtue of M itself being invertible and mapping whole curves to whole curves by definition of being a mixer), and the Möbius transformations are exactly the automorphisms of $\mathbb{C}$. □

5.4 $\mathbb{C}(t)$ Points on Biquadratics

This section looks at two kinds of biquadratics defined over $\mathbb{C}(t)$ that arise frequently and historically in planar integrable dynamical systems. These are defined below as biquadratics of type McMillan which is a subclass of biquadratics of type QRT. In particular it looks at what kind of points with coordinates in $\mathbb{C}(t)$ can lie on such biquadratics. The reason one would want to find (non-singular) points on these curves is so that theorem (4.1) can be applied - this theorem requires the existence of a point to work. This is because in the internal workings of the theory, one non-singular point is required to be mapped to $[0, 1, 0]$ in the Weierstrass setting. Thus we begin with an existence type theorem showing that a $\mathbb{C}(t)$ point exists on most QRT biquadratics then follow on with several more constructive theorems showing how to actually find such a point on fairly large classes of QRT biquadratics.

Definition 5.18. *Let $B(x, y) = ax^2y^2 + bx^2y + cxy^2 + dx^2 + exy + fy^2 + gx + hy + J$ be a biquadratic defined over the field $\mathbb{C}(t)$. Then B will be said to be of type McMillan if the only coefficient involving t is the constant term, i.e. $J = j - t$, and if $a \neq 0$ (all lower case letters are complex numbers).*

Definition 5.19. *Let $B(x,y) = ax^2y^2 + bx^2y + cxy^2 + dx^2 + exy + fy^2 + gx + hy + j$ with $a \neq 0$. Then B is said to be of type QRT if each lower case letter is affine in t. This means that upon solving $B = 0$ (uniquely) for t, the result is a ratio of biquadratics each defined over $\mathbb{C}$.*

Theorem 5.20. *Let B be a type QRT biquadratic satisfying condition Q (this constructed and technical condition is defined within the proof). Then there is a point of the form $(X(t), Y^*)$ on B, where X is fractional linear in t and Y^* is constant.*

Proof. We prove the result by finding an equation such that if Y^* satisfies it, then the theorem follows. Consider B as a quadratic in x, say $\alpha(y,t)x^2 + \beta(y,t)x + \gamma(y,t) = 0$, where α, β and γ are all linear in t and quadratic in y. Now if $\alpha = 0$, make the substitution that reverses the role of x and y and continue. If, after the substitution, $\alpha = 0$ again then B must be a conic and thus the result of theorem (4.1) won't apply regardless. So supposing $\alpha \neq 0$, we use the quadratic equation to solve for x to get $x = \frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$. Now for x to be in the field $\mathbb{C}(t)$ we require that the term under the square root sign be a perfect square. This term, $\beta^2 - 4\alpha\gamma$ is quadratic in t and quartic in y. So again we apply the quadratic equation to solve for t and consider its discriminant. We wish this discriminant to be 0, since then as a quadratic in t there is only one root and thus we have what we need - a perfect square. This new discriminant is degree 8 in y and also involves the coefficients of B. Define condition Q_1 to be the condition that this new discriminant is a genuine polynomial in y. Over the complex numbers this equation has at least one solution. Define Y^* to be a solution. Then this value of y makes $\beta^2 - 4\alpha\gamma$ a perfect square (supposing that upon substitution it is still quadratic in t; this is condition Q_2. Condition Q is a conjunction of Q_1 and Q_2) and thus makes x fractional linear in t. □

The way this proof works allows us to use any base points of the biquadratic which may already be known to hopefully find other points. Supposing a base point (x_b, y_b) is known, one can create a quadratic in x by substituting $y = y_b$ and knowing that this factorises into two linear factors, one being $(x - x_b)$, to find a second $\mathbb{C}(t)$ point on the curve. Thus an alternative way of performing the calculation involved in proving theorem (5.20) is to find a base point and then, fixing one of the coordinates, find the other point which must exist (and furthermore must lie in a field whose size is restricted by the coefficients of the biquadratic and the fixed ordinate).

Definition 5.21. *Let $I(x,y) = \frac{N(x,y)}{D(x,y)}$ be the integral of some integrable map. Then a point (x_0, y_0) is a base point of the system if $N(x_0, y_0) = D(x_0, y_0) = 0$, so that the height of the base point is ill-defined.*

Proposition 5.22. *For the type QRT biquadratic $B = (a_0+a_1t)x^2y^2+(b_0+b_1t)x^2y+\ldots+(j_0+j_1t)$ the y-coordinates of the base points are given by the solutions to the resultant (with respect to x) of the two polynomials $N(x) = a_0y^2x^2+b_0yx^2+\ldots+j_0$ and $D(x) = a_1y^2x^2+b_1yx^2+\ldots+j_1$*

Proof. Two polynomials simultaneously vanish when their resultant is zero. So treating the two bivariate polynomials as polynomials just in x with coefficients in y yields a resultant that is a degree 8 polynomial in y. The solutions to this are the y-coordinates of points where $N(x)$ and $D(x)$ will share a common root. □

While theorem (5.20) is largely non-constructive because the value Y^* will not be found in a useful exact algebraic form, it is sometimes possible to do so. Such a situation occurs in the example below.

Example 5.23. Let

$$\begin{aligned}B = x^2y^2 + (1+t)x^2y + (5+t)xy^2 + (1+t)x^2 + (1+t)xy+\\ (1+t)y^2 + (1+t)x + (1+t)y + (1+t).\end{aligned}$$

Now we substitute $y = Y$ into B and calculate the discriminant of the resulting quadratic in x. This discriminant is

$$\begin{aligned}D_1 = (Y^4 - 2Y^3 - 5Y^2 - 6Y - 3)t^2 + (6Y^4 - 6Y^2 - 12Y - 6)t+\\ (21Y^4 + 2Y^3 - Y^2 - 6Y - 3).\end{aligned}$$

Calculating the discriminant of this quadratic in t gives

$$D_2 = -16(3Y^2 - 13Y - 13)Y^4(Y^2 + Y + 1).$$

Let $y = \frac{13}{6} + \frac{5\sqrt{13}}{6}$, a zero of this discriminant. Then by construction, D_1 becomes a perfect square and the two x solutions are $x = -4$ (which gives a base point of B) and $x = -\frac{4(1+t)}{16+3t}$, a more conventional $\mathbb{C}(t)$ point. Any of these points can then be used in the algorithm for converting B to a Weierstrass equation.

This example is quite synthetic, as it was constructed solely to illustrate the method of theorem (5.20). For a more “real” example we consider the curve defined by the equation $B(x, y) = x^2y^2 - t(x^2 + y^2) - 2xy + 1 = 0$ which is considered in [30] and as an example in chapter 4. In that earlier paper we were unable to find a $\mathbb{C}(t)$ point on $B(x, y)$ although results suggested that one should exist. We avoided the problem at that time by switching t to t^2, whence a $\mathbb{C}(t)$ point could be found. However using this new method we can find one on the original curve.

Example 5.24. First we must find a base point of $B(x, y)$ considered as an integral. This entails solving, simultaneously, the equations

$$\begin{aligned}(xy - 1)^2 &= 0\\ x^2 + y^2 &= 0.\end{aligned}$$

One solution is $(\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}, \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}})$. So define $x^* = \frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}$ and substitute this into B giving the quadratic in y

$$(-t - i)y^2 - (1 - i)\sqrt{2}y + (ti + 1).$$

Solving this quadratic gives the constant (in t) value from the base point and a second answer of $-\frac{(1+i)(t-i)}{\sqrt{2}(t+i)}$. So one $\mathbb{C}(t)$ point on B is $(\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}, -\frac{(1+i)(t-i)}{\sqrt{2}(t+i)})$. If it is desired that both coordinates of the point vary with t (to make a picture more illustrative for example) one could perform a similar procedure with the y value from this new point - substitute it into the equation for B and solve for x, and choosing the other value. In general this second value may not vary with t but it is likely to.

Now that we have shown in a concrete manner the method of theorem (5.20) we will perform a similar exercise symbolically. Let us take a general QRT biquadratic and consider it as an integral (for finding base points). So, let $I(x, y) = \frac{N}{D} = \frac{a_0x^2y^2+b_0x^2y+c_0xy^2+d_0x^2+e_0xy+f_0y^2+g_0x+h_0y+j_0}{a_1x^2y^2+b_1x^2y+c_1xy^2+d_1x^2+e_1xy+f_1y^2+g_1x+h_1y+j_1}$ and recall that considering it as an integral is implicitly setting $I(x, y) = t$. We wish to find a point (x^*, y^*) so that both N and D vanish at that point. As mentioned above, we need to find the resultant of these two polynomials. To do this we write $N = y^2(a_0x^2 + c_0x + f_0) + y(b_0x^2 + e_0x + h_0) + (d_0x^2 + g_0x + j_0)$ and $D = y^2(a_1x^2 + c_1x + f_1) + y(b_1x^2 + e_1x + h_1) + (d_1x^2 + g_1x + j_1)$. To ease notation, we shall label the coefficient of y^i in N as N_i and the coefficient of y^i in D as D_i. Now the resultant of these two quadratics in y is given by the determinant of the Sylvester matrix, the Sylvester matrix being

$$\begin{pmatrix} N_2 & N_1 & N_0 & 0 \\ 0 & N_2 & N_1 & N_0 \\ D_2 & D_1 & D_0 & 0 \\ 0 & D_2 & D_1 & D_0 \end{pmatrix}$$

Calculating this determinant gives

$$N_2(N_2(D_0^2) - N_1(D_1D_0) + N_0(D_1^2 - D_0D_2)) - N_1(-N_1(D_2D_0) + N_0(D_2D_1)) +$$
$$N_0(-N_2(D_2D_0) + N_0(D_2^2))$$

where each N_i and each D_i are quadratics in x, thus the expression taken as a whole will generically be a degree eight polynomial in x. Finding a solution to this gives an x value for which N and D vanish simultaneously, and corresponding y values can be found by substitution. Now supposing that we have found, via this resultant, a base point (x^*, y^*), we can continue this algebraic treatment to find another point as a function of the coefficients in I.

Substituting $y = y^*$ into $I(x, y) = t$ and moving all terms onto one side gives

$$(a_0 + a_1t)x^2{y^*}^2 + (b_0 + b_1t)x^2y^* + (c_0 + c_1t)x{y^*}^2 + (d_0 + d_1t)x^2 +$$
$$(e_0 + e_1t)xy^* + (f_0 + f_1t){y^*}^2 + (g_0 + g_1t)x + (h_0 + h_1t)y^* + (j_0 + j_1t) = 0$$

which, considered as a quadratic in x is

$$x^2((a_0 + a_1t){y^*}^2 + (b_0 + b_1t)y^* + (d_0 + d_1t)) +$$
$$x((c_0 + c_1t){y^*}^2 + (e_0 + e_1t)y^* + (g_0 + g_1t)) +$$
$$((f_0 + f_1t){y^*}^2 + (h_0 + h_1t)y^* + (j_0 + j_1t)).$$

Let X be the second root to this quadratic, the first being x^*. Then the product of roots formula says that

$$X = \frac{\frac{(f_0+f_1t){y^*}^2+(h_0+h_1t)y^*+(j_0+j_1t)}{(a_0+a_1t){y^*}^2+(b_0+b_1t)y^*+(d_0+d_1t)}}{x^*},$$

whence we get the $\mathbb{C}(t)$ point (X, y^*). Let us turn now to the simpler curves of the McMillan type, where we can use simpler methods to find points on them.

Lemma 5.25. *Suppose $P = (X(t), Y(t))$ lies on $B(x, y) = 0$, a biquadratic of type McMillan. Then if $X(t)$ and $Y(t)$ are both polynomials, the degree of one of them must be 0.*

Proof. Suppose $deg(X) = d_X$ and $deg(Y) = d_Y$. Then the highest degree term of B is ax^2y^2 which has degree $2(d_X + d_Y)$. Without loss of generality, suppose $d_X \geq d_Y$. Since P lies on B, the coefficients of all powers of t must be zero. The next highest degree term is bx^2y with degree $2d_X + d_Y$. To have the coefficient of $t^{2(d_X+d_Y)}$ be equal to zero we must have that these two degrees are equal, that is, $2d_X + 2d_Y = 2d_X + d_Y$ which yields $d_Y = 0$. □

We can use this lemma to find a large class of type McMillan biquadratics with a point on them.

Theorem 5.26. *Let $B(x, y)$ be a type McMillan biquadratic. Let Y be a complex number such that $aY^2 + bY + d = 0$. Then the point $P = (-\frac{fY^2+hY+j-t}{cY^2+eY+g}, Y)$ lies on B.*

Proof. We proceed by substituting $y = Y$ into the equation for B and solving for x; from lemma (5.25) we know that this is the only way to proceed. The substitution gives $x^2(aY^2 + bY + d) + x(cY^2 + eY + g) + (fY^2 + hY + j - t) = 0$. Solving this with the quadratic equation gives a discriminant of

$$\Delta = (cY^2 + eY + g)^2 - 4(aY^2 + bY + d)(fY^2 + hY + j - t)$$

This discriminant is linear in t, which means that the only way it can be a perfect square is if the coefficient of t is zero. The reason we demand this be a perfect square (in the field $\mathbb{C}(t)$) is so that x is rational in t. This gives the condition $aY^2 + bY + d = 0$. Of course, this is the original coefficient of x^2 after substitution, so we must look to when, after substitution, the equation for B is linear. This yields the solution

$$x = -\frac{fY^2 + hY + j - t}{cY^2 + eY + g}$$

□

A similar theorem can easily be formulated assuming that the x-coordinate is constant rather than the y-coordinate. An example of a randomly chosen type McMillan biquadratic shall illustrate theorem (5.26).

Example 5.27. Let $B(x, y) = 2x^2y^2 + 3x^2y - x^2 - 4xy + 10y^2 + 6x - 7y + 9 - t$. Then we form the quadratic equation $2Y^2 + 3Y - 1$ and solve it, to get $Y = -\frac{3}{4} \pm \frac{\sqrt{17}}{4}$. Then we claim that the points $(-\frac{10Y^2 - 7Y + 9 - t}{-4Y + 6}, Y)$ lie on B. This is easy to check with any mathematics program or by hand.

We now look to polynomial points on type QRT biquadratics using similar methods.

Lemma 5.28. *Let $B(x, y)$ be a type QRT biquadratic containing a point $P = (X(t), Y(t))$ with polynomial coordinates. Then either $deg(X) \leq 1$ or $deg(Y) \leq 1$.*

Proof. Suppose that $deg(X) = d_X \geq deg(Y) = d_Y$. Consider the degree of the terms $\alpha = X^2(aY^2)$ and $\beta = X^2(bY)$ (it is obvious that these are the terms containing the highest degree of t). Now to have the point P lying on B, we need $B(X(t), Y(t))$ to be identically zero i.e. that the coefficient of each of the powers of t is zero. To achieve any cancellation of the highest degree terms of α and β we require $2d_X + deg(a) + 2d_Y = 2d_X + deg(b) + d_Y$ whence $d_Y = deg(b) - deg(a)$. Since $deg(a)$ and $deg(b)$ are either 0 or 1, the result follows. □

Note that this lemma does not, strictly speaking, take into account certain corner cases such as the case when $d_X = d_Y$ AND $deg(b) = 0$ AND $deg(c) = 1$ simultaneously. Nevertheless, by switching the roles of X and Y and following the same proof, the result still stands. The lemma also proves that if $deg(a) = 1$ then one coordinate must in fact be a constant. We can now use this lemma in a manner similar to the type McMillan case. Assuming again that one coordinate is constant leads to similar conclusions via similar methods as the type McMillan case. So instead we shall use the difference between the two cases and suppose that one coordinate is linear in t and construct a type QRT biquadratic on which such a point can lie.

Theorem 5.29. *Let $B(x, y) = (a_0 + a_1t)x^2y^2 + (b_0 + b_1t)x^2y^2 + \ldots + (j_0 + j_1t)$ be a type QRT biquadratic. Then if $a_1 = 0$ and $d_1^2a_0 - b_1(b_0d_1 - b_1d_0) = 0$, B admits a point whose y-coordinate is linear in t and is given by*

$$Y = -\frac{d_1^2}{b_0d_1 - b_1d_0}t - \frac{d_0d_1}{b_0d_1 - b_1d_0}.$$

The x-coordinate of this point is $X =$

$$\begin{gathered}
-(f_1 d_1^4)t^3 + (f_0 d_1^4 - h_1 d_1^3 b_0 + 2d_1^3 f_1 d_0 + h_1 d_1^2 d_0 b_1)t^2 + \\
(d_1^2 f_1 d_0^2 + j_1 d_1^2 b_0^2 + d_1 h_1 d_0^2 b_1 + h_0 d_1^2 d_0 b_1 - h_0 d_1^3 b_0 + j_1 d_0^2 b_1^2 - 2j_1 d_1 d_0 b_1 b_0 - h_1 d_0 b_0 d_1^2 + 2d_1^3 f_0 d_0)t + \\
(d_1^2 f_0 d_0^2 + j_0 d_0^2 b_1^2 - 2j_0 d_1 d_0 b_1 b_0 + j_0 d_1^2 b_0^2 + d_1 h_0 d_0^2 b_1 - h_0 d_0 b_0 d_1^2) \\
/ \\
(d_1^4 c_1)t^3 + (d_1^2 e_1 d_0 b_1 - d_1^3 e_1 b_0 + d_1^4 c_0 + 2d_1^3 c_1 d_0)t^2 + \\
(d_1^2 e_0 d_0 b_1 - d_1^2 e_1 d_0 b_0 - 2g_1 d_0 b_1 b_0 d_1 + g_1 d_0^2 b_1^2 + d_1^2 c_1 d_0^2 + g_1 b_0^2 d_1^2 + d_1 e_1 d_0^2 b_1 + 2d_1^3 c_0 d_0 - d_1^3 e_0 b_0)t + \\
(d_1 e_0 d_0^2 b_1 - d_1^2 e_0 d_0 b_0 - 2g_0 d_0 b_1 b_0 d_1 + d_1^2 c_0 d_0^2 + g_0 d_0^2 b_1 + g_0 b_0^2 d_1^2)
\end{gathered}$$

Proof. The technique for this proof is computational. Substituting $y = Y_0 + Y_1 t$ into B and forcing the coefficient of x^2 to be zero yields the conditions given above. □

Theorems (5.29) and (5.26) give large classes of biquadratics which contain a point with coordinates in $\mathbb{C}(t)$. The related problem of finding a particular biquadratic which contains a given point can also be dealt with in a computational fashion. One can easily choose a point $(X_0 + X_1 t, Y_0 + Y_1 t)$, substitute this into a full QRT type biquadratic, and then attain conditions on the coefficients of the biquadratic that must be satisfied. The higher degree in t the chosen point is, the more conditions there are to satisfy. The equations are all linear in the coefficients $a_0, a_1, \dots j_0, j_1$ and so solutions are easy to come by, up to a certain point. The equations are also homogeneous, meaning that if solutions are unique, then the only solution (since it certainly exists) will be the one with all coefficients equal to zero. So looking for interesting solutions means looking for underspecified systems or systems where the equations are linearly dependent.

We now turn back to the original problem, that of determining what $\mathbb{C}(t)$ points can lie on any type McMillan or type QRT biquadratics with a bit more complexity. Rather than supposing points are polynomial, we shall assume they are fractional linear. This lemma generalises lemma (5.25). A similar generalisation for lemma (5.28) remains elusive due to the excessive number of possible cases.

Lemma 5.30. *Any point P with fractional linear coordinates in t that lies on a McMillan type biquadratic must have at least one of its coordinates being a linear polynomial in t.*

Proof. Substituting a fractional linear point into a McMillan type biquadratic yields an equation with the term $-tX_d^2 Y_d^2$ where X_d is the degree of the denominator of the x-coordinate and Y_d is the degree of the denominator of the y-coordinate. If both X_d and Y_d have degree equal to 1, then this term is of degree 5, which no other terms in the equation reach (each being made up of exactly 4 factors chosen from X_n,X_d,Y_n and Y_d) thus there is no way to have the coefficient of t^5 being zero unless either X_d or Y_d have degree zero. This means that that particular coordinate is a linear polynomial. □

The last comment we will make in this section is that certain type QRT biquadratics are very easy to find several points on; this is useful when one is constructing biquadratics with the intention of having several points to work with. We have already seen that such biquadratics are useful when hoping to find elliptic curves of rank higher than one which in turn can be interesting for creating maps other than the standard QRT maps.

