Periodic orbits in systems of two non-rigid particles on the torus

Steven Gerald Maye
Department of Mathematics and Statistics
McGill University, Montreal

October 2008

A thesis submitted to McGill University
in partial fulfilment of the requirements of the degree of
Master of Science

Abstract

We establish a criterion on potential energy functions which, when satisfied, asserts the existence of an infinite number of periodic orbits in a dynamical system defined by two particles moving on the two-dimensional (or "flat") torus. The original system is reduced to that of a single point-mass moving about the torus, for which we find a continuum of trajectories satisfying a particular symmetry relation. Using a system of Poincaré maps, we obtain addition information about a particular subset of these trajectories in order to describe their behaviour in a linear portion of the space. Finally, we show under certain additional assumptions that, for any sufficiently large two-dimensional torus, a countably infinite subset of these trajectories are periodic.

Resumé

On établisse un critère sur les fonctions d'énergie potentielle qui, quand satisfait, affirme l'existence d'un nombre infini d'orbites d'un système dynamique de deux particules en se déplaçant autour du tore de deux dimensions. On réduit le système d'origine au système d'un point de masse qui se déplace autour du tore, duquel on trouve un continuum de trajectoires qui satisfait à un rapport particulier de symétrie. En utilisant un système de cartes de Poincaré, on apprend davantage d'un sous-ensemble particulier de ses trajectoires pour décrire leur comportement dans une partie linéaire de l'espace. Enfin, on démontre, selon certaines hypothèses, que pour n'importe quel tore assez grand de deux dimensions, un sous-ensemble infini dénombrable est périodique.

Acknowledgements

I would like to thank my supervisor, Paul Tupper, for suggesting this topic, and for allowing me to approach it from a decidedly analytic perspective. While my progress in this vein was sometimes halting, he readily applied his time and energy. Indeed, without his guidance and support, I could not have seen this document through to completion.

Contents

1 Introduction **9**

1.1 Overview 10

2 Preliminaries **13**

2.1 Hamiltonian Mechanics 13

2.2 Dynamical Systems and Ergodic Theory 19

2.2.1 Ergodicity for Hamiltonian Systems 24

3 Our Model and Its Reduction **27**

3.1 Two Point-Masses on the Torus 27

3.1.1 Initial Invariants 33

3.1.2 An Equivalence among Potential Functions 34

3.2 Dimension-Reducing Transformations 35

3.3 The Linear Component 37

3.4 The Non-linear Component 40

3.4.1 Existence of Periodic Orbits 40

3.4.2 Local Quadratures 43

3.4.3 Integrability and the Global Geometry 46

4 The Collision Map **49**

4.1 Constructing the Map 52

4.1.1 The x-Component 55

4.1.2 The M-Component 58

4.2 Fixed Points of the Components 59

5 Geometry and Periodic Orbits **63**

5.1 Deflections 63

5.2 Periodicity for a Special Case 66

5.3 A Criterion for Periodic Orbits 68

5.4 Endnote: Periodic Solutions and Ergodicity 68

A Derivations **71**

A.1 The Transformation $(M, E; r, \theta) \mapsto (p_x, p_y)$ 71

A.2 Equation (4.5): $(x, M) \mapsto \theta_0$ 72

Chapter 1

Introduction

The subject of this thesis arose from certain studies pertaining to the ergodicity of so-called billiard systems. A billiard is a dynamical system that consists of one or several circular masses moving in a given region of the plane. These masses behave in the manner of billiard balls: they undergo an instantaneous deflection when they encounter one another, or when they reach the boundary of the region, and otherwise move linearly according to their present velocity. In our case, this region is the two-dimensional torus and does not have a boundary as such; moreover, our masses are reduced to single points, and these point-masses instead exert a repulsive force on one another when sufficiently close, such that their behaviour is non-linear and non-trivial.

Such systems have long been important in terms of statistical mechanics. Consider, for example, the following quotation from Boltzmann:

> ...the molecules of gases are so far apart they no longer exert significant forces on one another; since the external forces acting on gases can usually be neglected, their molecules are indeed precisely in the state of the black and white balls mentioned earlier. [Bol, p. 24]

The molecules of a gas behave like billiard balls; however, to simulate these particles directly with an appropriately large number of point-masses would be impossible. Instead, we look to simpler systems, and try to ascertain what qualitative features they possess that might be shared with their higher-dimensional counterparts. Strictly speaking, the system we will study is not a traditional billiard: the torus has no boundary, nor is it a region of the plane. Moreover, the deflections of a traditional billiard are assumed to be instantaneous, while ours will have obvious duration. Yet our system aims to address the same physical phenomenon

as a traditional billiard: while a traditional billiard models one or more molecules confined in a particular container, two point-masses in motion on the torus may be seen as approximating an unbounded system of particles using a periodic domain, where each molecule is representative of an infinite lattice.

This model has been proven to admit non-ergodic flows given special choices of the geometry [Don2]. This work follows two other papers [Don1,DL] which deal with a single particle moving through a potential on the torus, the problem to which the corresponding mathematical system reduces. These papers, combined, gave sufficient and necessary conditions for the system to be ergodic independent of the geometry. Here, "geometry" refers to the dimensions of a rectangle in Euclidean space which, when appropriate edges are identified, gives rise to the torus in question. The results of [DL] showed certain systems to be ergodic for any geometry; the results of [Don1] showed that all other systems will not be ergodic for certain choices of the geometry. There are two obvious questions that remain open in the light of these results: (i) For non-ergodic systems, what proportion of the geometries result in non-ergodic flows? (ii) Can the results be generalized to deal with systems of three or more particles?

This project began in an attempt to answer the latter question but comes closer to addressing the former. Using a series of Poincaré maps, we address a particular family of symmetric trajectories, each of which may become periodic for certain choices of the geometry. Indeed, we show that, under certain conditions, there will be infinitely many periodic orbits for any geometry. While we will not address the stability of these orbits below, the maps were constructed in the hope that they would reveal a kind of stable behaviour for collisions between particles within the two-particle model, which could then be brought to bear on larger billiard systems. At any rate, I believe the Poincaré maps presented here are a valuable means to understanding the system, even if their merits are not entirely exploited.

1.1 Overview

Chapters 2 and 3 present the material necessary to understand and approach the problem broadly painted in this introduction. The former deals entirely with mathematical preliminaries; the latter formulates the physical problem in more precise mathematical terms, then manipulates that formulation to arrive at a more tractable mathematical problem. In particular, §3.2 shows how the origi-

nal system of differential equations may be reduced in dimensionality and then "uncoupled," leaving to two simpler systems—one linear, the other non-linear. The first describes linear motion on a torus and is trivially solvable; it is also ergodic for almost all initial conditions, although this is not sufficient to determine the ergodicity of the original system. The crux of the problem thus lies with the non-linear system, which is insolvable by current methods. Its properties are investigated in the remainder of Chapter 3.

The remainder of the thesis is devoted to finding symmetric trajectories for the non-linear system, since these may be periodic for certain choices of the geometry. While such trajectories are found initially in §3.4.1, they are there described by the point invariant under the symmetry in question; moreover, this point is in the non-linear region for all interesting trajectories. In Chapter 4, we use a series of Poincaré maps to find such trajectories, so that they are now given by a point in the linear region. This allows us to extend the trajectory in phase space—starting with either the point in question, or the point to which it is symmetric—so that, in Chapter 5, we can speculate as to the existence of periodic orbits in a more or less arbitrary geometry.