Lemma 5.31. *Let B be a type QRT biquadratic with $j = 0$. Then there are two points with coordinates in $\mathbb{C}(t)$ given by $(-\frac{g}{d}, 0)$ and $(0, -\frac{h}{f})$, as well as the point $(0, 0)$.*

Proof. Firstly it is clear that with the constant term being set to 0, the point $(0, 0)$ lies on the biquadratic. Now suppose $y = 0$, then the equation of the curve is $dx^2 + gx = 0$ which clearly has solutions $x = 0$ (taken into account above) and $x = -\frac{g}{d}$. Similarly the point $(0, -\frac{h}{f})$ lies on the biquadratic. □

Lemma (5.31) gives an easy way to find a $\mathbb{C}(t)$ point for very limited classes of type QRT biquadratics. For other type QRT biquadratics, we simply find a base point as described in proposition (5.22). For systems with a type McMillan biquadratic as the integral (or systems where the inverse of the preserved quantity is a type McMillan biquadratic) one can use theorem (5.26) to find a point on it.

Until now our examples all have foliations (whose individual curves are invariant) that were rational elliptic surfaces and in the literature this has consistently been the case also. However theorem (4.1) and its corollaries do not require this to be the case and so to finish the chapter, we apply our method to an example where the foliation is not a rational elliptic surface. It also turns out to be a map that must be left as being defined over $\mathbb{C}(t)$; there is no substitution $t = t(x, y)$ to make that reduces the map to a complex map.

Example 5.32. We consider example four from [29]. The map is

$$\begin{aligned} L : x' &= y \\ y' &= \frac{-144xy^2t^2 + 36xy^2t + 9xy^2 + 144xyt^2 - 18xyt + 36xy - 36xt^2 - 8xt + 36x + 144y^2t^2 -}{144y^2t^2 - 36y^2t - 9y^2 - 144yt^2 + 18yt - 36y + 36t^2 + 8t - 36} \\ &\frac{18y^2t + 36y^2 + 144yt^2 - 27yt + 126y - 108t^2 - 18t + 108}{144y^2t^2 - 36y^2t - 9y^2 - 144yt^2 + 18yt - 36y + 36t^2 + 8t - 36} \end{aligned}$$

and the curve that the map acts on is

$$\begin{aligned} B(x, y, t) =& (16t^2 - 4t - 1)x^2y^2 - 2(8t^2 - t + 2)(x^2y + xy^2) + 4(t^2 + \frac{2}{9}t - 1)(x^2 + y^2) - \\ & (16t^2 - 3t + 14)xy + 2(6t^2 + t - 6)(x + y) + (9t^2 + t - 9). \end{aligned}$$

In its original context, the authors make note of the point $(0, -\frac{3}{2})$ that lies on this curve, thus we use it as the identity for conversion to Weierstrass form. Performing the necessary calculations yields the Weierstrass form

$$\begin{aligned} & W(u, v, t) = u^3 + (-5461452t^2 - \frac{5347049929}{48}t^4 + 50781876t^3 + 2066688t^6 - 30632856t^5)u - \\ & 68527667853t^4 + 4912402896t^3 - \frac{354684217437701}{864}t^6 + 309576630067t^5 + 175220352t^8 - \\ & 192227166682t^7 + 2229755904t^9 + v^2. \end{aligned}$$

Some features to note regarding W is that there are no points of order 2 or 3 (and hence no points of any orders dividing 2 or 3); this is easy to check. Also, W does not satisfy the condition for being a rational elliptic surface, meaning that we cannot apply the results of [47] to give the $\mathbb{C}(t)$ group structure (4.18). Following the usual procedure, we can find the point Ω on W by whose translation L is conjugate to. In this case it is

$$\Omega = (-\frac{t(9792t^2 + 74909t - 16272)}{12}, 2t^2(7056t^2 - 244t - 1521)).$$

So in this example we have a curve where we cannot apply the results of this section, as the biquadratic's t dependence is not of type QRT or of type McMillan yet because of the fortunate happenstance of there being a simple point known to lie on it we can still apply the theory of chapter 4.

Chapter 6

Maps in Higher Dimensions

This chapter collects results arising from studying maps of dimension greater than two. It is separated into two sections, one containing a collection of remarks on various interesting three dimensional maps and the other an extension of a two dimensional reversibility detection test to maps of dimensions three and four.

6.1 Remarks on Integrability in Three Dimensions

The first map we consider is a three dimensional map that can be manipulated to yield a two dimensional map due to the simplicity of one of its integrals. It is included as a warning of the possibility that three dimensional maps can be degenerate in this sense which may yield unusual numerical results.

Example 6.1.

$$\begin{aligned} L_1 : x' &= y \\ y' &= z \\ z' &= -x - \frac{2yz}{y+z}. \end{aligned} \tag{6.1}$$

This map has two integrals, these being

$$I_1(x,y,z) = (xy + xz + yz)^2 \tag{6.2}$$

$$I_2(x,y,z) = (x+z)(xy + xz + yz - y^2). \tag{6.3}$$

To apply any algebraic geometric test for integrability, we need all varieties involved to be irreducible (recall theorem (2.58) requires varieties to be irreducible). In this case the integral I_1 will give problems as each level set can factor into $(xy+xz+yz-\sqrt{t})(xy+xz+yz+\sqrt{t})$. However we can avoid this problem simply by dealing with the map L_1^2 instead. This is because L_1 happens to bounce between each irreducible factor of $I_1(x,y)=t$, thus the square of the map leaves each factor fixed. So L_1^2 has $I_1'(x,y,z)=(xy+xz+yz)$ as an integral, which is clearly solvable for z as

$$z=\frac{K-xy}{x+y}$$

where K can be determined by putting any initial condition (x,y,z) into I_1'. We can similarly substitute this expression for z into each coordinate of the map and the second integral of the map. This process yields the map

$$\begin{aligned}\widetilde{L_1}:x' &= \frac{t-xy}{x+y}\\ y' &= \frac{xy^2-xt-2yt}{y^2+t}\\ z' &= \frac{y^3t+3yt^2-2xy^2t+x^2y^3-yx^2t+2xt^2}{-y^2t+t^2-4xyt+x^2y^2-x^2t}\end{aligned} \tag{6.4}$$

with the integral

$$\widetilde{I_2}=-\frac{(x^2+t)(y^2-t)}{x+y}. \tag{6.5}$$

We can construct a new map by projecting this information onto any plane of constant z (that is to say, we just remove the z coordinate from the map all together) to be left with a two dimensional map with a generically elliptic integral. However a parameter is introduced into this two dimensional map. The parameter comes from having to substitute for z using the first integral. In effect we typically have the following data for an initial condition in the three dimensional setting - the point (x,y,z). This data allows us to find $K=I(x,y,z)$ but similarly we could find z given K,x and y. So it is this K that appears as a parameter in the two dimensional "reduction" of the map. But we can leave K symbolic in both the map and the integral. It is quite interesting that starting with an integrable three dimensional map defined over $\mathbb{C}$ we can use one of the integrals to change this map into an integrable two dimensional map defined over $\mathbb{C}(t)$

A second example first exhibited in [24] is the third order difference equation given by

$$L:x_{n+3}=\frac{1}{x_n}\frac{p_3x_{n+1}x_{n+2}+p_4(x_{n+1}+x_{n+2})+p_5}{p_2x_{n+1}x_{n+2}+p_1(x_{n+1}+x_{n+2})+p_3}. \tag{6.6}$$

This can be reduced to a three dimensional dynamical system by the identifications $x_n = x, x_{n+1} = y, x_{n+2} = z$. The corresponding map has two independent integrals namely

$$H_1 = \frac{I_1 + I_2 - p_3 H_2}{p_4} \tag{6.7}$$
$$H_2 = \frac{I_1 I_2 - 4p_1 p_4 + 2p_3^2 + p_2 p_5}{p_1 p_4} \tag{6.8}$$

where I_1 and I_2 are both integrals of (L^2)[1] and are given by

$$I_1(x, y, z) = \frac{p_1 xyz + (p_3 y + p_4)(x + z) + p_4 y + p_5}{xz} \tag{6.9}$$
$$I_2(x, y, z) = \frac{(p_2 y + p_1)xz + (p_1 y + p_3)(x + z) + p_4}{y} \tag{6.10}$$

Since neither of these integrals are solvable for any of the three coordinates, we cannot apply a similar reduction procedure as in example (6.1) and yet calculating the orbit statistics for such a map (three dimensions, with two integrals) yields a picture similar to the statistics for a two dimensional map with one integral. This is due to the fact that the intersection between two two dimensional surfaces in three dimensions is generically a single curve. This point is important enough to belabour a little and give as a proposition.

Proposition 6.2. *Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be an infinite order birational map that preserves two families of algebraic surfaces $\mathcal{F}$ and $\mathcal{G}$ each of which foliates $\mathbb{R}^3$ by surfaces. Then the orbit lengths of the reduced map $\widetilde{L}$ over a finite field $\mathbb{Z}_p$ will, excluding exceptional cases such as singular or reducible level sets, lie under the upper Hasse-Weil bound with $g = 1$ (see theorem (2.58)).*

Proof. Since L preserves each surface in $\mathcal{F}$ and $\mathcal{G}$ it also preserves the intersection of any two such surfaces. Generically, the intersection of two two-dimensional surfaces in three dimensions is a curve (see, for example, [58]) so L preserves this curve. Since L is infinite order, Hurwitz' theorem (2.51) again tells us that this curve is either a conic or elliptic. Applying the Hasse-Weil bound to the curve to which any given orbit is confined gives the result. □

[1] Such objects are called 2-integrals. In general, integrals of the kth power of a map are called k-integrals.

Proposition (6.2) allows us to devise a test for three dimensional algebraic integrability using our simple two dimensional test with almost no extra work. The procedure followed is identical; reduce the map over a finite field to an action on a finite phase and measure all the orbit lengths. If the vast majority of these lie above the upper Hasse-Weil bound, the map cannot preserve the intersection of surfaces as above. The difference in three dimensions is that we are forced to consider the map's action "curve at a time". In chapter 4 we had theorem (4.8) which told us that in two dimensional integrable maps we could consider their action as a translation on a single Weierstrass equation defined over $\mathbb{C}(t)$ - this gave us a relation between the translative points on each preserved curve. In three dimensions we must typically stick to the less sophisticated idea that on each curve the map is conjugate to translation on a Weierstrass though the translative point on each curve may be unrelated to the translative points on other curves. A way to resolve this would be when the equations for the intersections of the level surfaces are able to be found algebraically. However this seems possible only in ideal circumstances. A second difference in three dimensions is that when we cannot find the equations of the curves being preserved, one cannot tell if the family of curves being preserved forms a rational elliptic surface. This is required for the structure theorem of [47] and the work of Duistermaat [14]. We can see proposition (6.2) in action as applied to the map from example (6.8) which we use in the next section on reversibility. This map is a three dimensional map with two integrals of motion and part of its phase space can be seen in figure (6.5). Figure (6.1) shows, taking the Monte Carlo style approach, normalised orbit lengths for the orbit of the point $[-2, -4, -7, 1]$ over finite fields $\mathbb{Z}_p$ with $1000 \leq p \leq 10000$.

Example 6.3. An example of a three dimensional map that preserves at least one surface can be created by considering one way of extending the QRT map quite naturally. Taking the view that the standard QRT map is the two dimensional map made by composing the two different involutions that each switch one ordinate while leaving the other fixed, one can see that a natural extension of this idea in n dimensions is to take the n different involutions that switch one ordinate whilst leaving the other $n - 1$ fixed and compose them together in some order. To create

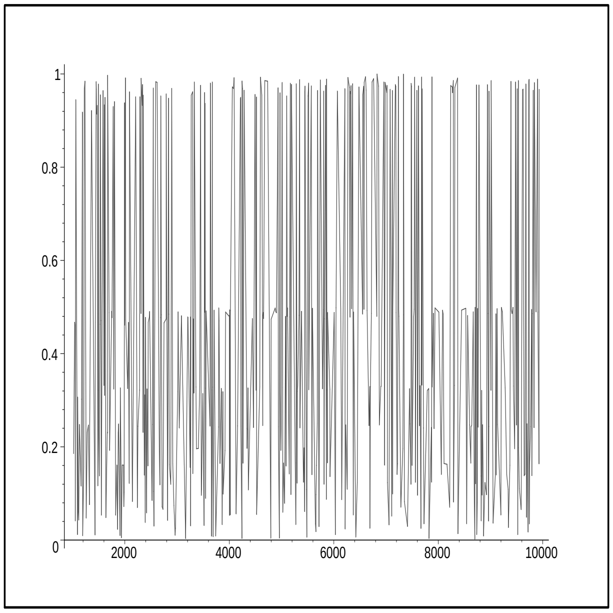

Figure 6.1: Normalised lengths for the orbit of the point $[-2, -4, -7, 1]$ under the map L from example (6.8) with parameters as given in that example considered over the finite field $\mathbb{Z}_p$. The primes used are $1000 \leq p \leq 10000$ and we deal with singularities in the map by considering only the forward orbit. Note that even though we only consider the forward orbits many normalised orbit lengths still approach the upper bound of one.

such a map we first consider a general triquadratic in xyz-space

$$T(x,y,z) = \sum_{0 \leq i,j,k \leq 2} a_{ijk} x^i y^j z^k. \tag{6.11}$$

Writing equation (6.11) as $X_2x^2 + X_1x + X_0$ we can construct an involution that leaves the y and z ordinates fixed while switching the x coordinate to the other root of the triquadratic. As usual, the sum of roots formula dictates that $x' + x = -\frac{X_1}{X_2}$ whence we get the map

$$\begin{aligned} G_x : x' &= -x - \frac{X_1}{X_2} \\ y' &= y \\ z' &= z \end{aligned} \tag{6.12}$$

and via similar techniques we can construct

$$\begin{aligned} G_y : x' &= x \\ y' &= -y - \frac{Y_1}{Y_2} \\ z' &= z \end{aligned} \tag{6.13}$$

and

$$\begin{aligned} G_z : x' &= x \\ y' &= y \\ z' &= -z - \frac{Z_1}{Z_2}. \end{aligned} \tag{6.14}$$

From these involutions we construct a map

$$L = G_x \circ G_y \circ G_z. \tag{6.15}$$

Certainly the map L leaves fixed the surface T but once again by having the coefficients $a_{ijk} = a_{ijk(1)}t + a_{ijk(0)}$ we can have T represent an entire family of surfaces each of which will be preserved by the map L. On the face of it then, a map L constructed in this fashion with the above affine coefficients is a three dimensional map with one family of surfaces that foliates three dimensional space. As it transpires, in [26], Iatrou had previously considered similar three dimensional maps that, by restricting some parameters where necessary, preserve two surfaces. In [12], Capel and Sahadevan introduced a four dimensional generalisation of the QRT maps via the method of composition of involutions.

6.2 Reversibility Detection

In this section we will be illustrating, numerically, that a test for detecting reversibility in maps of two dimensions first proposed in [61] continues to work in higher dimension. In fact, this test tests for reversibility in a map by testing for whether the map can be written as a composition of two involutions. Thus from now on we shall be dealing with this restricted version of reversibility, often referred to as r-reversibility in the literature. In [61], the test was devised and checked for polynomial maps of the plane; C.-M. Viallet has since confirmed the test for planar rational maps (private communication by J.A.G. Roberts). It is birational maps to which we shall be restricting our study. We shall later present evidence for a conjecture regarding the value of a parameter in this test.

For the background on this section we shall be drawing on two past papers. The second we shall be looking at is the aforementioned [61] by Roberts and Vivaldi which details a test for detecting reversibility in two dimensional maps. The first, [58], discusses the consequences of reversibility in three dimensions ($\mathbb{R}^3$) as well as the different types of reversibility in three dimensions.

6.2.1 Background

While the central topic of [58] is the type of local structures that reversible maps impose on phase space (recall that a map in two dimensions is reversible if it is conjugate to its own inverse) the part of it that is most useful here is the language and notation. It will soon become apparent that when $L = G \circ H$ is a reversible map, quantities of use are the dimensions of the fixed sets of the reversing symmetries themselves, i.e. the dimensions of $\mathrm{Fix}(G)$ and $\mathrm{Fix}(H)$. The notation introduced by Roberts and Lamb in [58] is to label each involution according to the dimension of its fixed set. Thus a Type 0 involution is an involution with a zero dimensional fixed set, a Type I involution has a one dimensional fixed set and a Type II involution a two dimensional fixed set and so forth. Extending this, we label reversible maps according to the types of the involutions that make them up. Thus a Type I-II map is made up of one Type I involution and one Type II involution and so forth. Due to the importance of these dimensions, Type 0 involutions are of comparatively little importance. From the point of view of [58] this is because of the underlying necessity for a symmetric orbit to at some point "land in" Fix(G) or Fix(H) which simply rarely happens for Type 0 involutions. From our point of view, Type 0 involutions will be of lesser consequence due primarily to computation concerns, explaining why no such involutions arise throughout this section.

Following the methodology of searching for signatures of properties of dynamical systems by studying how they act on finite phase spaces, [61] looks to detect reversibility in maps. Drawing conclusions from numerical data arising from maps of three dimensional space is a more complicated process than in two dimensions. One reason for this is that the reversibility (if any) of the map plays a large role in determining how the map decomposes a finite phase space into orbits. To explain this, we examine a corollary of Proposition 1 in [61]. Suppose that a map L is reversible and can be written as $L = G \circ H$ with G and H involutions and let Fix(M) denote the set of fixed points of any map M. Then

$$|Fix(G)| + |Fix(H)| = 2|\text{Symmetric cycles of L}|, \tag{6.16}$$

where the right hand side of equation (6.16) is counting (twice) the number of cycles of L that are invariant under G. The proof of the underlying proposition demands that the map L be an everywhere invertible map of a finite phase space. In fact the proposition (and hence the corollary) is a wholly combinatoric exercise with a natural language of permutations of finite sets. For our purposes however we will be considering it in the context of maps which are permutations when reduced to actions over finite phase spaces and that are the composition of two involutions. Manipulating equation (6.16) a little we get

$$|\text{Cycles of L}| \geq |\text{Symmetric cycles of L}| = \frac{1}{2}(|Fix(G)| + |Fix(H)|). \tag{6.17}$$

Now we can explain why the signature of reversibility in three dimensions is a little different than in two dimensions. Consider a set of the form Fix(G) with G being a two dimensional map, although the discussion will apply for any dimension. Finding the fixed points is a matter of solving the pair of equations

$$\begin{aligned} x &= G_x(x, y) \\ y &= G_y(x, y). \end{aligned} \tag{6.18}$$

The solutions to this pair of equations constitutes the intersection of two curves (in three dimensions, the intersection of three surfaces), these being defined by $x - G_x(x, y) = 0$ and $y - G_y(x, y) = 0$, and therefore the upper bound dictated by Bezout's theorem (the product of the degrees) exists for the number of solutions. The exception to this is when both curves have a factor in common so generically one would expect the fixed set of such a map to be finite. However it is often the case that for common involutions one or both equations of motion are very simple resulting in high dimensional fixed sets. For example in the case of the symmetric QRT maps, one of the involutions is simply

$$\begin{aligned} x' &= y \\ y' &= x. \end{aligned}$$

Solving equations (6.18) in this case gives the single line $y - x = 0$ so the fixed set would be just that line and hence have exactly p elements in the finite plane $\mathbb{Z}_p^2$. Supposing, then, that our two involutions each have p fixed points we see that this results in the number of cycles of their composition L having at least p cycles. Having a larger fixed set than this can only occur when each equation of motion can be factored into more than one component, each of which is in common with the other equations of motion. Certainly in examples that have arisen previously in the literature this does not happen, so we will assume that our maps have fixed sets that are either some finite set of points or a single d-dimensional set. This assumption will exclude instances where fixed sets are unions of distinct lines and similar situations. Thus under our assumptions, for a map of N dimensions the largest possible fixed set will be an $(N - 1)$-dimensional hypersurface.

In two dimensions the consequences of equation (6.17) are interesting as they can be compared to consequences from the Hasse-Weil bound. Recall the Hasse-Weil bound which, for two dimensional integrable maps over the field $\mathbb{Z}_p$, restricts orbits to a maximal length of $p + 2\sqrt{p} + 1$. In a phase space of p^2 points (the affine part of the projective plane is where the maps we are considering will be invertible) this imposes a lower bound on the number of cycles of $\frac{p^2}{p+2\sqrt{p}+1}$. To compare this with the lower bound from equation (6.17) we need to have some idea of the size of $|Fix(G)|$ and $|Fix(H)|$. On the face of it, this compares favourably with the lower bound on the number of cycles given by the Hasse-Weil bound for integrable reversible maps. The problem is that orbits in integrable maps only consume the upper Hasse-Weil bound on their own about half the time - it just as frequently takes several orbits all lying on the same level set to fill the Hasse-Weil bound. This results in there being a larger number of cycles in integrable maps than in reversible, non-integrable maps despite what the inequalities above would seem to suggest.

To give an idea of the kind of data presented in [61], we shall take three different birational maps of the plane. One will be integrable with the preserved foliation algebraic (and thus necessarily reversible; see theorem (4.8)[2]), one merely reversible and the third neither. They all must act as permutations of the space $\mathbb{Z}_p^2$. When performing numerical experiments we, for a fixed prime p, calculate all the cycles generated in the phase space $\mathbb{Z}_p^2$ and store the number of cycles. After finding this number for a suitable number of primes, we may plot a curve with the x-coordinate being the prime and the y-coordinate being the number of cycles. If the map in question is reversible, then this curve will lie above the curve $\frac{1}{2}(|Fix(G)|+|Fix(H)|)$. To compare to the integrable-reversible case and the non-integrable but reversible case we also conduct a similar procedure for a non-reversible (and hence necessarily non-integrable) map. Because we need maps that are invertible across the entire affine part of the plane we choose maps with denominators of the form z^2+1 and restrict our primes to be congruent to 3 modulo 4, as it is well known that in such finite fields $z^2+1=0$ has no solution in z.