Chapter 2

Preliminaries

2.1 Hamiltonian Mechanics

We begin by describing (in brief) the theory of Hamiltonian systems as they arise on smooth, connected manifolds. Once we have properly formulated the Hamiltonian system, we then review certain key results that are pertinent to the following chapters. However, before we can touch upon that material, we must recall the notion of a manifold itself—and define certain structures upon it, sufficient for the formulation of an ordinary differential equation.

Definition 1 *A n-dimensional manifold M is a paracompact Hausdorff space equipped with a family of charts $\{(U_i, \chi_i)\}_{i\in I}$ such that: (i) $\bigcup_{i\in I} U_i$ is an open cover of M; and (ii) each $\chi_i : U_i \to V_i$ is a homeomorphism between U_i and some open $V_i \subset \mathbb{R}^n$. The manifold is called* smooth *if, for all $i, j \in I$ with $U_i \cap U_j \neq \emptyset$, the change of co-ordinates $\chi_j \circ \chi_i^{-1} : V_i \to V_j$ is a smooth function.*

Every manifold M is "locally Euclidean" in the sense that, for each $x \in M$, there exists an open set U_i containing x that is homeomorphic to an open set in Euclidean space. These open sets U_i, when endowed with so-called "local co-ordinates" $V_i = \chi_i(U_i) \subset \mathbb{R}^n$, also allow us to perform certain basic computations that are more obviously defined in a metric space.

For instance, fix an $x \in M$, and choose $i \in I$ such that $x \in U_i$. Let $c(t)$ denote a curve in the local co-ordinate system $\chi_i(U_i)$ such that $c(t_0) = x$ for some $t_0 \in J \subset \mathbb{R}$. Suppose further that each component c_i of c is given by a smooth function $c_i : J \to \mathbb{R}$. To every such curve we can associate a tangent vector v at x, given by $v = \frac{d}{dt}c(t)\big|_{t=t_0}$, and the totality of such vectors is called the *tangent space* at x and denoted TM_x. It is elementary to show that, for each $x \in M$, TM_x

is a linear space isomorphic to $\mathbb{R}^n$. Next, we define the union of all such spaces

$$TM = \{(x, v) \mid x \in M, v \in TM_x\}.$$

This is called the *tangent bundle* of M, which we endow with a system of charts defined as follows: for every chart (U_i, χ_i) on M, we define an open set

$$U_i' = \{(x, v) \mid x \in U_i, v \in TM_x\}$$

and associate with it a homeomorphism $\chi_i' : (x, v) \mapsto (\chi_i(x), v)$, for all $(x, v) \in U_i$, where we identify v with its representation in $\mathbb{R}^n$. It follows that TM is a manifold of dimension $2n$. Moreover, since each χ_i' is as differentiable as the corresponding χ_i, TM will be smooth whenever M is itself smooth. Finally, insofar as TM is a smooth manifold, it too has a tangent space, which we will denote by $T(TM)$.

This provides us with enough structure to define the notion of a vector field in a non-Euclidean setting. An autonomous *vector field* v on a smooth manifold M is a smooth mapping $v : M \to TM$ such that, for each $x \in M$, $v(x) \in TM_x$. Here, smooth means that, for each x and $v(x)$, v is a smooth mapping between the local co-ordinate systems of M and TM. It so happens that, for all such vector fields v and for every point $x \in M$, there is a curve $c : \mathbb{R}_0^+ \to M$ such that (i) $c(0) = x$, and (ii) in local coordinates we have, for all $t_0 > 0$,

$$\frac{d}{dt}c(t)\bigg|_{t=t_0} = v(x) \quad \text{where} \quad x = c(t_0).$$

We will refer to $c(t)$ so constructed as an "integral curve" of the vector field v, starting at the point x, which makes $c(t_0)$ the value of the integral curve at time t_0. Now, for each $t \in \mathbb{R}_0^+$, we can define a mapping $\Phi_v^t : M \to M$ which sends each $x \in M$ to the integral curve of v starting at x, evaluated at time t. Thus, for instance, the integral curve $c(t)$ above satisfies $c(t) = \Phi_v^t x$ for all $t \in \mathbb{R}_0^+$. This family of maps is called the *flow* generated by v, and it satisfied the following semigroup properties:

$$\begin{aligned} &\text{(An Identity)} && \Phi_v^0 : x \mapsto x; \\ &\text{(Closure)} && \Phi_v^t \circ \Phi_v^s = \Phi_v^{t+s}, \quad \forall t, s \in \mathbb{R}^+. \end{aligned}$$

Note that the identity used to show closure also implies that the mappings which comprise a flow commute. If every integral curve can be extended backward in

time, as is the case whenever M is compact [Arn1, §35], then Φ_v^t is defined for all $t \in \mathbb{R}$. In that case, the second identity (with $s = -t$) gives the existence of inverses and the flow thus comprises a group. Finally, the expression $\Phi_v^t x$ can be shown to be continuous in its dependence on x, and is (by definition) once differentiable in t, with $\frac{d}{dt}\Phi_v^t x = v(x)$.

These ideas provide the framework for a generalization of Newtonian mechanics acting within alternative geometric structures. Consider a point-mass moving in an n-dimensional manifold M. At each time t, its position is given by some $q \in M$, and its velocity is given by a vector $\dot{q} \in TM_q$. Concatenating these gives a point in TM; moreover, if the momentum is so much as continuous in time, the curve it traces on M is the projection of another curve in TM. Now, it is clear that there are some smooth motions which cannot be given by an autonomous vector field on M—those which are self-intersecting, or where a trajectory doubles back on itself, for example—but there are certain of those motions that are still given by a projection of the corresponding motion in TM, where the latter is indeed given by a vector field $v : TM \to T(TM)$.

Example. Think, for instance, of the oscillation $q(t) = \sin t$ which defines a motion in $M = [0, 1]$. All but two points in that space are traversed repeatedly, with the same speed but different velocities, and so we cannot associate a unique tangent vector with each point. But $\dot{q}(t) = \cos t$, and the associated curve in TM is not self-intersecting; moreover, the curve $\{(\sin t, \cos t)\}_{t\in\mathbb{R}}$ is an integral curve of the vector field $v : (q, \dot{q}) \mapsto (\dot{q}, q)$, where $v : TM \to T(TM)$. ◇

Physically, such systems occur whenever the force acting on the point-mass depends only on its current position and momentum. In particular, this means that the law governing its acceleration does not change with time.

The natural state-space for any mechanical system is, thus, not the manifold M of possible positions but its tangent bundle TM. The underlying manifold M is here termed the *configuration space*, while TM is referred to as the *phase space*. If the force acting on such a system is independent of time, the motion of the system is given by the flow $\Phi_v^t : TM \to TM$ of an (autonomous) vector field $v : TM \to T(TM)$.

Definition 2 *A* Hamiltonian system *is a system of ordinary differential equations which, for some smooth $H : TM \to \mathbb{R}$, has the following form:*

$$\dot{q} = \nabla_p H, \quad \dot{p} = -\nabla_q H, \tag{2.1}$$

where $\nabla_x H$ represents the vector of partial derivatives of H, taken with respect to the elements of x. H is called the Hamiltonian function*, or simply* Hamiltonian*, and the equations (2.1) are called* Hamilton's equations.

In homage to the physical problems in which Hamiltonian systems most often appear, including the example above, the vector q is referred to as the "position" vector, while p is referred to as the "momentum" vector, and its components "[generalized] momenta." These metaphors are borne out by Hamiltonians of the form

$$H = \frac{1}{2}\|p\|^2 + F(U),$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. In that case, we clearly have $\dot{q}_i = p_i$ for each $i = 1 \ldots n$. If $(q(t), p(t))$ is a solution of the Hamiltonian system, then $p(t)$ gives the tangent vector to the (parameterized) curve defined by $q(t)$ at each point; in particular, it would be enough to solve for $q(t)$, and then find $p(t)$ by differentiation.