Example 6.4. Consider the QRT map given by

$$\begin{aligned} L : x' &= -x - \frac{y+1}{y^2+1} \\ y' &= -y - \frac{x'-1}{(x')^2+1}. \end{aligned} \tag{6.19}$$

This is the composition $G \circ H$ of the involutions

$$\begin{aligned} G : x' &= x \\ y' &= -y - \frac{x-1}{x^2+1} \end{aligned}$$

[2] While birational integrable maps of the plane are necessarily reversible, it is not true that other integrable planar maps must be reversible. For an example, refer to [4].

and

$$\begin{aligned} H: x' &= -x - \frac{y+1}{y^2+1} \\ y' &= y. \end{aligned}$$

This map has only denominators of the form $z^2 + 1$ and so we know that when considered over fields $\mathbb{Z}_p$ with $p \equiv 3 \mod 4$ both ordinates will remain defined. Therefore, the map acts as a permutation of the finite phase space $\mathbb{Z}_p^2$ and the result (6.17) applies. For primes congruent to 3 modulo 4 between 7 and 500, we count the number of cycles the map partitions the phase space into and plot the number of cycles versus prime in figure (6.2).

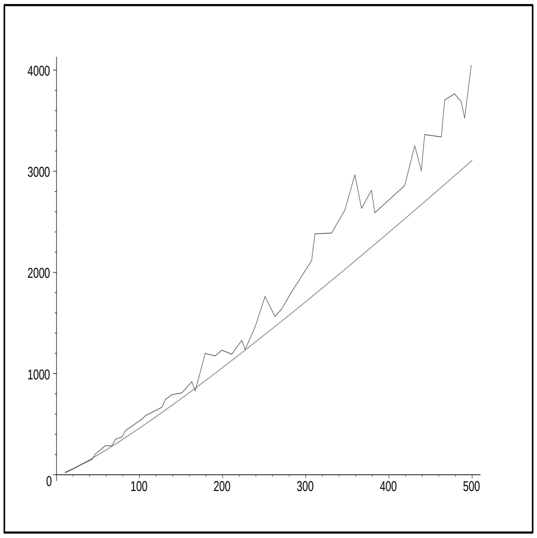

Figure 6.2: Number of cycles vs. prime for map (6.19) which is both integrable and (Type I-I) reversible. It is expected that the integrability creates significantly more cycles than predicted by the bound imposed by reversibility. The reference curve also plotted here is $y = x \log(x)$ which is the same reference curve used for integrable maps [61].

For this example, the dimensions of Fix(G) and Fix(H) is 1. This is obvious because the ordinate that is fixed in each involution yields a trivial relation in the system (6.18). Thus in the field $\mathbb{Z}_p$ the lower bound on the number of cycles in this case is $\frac{p+p}{2} = p$. However this lower bound is greatly surpassed by the actual number of cycles. This is due to the added integrability of the map.

Example 6.5. To break the integrability but preserve the reversibility of this map we will replace G with a completely different involution but leave H the same. The new involution G' we use (knowing that we must not introduce any new denominators) is

$$\begin{aligned} G' : x' &= -x + 1 \\ y' &= -y - (x - \frac{1}{2})^2. \end{aligned}$$

The construction of G' followed the following process. We know that transformations of one dimension of the form $x \to -x + a$ with a being constant are involutions, so we choose one dimension to be of that form. Suppose that the second ordinate has the form $y \to -y + f(x)$ then we find the second power of such a map and find a condition for the entire transformation to be an involution is that $f(x) = f(1 - x)$. Without having to revert to constants, $f(x) = (x - \frac{1}{2})^2$ is the simplest choice for this. The resulting map $G' \circ H$ is given by

$$\begin{aligned} x' &= x + \frac{y+1}{y^2+1} + 1 \\ y' &= -y - (x + \frac{y+1}{y^2+1} + \frac{1}{2})^2 \end{aligned}$$

which is Type 0-I as G' clearly has one fixed point only at $(\frac{1}{2}, 0)$. As it is Type 0-I, the lower bound on the number of cycles for a field $\mathbb{Z}_p$ is $\frac{p+1}{2}$. As in example (6.4) we find how the map decomposes the plane $\mathbb{Z}_p^2$ for the appropriate primes between 7 and 500 and graph the number of cycles against the prime.

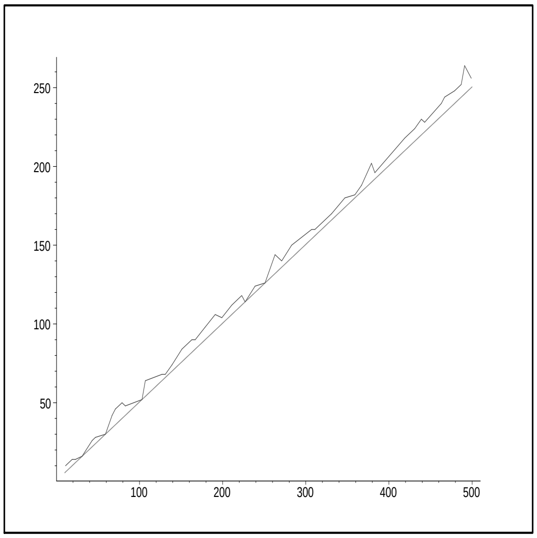

Figure 6.3: Number of cycles vs. prime for the map of example (6.5) which is (Type 0-I) reversible. The plotted lower bound of $y = \frac{x+1}{2}$ is quite strictly adhered to; the strictness of this adherence is yet to be explained.

A noteworthy feature is how strictly the actual data adheres to the predicted curve. In dynamical terms this is saying that over finite fields nearly all orbits are symmetric with respect to the reversing symmetry. This is in stark contrast to the real phase space where very few orbits are symmetric (see [42]).

For our last example we need to find a map that is definitely not reversible. The fact that this requires showing the map can not be decomposed into two involution in any way whatsoever makes it clear that we need some kind of technical result that gives a condition for a map not being reversible. Lemma (6.6) provides this.

Lemma 6.6. *Let L be a map with constant Jacobian determinant J_0 that has a fixed point. Then if $|J_0| \neq 1$, L can not be reversible.*

Proof. We shall show that if L is reversible and has a constant Jacobian determinant J_0 as well as a fixed point then it must be the case that $|J_0| = 1$. Suppose $LGL = G$ for some involution G (i.e. L is reversible). Let us consider the matrix of derivatives at a fixed point x. By the chain rule this yields the equation

$$dL(GLx)dG(Lx)dL(x) = dG(x).$$

Taking the determinants of these matrices and noting that $\det L = J_0$ is constant gives

$$J_0(\det dG(Lx))J_0 = \det dG(x).$$

However $Lx = x$ since x is a fixed point of L thus we are left with

$$J_0^2 = 1$$

and taking the modulus of each side gives the desired result. □

So to construct our map that is definitely not reversible we shall find a real, rational map that has constant Jacobian determinant different from ±1 with a fixed point.

Example 6.7. Consider the family of generalised Henon transformations in two dimensions defined by

$$\begin{aligned} x' &= y \\ y' &= -\delta x + h(y). \end{aligned} \tag{6.20}$$

This map has Jacobian matrix

$$J = \begin{pmatrix} 0 & 1 \\ -\delta & h'(y) \end{pmatrix}$$

which in turn has constant determinant

$$|J| = \delta$$

so as long as we choose $\delta \neq \pm 1$ we have satisfied the requirement on the Jacobian determinant. As for the fixed point, the map has a fixed point at any point (Y, Y) where Y is a solution to the equation

$$(1+\delta)y - h(y) = 0.$$

Thus a reasonable choice of $h(y)$ will ensure a fixed point. Let us take $\delta = \frac{4}{7}$ and $h(y) = y^2 + 1$ so that our final map is

$$\begin{aligned} x' &= y \\ y' &= -\frac{4}{7}x + y^2 + 1. \end{aligned} \tag{6.21}$$

Since this map is not reversible there is no imposed lower limit on the number of cycles this map generates in finite phase spaces. Indeed we see a marked difference between this and the two (reversible) other maps. Figure (6.4) plots the number of cycles against prime for the Henon map.

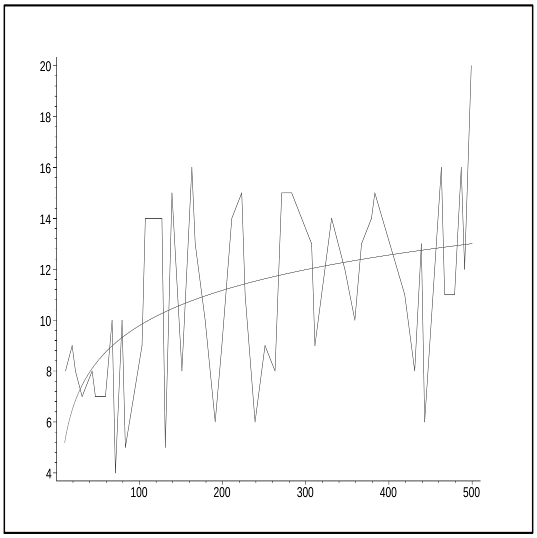

Figure 6.4: Number of cycles vs. prime for the map of example (6.7) which is not reversible. The plotted reference curve is $y = 2\log(x) + \gamma$ where γ is Euler's constant. This reference curve is the expected number of cycles for a random permutation of p^2 points, see [62].

6.2.2 A Menagerie of maps

Here we present the various three dimensional maps we will be using, along with their reversibility patterns and any integrals they may have. Briefly reviewing the various properties we can vary to get a spread of maps we have number of integrals (zero, one or two) and type of reversibility (Type $A - B$ where $A, B = 0, 1, 2$). Historically, reversible maps involving an involution of Type 0 are rare (and of less interest in any case, as described above) and such examples must be artificially constructed. For this reason as well as the fact that the numerics for such maps turn out to be closer to the numerics for a random permutation we stop at Type I-I maps. Note that the existence of an integral is a rare enough phenomenon that we know of no Type I-I maps in three dimensions that possess any integral whatsoever and similarly we know of no Type I-II map in three dimensions with two integrals. Thus this leaves us with the following classes to find examples from: Type II-II with two integrals, one integral and no integrals, Type I-II with one integral and no integral and Type I-I with no integral. Lastly, as a control map we should include a map that is definitely not reversible. This map will likely have no integral; again we know of no maps which possess an integral but are not reversible. We now present examples from these classes as a series of examples which discuss various things considered of relevance in each case.

Example 6.8. The integrable map (that is, two integrals) we shall consider will be from the family of maps from [59]. In general, these maps are alternating and orbits trace out two families of curves in real three dimensional space. In the construction of the map there are 21 parameters divided into 7 3-vectors. The map we consider is the three dimensional map created in the usual way from the recurrence relation

$$L : x_{n+3} = \frac{f_1(x_{n+1}, x_{n+2})x_n + f_2(x_{n+1}, x_{n+2})}{f_3(x_{n+1}, x_{n+2})x_n + \widetilde{f_1}(x_{n+2}, n_{n+1})} \tag{6.22}$$

with $\widetilde{f_1}$ being equal to f_1 with the switch $\alpha \leftrightarrow \beta$, these being two of the seven vectors of parameters. As for the definition of the polynomials f_i, this is done by defining them in terms of three 4×4 matrices:

$$\begin{aligned} f_1(x_{n+1}, x_{n+2}) &= \mathbf{x}_{n+1}^3 \cdot H_1 \mathbf{x}_{n+2}^3 \\ f_2(x_{n+1}, x_{n+2}) &= \mathbf{x}_{n+1}^3 \cdot H_2 \mathbf{x}_{n+2}^3 \\ f_3(x_{n+1}, x_{n+2}) &= \mathbf{x}_{n+1}^3 \cdot H_3 \mathbf{x}_{n+2}^3 \end{aligned}$$

with the usual notation that $\mathbf{z}^i$ is the $i+1$ dimensional vector $(z^i, z^{i-1}, \dots, 1)^T$. The entries of the matrices H_i are given in terms of determinants of matrices built up by the length three vector parameters $\alpha, \beta, p_1, p_2, p_3, p_4$ and p_5 with the convention that $\alpha = (\alpha^0, \alpha^1, \alpha^2), p_1 = (p_1^0, p_1^1, p_1^2)$ and so forth. Let us denote determinants of three such vectors in the manner dictated by the following example.

$$34\beta := \det \begin{pmatrix} p_3^0 & p_3^1 & p_3^2 \\ p_4^0 & p_4^1 & p_4^2 \\ \beta^0 & \beta^1 & \beta^2 \end{pmatrix}.$$

Now we can define the matrices H_i as

$$H_1 = \begin{pmatrix} 123 & 124 & 243 & 143 \\ 124 + 23\alpha & 24\alpha + 13\alpha + 125 & 253 + 14\alpha & 153 \\ 13\alpha + 24\alpha & 134 + 25\alpha + 2(14\alpha) & 254 + 15\alpha + 34\alpha & 154 \\ 14\alpha & 15\alpha + 34\alpha & 154 + 35\alpha & 354 \end{pmatrix}$$

$$H_2 = \begin{pmatrix} 0 & 31\beta & 143 + 41\beta & 43\beta \\ * & 14\alpha + 41\beta + 3\alpha\beta & 153 + 51\beta + 4\alpha\beta & 53\beta \\ * & * & 5\alpha\beta & 345 + 54\beta \\ * & * & * & 0 \end{pmatrix}$$

$$H_3 = \begin{pmatrix} 0 & 123 + 21\beta & 23\beta & 13\beta \\ * & 2\alpha\beta & 234 + 24\beta + 1\alpha\beta & 134 + 14\beta \\ * & * & 41\alpha + 14\beta + 3\alpha\beta & 34\beta \\ * & * & * & 0 \end{pmatrix},$$

where the missing entries of H_2 and H_3 can be filled out using the relation $\widetilde{H}_j^T = -H_j$ i.e. taking the tranpose of H_2 (or H_3) and then switching $\alpha \leftrightarrow \beta$ has the same effect as taking the negative of H_2 (or H_3). With this rather lengthy set up we are now poised to say a few things. A fact that is not obvious, but given in [59], is that when $\alpha = \beta$, the map L becomes integrable and measure-preserving. It leaves fixed two families of surfaces and hence leaves fixed the intersections of these families. More generally when $\alpha \neq \beta$ the map L is an alternating map (i.e. it acts differently on odd numbered iterates than it does on even numbered iterates) and the integrals become 2-integrals (i.e. quantities preserved every two iterates). Unless otherwise mentioned, our parameter choices are:

$$\begin{aligned} p_1 &= (1,2,3) \\ p_2 &= (4,5,6) \\ p_3 &= (7,8,9) \\ p_4 &= (1,2,4) \\ p_5 &= (1,2,5) \\ \alpha &= (4,5,7) \\ \beta &= (4,5,7), \end{aligned}$$

which were chosen in such a way that the preserved curves (the intersections of the two preserved surfaces) are clearly visible whilst maintaining some simplicity in the form of the map but not so simple as to have parallel vectors leading to some uninteresting determinants. A small piece of the phase space portrait of this map can be seen in figure (6.5).

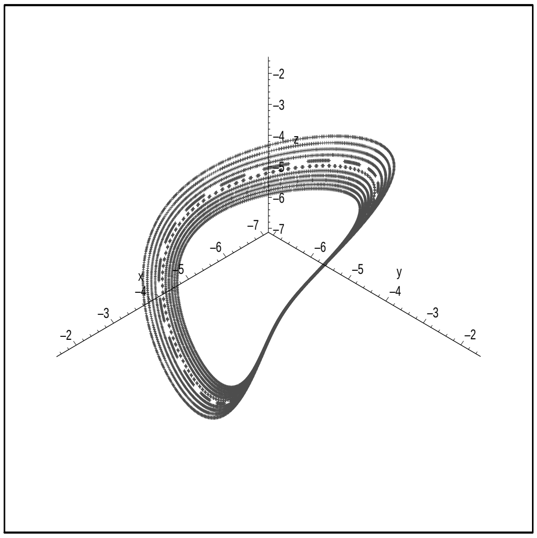

Figure 6.5: Part of the phase space portrait generated by the map of example (6.8). The orbits displayed trace out curves that are the intersections of two surfaces.

It is easy to calculate that L with these paramater choices is type II-II reversible but as in the two dimensional case the signature given by the lower bound on the number of cycles of $\frac{p^2+p^2}{2}$ is expected to be washed out by the integrability. A problem with this map is that we are unable to employ a trick to make the map a permutation of the affine part of some finite planes. Hence when we count the number of orbits we may get a low number, depending on how many orbits are not acting as a permutation of the affine part of the finite plane in question.

Example 6.9. The family of maps from [18] from which we draw a type II-II example has the general form

$$\begin{aligned} L_{GM} : x' &= y \\ y' &= z \\ z' &= x + \frac{(y-z)(\alpha - \beta yz)}{1 + \gamma(y^2+z^2) + \beta yz + \delta y^2 z^2}. \end{aligned} \tag{6.23}$$

It is noted that such a map always possess an integral of motion, namely

$$\begin{aligned}\Phi(x,y,z) = x^2+y^2+z^2+\alpha(xy+yz-zx)+\gamma(x^2y^2+y^2z^2+z^2x^2)+ \\ \beta(x^2yz+z^2xy-y^2zx)+\delta x^2y^2z^2.\end{aligned} \tag{6.24}$$

Unnoted but readily calculable is that L_{GM} is reversible, being the composition $G_{GM} \circ H_{GM}$ with

$$\begin{aligned} H_{GM} : x' &= z \\ y' &= y \\ z' &= x \end{aligned} \tag{6.25}$$

and

$$\begin{aligned} G_{GM} : x' &= y \\ y' &= x \\ z' &= z + \frac{(x-y)(\alpha-\beta xy)}{1+\gamma(x^2+y^2)+\beta xy+\delta x^2y^2}. \end{aligned} \tag{6.26}$$

Finding the fixed sets of these two involutions is not difficult. It is clear that $Fix(H_{GM}) = \{(x,y,x)|x,y \in K\}$ while for G_{GM}, one has just to notice that we need $x = y$ from the first two ordinates which then trivialises the equation from the third ordinate. This results in $Fix(G_{GM}) = \{(x,x,z)|x,z \in K\}$. As these two sets are both two dimensional, the map L_{GM} is indeed type II-II as desired and the lower bound inherited from equation (6.17) is $\frac{p^2+p^2}{2} = p^2$. The last property we must ensure in L_{GM} before performing the numerics over finite phase spaces is that it is properly invertible. Since this map is rational we are not automatically assured of this. So using the same method as in the two dimensional case, we force all denominators in the map to be of the form Z^2+1. In this case this is achieved by taking $\alpha = 1, \beta = 0, \gamma = 0, \delta = 1$ under which substitutions the denominator of L_G becomes y^2z^2+1. Now we restrict ourselves to primes p congruent to 3 modulo 4 and the map acts as a permutation on the finite spaces $\mathbb{Z}_p^3$.

Example 6.10. A second Type II-II map we use is only a small variation of the previous example. Rather than taking $\alpha = 1$ ($\beta = 0, \gamma = 0, \delta = 1$) in equation (6.23) we instead take $\alpha = 2$. The reason for the inclusion of this map was solely to get some idea of the universality of the conjecture in the next subsection and therefore we do not give the number of cycles versus prime plot for this map.

Example 6.11. The Fibonacci trace map, so called as it can be derived from the traces of a sequence of matrices, is a well known example of a three dimensional map with one integral and has been studied in [54] and [58] as well as noted in [18] as arising in a physical setting from Schrödinger's equation. A more general form of the trace map we are interested in is

$$\begin{aligned} L_T : x' &= y \\ y' &= z \\ z' &= -x + h(y,z) \end{aligned} \tag{6.27}$$

with $h(y,z) = h(z,y)$. Such a symmetry condition is imposed to ensure that the map L_T is reversible, being made up of the composition of the two involutions $G_T \circ H_T$ with

$$\begin{aligned} H_T : x' &= z \\ y' &= y \\ z' &= x \end{aligned} \tag{6.28}$$

and

$$\begin{aligned} G_T : x' &= y \\ y' &= x \\ z' &= -z + h(y,x). \end{aligned} \tag{6.29}$$

With these two involutions explicitly given it is easy to see that $Fix(H_T)$ is the plane $\{(x,y,x)|x,y \in K\}$ and $Fix(G_T)$ is the curve $\{(x,x,\frac{h(x,x)}{2}|x \in K\}$ with K being an appropriate field over which the map is defined. Being a polynomial map that is clearly invertible (being made up as it is of two involutions), L_T is invertible on the affine part of the plane and thus acts as a permutation of the affine part of the finite space $\mathbb{Z}_p^3$ for all choices of prime p. Checking the cardinality of $Fix(H_T)$ and $Fix(G_T)$ in such a finite phase space is easy and we get that $|Fix(H_T)| = p^2$ and $|Fix(G_T)| = p$ and so the lower bound on the number of cycles from equation (6.17) is $\frac{p^2+p}{2}$. To ensure that L_T has one integral of motion we set $h(y,z) = 2yz$; this choice makes L_T the actual Fibonacci trace map.