If $(q(t), p(t))$ is a solution of system (2.1), then we clearly have

$$\frac{d}{dt}H(q(t), p(t)) = \sum_i (\dot{q}_i H_{q_i} + \dot{p}_i H_{p_i}) = 0.$$

Thus the Hamiltonian is preserved under the action of the flow. Such a function is known as an *invariant* or *first integral* of the system, and many Hamiltonian systems (including the one studied here) possess additional functions with this property. Note that every value of the first integral defines a surface in phase-space, and each trajectory must be contained in the surface whose value is given by the trajectory's initial condition. In particular, if a system admits multiple first integrals, then each of its trajectories must be contained in an intersection of the corresponding level surfaces.

Definition 3 *A system of n first-order ordinary differential equations is said to be* integrable by quadratures*, or simply* integrable*, if it has $n-1$ independent first integrals.*

The integrals $I_i : TM \to \mathbb{R}$ are *independent* if their differentials are linearly independent on each set

$$TM_z := \{x \in TM \mid I_i(x) = z_i,\ i = 1, \ldots, n\}$$

—that is, on the intersection of their level surfaces. For such a system, the intersection of all known level surfaces define a space of degree one—that is, a curve—in phase space. Thus, we can find the points comprising each of its trajectories. In the case of Hamiltonian systems, a stronger result holds:

Theorem 1 *Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be the Hamiltonian corresponding to a given system of ordinary differential equations. If H has n independent first integrals, then it is integrable by quadratures. [Arn2, §49]*

The integrals of a Hamiltonian system can be linked with the broader notion of "symmetries":

Definition 4 *A group of transformations $\{g_i\}_{i\in I}$ is called a symmetry of a system of ordinary differential equation if, for every solution $f(t)$, the function $g_i \cdot f(t)$ is also a solution.*

The group action $g_i \cdot f$ may be any transformation on the points $(f(t), t)$ that comprise the solution. The most obvious use of a symmetry is to generate new solutions from an existing one: once we solve a system for the initial condition $f(0)$, we will have also solved the system for each $g_i \cdot f(0)$. For instance, if $g_i : TM \to TM$ acts only on the phase space, then we have the relation

$$g \cdot \Phi^t(x_0) = \Phi^t(g \cdot x_0). \tag{2.2}$$

But symmetries may be revealing in other ways. Hamiltonian systems have both continuous and discrete symmetries, and these differ in their derivation and applications. Continuous symmetries may be shown to arise from first integrals of a Hamiltonian system. Each integral I_i can itself be considered as a Hamiltonian on the same phase space. It thus gives rise to a (generally distinct) system of differential equations, whose solutions are obviously a group of transformations on the space, and their group action is defined by composition with the original flow—i.e., $g_i \cdot f := g_i \circ f$.

Theorem 2 *Let $H : M \to \mathbb{R}$ and $I : M \to \mathbb{R}$ denote Hamiltonian functions. If I is a first integral of H, then the flow generated by I corresponds to a continuous symmetry of H. [Olv, §6.3]*

Thus, every first integral of the system gives rise to a symmetry. In the case of autonomous equations, the converse is also true:

Theorem 3 *Let $H : M \to \mathbb{R}$ and $I : M \to \mathbb{R}$ denote Hamiltonian functions. If the flow generated by I corresponds to a continuous symmetry of H, then I is a first integral of H.*

This is a variant of Noether's Theorem, adapted for autonomous Hamiltonian systems;[1] together with the previous theorem, it asserts a correspondence between invariant functions and symmetries, which is useful because certain symmetries are more obvious than their integral counterparts (for an example, see §3.4.3 below). Finally, by Theorem 1, knowing n continuous symmetries of a Hamiltonian system with dimension $2n$ would allow us to solve the system.

Discrete symmetries do not allow us to construct such solutions, but they can give us important insights into a system's behaviour. Also, unlike the continuous symmetries considered above, their group actions generally affect the temporal variable. One important example, which applies to all Hamiltonian systems, is reversability:

Definition 5 *Let $\rho : M \to M$ be an invertible linear transformation. The flow Φ is called ρ-reversible if, for each $\Phi^t : M \to M$,*

$$\rho \circ \Phi^t = \Phi^{-1} \circ \rho.$$

While the symmetry group is not explicitly stated in the definition, it is of the same form as equation (2.2), excepting the reflection of the temporal variable. Since we will make extensive use of such symmetries later (see §3.4.1), the following material deserves mention:

Definition 6 *Let $\rho : M \to M$ be an invertible linear transformation. A vector field $v : M \to TM$ is called ρ-reversible if*

$$\rho \circ v(x) = -v \circ \rho(x), \quad \forall x \in M.$$

Proposition 1 *The flow Φ arising from a ρ-reversible vector field is ρ-reversible in the sense of Definition 5. [HLW, § V.1]*

[1] This is a reworded version of the theorem as it appears in [KH, §5.5]; a non-autonomous version is given in [Olv, §6.3]. Although these are both stated for Hamiltonian systems, Noether's theorem is more often formulated within the framework of Lagrangian mechanics.

2.2 Dynamical Systems and Ergodic Theory

Definition 7 *A* dynamical system *is a pair* $(\Phi^t; X)$ *consisting of a topological space* X*, whose points are known as "states"; and a semi-group of transformations* $\Phi^t : X \to X$*.*

In general, we are concerned with investigating the ergodicity of certain smooth, continuous-time dynamical systems described by ordinary differential equations. In this case, the semi-group in question has the index set $t \in \mathbb{R}$, and each element Φ^t is a mapping between the set of initial conditions $x \in X$ and their corresponding solutions at a fixed time t, known as a *flow*. Later, we will also need to consider certain discrete-time dynamical systems, whose groups of transformations will be indexed by $n \in \mathbb{Z}$; however, we are not directly interested in the ergodicity of such systems, and the definitions that follow address only the continuous-time case.

For what follows, we will fix X and Φ^t such that: (1) X is compact and Hausdorff; (2) each $\Phi^t : X \to X$ is continuous, and $\Phi^t(x)$ varies continuously with t for fixed x. Now, suppose we have a function $f : X \to \mathbb{R}$, which assigns a real number to each state in X. For each solution curve $\{\Phi^t(x)\}_{t\in\mathbb{R}^+}$, we can then attempt to average the value of f along the length of the curve as $t \to \infty$:

Definition 8 *Let* $\Phi^t : X \to X$ *be a group of transformations, and* $f : X \to \mathbb{R}$*. We define the* time average *of* f *as*

$$I_x(f) := \lim_{T\to\infty} \frac{1}{T} \int_0^T f(\Phi^t x)\, dt,$$

wherever it exists.

For fixed x, each function f gives rise to a time average giving asymptotic information about the behaviour of $\Phi^t x$. In particular, for each set $U \subset X$, we can compute the time average of the characteristic function χ_U, which gives the so-called "asymptotic density" of $\Phi^t(x)$ in U; that is, the fraction of time for which we have $\Phi^t(x) \in U$ as $t \to \infty$. Given a trajectory $\{\Phi^t x\}_{t\in\mathbb{R}^+}$, we could thus consider the time average of each χ_U, and in doing so describe the trajectory's asymptotic distribution. It is, however, unclear from the above definition that any non-trivial time averages exist.

We begin by considering a restriction. Let $C(X)$ be the set of all continuous functions mapping X to $\mathbb{R}$, and consider for the moment I_x acting on only those $f \in C(X)$. The time average then, for each x, defines a real-valued functional

$I_x : C(X) \to \mathbb{R}$, which has the following properties:[2]

$$\begin{aligned}
\text{(Linearity)} \qquad & I_x(\alpha f + \beta g) = \alpha I_x(f) + \beta I_x(g), \forall \alpha, \beta \in \mathbb{R}; \\
\text{(Boundedness)} \qquad & |I_x(f)| \leq \sup_{y \in X} |f(y)|; \\
\text{(Positivity)} \qquad & I_x(f) \geq 0 \text{ whenever } f \geq 0, \text{ and } I_x 1 = 1; \\
\text{(Invariance under } \Phi) \qquad & I_x(f \circ \Phi^s) = I_x(f), \forall s \in \mathbb{R}.
\end{aligned}$$

Here, boundedness depends on the topological properties of X, and invariance follows because truncating the curve does not alter the limit. Since I_x is a positive, bounded linear functional, we can apply a particular variant of the Riesz Representation Theorem:

Theorem 4 *Let X be a compact Hausdorff space. Then for each positive bounded linear functional J on $C^0(X)$ there exists a unique finite Borel measure μ such that $J(f) = \int_X f \, d\mu$ for all $f \in C(X)$.*[3]

Thus, for each x there exists a μ_x such that

$$I_x(f) = \int_X f \, d\mu_x, \quad \forall f \in C(X). \tag{2.3}$$

It can be shown [KH, §4.1.a] that the Φ-invariance of I_x implies that μ_x is also Φ-invariant; that is, $\mu(\Phi^t A) = \mu(A)$ for all $t \in \mathbb{R}$. Moreover, noting that $I_x(f \circ \Phi^s) = I_{\Phi^s x}(f)$, we have $I_{\Phi^s x}(f) = I_x(f)$ for all $s \in \mathbb{R}$, and thus any two points from the same solution curve give rise to the same Φ-invariant measure. Equation (2.3) therefore tells us that, for each trajectory, there exists a unique Φ-invariant Borel probability measure for which all time averages can be replaced with integrals over the total phase space, taken with respect to said measure.

What is most interesting about this representation is that it assigns a measure on the entire space based on the behaviour of a single trajectory. This "extension" may often be trivial:

Example. If $\Phi^t x$ is a periodic trajectory with period T, then, supposing our σ-algebra is sufficiently refined,[4] we can define a measure μ_x such that the following

[2]This treatment of the time-average (as a linear functional) is presented in [KH, §4.1.a] for the case of discrete maps.

[3]This is a slight modification of Theorem A.2.7 in [KH], which they state for a general $f \in C(X)$; immediately after stating the general result, they note that the result above applies in the case of positive functionals.

[4]For example, if we choose the Borel σ-algebra, generated by the class of all closed sets in the topology of X.

hold:

1. $\mu_x(X \setminus \{\Phi^t x\}_{t\in\mathbb{R}}) = 0$;
2. $\mu_x(\{\Phi^t x \mid t \in [t_1, t_2]\}) = \dfrac{t_2 - t_1}{T}$, whenever $0 \le t_1 \le t_2 \le T$.

Clearly, equation (2.3) holds for this measure—since it assigns spatial weights based on the frequency with which each section of the trajectory is traversed in time—and by Theorem 4 it must by the *only* measure for which equation (2.3) holds. Unfortunately, it tells us nothing about the behaviour of the system outside this trajectory, and nothing about the trajectory which could not be deduced from periodicity alone. ◇

Perhaps unsurprisingly, more is needed to assert the relevance of such a measure to the entire space.

One obvious criterion, to that end, would be if μ_x was identical for all $x \in X$. Then, all time averages would be independent of the trajectory in question—in particular, all trajectories would have the same asymptotic distributions. This idea forms a basis for ergodic theory, which is typically formulated rather differently, in terms of invariant sets and functions.

Definition 9 *A function f is said to be* Φ-invariant *if $f(\Phi^t x) = f(x)$ for all $t \in \mathbb{R}$, for each $x \in X$. A set $U \subset X$ is said to be Φ-invariant if its indicator function χ_U is invariant; or, equivalently, if $\Phi^t U = U$ for all $t \in \mathbb{R}$.*

A Φ-invariant function is thus constant on the entirety of each trajectory $\{\Phi^t_x\}_{t\in\mathbb{R}}$; each such trajectory is a Φ-invariant set; and each Φ-invariant set is necessarily a union of these trajectories, since for Φ-invariant $U \subset X$:

$$U = \bigcup_{t\in\mathbb{R}} \Phi^t U = \bigcup_{t\in\mathbb{R}} \left\{ \Phi^t \bigcup_{x\in U} \{x\} \right\} = \bigcup_{t\in\mathbb{R}} \bigcup_{x\in U} \Phi^t\{x\} = \bigcup_{x\in U} \{\Phi^t x\}_{t\in\mathbb{R}}.$$

The notion of ergodicity restricts the nature of such sets as they occur in a given measure space (X, μ), where μ is some Φ-invariant measure.

Definition 10 *Let $\mu : \sigma(X) \to \mathbb{R}^+$ be a Φ-invariant Borel probability measure. A dynamical system $(\Phi^t; X, \mu)$ is said to be* ergodic *if every Φ-invariant set U has measure $\mu(U) = 0$ or $\mu(U) = 1$.*

Thus, for an ergodic system, any trajectory or union of trajectories has full or no measure. Note that, here, we have the measure in hand, and we apply it to the

totality of the system: While the earlier measures μ_x are obvious candidates for ergodicity, there is no guarantee that (i) they will be Φ-invariant over the entire space, or (ii) that the measure of every Φ-invariant set will be either zero or one.

Now, ergodicity can be brought to bear on the calculation and validity of time averages. To see this, we begin by considering the set of Φ-invariant functions on an ergodic system $(\Phi^t; X, \mu)$:

Lemma 1 *A dynamical system $(\Phi^t; X, \mu)$ is ergodic if and only if any Φ-invariant function is constant almost everywhere.*

Proof. We present a slightly modified version of the proof in [CFS]. For any Φ-invariant function f, the set $U_a := \{x \in X \mid f(x) < a\}$ is also invariant for any $a \in \mathbb{R}$. Thus $\mu(U_a)$ is zero or one for each a, while $\mu(U_a) \leq \mu(U_b)$ whenever $a < b$. By defining $\tilde{a} := \inf\{a \in \mathbb{R} \mid \mu(U_a) = 1\}$, we immediately see $f(x) = \tilde{a}$ almost everywhere, and the first direction is proven.

For the other direction, let A be some Φ-invariant set. By definition, its indicator function χ_A is also Φ-invariant, and by hypothesis this function must be constant almost everywhere; yet this function only takes on the values $\{0, 1\}$. Thus, either $\chi_A = 1$ almost everywhere, and $\mu(A) = 1$, or $\chi_A = 0$ almost everywhere, and $\mu(A) = 0$. This completes the proof. □

If the time average $I_x(f)$ is well-defined for all (or almost all) $x \in X$, it can be considered for fixed f as a function over the phase space, and we thus write $f_\Phi(x) := I_x(f)$. Since each time average is Φ-invariant, it then follows that the time average must be constant over the entire space, except possibly a set of measure zero. Thus, ergodicity ensures that time averages are the same for trajectories beginning at almost any $x \in X$.