Breaking the integral of motion but retaining the reversibility types possessed by the maps in examples (6.11) and (6.9) is easily achieved by altering one of the involutions in each map.

Example 6.12. While perturbing map (6.27) in such a way to retain reversibility but destroy the integral of motion is not hard to do from first principles, it has already been done in [58] where it is noted that taking $h(y,z) = 2yz + \epsilon(y^2 + z^2)$ gives the desired effect. Define

$$\begin{aligned} L_{pT} : x' &= y \\ y' &= z \\ z' &= -x + 2yz + \frac{5}{7}(y^2 + z^2). \end{aligned} \tag{6.30}$$

In this case we have the same cycle lower bound of $\frac{p^2+p}{2}$ so comparing the data from this map with that from example (6.11) may show differences between Type I-II maps with one integral and maps with no integral.

Example 6.13. For consistency's sake, we perturb map (6.23) in a similar way. By adding the term $\frac{5}{7}(y^3 - z^3)$ to the z-coordinate of the map we achieve the desired result of destroying the integral of motion whilst preserving the reversibility and its type of II-II. The new map, then, is

$$\begin{aligned} L_{pG} : x' &= y \\ y' &= z \\ z' &= x + \frac{(y-z)}{1 + y^2 z^2} + \frac{5}{7}(y^3 - z^3). \end{aligned} \tag{6.31}$$

The reason we must perturb this map with the term $y^3 - z^3$ instead of $y^2 + z^2$ as in the case of the trace map is that to remain reversible in the same manner, the "complicated part" of z' (that is, if we wrote $z' = x + F(y,z)$, the function F) must be an odd function. The simplest odd function that destroys the integral of motion is just this $y^3 - z^3$.

The next two examples are taken from [58] where they were used to illustrate local structures in phase space coming from Type I-II and Type II-II maps. These two maps were studied here as some more distinct examples from their reversibility categories as evidence for the conjecture in the next subsection and therefore we do not give number of cycles plots as for most of the other examples.

Example 6.14. Extracted from Example 1 of the appendix of [58], this map is given as an example of a Type I-II map. It has equation

$$\begin{aligned} x' &= (k-y)(1+(y'-1)^2) \\ y' &= \frac{x+e(2y-k)(z+e(y-k))}{1+(y+1-k)^2} \\ z' &= -z+e(k-2y) \end{aligned} \tag{6.32}$$

where e is a parameter that ensures the map deviates from what is essentially a two dimensional map. We use parameter choices $e = 2, k = -1$. Notice that, conveniently, all denominators are of the form $(1+Z^2)$.

Example 6.15. Extracted from Example 2 of the appendix of [58], this map is the Type II-II counterpart of example (6.14). It has equation

$$T_\epsilon^{-1} \circ H \circ T_\epsilon \circ G \tag{6.33}$$

where

$$\begin{aligned} T_\epsilon : x' &= x+\epsilon F(y) \\ y' &= y \\ z' &= z+\epsilon K(y), \end{aligned} \tag{6.34}$$

$$\begin{aligned} H : x' &= y+f(x,z) \\ y' &= x-f(x',z) \\ z' &= z \end{aligned} \tag{6.35}$$

and

$$\begin{aligned} G : x' &= x \\ y' &= -y \\ z' &= z. \end{aligned} \tag{6.36}$$

The parameters taken are $\epsilon = 1, f(x,z) = xz-(1-z)x^2, F(y) = y$ and $K(y) = y(y^2-1)$.

To find a map that is reversible and of Type I-I we must actively construct one. We begin with a lemma to aid in the construction of such a map.

Lemma 6.16. *Let L be an involution with a d dimensional fixed set. Then $M = G \circ L \circ G^{-1}$ is also an involution with a d dimensional fixed set.*

Proof. That order is a conjugacy invariant is well known and easily checked in this case by the verification

$$M^2 = (GLG^{-1})(GLG^{-1}) = GL^2G^{-1} = GG^{-1} = id.$$

That the size of fixed sets is preserved is similarly easy to prove. Suppose that P is a fixed point of L. Then consider $M \circ G(P) = G \circ L(P) = G(P)$ so that $G(P)$ is a fixed point of M. Since G is invertible, it acts as a bijection between $\mathrm{Fix}(L)$ and $\mathrm{Fix}(M)$ whence in particular they have the same dimension. □

Lemma (6.16) tells us that we can construct a Type I-I map by taking two simple involutions with one dimensional fixed sets, alter them by taking a conjugate map using some (different) non-linear conjugacies, then the map made up of the composition of the two resulting involutions will be a non-linear Type I-I reversible map.

Example 6.17. Let G_C and H_C be the involutions

$$\begin{aligned} G_C : x' &= y \\ y' &= x \\ z' &= -z + x^2 + y^2 \end{aligned} \tag{6.37}$$

and

$$\begin{aligned} H_C : x' &= z \\ y' &= -y \\ z' &= x. \end{aligned} \tag{6.38}$$

Of these maps, G_C has been constructed by taking a conjugacy of the simple involution

$$\begin{aligned} g_C : x' &= x \\ y' &= -y \\ z' &= -z. \end{aligned} \tag{6.39}$$

with the invertible non-linear map

$$\begin{aligned} \alpha_C : x' &= \frac{1}{2}(x-z) \\ y' &= \frac{1}{2}(x+z) \\ z' &= -y + x'^2 \end{aligned} \tag{6.40}$$

in a manner such that $G_C = \alpha_C g_C \alpha_C^{-1}$. Because the resulting G_C and H_C are both polynomial maps, they have no singularities in the affine part of the projective plane. Therefore we can apply the usual lower bound to the number of cycles that $L = G_C \circ H_C$ decomposes this part of the plane into. In this case, the bound is $\frac{p+p}{2} = p$.

Our last example will not be reversible and will have no integrals of motion.

Example 6.18. Consider the family of generalised Henon transformations from example (6.7) defined by

$$\begin{aligned} x' &= y \\ y' &= -\delta x + h(y). \end{aligned} \tag{6.41}$$

For $\delta \neq 1$ and h a polynomial of degree greater than one, this map is not reversible (see [61]) and we shall use its natural extension to three dimensions. Consider that as a recurrence relation we may write equation (6.41) as

$$x_{n+2} + \delta x_n = h(x_{n+1}).$$

It is not too difficult to see that the extension of this recurrence relation is simply

$$x_{n+3} + \delta x_n = h(x_{n+1}, x_{n+2}). \tag{6.42}$$

This gives the three dimensional map

$$\begin{aligned} x' &= y \\ y' &= z \\ z' &= -\delta x + h(y,z). \end{aligned} \tag{6.43}$$

The reason that this fails to be reversible is again because the Jacobian determinant is constant and equal to $-\delta$ and has a fixed point for reasonable choices of $h(y, z)$. In this case we choose $h(y, z) = 2yz$ and $\delta = 2$. It is quite uncanny how such a small deviation from the trace map will make quite a large difference in the statistics.

Table (6.1) lists the above example maps with their relevant properties.

Abbreviation	Map	Reversibility	Integrals
3DQRT	Ex. (6.8)	II-II	2
G-M	$x' = y,\ y' = z,\ z' = x + \frac{y-z}{1+y^2z^2}$ Ex. (6.9)	II-II	1
G-M 2	$x' = y,\ y' = z,\ z' = x + 2\frac{y-z}{1+y^2z^2}$ Ex. (6.10)	II-II	1
Trace	$x' = y,\ y' = z,\ z' = -x + 2yz$ Ex. (6.11)	I-II	1
Pert. trace	$x' = y,\ y' = z,\ z' = -x + 2yz + \frac{5}{7}(y^2 + z^2)$ Ex. (6.12)	I-II	0
Pert. G-M	$x' = y,\ y' = z,\ z' = x + \frac{y-z}{1+y^2z^2} + \frac{5}{7}(y^3 - z^3)$ Ex. (6.13)	II-II	0
J. I-I	$x' = -y,\ y' = z,\ z' = -x + z^2 + y^2$ Ex. (6.17)	I-I	0
Henon 3D	$x' = y,\ y' = z,\ z' = -2x + 2yz$ Ex. (6.18)	None	0
R-L 1	$x' = (-1 - y)(1 + (y' - 1)^2),\ y' = \frac{x+2(2y+1)(z+2(y+1))}{1+(y+2)^2}$	I-II	0
	$z' = -z + 2(-1 - 2y)$ Ex. (6.14)		
R-L 2	Ex. (6.15)	II-II	0

Table 6.1: Some three dimensional maps and their properties. Note that while the fourth column is listed as "Integrals" it would be more strictly accurate to say "Known Integrals". However the numerical evidence suggests that the number given is correct anyway.

Figure (6.6) shows, as a function of prime, the number of cycles generated by each of these maps in the affine phase space ${\mathbb{Z}_{|}}^{3}$. Only primes for which the map acts as a permutation are included. An enlarged part of the bottom of the plot can be seen in figure (6.7).

6.2.3 Numerical Data

To present data to support the extension of the results from [61] the most obvious way to proceed is to calculate the same observables calculated to support the original hypotheses in two dimensions. To do this we will be decomposing single phase spaces $\mathbb{Z}_p^3$ under the map and storing lists that contain data of the form [cycle length l, number of cycles with length l]. Armed with such lists we are capable of finding a lot of interesting observables including number of points in (periodic) cycles, number of cycles, average cycle length, maximum cycle length and cumulative

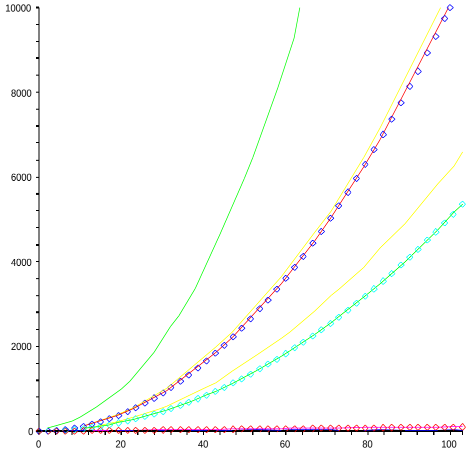

Figure 6.6: The number of cycles generated by (from the top to bottom; see table (6.1)) 3DQRT, G-M, Pert. G-M, Trace, Pert. trace, J. I-I and Henon 3D. The curve made up of boxes are the reference curves $y = x^2$, $y = \frac{x^2+x}{2}$ and $y = x$. Pert. G-M (Type II-II) and Pert. trace (Type I-II) pass through their reference curves almost exactly. An enlarged portion of the bottom of the plot can be seen in figure (6.7).

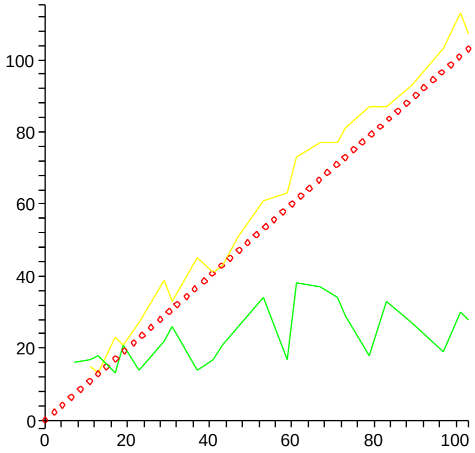

Figure 6.7: Shows the number of cycles generated by J. I-I and Henon 3D maps along with the reference curve $y = x$.

frequency distributions. It is this last observable upon which a conjecture was made in [61]. Also being the most complicated it warrants an explanation. Recall first the definition

$$D_p(x) = \frac{|\{y \in C_p : T(y) \leq rx\}|}{|C_p|} \tag{6.44}$$

where C_p is the set of periodic points we are considering, $T(y)$ is the period of the point y and r is the normalising factor being used. When we plot such a distribution we will have, on the x-axis, normalised cycle length $\frac{T(y)}{r}$ while the y-axis counts the proportion of (periodic) phase space consumed in cycles of normalised length at most x. Because we are counting only the periodic points in phase space, the maximum value of y is one. The normalising factor (by which all orbit lengths are divided) one should use is an interesting question; it shall be argued that it only makes a cosmetic difference. For now we present these observables for the various maps which we use as our three dimensional reversibility examples. Firstly we have data that were collected using the phase space $\mathbb{Z}_{103}^3$ so that the maps we use are permutations. Secondly we use the phase space $\mathbb{Z}_{101}^3$ so that many of the maps are not invertible. The challenge here is to see if running tests on the small periodic parts of phase space return equally valid results as when the tests are run over entirely periodic phase spaces. Notice that only the plot of the cumulative frequency distribution gives a meaningful spread of data for any particular phase space decomposition; the other observables mentioned earlier are just a single number per phase space decomposition. Therefore to see trends and fits to predictions, it is necessary for these other observables to look at many phase space decompositions rendering them less efficient.

It was conjectured in [61] that cumulative frequency distributions of the type we plot fit, for two dimensional reversible maps, the distribution $y = 1 - e^{-x}(1 + x)$ assuming a normalising factor $r = p$. Through reviewing the numerical data we generated and some appropriate guesswork, a stronger conjecture presented itself.

Conjecture 6.19. *Suppose L is a three dimensional reversible map acting as a permutation on P points that decomposes those P points into N cycles. Suppose also that L has a single family of reversing symmetries. Then the cumulative frequency* *distribution* (6.44) *satisfies*

$$\lim_{p\to\infty} D_p(x) = 1 - e^{-ax}(1+ax) \tag{6.45}$$

with $a = 1$ *assuming a normalising factor* $r = \frac{P}{N}$ *i.e. a normalising factor equal to the average cycle length. Indeed, more generally one may take any normalising factor* r *and arrive at the distribution* (6.45) *but with*

$$a = \frac{rN}{P}. \tag{6.46}$$

It is easy to see how this conjecture really required work in three dimensions to formulate. In two dimensions, previously studied reversible maps have always been of Type I-I, resulting in a minimum of p cycles. In a phase space of p^2 points (with the observation that this minimum seems to always be adhered to), this means the normalising factor given by conjecture (6.19) is p, which was the natural normalising factor taken by Roberts and Vivaldi in [61]. The value p was chosen as the square root of the periodic phase space (for maps which were permutations on the whole space this happens to be p). While it is perhaps a little unnatural to present conjecture (6.19) before any numerical evidence either supporting or opposing it, something is required to inform the way we present our data. We can now choose the normalising factor when presenting the distributions to be the average cycle length and then test the conjecture by checking how well the actual data fits the predicted curves and if the fit improves as larger primes are taken.

Figure (6.8) shows three different plots all for $p = 103$. The first contains the cumulative frequency distribution plots for all Type II-II reversible maps normalised by average cycle length, the second shows the same for all Type I-II maps, the third shows the same for the lone Type I-I map. Also on each plot is the reference curve $y = 1 - e^{-ax}(1 + ax)$ with $a = 1$.

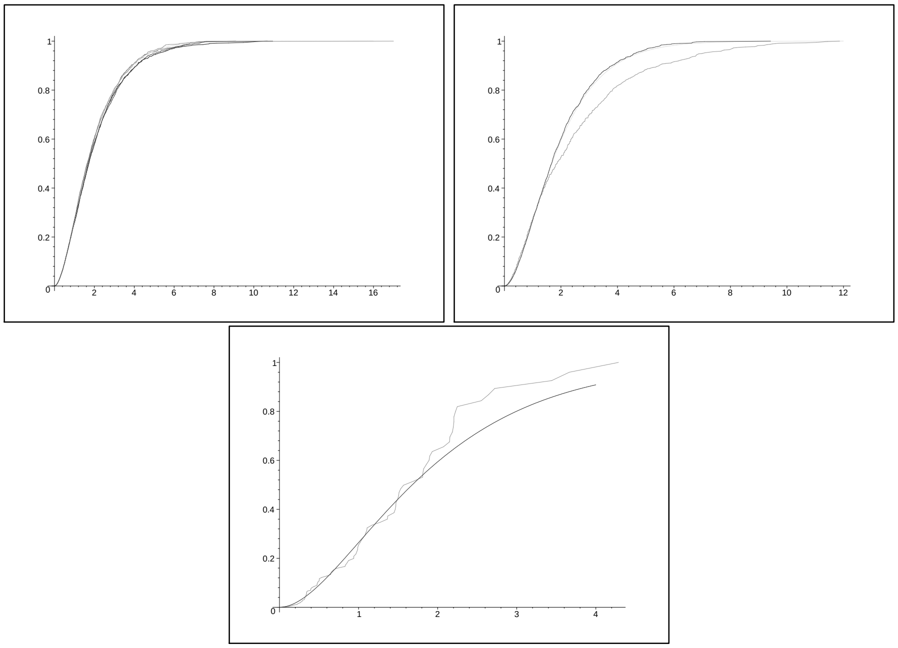

Figure 6.8: Cumulative frequency distributions for reversible maps, $p = 103$. Clockwise from top left, the first contains distributions for type II-II maps G-M,G-M 2, Pert. G-M and R-L 2, the second contains distributions for Type I-II maps Trace and Pert. trace and the third contains the distribution for the map J. I-I.

Figure (6.8) is a little coarse to see how far each curve is deviating from the predicted curve. To show this with more refinement we also have a series of plots, one for each map/phase space combination that has as its y data the ratio $\frac{D_p(x)}{1-e^{-x}(1+x)}$ and as its x data the usual normalised cycle length. Figure (6.9) shows this deviance from the predicted values for the Type II-II maps, figure (6.10) shows this deviance from the predicted values for the Type I-II maps and figure (6.11) shows this deviance from the predicted values for the Type I-I map in three different cases: in the first no data points have been omitted, in the second one the first data point has been omitted and in the third the first two data points have been omitted.

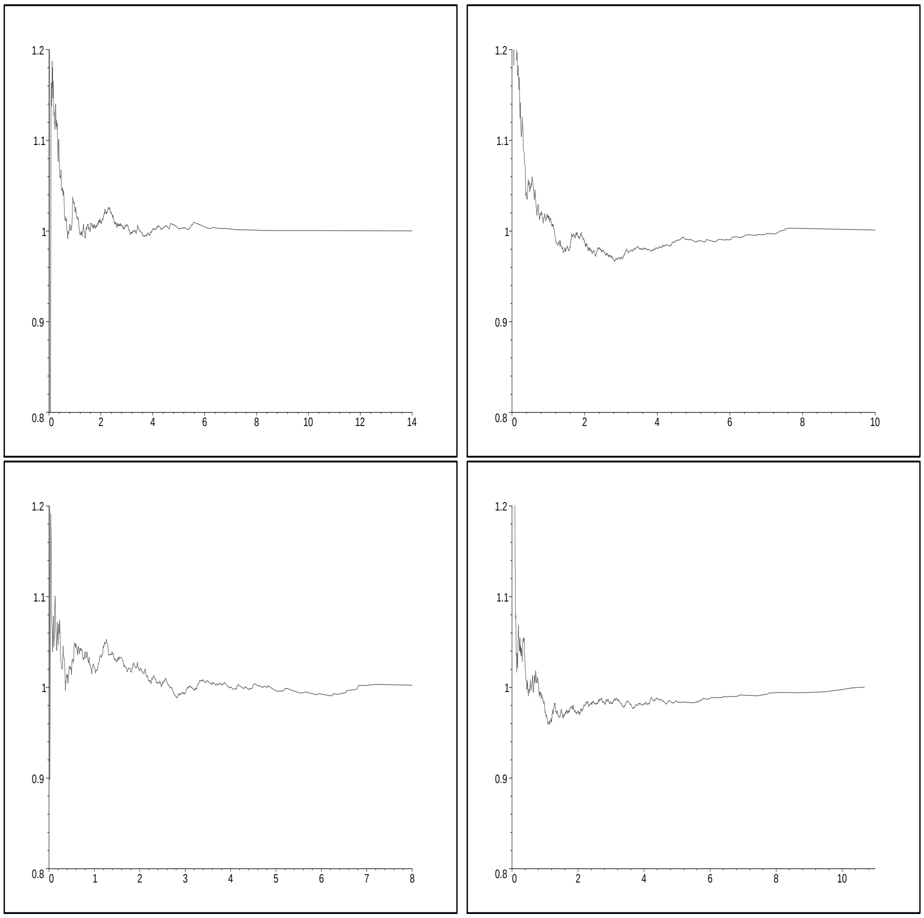

Figure 6.9: Ratio of actual data to predicted value for type II-II maps,$p = 103$. Clockwise from top left they are G-M, G-M 2, R-L 2 and Perturbed G-M.