In fact, more is true. While to this point we have guaranteed existence for only those time averages arising from continuous functions, the result may be stated for all μ-measurable functions on X. This result was first proven by Birkhoff, and is thus referred to as the Birkhoff Ergodic Theorem:

Theorem 5 *Let $\Phi^t : (X, \mu) \to (X, \mu)$ be a measure-preserving transformation of (X, μ), and suppose $f : (X, \mu) \to \mathbb{R}$ is μ-measurable. Then for μ-almost every $x \in X$ the following time average exists:*

$$f_\Phi(x) = I_x(f) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(\Phi^t x)\, dt.$$

Thus the time average is well-defined almost everywhere, and constant almost everywhere when the system is ergodic. The proof of Birkhoff's theorem is technical, and while alternate proofs have arisen, they are all too lengthy for inclusion here. (See [CFS] and [KH] for two such proofs.) However, three important properties of $f_\Phi(x)$ arise within these proofs: the time average is μ-measurable; it is invariant under Φ, as we showed earlier for continuos f; and the following identity holds—

$$\int_X f_\Phi \, d\mu = \int_X f \, d\mu.$$

This last point brings us to the most salient feature of ergodicity: Since f_Φ is Φ-invariant, it must be constant on any set of full measure, thus $f_\Phi(x) = \int_X f_\Phi \, d\mu$ almost everywhere, and

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\Phi^t x)\, dt = \int_X f \, d\mu \tag{2.4}$$

almost everywhere. The expression on the left is generally known as the *space average* of f; thus, for an ergodic system, time and space averages are said to agree at almost every point of (X, μ).

This property, as presented in equation (2.4), is useful for studying the long-term dynamics of a system; but, to use it, one must first know a Φ-invariant measure for which the system is ergodic. And yet, knowing just any ergodic measure is not enough. In fact, a system may be ergodic for more than one measure without any of these producing meaningful space averages.

Example. Consider, for instance, a system that contains two periodic trajectories, denoted $\Phi^t x_1$ and $\Phi^t x_2$. For the i-th periodic trajectory, we can define a measure μ_{x_i} according to (1.) and (2.) from the previous example, and the system will, in each case, be ergodic with respect to that measure: Every invariant set either contains $\Phi^t x_i$, and hence has full measure, or does not, and hence has no measure. Note that time and space averages will also hold "almost everywhere" with respect to each μ_i—but this tells us nothing about $X \setminus \Phi^t x_i$, which may be almost all of the space in the sense familiar from Lebesgue measure. ◇

This behaviour is exemplary of a more general result, which qualifies the set of measures for which a given system may be ergodic. Let μ_1 and μ_2 define measures on the same state-space X, and recall that μ_1 is said to be *absolutely continuous*

with respect to μ_2 if, for every measurable $A \subset X$,

$$\mu_2(A) = 0 \;\Rightarrow\; \mu_1(A) = 0;$$

recall also that μ_1 and μ_2 are said to be mutually singular if there are disjoint measurable sets U and V such that $\mu_1(U) = \mu_2(V) = 1$.

Theorem 6 *Let $(\Phi^t; X)$ be a dynamical system, and let μ_1 and μ_2 be Φ-invariant Borel probability measures. If $(\Phi^t; X, \mu_1)$ is ergodic and μ_2 is absolutely continuous with respect to μ_1, then $\mu_1 = \mu_2$. Otherwise, if both $(\Phi^t; X, \mu_1)$ and $(\Phi^t; X, \mu_2)$ are ergodic, then either $\mu_1 = \mu_2$, or μ_1 and μ_2 are mutually singular.*

A proof of this result is provided in [CFS] for the discrete case, but can be easily modified. Note that, if we considered only those ergodic measures with respect to which Lebesgue measure is absolutely continuous, then the second property gives $\mu_1 = \mu_2$ except on sets with Lebesgue measure zero. Additionally, all sets with positive Lebesgue measure would contribute in the calculation of space averages.

2.2.1 Ergodicity for Hamiltonian Systems

Recall that the flow of a Hamiltonian system preserves the corresponding function $H : TM \to \mathbb{R}$. If the system were ergodic, then by Lemma 1 it would be constant everywhere. But this is only the case for the null flow, and there are obviously many other choices of the Hamiltonian. Thus we have outright that no Hamiltonian system can be ergodic when considered on the entirety of its phase-space; it could however, be ergodic when the system is restricted to a particular energy surface $H = E$ for some fixed $E \in \mathbb{R}$. Of course, the same argument could be applied to any other first integrals of the system, if such functions exist. Thus, when investigating the ergodicity of a Hamiltonian system, it is necessary to fix all such functions and consider the flow on the intersection of their respective surfaces.

It should be noted that every Hamiltonian system has a natural invariant measure: by the result of Liouville, every Hamiltonian system preserves volume in phase space [Arn2]. The corresponding measure then induces a natural measure on the aforementioned surface, which is known as the *conditional Liouville measure*. Since this paper deals only with the construction of orbits, and not the ergodicity of the system, we note only that this measure is absolutely continuous

with Lesbegue measure. The interested reader is directed to [DLT] for details of the measure's construction.

Chapter 3

Our Model and Its Reduction

3.1 Two Point-Masses on the Torus

Our model begins with a rectangular domain $\Omega = [-\alpha, \alpha] \times [-\beta, \beta] \subset \mathbb{R}^2$, for some α and $\beta \in \mathbb{R}^+$. We then identify opposite edges of Ω such that, topologically speaking, Ω is equivalent to the 2-torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$. Moreover, just as $\mathbb{S}^1$ may be identified with $\mathbb{R}/\mathbb{Z}$ in the usual way, our $\Omega = \mathbb{T}^2$ is equivalent to $\mathbb{R}^2/\mathbb{Z}^2$, and so it can be endowed with a quasi-intuitive algebraic structure:

$$(s_1, t_2) \bmod 1 + (s_2, t_2) \bmod 1 = (s_1 + s_2, t_1 + t_2) \bmod 1.$$

Formally, the interior of $\Omega =: \tilde{\Omega}$ may be used to define a chart or local co-ordinate system on the torus with homeomorphism $\chi_I : U \to \tilde{\Omega} \subset \mathbb{R}^2$ such that

$$\chi_0 : (s, t) \bmod 1 \mapsto ((2s - 1)\alpha, (2t - 1)\beta), \quad \text{for } s, t \in [0, 1).$$

However, such a map cannot be constructed for Ω itself, since the points belonging to its boundary $\partial\Omega$ do not correspond to unique points of $\mathbb{T}^2$. Other charts are therefore necessary to endow the torus with the structure of a manifold.

These may be chosen in various ways. In particular, using the algebraic structure of $\mathbb{T}^2$, we can rotate U via addition of some $(r_1, r_2) \in \mathbb{T}^2$, and consider the resulting set

$$U_{r_1,r_2} := \{(s, t) + (r_1, r_2) \mid (s, t) \in U\}$$

together with a homeomorphism $\chi_{r_1,r_2} : U_{r_1,r_2} \to \tilde{\Omega}$ such that

$$\chi_{r_1,r_2}(s, t) := \chi_0(s - r_1, t - r_2).$$

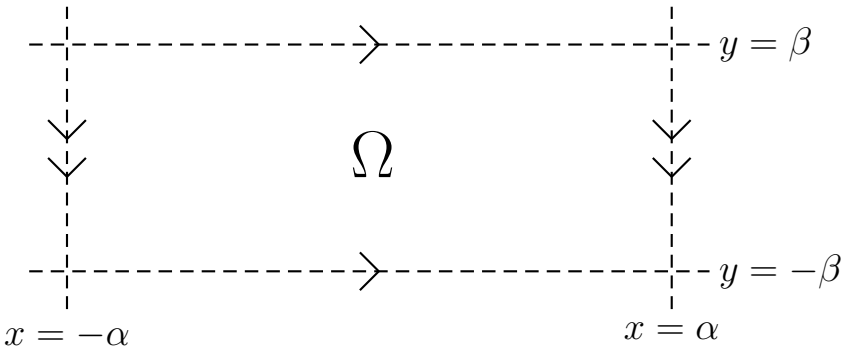

Figure 3.1: The domain $\Omega \subset \mathbb{R}^2$, which forms the basis for our charts. We can also use Ω to construct the torus by identifying the edges marked with arrows.

Now, let $q = (r_1, r_2)$ denote a point in $\mathbb{T}^2$. For each such q we in particular have $q \in U_{r_1,r_2} =: U_q$, with the corresponding homeomorphism $\chi_q := \chi_{r_1,r_2}$. Next consider the overlap $U_{r_1,r_2} \cap U_{r_3,r_4}$. If $(r_1, r_2) \neq (r_3, r_4)$, this set will be disconnected, although it can be written as a union of between two and four connected components; moreover, it is easy to see that each connected component undergoes a linear transformation between co-ordinate systems. (See figure 3.2.) In particular, for each component we can choose $r_i \in \mathbb{R}$ such that (i) $r_i \bmod 1 = r_i \in \mathbb{S}_1$ and (ii) on the overlap we have

$$\chi_{r_3,r_4} \circ \chi_{r_1,r_2}^{-1} : (x, y) \mapsto (x + 2\alpha(r_1 - r_3), y + 2\beta(r_2 - r_4))$$

for all (x, y) in the chosen set. Equivalently, we may write this

$$\chi_{r_3,r_4} \circ \chi_{r_1,r_2}^{-1} : (x, y) \mapsto (x, y) - \chi_{r_1,r_2}(r_3, r_4) + 2(\alpha N, \beta M) \tag{3.1}$$

for $N, M \in \{0, \pm 1\}$, where N and M may vary between the connected components of $\chi_{r_1,r_2}(U_{r_1,r_2})$. The torus $\mathbb{T}^2$ is thus a smooth manifold when endowed with the family of charts $\{(U_q, \chi_q)\}_{q \in \mathbb{T}^2}$. Additionally, note that each of these charts contains all of $\mathbb{T}^2$ except a trivial set.

We will use these charts to define a metric on $\mathbb{T}^2$. If q_1 and q_2 are two points on the torus, we define

$$d(q_1, q_2) = \inf_{q \in \mathbb{T}^2} \{ \|\chi_q(q_1) - \chi_q(q_2)\| \mid q_1, q_2 \in U_q \} . \tag{3.2}$$

To show that this is indeed a metric, we will use the following lemma:

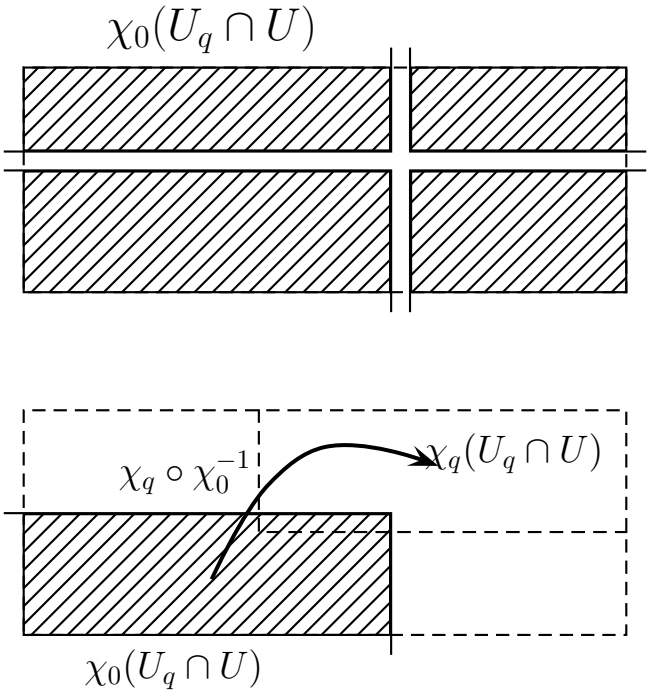

Figure 3.2: Top: The set $U_q \cap U$, for $q \approx (\frac{2}{3}, \frac{2}{3}) \in \mathbb{R}^2/\mathbb{Z}^2$, as it appears in the co-ordinates of the original $\tilde{\Omega} = \chi_0(U)$. Note that although their intersection is disconnected, it can be written as the union of four connected sets. Bottom: The change of co-ordinates maps moves the top-right corner of U_q to the top-right corner of Ω. Within each connected component, the map is linear.

Lemma 2 *The infimum in equation (3.2) is achieved for $q = q_1$ whenever $q_2 \in U_{q_1}$.*

Proof. Let $q \in \mathbb{T}^2$ be such that $q_1, q_2 \in U_q$. In local co-ordinates $\chi_q(U_q)$, we have the distance $\|\chi_q(q_2) - \chi_q(q_1)\|$. Whereas in the local co-ordinates $\chi_{q_1}(U_{q_1})$ we have

$$\|\chi_{q_1}(q_2) - \chi_{q_1}(q_1)\| = \|\chi_{q_1}(q_2)\| = \|\chi_q(q_2) - \chi_q(q_1) + 2(\alpha N, \beta M)\|$$

using the transformation (3.1). Define

$$D(N, M) := \|\chi_q(q_2) - \chi_q(q_1) + 2(\alpha N, \beta M), \|$$

and note that the distance in $\chi_q(U_q)$ is achieved for $N = M = 0$. It is now sufficient to show that the values of N and M which minimize D are indeed those of transformation (3.1), for which

$$\chi_q(q_2) - \chi_q(q_1) + 2(\alpha N, \beta M) = \chi_{q_1}(q_2) \in \chi_{q_1}(U_{q_1}).$$

Define $\Delta := \chi_q(q_2) - \chi_q(q_1)$ and let π_i denote the projection onto the i-th component. It is clear that $|\pi_1(\Delta)| < 2\alpha$ and $|\pi_2(\Delta)| < 2\beta$, and that

$$D(N, M) = \left(|\pi_1(\Delta) - 2\alpha N|^2 + |\pi_2(\Delta) - 2\beta M|^2\right)^{\frac{1}{2}}.$$

Now, for every fixed $M \in \mathbb{Z}$, D is minimized by

$$N = \begin{cases} -1 & \text{if } \pi_1(\Delta) > \alpha \\ 0 & \text{if } -\alpha < \pi_1(\Delta) < \alpha \\ 1 & \text{if } \pi_1(\Delta) < -\alpha \end{cases}$$

And for every fixed $N \in \mathbb{Z}$, D is minimized by

$$M = \begin{cases} -1 & \text{if } \pi_2(\Delta) > \beta \\ 0 & \text{if } -\beta < \pi_2(\Delta) < \beta \\ 1 & \text{if } \pi_2(\Delta) < -\beta \end{cases}$$

It is now easy to verify that

$$\Delta + 2(\alpha N, \beta M) \in \chi_{q_1}(U_{q_1}) = (-\alpha, \alpha) \times (-\beta, \beta),$$

and, given the size of $(2\alpha, 0)$ and $(0, 2\beta)$, that there is only one such point $\chi_{q_1}(q_2)$.