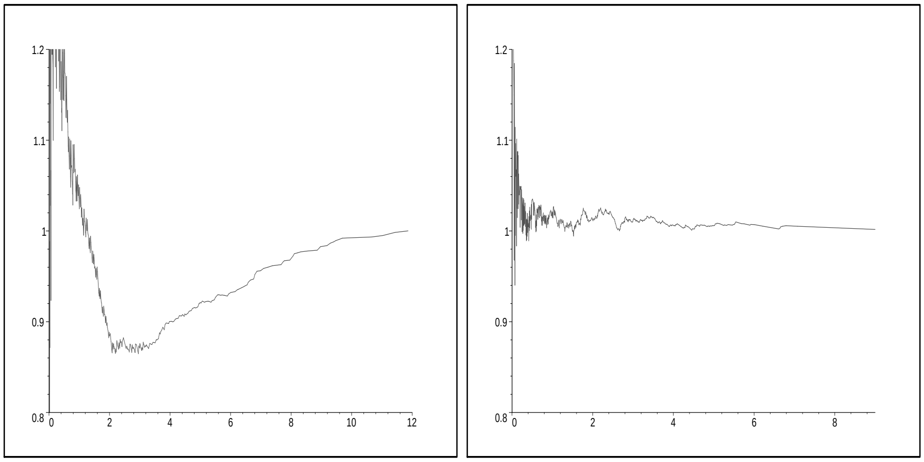

Figure 6.10: Ratio of actual data to predicted value for Type I-II maps,$p = 103$. From left to right they are Trace and Perturbed Trace.

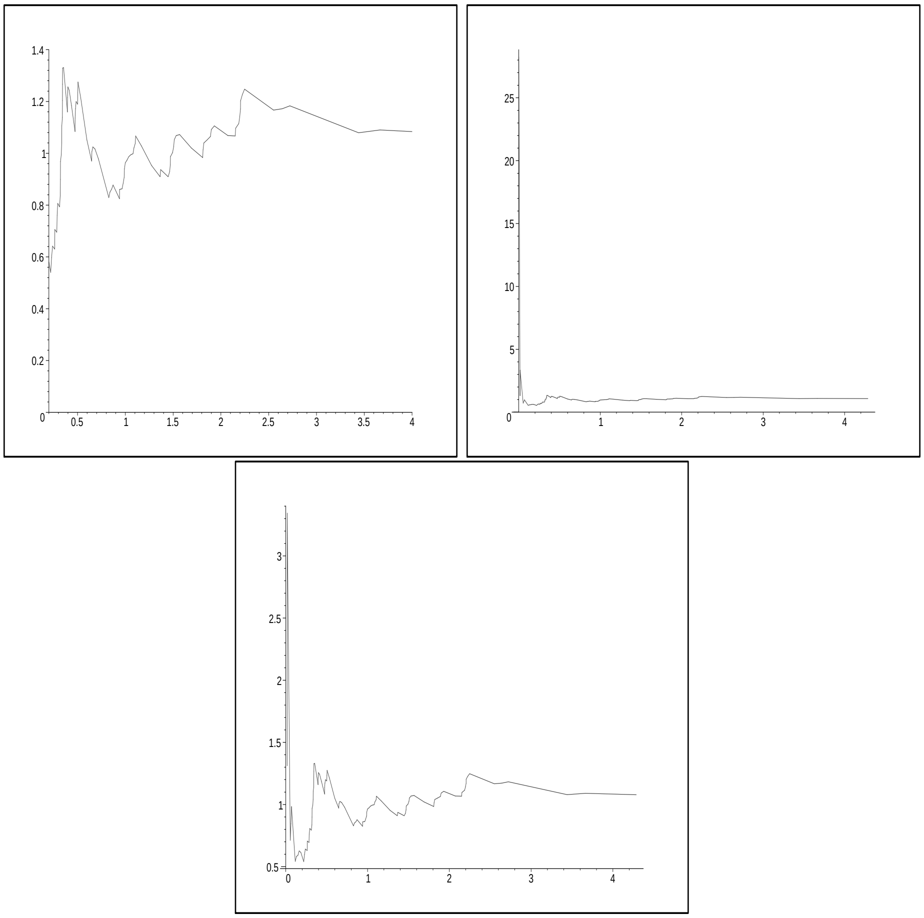

Figure 6.11: Ratio of actual data to predicted value for Type I-I map,$p = 103$. The map in all three is J. I-I. Clockwise from top left we ignore no data points, the first data point and the first two data points respectively. This removes contributions from incredibly short orbits (fixed points and two cycles).

Figures (6.8) through to (6.11) all arise from studying the maps with $p = 103$. As mentioned, certain parameters in these maps were chosen so that when a prime was chosen congruent to three modulo four, the map would act as a permutation on the affine part of the relevant phase space and never have orbits that “leak off” to infinity. As evidence that reversibility also has a signature when the map does not act as such a permutation, we present the same data for $p = 101$. Note that we still only look at the portion of the phase space where the map does act as a permutation, the statement being supported by this data is that the orbits we lose in such cases do not affect the statistics we observe. Figure (6.12) shows the cumulative frequency distribution plots; the first is for the type II-II maps, the second for the

Type I-II maps. Figures (6.13) and (6.14) show, respectively, the ratio of the actual distribution values and the predicted values for type II-II and Type I-II maps in the same order as for $p = 103$.

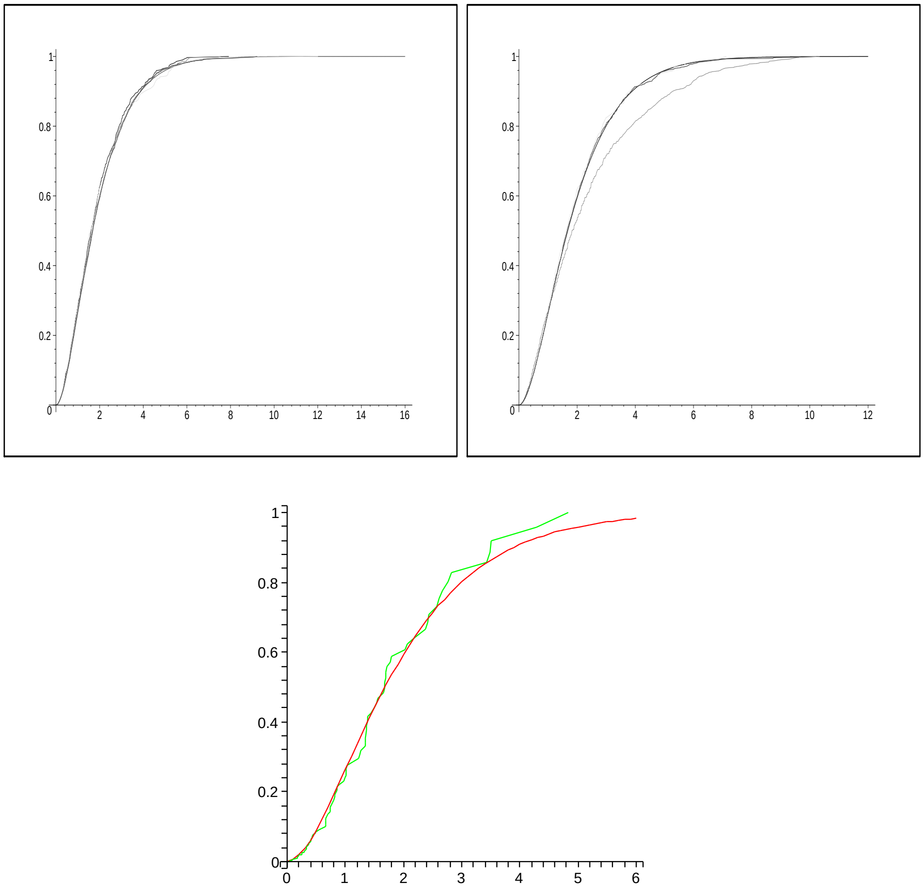

Figure 6.12: Cumulative frequency distributions for reversible maps, $p = 101$. Clockwise from top left, the first contains distributions for type II-II maps G-M,G-M 2, perturbed G-M and R-L 2, the second contains distributions for Type I-II maps Trace and Perturbed Trace and the third contains the distribution for the map J. I-I.

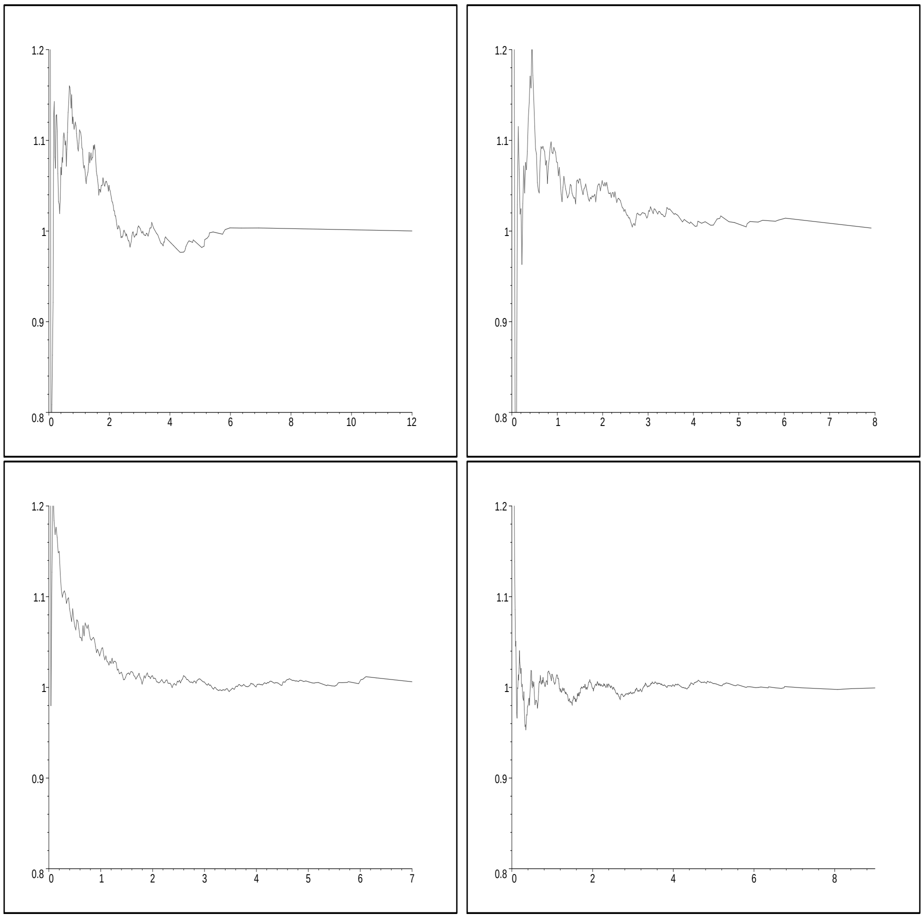

Figure 6.13: Ratio of actual data to predicted value for type II-II maps,$p = 101$. Clockwise from top left they are G-M, G-M 2, R-L 2 and Perturbed G-M.

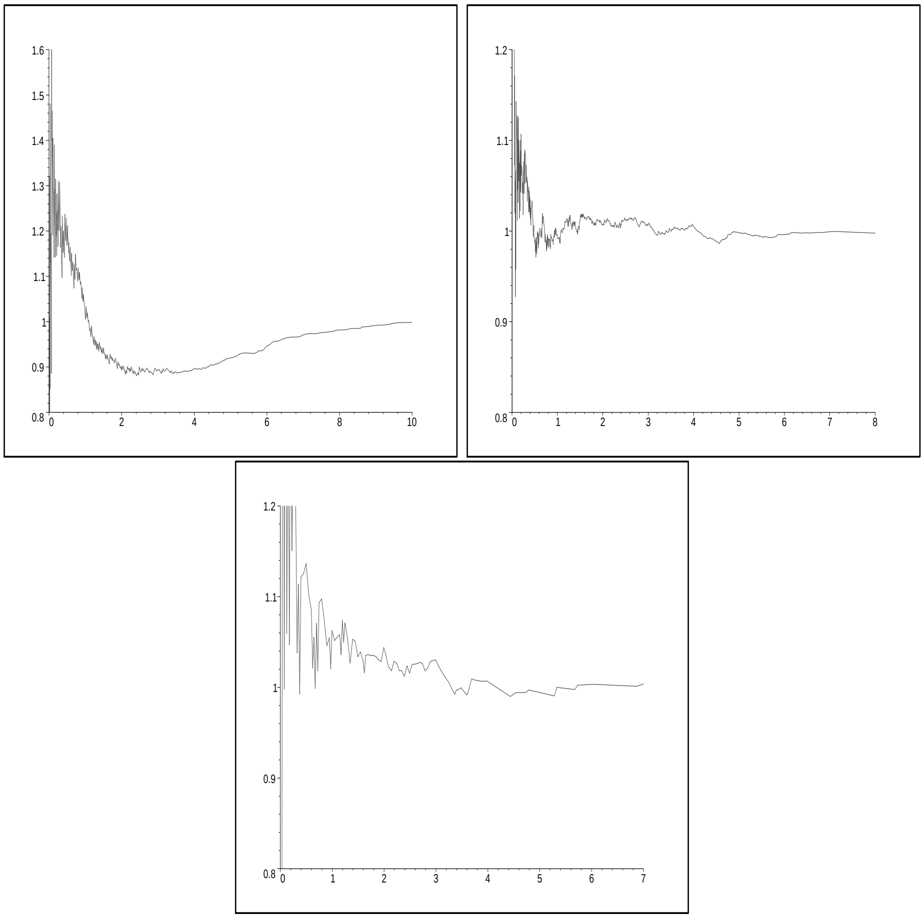

Figure 6.14: Ratio of actual data to predicted value for Type I-II maps,$p = 101$. Clockwise from the top left they are Trace, Perturbed Trace and R-L 1.

While this cumulative frequency distribution of normalised orbit lengths remains the most thorough way to look for signatures of time-reversibility, a single number statistic that may be of use is the average cycle length. Note that it is exactly this number by which we normalised cycle lengths to generate uniform distributions. Thus it might be expected to differentiate between different types of reversibility. Table (6.2) shows the average cycle lengths for the different maps for which we have shown the cumulative frequency distribution data for primes 103 and 101. The numbers in this table tells us that average cycle length is not a reliable statistic when we are forced to cut out parts of the phase space. The reason for this, which had been observed previously as far back as [60], is that aperiodic orbits generally consume an inordinately large proportion of phase space. So while maps with aperiodic orbits lower the number of periodic points a great deal, they do this without lowering the number of cycles a great deal. This results in a much lower average cycle length even though it does not much change the cumulative frequency distribution as the periodic orbits are still distributed in a similar way as in the fully periodic case.

Map	$p = 103$	$p = 101$
Trace	165.9	164.3
Pert. Trace	203.8	199.1
R-L 2	102.9	100.9
G-M	99.3	50.3
G-M 2	99.2	49.8
Pert. GM	102.9	50.1
R-L 1	203.9	40.1
J I-I	10212.4	9117.7

Table 6.2: Average cycle lengths for $p = 103$ and $p = 101$.

Conjecture (6.19) claims, however, that when plotting the cumulative frequency distribution of orbit lengths of reversible maps, the curve should fit $y = 1 - e^{-ax}(1 + ax)$ with $a = \frac{rN}{P}$ where N is the number of cycles considered, P is the number of points considered and r is the normalising factor used. Let us see how this pans out if one uses the normalising factor proposed in [61]. As mentioned earlier, Roberts and Vivaldi used the nth root of the number of points decomposed where n is the dimension of the map. For n-dimensional polynomial and other maps which only generate periodic orbits this is just the nth root of the total number of points in the finite phase space $\mathbb{Z}_p^n$ which is just p. The first hint that this could be improved came from the data from the map J. I-I of example (6.17). This map, being Type I-I reversible, has rather larger orbits such that they are comparable to p^2 in length rather than just p. So normalising all lengths by p gives a cumulative distribution that is far from a distribution of the form (6.45) with $a = 1$ which was the only case that arose in [61]. However in this case, even when normalising by a more appropriate number, which is around p^2, the fit for this map is still fairly loose although the trend of conformity is noticeable the looseness appears to come from the fact that orbits for this Type I-I are typically large. This leads to a small number of orbits and hence a small data set. Presumably by increasing the prime to larger values better

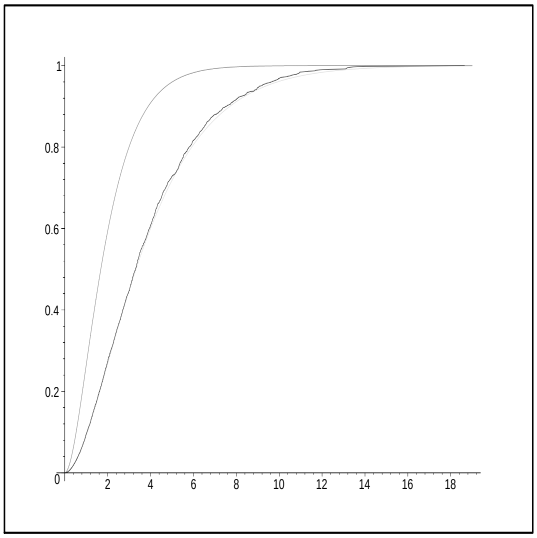

Figure 6.15: The distribution for the perturbed trace map with normalising factor of 103 (i.e. $r = 103$) and the theoretical distributions from equation (6.45) with $a = 1$ (the lone top curve) and $a = \frac{rN}{P}$. We see that using $a = \frac{rN}{P}$ gives a tight fit to the data while $a = 1$ gives a curve far from the data.

conformity to the predicted distribution would ensue. Since the Type I-I map is still quite loose at this prime size, to illustrate our point here we use the perturbed trace map. Considering the data set given by setting $p = 103$ and normalising factor $r = 103$, we see how this overlays the distribution (6.45) with $a = \frac{rN}{P}$ while it is quite different from the same distribution with $a = 1$. These three curves are seen in figure (6.15).

To calculate what a should be used, one needs to know only the number of periodic points and the reversibility type of the map. From the reversibility type of the map, we can get a lower bound on the number of cycles which is adhered to quite strictly in most cases and thus can use this lower bound for an estimation of the number of cycles which in this case is just $\frac{p+p^2}{2}$. Also in this case since we have a polynomial map we know the entire phase space (p^3 points) is periodic so with a normalising factor of p we can put these three values into equation (6.46) to get $a = \frac{p^3+p^2}{2p^3} = \frac{1}{2} + \frac{1}{2p}$ with $p = 103$.

6.2.4 Proportion of symmetric points and multiple reversing symmetries

One can see, for example, from the second plot in figure (6.8) that the trace map does not adhere as well to the predicted distribution for its reversibility type as the other maps studied. In this section we attempt to explain this inconsistency and give reason to believe that this is a fairly exceptional situation. An oddity with the trace map can be seen as far back as figure (6.6) where the number of cycles generated by the trace map is significantly greater than the minimum number implied by the reference curve $y = \frac{x^2+x}{2}$ for its reversibility type. The two other maps that exhibit similar behaviour are the QRT3D map and the G-M map. Firstly we shall explain the situation for the QRT3D map as it is due to a different reason than in the other two cases.

A theorem due to De Vogelaere (given in [58]) tells us that if a map is reversible and can be written $L = G \circ H$ then any symmetric (under G) orbit of even length is such that exactly two points of it are from $Fix(G)$ or $Fix(H)$ and any symmetric orbit of odd length is such that exactly one point of it is from $Fix(G)$ and one point is from $Fix(H)$. Let us examine the consequences of this in the case of an integrable map i.e. where each orbit is confined to elliptic curves (and the same elliptic curves are made up of orbits of the same length). Suppose a particular level set C is comprised of orbits of even length. The intersection of C with $Fix(G)$ is generically finite, as is the intersection of C with $Fix(H)$. Specialising a little to get some idea of the cardinality of these intersections, in the case of a Type I-I reversible map in two dimensions where the integral of motion is biquadratic the number of real intersections is at most 2. This leaves at most 2 symmetric orbits per level set and we have seen in the past that, typically, level sets decompose into considerably more than 2 orbits and each of these orbits except 2 must necessarily be asymmetric. While the intersection calculations in three dimensions are less simple, the objects are the same - the level sets are still elliptic curves, the fixed sets of the involutions are either points, lines or planes and so forth. Thus by analogy we expect that for integrable three dimensional maps there will be more asymmetric orbits than symmetric orbits.

One reason this line of reasoning fails to work for three dimensional maps with only one integral is that the intersection of each level set (likely to be a surface) with $Fix(G)$ is not limited to being of finite size. Yet it is the maps with one integral that have more asymmetric orbits than their perturbed counterparts with no such integral. So while this gives one possible explanation - that the additional restriction imposed by the single integral creates more orbits - more likely is that the additional structure enforced upon the phase space decompositions by these maps having an additional family of reversing symmetries creates these extra asymmetric orbits.