This completes the proof. □

The lemma applies only for $q_2 \in U_{q_1}$; the remaining points present a formal problem. To resolve this, we extend the local co-ordinates $\chi_{q_1}(U_{q_1}) = \tilde{\Omega}$ to the whole of $\Omega = \tilde{\Omega} \cup \partial\Omega$, and extend the map $\chi_{q_1}^{-1} : \Omega \to \mathbb{T}^2$ such that it is continuous. Recall from our original construction of $\mathbb{T}^2$ that the points $\mathbb{T}^2 \setminus U$ corresponded to the boundary $\partial\Omega$. Since U_{q_1} is simply a rotation of U, the same holds here, and we see that $\chi_{q_1}^{-1}$ is onto $\mathbb{T}^2$. Finally, while points $q_2 \notin U_{q_1}$ do not have a unique pre-image in Ω, it is enough to note that, because of the way boundary points have been identified, they have the same norm in $\chi_{q_1}(\mathbb{T}^2)$.

It is thus possible to identify any $d(q_1, q_2)$ with its norm in the closure of $\chi_{q_2}(U_{q_2})$. We will use this fact to prove the following:

Proposition 2 *Equation (3.2) defines a metric on* $\mathbb{T}^2$.

Proof. The property $d(q_1, q_2) \geq 0$ follows readily from the definition; that $d(q_1, q_2) = 0$ if and only if $q_1 = q_2$ is clear from Lemma 2. Now consider the sum $d(q_1, q_2) + d(q_2, q_3)$. In extended local co-ordinates $\chi_{q_2}(U_{q_2})$ we have

$$\begin{aligned} d(q_1, q_2) + d(q_2, q_3) &= \|\chi_{q_2}(q_1)\| + \|\chi_{q_2}(q_3)\| \\ &\geq \|\chi_{q_2}(q_1) - \chi_{q_2}(q_3)\|. \end{aligned}$$

But by the definition of $d(\,\cdot\,,\,\cdot\,)$, we have $\|\chi_{q_2}(q_1) - \chi_{q_2}(q_3)\| \geq d(q_1, q_3)$. It follows that $d(q_1, q_2) + d(q_2, q_3) > d(q_1, q_3)$ and equation (3.2) defines a metric on $\mathbb{T}^2$. □

Thus, our torus is now both a smooth manifold and a metric space.

Having properly constructed the torus, we may now define our dynamical system upon it. Consider two identical point-masses moving in $\mathbb{T}^2$ with position co-ordinates $q_1(t), q_2(t)$. Their motion is governed by a potential energy $\mathcal{U} : \mathbb{T}^2 \times \mathbb{T}^2 \to \mathbb{R}$ which depends only on the distance between these points; we thus write $\mathcal{U}(q_1, q_2) = U(d(q_1, q_2))$ for some $U : \mathbb{R}^+ \to \mathbb{R}$. Using this notation, we will now make additional assumptions regarding our potential:

$$U \text{ is at least once differentiable;} \tag{3.3}$$

$$\exists R \in \mathbb{R}^+ \text{ such that } R < \min\{\alpha, \beta\} \text{ and } U(r) = 0, \forall r > R. \tag{3.4}$$

We further assume that there exists an $\xi \in \mathbb{R}^+$ and a subinterval $I \subset [0, \xi) \subset \mathbb{R}$

of positive measure, such that the following properties hold:

$$U'(r) \leq 0, \quad \forall r \in [0, \xi), \tag{3.5}$$

$$U'(r) < 0, \quad \forall r \in I \subset [0, \xi), \tag{3.6}$$

$$\text{and} \qquad \inf_{r \in [0,\xi)} U(r) \geq \sup_{r \in [\xi,\infty)} U(r), \tag{3.7}$$

One the one hand, these assumptions are mathematical conveniences, but they are also made in accordance with the physical model. Equation (3.4), for instance, formalizes the idea that the molecules of gases, when sufficiently far apart, no longer exert significant forces on one another—to paraphrase Boltzmann's discussion of gases quoted earlier. Similarly, as we will see in §3.4.2, the final criteria (3.5–3.7) ensure that all high-energy collisions result in deflections. (For low energies, it is possible for the particles to "orbit" one another.) In mathematical terms, these criteria ensure that U is strictly decreasing on some non-trivial interval $I \subset \mathbb{R}^+$, and that each $U(r)$ has a unique pre-image whenever $r \in I$. While there are continuous differentiable functions that satisfy this criterion but do not satisfy equations (3.5–3.7), the latter is enough to ensure the former.

Example. A simple potential function satisfying (3.3–3.7) is of the form

$$U(d(q_1, q_2)) = \frac{1}{2} k^2 (2R - d(q_1, q_2))_+^2, \tag{3.8}$$

for some $R, k \in \mathbb{R}$, where $(\,\cdot\,)_+ := \max\{\,\cdot\,, 0\}$. ◇

Now, the momentum of each point-mass is given by $p = m\dot{q}$, where $m \in \mathbb{R}$ is the "mass" associated with each point. The kinetic energy for the system is then given by $\mathcal{T} = \frac{1}{2m_1} p_1^2 + \frac{1}{2m_2} p_2^2$. Since we are interested in identical particles, we will assume without loss of generality that $m_1 = m_2 = 1$. The total mechanical energy $H = \mathcal{T} + \mathcal{U}$ will now be our Hamiltonian function. However, in formulating this problem as a Hamiltonian system, we will concatenate each pair of position and momentum vectors to form $q \in \mathbb{T}^4$ and $p \in \mathbb{R}^4$. The above description thus corresponds to a Hamiltonian of the form

$$H(q, p) = \frac{1}{2} \|p\|^2 + U(d(q_1, q_2)). \tag{3.9}$$

As in a classical mechanical system, the total "mechanical energy" is clearly preserved, such that the kinetic energy decreases as U increases. More precisely, we

expect the point-masses to move with constant momentum when $d(q_1, q_2) > R$, since this gives $U = 0$; and we expect the point-masses, when they subsequently encounter each other, will undergo a change in kinetic energy analogous to an elastic collision. This motion, of course, is formally described using Hamilton's equations—

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} = p, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} = -\nabla_q U(d(q_1, q_2)) \tag{3.10}$$

—which define a vector field on the whole of $T\mathbb{T}^2$.

As we mentioned earlier, the closure of $\tilde{\Omega}$ can be mapped onto the torus. For this reason, we can work almost entirely within any system of local co-ordinates $\chi_q(U_q)$: solutions may be solved within the local co-ordinate system, then extended continuously beyond their boundary $\partial\Omega$, whereupon the solution can again be identified with a point in $\tilde{\Omega}$. Using these local co-ordinates, we may speak of a point $q = (x, y) \in \Omega$ as having "x" and "y" co-ordinates; similarly, the momentum of a particle moving in Ω may be said to have "x" and "y" components. In order to highlight the geometry of the problem, we will speak of the position of our particles as points on the torus; but the distinction between x and y co-ordinates that is natural in Ω will prove beneficial in describing certain classes of trajectories in the systems under consideration.

3.1.1 Initial Invariants

In addition to the total energy H, the system has two further invariants that arise from the total momentum of the system. To see this, note

$$\begin{aligned}
\frac{dp_1}{dt} &= -\nabla_{q_1} U(d(q_1, q_2)) \\
&= -U'(s)|_{s=d(q_1,q_2)} \cdot \nabla_{q_1} d(q_1, q_2) \\
&= -U'(s)|_{s=d(q_1,q_2)} \cdot -\nabla_{q_2} d(q_1, q_2) \\
&= \nabla_{q_2} U(d(q_1, q_2)) \;=\; -\frac{dp_2}{dt}
\end{aligned}$$

by the chain rule, and hence

$$\frac{d}{dt}(p_1 + p_2) = 0. \tag{3.11}$$

Projecting this into the x and y components of phase space then yields two first integrals of (3.10)—in addition to the Hamiltonian itself—but no additional first

integrals are known to exist; hence, the system will be presumed un-integrable. That said, in §3.2 we will reduce the dimension of the phase-space under consideration via a change of co-ordinates, which uses the energies implicit in (3.11) to eliminate two components of the momenta.