Let us look first at the G-M map, whose equation can be seen in table (6.1). It is fairly obvious that this map commutes with the involution $(x, y, z) \to (-x, -y, -z)$. The existence of this symmetry, along with the already known reversing symmetry, gives rise to a second independent reversing symmetry care of proposition (3.20).

Upon calculating the fixed sets of the two involutions that make up the G-M map with this new reversing symmetry, one finds that it exhibits Type I-I reversibility. So the G-M map is both Type I-I reversible and Type II-II reversible. Certainly the number of cycles it generates must satisfy the larger lower bound given by the Type II-II reversibility but it would perhaps not be a surprise if the existence of this Type I-I reversibility structure drove up the number of asymmetric orbits (that is, asymmetric under the Type II-II regime). Due to the larger size of the fixed sets in the Type II-II regime, we see more symmetric orbits than under the Type I-I regime. Table (6.3) shows the proportion (to two significant figures) of phase space consumed in (a)symmetric orbits under each regime for the phase space Z_{83}^3.

	Symmetric II-II	Asymmetric II-II
Symmetric I-I	0.02	0.00023
Asymmetric I-I	0.97	0.011

Table 6.3: Proportion of phase space consumed in (a)symmetric orbits for the two types of reversibility exhibited by the G-M map working over $p = 83$.

The case of the trace map is more complicated. The trace map does not actually have an extra symmetry that in turn implies the existence of a second reversing symmetry. Rather it possesses a family of almost-symmetries. The group of size four made up of the identity and the three maps $\sigma_1 : (x, y, z) \to (x, -y, -z)$, σ_2 :

$(x, y, z) \to (-x, y, -z)$ and $\sigma_3 : (x, y, z) \to (-x, -y, z)$ is such that if L is the trace map (again, see table (6.1)) then $\sigma_i \circ L = L \circ \sigma_{i+1}$ with the subscripts wrapping around in the obvious way. The consequences of these almost-symmetries are unclear in this context; in the literature they are called k-symmetries (see, for example, [36] which concentrates on when the group of these k-symmetries is cyclic and references therein). It would certainly fit with the numerical patterns seen as well as comments from [55] that it would impose more structure onto the map's orbits, structure which could disturb an otherwise tight fit to the lower bound of orbits expected for maps with a single reversing symmetry. However, the details have not been explored. It is certainly true that if we consider the cube of the trace map, these almost-symmetries become symmetries and we can create a similar table to table (6.3) except here we would need a table of more than two dimensions to account for each independent reversing symmetry. Instead we just write the numbers and which symmetry patterns they lie in. The four different reversing symmetries in this case are $G : (x, y, z) \to (z, y, x)$, $\Sigma_1 : (x, y, z) \to (z, -y, -x)$, $\Sigma_2 : (x, y, z) \to (-z, y, -x)$ and $\Sigma_3 : (x, y, z) \to (-z, -y, x)$. Respectively, the reversibility types of these are I-II, 0-II, I-II and 0-I. We find that decomposing $\mathbb{Z}_{103}^3$ under the cube of the trace map yields orbits such that as proportions,

- 0.255 of the phase space is not symmetric under any of the reversing symmetries,
- 0.344 of the phase space is symmetric under only G,
- 0.344 of the phase space is symmetric under only Σ_2,
- 0.0567 of the phase space is symmetric under both G and Σ_2 only,
- 103 points are symmetric under all four reversing symmetries and
- no points are symmetric under any other combination of the reversing symmetries.

The fact that there are two different Type I-II reversibility regimes competing here is a likely cause of abnormality and is certainly a suspect for the unusual statistics of the trace map itself, though nothing pointing to this as a sufficient condition for the statistics has been found.

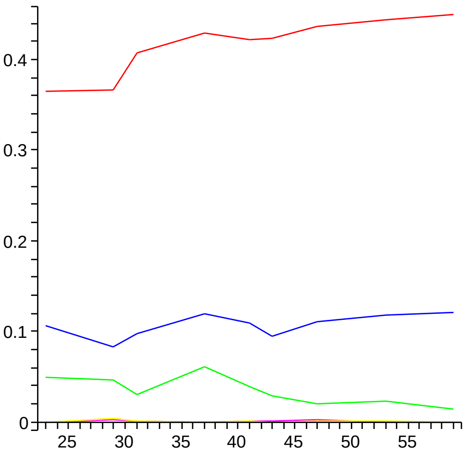

Figure 6.16: Proportion of periodic phase space consumed by orbits that are not symmetric under the reversing symmetry of the map with prime on the x-axis. From top to bottom the maps are QRT3D, Trace and G-M while Pert. G-M and Pert. trace are both close to zero.

Given that such maps exist for which the number of asymmetric cycles is significant the question, assuming that it is these asymmetric cycles that result in larger variations from predicted distributions, becomes whether such maps are common and if the proportion of phase space consumed in asymmetric cycles increases, decreases or remains constant as prime increases. In [61] a trend was noticed that this proportion was inversely proportional to increasing prime. Figure (6.16) shows, as a proportion of the phase space consumed in periodic cycles, the number of points consumed in asymmetric cycles with respect to the reversing symmetry discussed in the individual examples of section 6.4.2. We do not see the inverse relation in this figure, however the size of the prime is rather small due to the extra dimension involved compared to [61]. The vital thing to take from figure (6.16) is that for the ordinary maps, i.e. the maps without this extra symmetry structure as discussed, the proportion of points in asymmetric orbits is extremely small. The existence of two families of reversing symmetries is a rare occurrence, rare enough that checking for distribution conformity remains a useful numerical check for reversibility.

6.2.5 Detecting Reversibility in Four Dimensions

As part of confirming that our extension of the reversibility test to three dimensions does in fact have greater applicability, we turn our attention to some four dimensional reversible maps. To find such examples we look to symplectic maps whose general theory is beyond the scope of this thesis but where it is easy to find four dimensional maps. As with the three dimensional maps, we first present the maps we will be using, followed by the numerical data gathered using them.

Example 6.20. The first four dimensional reversible map we look at is a generalisation of the so-called standard map in symplectic map theory which we take from [46]. The standard map is a two dimensional map but can be generalised to $2n$ dimensions as

$$\begin{aligned} q' &= q + p - \nabla V(q) \\ p' &= p - \nabla V(q), \end{aligned} \tag{6.47}$$

where in physical context $q \in \mathbb{T}^n$, $p \in \mathbb{R}^n$ and $V(q)$ is a periodic potential. However for our purposes, which is just to find a reversible map, we take $V(q)$ to be polynomial and $n = 2$. This forces the map itself to be polynomial which in turn ensures that no cycles will leave the affine part of finite phase spaces. We can also allow p, q to be from different vector spaces. The specific case of map (6.47) we use for our numerical experiments comes by setting $V(q_1, q_2) = q_1^3 + 2q_2^3 - q_1^2 q_2 + 3q_2 + 5$. This has partial derivatives

$$\begin{aligned} \frac{\partial V}{\partial q_1} &= 3q_1^2 - 2q_1 q_2 \\ \frac{\partial V}{\partial q_2} &= 6q_2^2 - q_1^2 + 3. \end{aligned} \tag{6.48}$$

With appropriate guesswork, one can calculate that one way of decomposing the generalised standard map into involutions is $G \circ H$ with

$$\begin{aligned} G : q' &= q - p \\ p' &= -p \\ H : q' &= q \\ p' &= -p + \nabla V(q). \end{aligned} \tag{6.49}$$

Denoting $q = (y, z)$ and $p = (w, x)$, as well as switching the order in which p and q appear in the map, this leaves as our final mapping

$$\begin{aligned} L : w' &= w - 3y^2 + 2yz \\ x' &= x - 6z^2 + y^2 - 3 \\ y' &= y + w' \\ z' &= z + x'. \end{aligned} \tag{6.50}$$

Following the same classification as before, this map is Type II-II since in both G and H of (6.49) the fixed set leaves q free and determines p as a function of q.

The second example we use is a map constructed specifically to be of Type I-III.

Example 6.21. Here we again want to create a polynomial map to ensure that the map acts as a permutation on the affine part of finite phase spaces. Therefore we will take two simple involutions, one which flips the sign of one ordinate and one which flips the sign of three ordinates and sandwich each between an arbitrary polynomial map and its inverse. As earlier, this process will retain the reversibility type but the composition will be an infinite order map. Define

$$\begin{aligned} \phi_1 : w' &= x \\ x' &= -y + 2x^2 \\ y' &= z - \frac{y}{y^2+1} \\ z' &= -w + 2x + yz, \end{aligned} \tag{6.51}$$

$$\begin{aligned} \phi_2 : w' &= -w + x \\ x' &= -z + 2y \\ y' &= -w + x + y \\ z' &= -z + w, \end{aligned} \tag{6.52}$$

$$\begin{aligned} G_1 : w' &= w \\ x' &= x \\ y' &= -y \\ z' &= z \end{aligned} \tag{6.53}$$

and

$$\begin{aligned} H_1 : w' &= w \\ x' &= -x \\ y' &= -y \\ z' &= -z. \end{aligned} \tag{6.54}$$

Then letting $G = \phi_1 \circ G_1 \circ \phi_1^{-1}$ and $H = \phi_2 \circ H_1 \circ \phi_2^{-1}$ gives us, respectively, the type III and I involutions to make up the map $L = G \circ H$ that we use in our numerical experiments. The map L is rational, with a denominator chosen such that we can turn on and off leaking of affine orbits to the projective line by alternating between primes congruent to 1 and 3 modulo 4 (recall that if $p \cong 3 \mod 4$ then $y^2 + 1 = 0$ can have no solution in y).

Generating data with these maps runs into the problem that small four dimensional phase spaces contain a lot of points. In fact, for two dimensional maps, we typically went up to $p = 1000$, which corresponds to $p = 100$ for three dimensional maps and finally $p = 31$ in four dimensions. Most statistics we consider are dependent on single phase spaces; in figures (6.17) and (6.18) we have the usual cumulative frequency distribution for the type II-II map (6.50) and Type I-III map of example (6.21) respectively. Also plotted is the theoretical distribution (6.45) with $a = 1$, since we normalise by the average cycle length. Indeed, the average cycle lengths in the case of $p = 31$ (the case displayed in the figures) are 957 and 61.9. Compared to the expected numbers of $\frac{31^4}{31^2} = 961$ and $2\frac{31^4}{31^3+31} = 61.9$, this suggests that again the number of asymmetric cycles is small. Our final figure, (6.19) shows the theoretical and actual distributions for the Type I-III map of example (6.21) for $p = 29$. For this prime we are forced to exclude a large amount of the phase space for the map to remain a permutation; nevertheless the agreement between the theoretical and actual distributions is remarkable.

6.3 Future Directions

Throughout section (6.2) we have been demonstrating two things. Firstly that the two dimensional test for reversibility given in [61] remains valid in higher dimensions

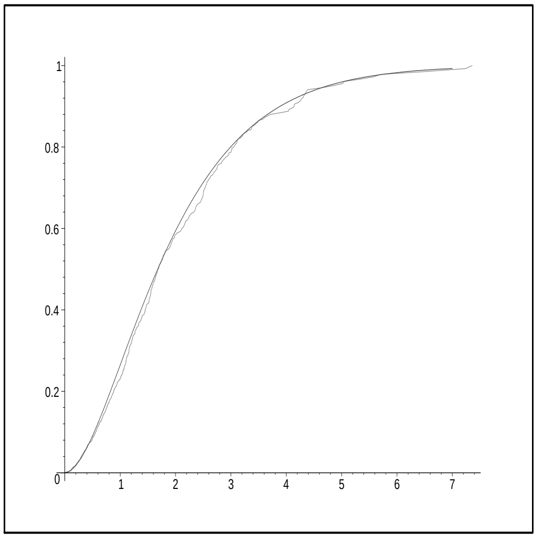

Figure 6.17: Cumulative distribution frequency for the generalised standard map (6.50) with $p = 31$. We use the average cycle length to normalise the orbit lengths. Also shown is the theoretical distribution (6.45) with $a = 1$.

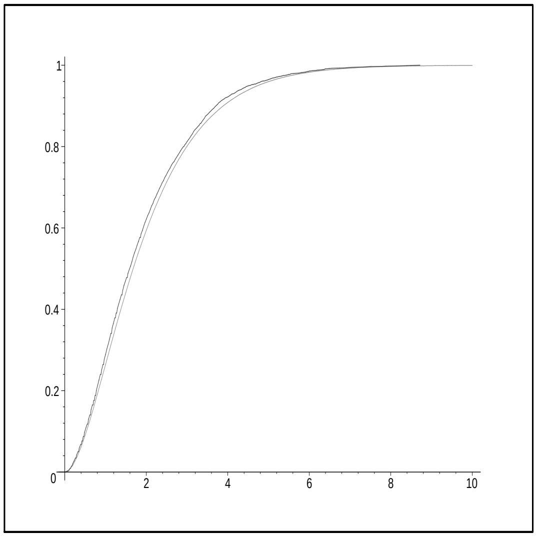

Figure 6.18: Cumulative distribution frequency for the Type I-III map of example (6.21) with $p = 31$. We use the average cycle length to normalise the orbit lengths. Note that for this prime all affine orbits remain in the affine part of the plane. Also shown is the theoretical distribution (6.45) with $a = 1$.

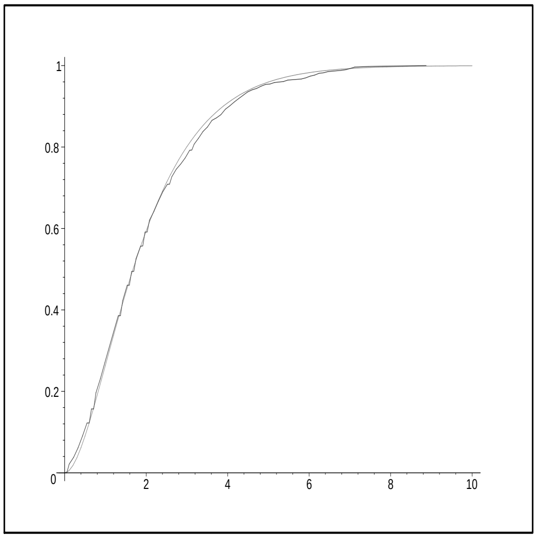

Figure 6.19: Cumulative distribution frequency for the Type I-III map of example (6.21) with $p = 29$. We use the average cycle length (which is around 18.25) to normalise the orbit lengths. Note that for this prime not all affine orbits remain in the affine part of the plane. To work around this, we count in our statistics only those that do remain in the affine plane, hence the short average cycle length. This prime's data was shown to exhibit the universality of the distribution even when counting only part of the phase space, in this case the part that remains in the affine part of the plane. Also shown is the theoretical distribution (6.45) with $a = 1$.

in so far as reversible three (and to a lesser extent four) dimensional maps possess orbit length distributions that are remarkably different to non-reversible maps and secondly that the distribution given in conjecture (6.19) is the appropriate distribution to use as the benchmark for detecting reversibility assuming an absence of a second independent family of reversors. What we have not done is delve into the reason behind a universal distribution existing nor the reason for this particular distribution's form. Thus a highly relevant direction for future study in this area is to answer these two questions. Indeed this is following a similar path taken with the integrability detection method - first the numeric test using the Hasse-Weil bound was developed then the theoretical framework behind it was developed further with the proof of theorem (4.1) and its corollaries. In this case the theoretical grounding for the test is combinatoric in nature, thus this is where we should look for the theory to back up the test for reversibility. There are preliminary reasons to suggest that considering maps as random walks through space phase with the additional conditions necessary for reversibility suffices to give the distribution we find. These conditions, due to theorems of DeVogelaere given in [58], include that symmetric cycles of even length must contain two points from either $\mathrm{Fix}(G)$ or $\mathrm{Fix}(H)$ (where G and H are, as usual, involutions making up a reversible map $L = G \circ H$) while odd symmetric cycles must contain one point from $\mathrm{Fix}(G)$ and one point from $\mathrm{Fix}(H)$. Thus over finite phase spaces this condition is saying that symmetric cycles either loop through both fixed sets, or return to one fixed set. Knowing the sizes of these fixed sets allows us to determine a distribution for the orbit lengths based solely on the premise that the map is a random walk with the caveat that all cycles (recalling that in practice nearly all cycles are symmetric) either return to one fixed set or contain one point from each fixed set. Note that when one fixed set is of a different size this is actually quite strong. Take an example in three dimensions with $\mathrm{Fix}(G)$ having dimension one and $\mathrm{Fix}(H)$ having dimension two. We can see from this that an assumption of uniformity automatically implies that odd symmetric orbits must be relatively rare since over a finite field there are only p possible ways to construct an odd symmetric orbit - one for each member of $\mathrm{Fix}(G)$. Certainly this has been observed in preliminary experiments performed which will be expanded in later work. A second thing to determine is whether the argument is, as it appears at first glance, independent of the map's dimension. The important quantities are the dimension of the involution's fixed sets relative to the dimension of the whole space. Apart from this dependence, the combinatoric nature of the theory behind this work pays no heed to dimension. Examining this in depth will require a plethora of examples from high dimensions which are rarely considered in the literature.

Appendix A

Appendix: MAPLE Code

This appendix contains the MAPLE code (with some documentation) used most in the production of the numerics and examples of the thesis. An electronic version is available by contacting me directly at daneshj@optusnet.com.au

```
#######################################################################
#######################################################################

## Windows and Hasse-Weil things
## forbidden gives how much of [0,1] is forbidden
## HW = upper HW bound, lHW = lower
## biggestN is the biggest n you should bother looking for;
## above this n the allowable windows overlap

HW := p -> evalf(p + 2* sqrt(p) + 1):
lHW := p -> evalf(p - 2*sqrt(p) + 1):
forbidden := p -> 1 - evalf(sum(4*sqrt(p)/HW(p)/n, n=1..floor(HW(p)))):
biggestN := p -> floor( sqrt(p) / 4 + 1/4/sqrt(p) - 1/2 ):
window := (n,p) -> evalf(1/n - 4*sqrt(p) / n / HW(p)):

#######################################################################
#######################################################################
#######################################################################

## projectivize takes two arguments, first being a map as a list
## [x'[1](x[1],x[2],...), x'[2](x[1],x[2],...), ... , x'[n](x[1],...)]
## Note that it is necessary to use x[i] as the ordinates
## the string is either "aut" for autonomous maps or "nonaut" for
## non autonomous maps. Returns projective version of said map with
## H being the projective variable.

projectivize := proc(map,str)
   local map2,l,i,map3;

   if str = "nonaut" then
      map2 := map;
      for i from 1 to nops(map) do
         map2[i] := subs(n=x[nops(map)+1],map[i]);
      od;
      map2 := [ op(map2), x[nops(map)+1] + 1 ];
      return(projectivize(map2,"aut"));
   fi;
   map2 := map;
   for i from 1 to nops(map) do
      map2 := subs(x[i] = X[i]/H, map2);
   od;
   map2 := normal(map2);
   l := lcm(denom(map2[1]),denom(map2[2]));
   for i from 3 to nops(map) do
      l := lcm(l,lcm(denom(map2[1]),denom(map2[i])));
   od;
   for i from 1 to nops(map) do
      map2[i] := map2[i] * l;
   od;
   map2 := [op(map2),l];
   return(map2);
end proc:

######################################################################
######################################################################

## projNorm takes in a projective numerical point [1,2,3,...] and a
## prime and returns the normalized version of that point. Used to
## compare points.

projNorm := proc(point,p)
   local i,g,point2;
   g := gcd(point[1],point[2]);
   for i from 3 to nops(point) do
      g := gcd(g,point[i]);
   od;
   if g = 0 then
      return(point mod p):
   fi:
   if g mod p = 0 then
      return(projNorm(point/g,p)):
   fi:
   point2 := point / g;
   for i from nops(point2) to 1 by -1 do
      if point2[i] mod p <> 0 then
         return(point2 / point2[i]) mod p;
      fi;
   od;
end proc:

## Mostly internal function to find the iterate of a projective point
## under a projective map modulo a prime p

iterate := proc(map, pointun, p)
   local i,genesis, newPoint, point;
   point := projNorm(pointun,p);

   genesis := subs(H=point[nops(point)],map);
   for i from 1 to nops(point)-1 do
      genesis := subs(X[i]=point[i],genesis);
   od;
   newPoint := projNorm(genesis,p);
   return(newPoint);
end proc:
```

```
## Uses iterate to find the forward orbit under a map
## of a point modulo a prime p. Returns two outputs, firstly the orbit
## secondly an indicator 1 is the orbit is periodic (i.e. it wraps
## around) or 0 is the orbit was somehow aperiodic (black hole or
## only partially wraps around).

fwdOrbit := proc(map,point,p)
   local nextPoint, o;

   o := [projNorm(point,p)];
   nextPoint := iterate(map,point,p);
   while not member(nextPoint,o) do
      o := [op(o),nextPoint];
      nextPoint := iterate(map,nextPoint,p);
   od;
   if projNorm(point,p) = nextPoint then
      return([o,1]):
   else
      return([o,0]):
   fi;
end proc:

## Taking in a map and its inverse will generate the forward and
## backwards orbit of a point using fwdOrbit twice, concatenating
## them to make one single orbit.


dblOrbit := proc(map,invmap,point,p)
   local of,ob,o;
   of := fwdOrbit(map,point,p)[1];
   ob := ListTools[Reverse](fwdOrbit(invmap,point,p)[1]);
   o := [op(ob),op(of[2..(nops(of))])];
   return(ListTools[MakeUnique](o));
end proc:

################################################################

## Generates a list with elements [prime,orbit length]
## for a single point modulo many primes and write it to a file
## writeFile should be a string.
## method 1 = Periodic orbits only (singularity in map or no?)
##        2 = Forward orbits regardless of whether it goes to 0
##        3 = Full orbits, from 0 to 0 (or periodic, as case may be)

findLengths := proc(map,invMap,point,minPrime,maxPrime,method,writeFile)
   local i,p,l,o,d,zeroV,isPeriodic,f;
   d := [];
   zeroV := point;
   for i from 1 to nops(point) do
      zeroV[i] := 0;
   od;
   for p from minPrime to maxPrime do
      if isprime(p) and (method=1) then
         o := fwdOrbit(map,point,p):
         if o[2]=1 then
    f := fopen(writeFile,APPEND);
            l := nops(o[1]):
    fprintf(f,`[%a, %a], \n`, p, l);
    fclose(f);
#            d := [op(d), [p,l / NF(p)] ]:
         else
    f := fopen(writeFile,APPEND);
    fprintf(f,`### %a NOT PERIODIC ### \n`, p);
    fclose(f);
 fi;
      elif isprime(p) and (method = 2) then
         o := fwdOrbit(point,map,p):
         if member(zeroV,o[1]) then
            l := nops(o[1]) - 1;
         else
            l := nops(o[1]);
         fi;
 f := fopen(writeFile,APPEND);
 fprintf(f,`[%a, %a], \n`, p, l);
 fclose(f)
#         d := [ op(d), [p, l / NF(p)] ]:
      elif isprime(p) and (method = 3) then
         o := dblOrbit(map,invMap,point,p):
         if member(zeroV,o) then
            l := nops(o) -1;
         else
            l := nops(o);
         fi;
 f := fopen(writeFile,APPEND);
 fprintf(f,`[%a,%a], \n`, p, l);
 fclose(f);
#         d := [op(d), [p,l/ NF(p)]]:
      fi;
   od;
   return(d);
end proc:

## normalizes the lengths in a list returned by the above function

normalizeLengths := proc(ll)
   local n,i,p,ll2;
   ll2 := [];
   n := nops(ll);
   for i from 1 to n do
      p := ll[i][1];
      ll2 := [ op(ll2) ,  [ p, ll[i][2] / NF(p)] ];
   od;
   return(ll2);
end proc:

######################################################################
## These procedures take in the kind of list i plot ie
## [[p,normalisedOrbitLength] ]
######################################################################

## Finds the mean normalized length

meanNL := proc(L)
   local i, s;
   s := 0;
   for i from 1 to nops(L) do
      s := s + L[i][2];
   od;
   return(s/nops(L));
end proc:

## Internal function
```

```
mySort := proc(a,b)
   local bo;
   if a[2] <= b[2] then
      return(true)
   else
      return(false)
   fi;
end proc:

## Finds the median normalized length

medianNL := proc(L)
   local L2;
   L2 := sort(L,mySort);
   return(L2[ceil(nops(L2)/2)][2])
end proc:

## Frequency histogram from data list

freqHisto := proc(L)
   local i, NUMPOINTS, newL;

   NUMPOINTS := nops(L);
   newL := [L[1][2]];
   for i from 2 to NUMPOINTS do
      newL := [op(newL), L[i][2]];
   od;
   stats[statplots,histogram](newL,area=count,numbars=20);
end proc:

## Returns a list that, when plotted, is a cumulative frequency
## distribution of the normalized orbit lengths from a data list

cumuFreq := proc(L)
   local i, NUMPOINTS, newL, plotThis;

   NUMPOINTS := nops(L);
   if type(L[1],float) then
      newL := L:
   else
      newL := [L[1][2]];
      for i from 2 to NUMPOINTS do
         newL := [op(newL), L[i][2]];
      od;
      newL := sort(newL);
   fi:
   plotThis := [ [0,0] ];
   for i from 1 to NUMPOINTS do
      plotThis := [ op(plotThis), [newL[i], (i-1)/NUMPOINTS], [newL[i], i/NUMPOINTS]];
   od;
   plotThis := [op(plotThis), [newL[NUMPOINTS], NUMPOINTS/NUMPOINTS], [newL[NUMPOINTS], 0]];
   return(plotThis);
end proc:


##cumuFreq2 wants a list
## [ [normalizedLength, #times this length occurs] ]

## impCumuFreq2 generates a cumulative frequency distribution from
## a list of (sorted) orbit lengths. Used to do the same thing as
## cumuFreq but for when you decompose an entire phase space for one
## fixed prime.

impCumuFreq2 := proc(L,p)
   local l2, i, l3:
   l2 := []:
   for i from 1 to nops(L) do
      if isInList(L[i],l2) then
         l2 := setCounterPlus(L[i],l2):
      else
         l2 := addNewCounter(L[i],l2):
      fi:
   od:
   l3 := map(weirdDivide,l2,HW(p)):
   return(cumuFreq2(l3)):
end proc:

## lots of internal functions.

isInList := proc(l,L)
   local i:
   for i from 1 to nops(L) do
      if L[i][1] = l then
         return(true):
      fi:
   od:
   return(false):
end proc:

addNewCounter := proc(l,L)
   local l2:
   l2 := [ op(L), [l,1] ]:
   return(l2):
end proc:

setCounterPlus := proc(l,L)
   local l2, i:
   for i from 1 to nops(L) do
      if L[i][1] = l then
         l2 := [ op(L[1..i-1]) , [l,L[i][2]+1], op(L[(i+1)..nops(L)]) ]:
      fi:
   od:
   return(l2):
end proc:

weirdDivide := proc(sl,n)
   local stuff:
   stuff := [ sl[1]/n, sl[2] ]:
   return(stuff):
end proc:

## Constructs a cumulative frequency distribution for lists of the
## form [ [normalized orbit length, # times this length occurs] ].

cumuFreq2 := proc(L)
   local l2, NUMPOINTS, plotThis,i:
   l2 := sort(L,lengthSort):
   NUMPOINTS := sum(l2[i][2]*l2[i][1],i=1..nops(l2)):
   plotThis := [ [0,0], [ l2[1][1], 0] ]:
   plotThis := [ op(plotThis) , [ l2[1][1], l2[1][1]*l2[1][2]/NUMPOINTS] ]:
   for i from 2 to nops(l2) do
```

```
      plotThis := [ op(plotThis), [ l2[i][1], plotThis[nops(plotThis)][2] ] ]:
      plotThis := [ op(plotThis), [ plotThis[nops(plotThis)][1], plotThis[nops(plotThis)][2] + l2[i][1]*l2[i][2]/NUMPOINTS] ]:
   od:
   return(plotThis):
end proc:

## Internal function

lengthSort := proc(l1,l2)
   if l1[1] <= l2[1] then
      return(true):
   else
      return(false):
   fi:
end proc:

## Converts a list with just the orbits of a phase space under a map
## into a list of the form [ [length, # times length occurs] ]

orbitsToLengthsList := proc(o)
   local newList, lengths,i, newerList:
   lengths := sort(map(nops,o)):
   newList := Array(1..lengths[nops(lengths)]):
   for i from 1 to nops(lengths) do
      newList[lengths[i]] := newList[lengths[i]]+1:
   od:
   newerList := []:
   for i from 1 to lengths[nops(lengths)] do
      if newList[i] > 0 then
         newerList := [ op(newerList), [i, newList[i]] ]:
      fi:
   od:
   return(newerList);
end proc:

## Use normLengthsList to normalize the lengths outputted by the
## function orbitsToLengthsList

normLengthsList := proc(L,p)
   return(map(myDivide,L,p)):
end proc:

## Internal function

myDivide := proc(l,p)
   return( [ l[1]/HW(p), l[2] ] ):
end proc:

######################################################################
## Take in a table that stores all the level sets at the appropriate## indices of the table
######################################################################

normalizeCycleLengths := proc(T, n)
   local L2, div:
   L2 := ListTools[Flatten]([entries(T)]):
   div := a -> a/n:
   L2 := map(div,L2):
   return(sort(L2)):
end proc:

normalizeHeightLengths := proc(T,n)
```

```
   local L2, div:
   L2 := ListTools[Flatten](map(mySum,ListTools[Flatten]([entries(T)],1))):
   div := a -> a/n:
   L2 := map(div,L2):
   return(sort(L2)):
end proc:

mySum := proc(L)
   return(sum(L[i],i=1..nops(L))):
end proc:

lengthsToFreq := proc(L)
   local i, plotList;
   plotList := [];
   for i from 1 to nops(L) do
      plotList := [ op(plotList), [ L[i], cI(L, L[i]) ] ]:
   od:
   return(plotList):
end proc:

cI := proc(L,n)
   local i, count;
   count := 0:
   for i from 1 to nops(L) do
      if L[i] = n then
         count := count+1:
      fi;
   od:
   return(count):
end proc:

###### End stats procedures #############################


######################################################################
###### For studying single phase space cycle decompositions ##########
######################################################################

#genSpace := proc(n,p) ## n = dimension, p = prime
#   local i,base, j, newBase;
#   if n = 0 then
#      return([[]]);
#   else
#      base := genSpace(n-1,p);
#      newBase := [];
#      for i from 1 to nops(base) do
#         for j from 0 to p-1 do
#            newBase := [ op(newBase), [ op(base[i]), j] ];
#         od;
#      od;
#   fi;
#   return(newBase);
#end proc:

#genProjSpace := proc(n,p)
#   local base;
#   base := map(projNorm,genSpace(n,p),p):
#   base := ListTools[MakeUnique](base);
#   base := sort(base,planeSort);
#   return(base[2..nops(base)]);
#end proc;
```

```
## Takes in a map and prime p and tests random points
## from the finite projective space to see if that
## point’s orbit does not lie within one of the first three
## allowable windows

monteCarloTester := proc(m,p)
   local d, pm,i,point,o,currentOrbit,space:
   d := nops(m):
   space := genProjSpace(d,p):
   pm := projectivize(m,"aut"):
   i := 0:
   while nops(space) > 0 do
      i := i+1:
      point := randomPoint(space):
      o := fwdOrbit(pm,point,p):
      if o[2] = 1 then
 currentOrbit := o[1]:
 if isBad(nops(currentOrbit),3,p) then
    return(i):
 else
    continue:
 fi:
      else
         continue:
      fi:
      space := convert( ‘minus‘(convert(space,set),convert(currentOrbit,set)) , list):
   od:
   return(["Not integrable", i]):
end proc:

randomPoint := proc(space)
   local i, point,l:
   randomize():
   i := rand(nops(space))():
   point := space[i]:
   return(point):
end proc:

genProjSpace := proc(n,p)
   local space,i,j,k;
   if n=1 then
      space := [ seq([i,1],i=0..p-1),[1,0]]:
   elif n=2 then
      space := [ seq( seq( [i,j,1], i=0..p-1), j=0..p-1) , seq( [i,1,0], i=0..p-1), [1,0,0] ]:
   elif n=3 then
      space := [ seq( seq( seq( [i,j,k,1], i=0..p-1), j=0..p-1), k=0..p-1), seq( seq( [i,j,1,0], i=0..p-1), j=0..p-1), seq( [i,1,0,0], i=0..p-1), [1,0,0,0]
   else
      space := []:
   fi:
   return(space):
end proc:

genAffineSpace := proc(n,p)
   local space:
   if n = 1 then
      space := [seq([i,1], i=0..p-1)]:
   elif n=2 then
      space := [seq(seq([i,j,1],i=0..p-1),j=0..p-1)]:
   elif n=3 then
      space := [seq(seq(seq([i,j,k,1],i=0..p-1),j=0..p-1),k=0..p-1)]:
```

```
   elif n=4 then
      space := [seq(seq(seq(seq([h,i,j,k,1],h=0..p-1),i=0..p-1),j=0..p-1),k=0..p-1)]:
   else
      space := []:
   fi:
   return(space):
end proc:

planeSort  := proc(a,b)
   local i;
   for i from 1 to nops(a) do
      if a[i] < b[i] then
         return(true);
      elif a[i] > b[i] then
         return(false);
      fi;
   od;
   return(true);
end proc:

## Finds all the orbits (periodic and non) arising from an
## input set of initial conditions (inspace)

allOrbits := proc(map,invmap,p,inspace)
   local space,currentOrbit,orbits,pmap,pinvmap;
   orbits := [];
   space := inspace;
   pmap := map;
   pinvmap := invmap;
   while nops(space) > 0 do
      currentOrbit := dblOrbit(pmap,pinvmap,space[1],p);
      orbits := [op(orbits), currentOrbit];
      space := convert( ‘minus‘(convert(space,set),convert(currentOrbit,set)) , list);
   od;
   return(orbits);
end proc:

## Same as above, but excludes all nonperiodic orbits

periodicOrbits := proc(pmap,invmap,p,inspace)
   local space,currentOrbit,orbits,o;
   orbits := [];
   space := inspace;
   while nops(space) > 0 do
      o := fwdOrbit(pmap,space[1],p);
      currentOrbit := o[1];
      if o[2] = 1 then
         orbits := [ op(orbits), currentOrbit ];
      fi;
      space := convert( ‘minus‘(convert(space,set),convert(currentOrbit,set)) , list);
   od;
   return(orbits);
end proc:

## Same as above, but shows only aperiodic orbits

aperiodicOrbits := proc(pmap,invmap,p,inspace)
   local space,currentOrbit,orbits,o;
   orbits := [];
   space := inspace;
   while nops(space) > 0 do
```

```
      o := fwdOrbit(pmap,space[1],p);
      if o[2] <> 1 then
         currentOrbit := dblOrbit(pmap, invmap,space[1],p):
         orbits := [ op(orbits), currentOrbit ];
      else
         currentOrbit := o[1]:
      fi:
      space := convert( ‘minus‘(convert(space,set),convert(currentOrbit,set)) , list);
   od;
   return(orbits);
end proc:

## Not very tested, but ostensibly takes in a list of integrals
## with variables X[1],X[2],...,H and a list of orbits and returns
## an array where the i’th entry is the indices of all the orbits
## It works, but lots can go wrong with it too.

sortByHeights := proc(integrals, orbits, p)
   local i,j, cOrbit, testPoint, intValues, indexArray, oldEntry;

   if nops(integrals) = 1 then
      indexArray := Array(1..p+2,fill=[]);
   elif nops(integrals) = 2 then
      indexArray := Matrix(1..p+2,1..p+2);
      for i from 1 to p+2 do
         for j from 1 to p+2 do
    indexArray[i,j] := [];
 od;
      od;
   elif nops(integrals) = 3 then
      indexArray := Array(1..p+2,1..p+2,1..p+2,fill=[]);
   else
      return("Sorry, too many integrals to be supported by this version");
   fi;

   for i from 1 to nops(orbits) do
      cOrbit := orbits[i];
      testPoint := cOrbit[ceil(nops(cOrbit)/2)];
      intValues := [ seq( evalOnIntegral(testPoint,integrals[j],p) , j=1..nops(integrals)) ];
      oldEntry := indexArray[op(intValues)];
      indexArray[op(intValues)] := [ op(oldEntry), i ];
   od;
   return(indexArray);
end proc;

## Internal, but returns the height of a point on an integral

evalOnIntegral := proc(testPoint, integral, p)
   local i, projCurve, currentDenom, currentNumer,someNum;

   projCurve := projectivizeCurve(integral,nops(testPoint)-1);
   currentDenom := denom(projCurve);
   currentNumer := numer(projCurve);
   for i from 1 to nops(testPoint)-1 do
      currentDenom := subs(X[i] = testPoint[i], currentDenom) mod p;
      currentNumer := subs(X[i] = testPoint[i], currentNumer) mod p;
   od;
   currentDenom := subs(H=testPoint[nops(testPoint)], currentDenom) mod p;
   currentNumer := subs(H=testPoint[nops(testPoint)], currentNumer) mod p;
   if currentDenom mod p = 0 and currentNumer mod p = 0 then
      return(p+2);
   elif currentDenom mod p = 0 then
      return(p+1)
   elif currentNumer mod p = 0 then
      return(p);
   else
      someNum := modp(simplify(modp(currentNumer/currentDenom,p)),p);
      if someNum = 0 then
         return(p);
      else
         return(someNum mod p):
      fi;
   fi;
end proc;

## Used to get curves into the form desired by the above functions

projectivizeCurve := proc(curve,dim)
   local i, projCurve;
   projCurve := curve;
   for i from 1 to dim do
      projCurve := subs(x[i] = X[i]/H, projCurve);
   od;
   return(simplify(projCurve));
end proc;

## Takes the indices and orbits from sortByHeights and ends up
## returning what lengths are at each height. Was good for
## checking equidistribution conjecture

indexToLengths := proc(indexArray,orbits,integrals)
   local i,j,n,lengthArray,iList,newList;
   n  := nops(integrals);
   lengthArray := [];
   if n = 1 then
      for i from 1 to ArrayNumElems(indexArray) do
         iList := indexArray[i];
 newList := [ seq( nops(orbits[iList[j]]) , j=1..nops(iList)) ];
         lengthArray := [ op(lengthArray), newList ];
      od;
   elif n = 2 then
   lengthArray := Matrix(op(bh)[1],op(bh)[2],[]):
   for i from 1 to op(indexArray)[1] do
      for j from 1 to op(indexArray)[2] do
         iList := indexArray[i,j]:
 lengthArray[i,j] := [ seq(nops(orbits[iList[k]]),k=1..nops(iList)) ]:
      od:
   od:
   elif n=3 then
   fi;
   return(lengthArray);
end proc:

## Makes results of the last function readable by humans

lengthsToPrintable := proc(lengthsArray, integrals,writeFile)
   local i,j,f,istr,jstr, MAXINDEX;
   f := fopen(writeFile,APPEND);
   fprintf(f,`##Comments here, map name, prime, etc \n`):
   fprintf(f,`T := table([ \n`):
   if nops(integrals) = 1 then
      for i from 1 to nops(lengthsArray) do
```

```
         if i = nops(lengthsArray) then
    fprintf(f,`%a \n ]):`,lengthsArray[i]);
 elif i = nops(lengthsArray) - 1 then
    fprintf(f,`%a , \n`, lengthsArray[i]);
 else
    if nops(lengthsArray[i]) > 0 then
               fprintf(f,`%a, \n`, lengthsArray[i]);
            else
       fprintf(f,` [], \n`):
    fi;
 fi;
      od;
   elif nops(integrals) = 2 then
      MAXINDEX := op(1,lengthsArray)[1]:
      for i from MAXINDEX to 1 by -1 do
         for j from MAXINDEX to 1 by -1 do
            if i = MAXINDEX then
       istr := NaN;
    elif i = MAXINDEX - 1 then
       istr := Inf;
    else
       istr = i;
    fi;
    if j = MAXINDEX then
       jstr := NaN;
    elif j = MAXINDEX - 1 then
       jstr := Inf;
    else
       jstr := j;
    fi;
    if nops(lengthsArray[i,j]) > 0 then
       fprintf(f,` %a , %a | %a \n`, istr, jstr, lengthsArray[i,j]);
    fi;
 od;
      od;
   else
      print("Babow");
   fi;
   fprintf(f,`##Modify this file`):
   fclose(f);
   return(NULL);
end proc:

## Finds where there were problems with equidistribution.

returnNonEqui := proc(lengthsArray, integrals, writeFile)
   local i,j,f,heightOrbitLength;
   f := fopen(writeFile, APPEND);
   if nops(integrals) = 1 then
      for i from nops(lengthsArray) to 1 by -1 do
         if nops(lengthsArray[i]) > 0 then
            heightOrbitLength := lengthsArray[i][1];
    for j from 2 to nops(lengthsArray[i]) do
       if heightOrbitLength <> lengthsArray[i][j] then
          fprintf(f,` Problem at %a \n`, i);
                  break;
       fi;
    od;
 fi;
      od;
   fi;
   fclose(f):
   return(NULL);
end proc:

checkNormedLengths := proc(l,p)
   local i, c, MAX:
   MAX := floor((sqrt(p)-2)/4):
   c := 0:
   for i in l do
      if isBad(i,MAX,p) then
         c := c+1:
      fi:
   od:
   return(c / nops(l)):
end proc:

## Internal function.

isBad := proc(x,max,p)
   local i:
   if x > 1 then
      return(true)
   fi:
   for i from 1 to max do
      if x > evalf(1/(i+1)) and x < evalf((1/i - 4 * sqrt(p)/ i / HW(p))) then
         return(true):
      fi:
   od:
   return(false):
end proc:

## Turns a list of lengths into a list of
## [ [period,number times period occurred] ]

lengthsToPerFreq := proc(L)
   local A,N,uL,i,j:
   uL := sort(ListTools[MakeUnique](L)):
   N := nops(uL):
   A := array(1..N):
   for i from 1 to N do
      A[i] := [uL[i],0]:
   od:
   for i from 1 to nops(L) do
      j := getIndex(L[i],A):
      A[j][2] := A[j][2]+1:
   od:
   return(A):
end proc:

getIndex := proc(l,A)
   local i:
   for i from 1 to nops([entries(A)]) do
      if A[i][1] = l then
         return(i):
      fi:
   od:
   return(0)
end proc:

## these two take in period frequency lists
## [ [period,number times period occurred] ]
```

```
totalPoints := proc(L)
   local s,i:
   s := 0:
   for i from 1 to nops(L) do
      s := s + L[i][1]*L[i][2]:
   od:
   return(s):
end proc:

perFreqsToCumuFreq := proc(L,NF)
   local plotList,tp,i,dataX,dataY:
   tp := totalPoints(L):
   plotList := [ [0,0] ]:
   for i from 2 to nops(L)+1 do
      dataX := evalf(L[i-1][1]/NF):
      dataY := evalf(L[i-1][1]*L[i-1][2]/tp) + plotList[i-1][2]:
      plotList := [ op(plotList), [dataX,dataY] ]:
   od:
   return(plotList):
end proc:

numCycles := proc(perfreqs)
   local num,i:
   num := 0:
   for i from 1 to nops(perfreqs) do
      num := num+perfreqs[i][2]:
   od:
   return(num):
end proc:


########################################################################
## Singularity Confinement Procedures
## (A little too old and undocumented to be propery commented)
########################################################################

paramIterate := proc(m,point, str)
   local i,newPoint,g;
   newPoint := subs(H=point[nops(point)], m);
   for i from 1 to nops(m) - 1 do
      newPoint := subs(X[i] = point[i], newPoint);
   od;
   if str = "norm" then
      return(removeGCD(newPoint)[1]);
   elif str = "nonorm" then
      return(newPoint);
   elif str = "gcd" then
      return( [newPoint, removeGCD(newPoint)[2]] );
   fi;
end proc:

removeGCD := proc(point)
   local i,g,newPoint;
   newPoint := point;
   g := gcd(point[1], point[2]);
   for i from 3 to nops(point) do
      g := gcd(g, newPoint[i]);
   od;
   for i from 1 to nops(point) do
      newPoint[i] := simplify(point[i] / g);
   od;
   return([newPoint,g]);
end proc:

## map, initial condition, how many iterates to take, what coordinate to store, norm/nonorm/gcd

getOrdinate := proc(m,ic,maxlength,i,str)
   local j, polys, currentPoint;
   currentPoint := ic;
   polys := [currentPoint[i]];
   for j from 1 to maxlength do
      currentPoint := paramIterate(m, currentPoint,str);
      polys := [ op(polys), currentPoint[i] ];
   od;
   return(polys);
end proc:

## use map(degree, polys, {var1,var2,...}); to turn polys into degrees

timeToDeath := proc(m,ic,maxlength,p)
   local i, currentPoint;
   currentPoint := projNorm(ic,p);
   for i from 0 to maxlength do
      if currentPoint[nops(currentPoint)] = 0 then
            return(i);
      fi;
      currentPoint := iterate(m,currentPoint,p);
   od;
   return(infinity);
end proc:

lineDeath := proc(m,ic,maxlength,p,writeFile)
   local init, currentTime, allTimes, i, f;
   allTimes := [];
   for i from 0 to p-1 do
      init := subs(epsilon=i,ic);
      currentTime := timeToDeath(m,init,maxlength,p);
      if currentTime <> infinity then
        f := fopen(writeFile,APPEND);
 fprintf(f,`[%a, %a], \n`, i, currentTime);
 fclose(f)
#         allTimes := [ op(allTimes), [i, currentTime] ];
      fi;
   od;
   return(allTimes);
end proc:

plotDeaths := proc(L,maxLength)
   local i,j, d, counter;
   d := [];
   for i from 0 to maxLength do
      counter := 0;
      for j from 1 to nops(L) do
         if L[j][2] = i then
    counter := counter + 1;
         fi;
      od;
      d := [ op(d), [i, counter] ];
   od;
   return(d,labels=["Time until infinity","Number of points"]);
   plot(d,labels=["Time until infinity","Number of points"]);
```

```
end proc:

pointsExploding := proc(L,p)
   local i, NUMPOINTS, newL, plotThis;

   NUMPOINTS := nops(L);
   newL := [L[1][2]];
   for i from 2 to NUMPOINTS do
      newL := [op(newL), L[i][2]];
   od;
   newL := sort(newL);
   plotThis := [ [0,0] ];
   for i from 1 to NUMPOINTS do
      plotThis := [ op(plotThis), [newL[i], (i-1)/p], [newL[i], i/p]];
   od;
   plotThis := [op(plotThis), [newL[NUMPOINTS], NUMPOINTS/p], [newL[NUMPOINTS], 0]];
   return(plotThis);
end proc:

## returns a list of the form [ [i, n(i) ] where i=step #, n(i) = difference in roots of
## the projective ordinate not normed and normed / normed.

rootsOfZ := proc(map,ic,maxlength,p)
   local i,polysN,polysNN, r;
   r := [];
   polysN := getOrdinate(map,ic,maxlength,3,"norm"):
   polysNN := getOrdinate(map,ic,maxlength,3,"nonorm"):
   for i from 1 to nops(polysN) do
      r := [ op(r), [i, (nops([msolve(polysNN[i],p)]) - nops([msolve(polysN[i],p)])) / (nops([msolve(polysN[i],p)])+1)]]:
   od;
   return(r):
end proc;

######################################################################
## New set of procedures for testing symmetry etc
######################################################################

## Finds the orbits from a given set of initial conditions
## that are symmetric under the given symmetry (which should be
## a projective map)

symmetricOrbits := proc(pmap,invmap,p,inspace,sym)
   local space,currentOrbit,orbits,o;
   orbits := [];
   space := inspace;
   while nops(space) > 0 do
      o := symFwdOrbit(pmap,space[1],p,sym);
      currentOrbit := [o[1],o[3]];
      if o[2] = 1 then
         orbits := [ op(orbits), currentOrbit ];
      fi;
      space := convert( ‘minus‘(convert(space,set),convert(currentOrbit[1],set)) , list);
   od;
   return(orbits);
end proc:

## Finds the forwards orbit of a point and indicates if it is
## (non)periodic and (a)symmetric

symFwdOrbit := proc(map,point,p,sym)
   local nextPoint, o, isSym;
   isSym := 0:
   o := [projNorm(point,p)];
   nextPoint := iterate(map,point,p);
   if iterate(sym,point,p) = projNorm(point,p) then
      isSym := 1:
   fi:
   while not member(nextPoint,o) do
      o := [op(o),nextPoint];
      if isSym = 0 then
         if iterate(sym,point,p) = nextPoint then
            isSym := 1:
         fi:
      fi:
      nextPoint := iterate(map,nextPoint,p);
   od;
   if projNorm(point,p) = nextPoint then
      return([o,1,isSym]):
   else
      return([o,0,isSym]):
   fi;
end proc:

## Turns things from symmetricOrbits into lengths

symOrbitsToLengths := proc(L)
   local stuff:
   return(map(myNops,L) ):
end proc:

myNops := proc(L)
   return( [nops(L[1]),L[2]] ):
end proc:

## Turns the lengths from symOrbitsToLengths into a period
## frequency list

symLengthsToPerFreq := proc(L)
   local A,N,uL,i,j:
   uL := sort(ListTools[MakeUnique](L,1,symEquality),symSort):
   N := nops(uL):
   A := array(1..N):
   for i from 1 to N do
      A[i] := [uL[i][1],0,0]:
   od:
   for i from 1 to nops(L) do
      j := symGetIndex(L[i],A):
      if L[i][2] = 1 then
         A[j][2] := A[j][2]+1:
      else
         A[j][3] := A[j][3]+1:
      fi:
   od:
   return(A):
end proc:

symEquality := proc(L1,L2)
   if L1[1]=L2[1] then
      return(true):
   else
      return(false):
   fi:
```

```
end proc:

symSort := proc(L1,L2)
   if L1[1] <= L2[1] then
      return(true):
   else
      return(false):
   fi:
end proc:

symGetIndex := proc(l,A)
   local i:
   for i from 1 to nops([entries(A)]) do
      if A[i][1] = l[1] then
         return(i):
      fi:
   od:
   return(0)
end proc:

## Returns proportions of points consumed in symmetric orbits to total
## points and points consumed in asym orbits to total points

symmetricProportions := proc(L)
   local tP,sP,aP:
   tP := symTotalPoints(L):
   sP := symSymPoints(L):
   aP := symAsymPoints(L):
   return([sP/tP,aP/tP]):
end proc:

symTotalPoints := proc(L)
   local currentS,i:
   currentS := 0:
   for i from 1 to nops(L) do
      currentS := currentS + L[i][1]*(L[i][2]+L[i][3]):
   od:
   return(currentS):
end proc:

symSymPoints := proc(L)
   local currentS,i:
   currentS := 0:
   for i from 1 to nops(L) do
      currentS := currentS + L[i][1]*L[i][2]:
   od:
   return(currentS):
end proc:

symAsymPoints := proc(L)
   local currentS,i:
   currentS := 0:
   for i from 1 to nops(L) do
      currentS := currentS + L[i][1]*L[i][3]:
   od:
   return(currentS):
end proc:
```

Bibliography

[1] M.J. Ablowitz, R.G. Halburd, and B. Herbst. On the extension of the Painlevé property to difference equations. *Nonlinearity*, 13(3):889–905, 2000.

[2] V. I. Arnol′d, V. V. Kozlov, and A. I. Neĭshtadt. *Dynamical systems. III*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1988. Translated from the Russian by A. Iacob.

[3] V.I. Arnol′d. Dynamics of complexity of intersections. *Bol. Soc. Brasil. Mat. (N.S.)*, 21(1):1–10, 1990.

[4] V.I. Arnol′d and M.B. Sevrjuk. Oscillations and bifurcations in reversible systems. In R.Z. Sagdeev, editor, *Nonlinear Phenomena in plasma physics and Hydrodynamics*, pages 31–64. Mir Publishers, 1986.

[5] M. Baake and J.A.G. Roberts. Reversing symmetry group of $\mathrm{Gl}(2, \mathbf{Z})$ and $\mathrm{PGl}(2, \mathbf{Z})$ matrices with connections to cat maps and trace maps. *J. Phys. A*, 30(5):1549–1573, 1997.

[6] M. Baake and J.A.G. Roberts. Symmetries and reversing symmetries of toral automorphisms. *Nonlinearity*, 14(4):R1–R24, 2001.

[7] M. Baake and J.A.G. Roberts. Symmetries and reversing symmetries of polynomial automorphisms of the plane. *Nonlinearity*, 18(2):791–816, 2005.

[8] M. Baake and J.A.G. Roberts. The structure of reversing symmetry groups. *Bull. Austral. Math. Soc.*, 73(3):445–459, 2006.

[9] M.P. Bellon. Algebraic entropy of birational maps with invariant curves. *Lett. Math. Phys.*, 50(1):79–90, 1999.

[10] M.P. Bellon and C.-M. Viallet. Algebraic entropy. *Communications in Mathematical Physics*, 204:425–437, 1999.

[11] F. Beukers and R. Cushman. Zeeman's monotonicity conjecture. *J. Differential Equations*, 143(1):191–200, 1998.

[12] H.W. Capel and R. Sahadevan. A new family of four-dimensional symplectic and integrable mappings. *Phys. A*, 289(1-2):86–106, 2001.

[13] D.A. Cox. Mordell-Weil groups of elliptic curves over $\mathbf{C}(t)$ with $p_g = 0$ or 1. *Duke Math. J.*, 49(3):677–689, 1982.

[14] J.J. Duistermaat. *The QRT map as an automorphism of a rational elliptic surface.* Preprint. Utrecht University, February 2007.

[15] A. Dujella. High rank elliptic curves with prescribed torsion; http://web.math.hr/ duje/tors/tors.html, Last accessed: 23/2/2006 Last modified: 2006.

[16] J. Esch and T.D. Rogers. The screensaver map: dynamics on elliptic curves arising from polygonal folding. *Discrete Comput. Geom.*, 25(3):477–502, 2001.

[17] J. Franks. Rotation numbers and instability sets. *Bull. Amer. Math. Soc. (N.S.)*, 40(3):263–279 (electronic), 2003.

[18] A. Gómez and J.D. Meiss. Volume-preserving maps with an invariant. *Chaos*, 12(2):289–299, 2002.

[19] G.R. Goodson. Inverse conjugacies and reversing symmetry groups. *Amer. Math. Monthly*, 106(1):19–26, 1999.

[20] B. Grammaticos, A. Ramani, and V. Papageorgiou. Do integrable mappings have the Painlevé property? *Physical Review Letters*, 67(14):1825–1828, September 1991.

[21] R.G. Halburd. Diophantine integrability. *J. Phys. A*, 38(16):L263–L269, 2005.

[22] R. Hartshorne. *Algebraic Geometry.* Springer-Verlag New York, 1977.

[23] J. Hietarinta and C.-M. Viallet. Singularity confinement and degree growth. In *SIDE III—symmetries and integrability of difference equations (Sabaudia, 1998)*, volume 25 of *CRM Proc. Lecture Notes*, pages 209–216. Amer. Math. Soc., Providence, RI, 2000.

[24] R. Hirota, K. Kimura, and H. Yahagi. How to find the conserved quantities of nonlinear discrete equations. *J. Phys. A : Math. Gen.*, 34:10377–10386, 2001.

[25] A. Iatrou. Real jacobian elliptic function parametrizations for a genuinely asymmetric biquadratic curve. http://arxiv.org/pdf/nlin.SI/0306051 (PDF), June 2003.

[26] A. Iatrou. Three dimensional integrable mappings . http://arxiv.org/pdf/nlin.SI/0306052 (PDF), June 2003.

[27] A. Iatrou and J.A.G. Roberts. Integrable mappings of the plane preserving biquadratic invariant curves. *J. Phys. A*, 34(34):6617–6636, 2001.

[28] A. Iatrou and J.A.G. Roberts. Integrable mappings of the plane preserving biquadratic invariant curves ii. *Nonlinearity*, 15(2):459–489, 2002.

[29] A. Iatrou and J.A.G. Roberts. Integrable mappings of the plane preserving biquadratic invariant curves iii. *Physica A*, 326:400–411, 2003.

[30] D. Jogia, J.A.G. Roberts, and F. Vivaldi. An algebraic geometric approach to integrable maps of the plane. *J. Phys. A.*, 39(5):1133–1149, 2006.

[31] N. Joshi, B. Grammaticos, T. Tamizhmani, and Ramani A. From integrable lattices to non-qrt mappings. *Letters in Mathematical Physics*, 78:27–37, 2006.

[32] K. Kimura, H. Yahagi, R. Hirota, A. Ramani, B. Grammaticos, and Y. Ohta. A new class of integrable discrete systems. *J. Phys. A*, 35(43):9205–9212, 2002.

[33] V. L. Kocić and G. Ladas. *Global behavior of nonlinear difference equations of higher order with applications*, volume 256 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.

[34] M. R. S. Kulenović and G. Ladas. *Dynamics of second order rational difference equations*. Chapman & Hall/CRC, Boca Raton, FL, 2002. With open problems and conjectures.

[35] J. S. W. Lamb. Reversing symmetries in dynamical systems. *J. Phys. A*, 25(4):925–937, 1992.

[36] J.S.W. Lamb and G.R.W. Quispel. Cyclic reversing k-symmetry groups. *Nonlinearity*, 8(6):1005–1026, 1995.

[37] J.S.W. Lamb and J.A.G. Roberts. Time-reversal symmetry in dynamical systems: a survey. *Phys. D*, 112(1-2):1–39, 1998. Time-reversal symmetry in dynamical systems (Coventry, 1996).

[38] S. Lang. *Algebra*. Graduate Texts in Mathematics. Springer-Verlag New York, revised third edition edition, 2002.

[39] S. Lang and Weil A. Number of points of varieties in finite fields. *American Journal of Mathematics*, 76:819–827, 1954.

[40] H. Blaine Lawson, Jr. Foliations. *Bull. Amer. Math. Soc.*, 80:369–418, 1974.

[41] R.C. Lyness. Note 1847. *Math. Gaz.*, 29:231–233, 1945.

[42] R. S. MacKay. *Renormalisation in area-preserving maps*, volume 6 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co. Inc., River Edge, NJ, 1993.

[43] R. S. MacKay and J. D. Meiss, editors. *Hamiltonian dynamical systems*. Adam Hilger Ltd., Bristol, 1987.

[44] B. Mazur. Rational points on modular curves. In *Modular functions of one variable, V (Proc. Second Internat. Conf., Univ. Bonn, 1976)*, pages 107–148. Lecture Notes in Math., Vol. 601. Springer, Berlin, 1977.

[45] E.M. McMillan. A problem in the stability of periodic systems. In E. Britton and H. Odabasi, editors, *Topics in Modern Physics. A tribute to E. U. Condon*, pages 219–244. Colorado University Press, Boulder, 1971.

[46] J.D. Meiss. Symplectic maps; http://amath.colorado.edu/pub/dynamics/papers/sympmaps.pdf, Last accessed: 19/12/2007 Last modified: July 2002.

[47] K. Oguiso and T. Shioda. The Mordell-Weil lattice of a rational elliptic surface. *Comment. Math. Univ. St. Paul.*, 40(1):83–99, 1991.

[48] G.R.W. Quispel. An alternating integrable map whose square is the QRT map. *Phys. Lett. A*, 307(1):50–54, 2003.

[49] G.R.W. Quispel and J.A.G. Roberts. Reversible mappings of the plane. *Phys. Lett. A*, 132(4):161–163, 1988.

[50] G.R.W. Quispel, J.A.G. Roberts, and C.J. Thompson. Integrable mappings and soliton equations. *Phys. Lett. A*, 126(7):419–421, 1988.

[51] G.R.W. Quispel, J.A.G. Roberts, and C.J. Thompson. Integrable mappings and soliton equations. II. *Phys. D*, 34(1-2):183–192, 1989.

[52] K.V. Rerikh. Algebraic-geometry approach to integrability of birational plane mappings. Integrable birational quadratic reversible mappings. I. *J. Geom. Phys.*, 24(3):265–290, 1998.

[53] J.A.G. Roberts. *Order and chaos in reversible dynamical systems.* PhD thesis, University of Melbourne, 1990.

[54] J.A.G. Roberts. Escaping orbits in trace maps. *Phys. A*, 228(1-4):295–325, 1996.

[55] J.A.G. Roberts and M. Baake. Trace maps as 3D reversible dynamical systems with an invariant. *J. Statist. Phys.*, 74(3-4):829–888, 1994.

[56] J.A.G. Roberts and M. Baake. Symmetries and reversing symmetries of area-preserving polynomial mappings in generalised standard form. *Phys. A*, 317(1-2):95–112, 2003.

[57] J.A.G. Roberts, D. Jogia, and F. Vivaldi. The Hasse-Weil bound and integrability detection in rational maps. *J. Nonlinear Math. Phys.*, 10(suppl. 2):166–180, 2003.

[58] J.A.G. Roberts and J.S.W. Lamb. Self-similarity of period-doubling branching in 3-D reversible mappings. *Phys. D*, 82(4):317–332, 1995.

[59] J.A.G. Roberts and G.R.W. Quispel. Creating and relating three-dimensional integrable maps. *J. Phys. A*, 39(42):L605–L615, 2006.

[60] J.A.G. Roberts and F. Vivaldi. Arithmetical method to detect integrability in maps. *Phys. Rev. Lett.*, 90(3):034102, 4, 2003.

[61] J.A.G. Roberts and F. Vivaldi. Signature of time-reversal symmetry in polynomial automorphisms over finite fields. *Nonlinearity*, 18:2171–2192, 2005.

[62] L. A. Shepp and S. P. Lloyd. Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.*, 121:340–357, 1966.

[63] J. Silverman. *The Arithmetic of Elliptic Curves.* Graduate Texts in Mathematics. Springer-Verlag New York, 1986.

[64] J. Silverman and J. Tate. *Rational Points on Elliptic Curves.* Springer-Verlag New York, 1992.

[65] T. Tsuda. Integrable mappings via rational elliptic surfaces. *J. Phys. A*, 37(7):2721–2730, 2004.

[66] M. van Hoeij. An algorithm for computing the weierstrass normal form. *ISSAC '95 Proceedings*, pages 90–95, 1995.

[67] A.P. Veselov. Integrable maps. *Russian Math. Surveys*, 46:1–51, 1991.

[68] C.-M. Viallet, B. Grammaticos, and A. Ramani. On the integrability of correspondences associated to integral curves. *Phys. Lett. A*, 322(3-4):186–193, 2004.

[69] L. Washington. *Elliptic Curves: Number Theory and Cryptography.* Boca Raton; London: Chapman and Hill/CRC, 2003.