3.1.2 An Equivalence among Potential Functions

In equation (3.8), we introduced a sample potential which dependent on a multiplicative constant k^2. Consider now an arbitrary potential $U : \mathbb{R}^+ \to \mathbb{R}$ satisfying at least (3.3) and (3.4), and define the related potential $\tilde{U} = k^2 U$ for an arbitrary $k \in \mathbb{R}$. Now suppose we are given a set of initial conditions $(q(0), p(0))$ for which $d(q_1, q_2) \geq R$. Given the energy $\tilde{U}$, each trajectory in phase space is of course determined by the equations

$$\dot{q} = p, \tag{3.12}$$

$$\dot{p} = -\nabla_q \tilde{U}(d(q_1, q_2)). \tag{3.13}$$

Now, we can employ the following scalar transformation of coordinates in t and p, while leaving q unchanged—

$$t \mapsto t/k$$
$$p \mapsto kp$$

—which leaves $\dot{q} = p$, and on the other hand gives

$$\dot{p} = -\nabla_q \left(\frac{1}{k^2} \tilde{U}(d(q_1, q_2)) \right) = -\nabla_q \left(U(d(q_1, q_2)) \right).$$

But these equations are Hamilton's equations for a system with potential energy U; indeed, every trajectory $(q(t), p(t))$ described by equations (3.12–3.13) may be mapped to the trajectory $(q(kt), \frac{1}{k} p(kt))$ of the original system, which is identical in configuration space and isometric otherwise. The potential energy U can therefore be altered by a scalar factor without affecting the behaviour of the system as a whole; however, as we will see, the dynamics of the system are otherwise dependent on the graph of U in $\mathbb{R}^+ \times \mathbb{R}$.

3.2 Dimension-Reducing Transformations

A dimension-reducing transformation for the two-particle problem occurs in the work of Sinai [Sin]. In terms of the spatial co-ordinates, his transformation takes the following form:

$$Q_1 = q_1 + q_2, \tag{3.14}$$
$$Q_2 = q_1 - q_2; \tag{3.15}$$

—or, written using a matrix over the field of 2-vectors,

$$\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}. \tag{3.16}$$

Sinai's transformation is a four-to-one cover of the original configuration space. To see this, begin by considering the system (3.14–3.15) for points $q_1, q_2, Q_1, Q_2 \in \mathbb{S}^1$. Now suppose we are given Q_1 and Q_2, while q_1 and q_2 remain unknown, and then replace the known quantities with arbitrary lifts into $\mathbb{R}$; we will denote these lifts by $Q_1 + N$ and $Q_2 + M$ respectively, such that $Q_1, Q_2 \in [0, 1)$ and $N, M \in \mathbb{Z}$. The resulting algebraic system in $\mathbb{R}$ has solutions

$$q_1 = \frac{1}{2}(Q_1 + Q_2) + \frac{1}{2}(N + M)$$
$$q_2 = \frac{1}{2}(Q_1 - Q_2) + \frac{1}{2}(N - M)$$

For each choice of N and M, these solutions can be projected back into $\mathbb{S}_1$, which gives two distinct solutions depending on whether $N \pm M$ is even. Since our system is in fact over $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$, this ambiguity exists in determining each co-ordinate, and hence each transformed point is the image of four points on the original torus.

More recently, Donnay [Don2] has presented three different dimension-reducing transformations for the above system, two of which are also suitable for particles of unequal mass. The first of these is applicable only when the ratio of these masses is rational; in the case of equal masses, this particular transformation reduces to that of Sinai. More interesting is the third transformation, which he derives for particles of equal mass and zero total momentum—and which follows from simple, geometric considerations. Again, we are concerned with particles of equal, unitary mass; for such a system, this transformation would amount to

considering the position of the midpoint between the two particles, together with the displacement of the first particle from that midpoint:

$$\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}. \tag{3.17}$$

This is only a scalar multiple of the transformation employed by Sinai, but it has the benefit of an obvious geometric interpretation, and (incidentally) the present matrix is the inverse of the former.

Of course, either of these transformations are incomplete without a corresponding transformation of the momentum variables. The overarching goal is to find a transformation $L : (q_1, q_2, p_1, p_2) \mapsto (Q_1, Q_2, P_1, P_2)$ which is symplectic, or "canonical," and so would be measure preserving. To simplify matters, we will assume L is linear, and that it treats position and momentum variables independently; that is,

$$L = \begin{bmatrix} L_Q & 0 \\ 0 & L_P \end{bmatrix},$$

for suitable 4×4 matrices $L_Q : (q_1, q_2) \mapsto (Q_1, Q_2)$ and $L_P : (p_1, p_2) \mapsto (P_1, P_2)$. Now, for a linear transformation, symplecticity requires $L^T J L = J$, where

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

A quick calculation reveals that, for L as described, this holds if and only if $L_Q L_P = I$, and so $L_P = L_Q^{-1}$. If L_Q is chosen as in (3.17), then this gives L_P such that

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \tag{3.18}$$

—again, written over the space of 2-vectors. We note that our choice of position co-ordinates gives $\dot{P}_1 = 0$ by equation (3.11). As shown in [HLW, §VI.2], the symplectic transformation of a Hamiltonian system is again a Hamiltonian, defined via the function $\tilde{H}(Q, P) = H(L^{-1}(Q, P))$; explicitly, this gives

$$\tilde{H} = \frac{1}{4}\|P_1\|^2 + \frac{1}{4}\|P_2\|^2 + U(2d(0, Q_2)). \tag{3.19}$$

Note that this transformed system is separable: letting

$$H_1(Q_1, P_1) = \frac{1}{4}\|P_1\|^2 \tag{3.20}$$

$$\text{and} \quad H_2(Q_2, P_2) = \frac{1}{4}\|P_2\|^2 + U(2d(0, Q_2)), \tag{3.21}$$

we can write $\tilde{H} = H_1 + H_2$; each co-ordinate pair (Q_1, P_1) and (Q_2, P_2) evolves independently of the other, and thus the resulting system is, in fact, the product of two lower-dimensional systems. The system defined by H_1, which describes the movement of the midpoint, may be integrated trivially; however, the system that arises from H_2, describing the displacement vector, exhibits more complicated behaviour.

3.3 The Linear Component

Let (q_1, q_2) and (p_1, p_2) denote the components of Q_1 and P_1 respectively. For the system defined by $H_1(Q_1, P_1) = \frac{1}{4}\|P_1\|^2$, Hamilton's equations give

$$\dot{q}_i = \frac{1}{2}p_i, \qquad \dot{p}_i = 0; \tag{3.22}$$

for $i = 1, 2$. Each p_i is thus an invariant of the system, while

$$q_i(t) = q_i(0) + \frac{1}{2}p_i(0)\, t, \quad i = 1, 2 \tag{3.23}$$

for all $t \in \mathbb{R}^+$. Because each momentum is an invariant, the flow cannot be ergodic on the total phase space; however, it may still be ergodic when restricted to configuration space.

We begin by considering the ergodicity of an arbitrary discrete translation, $T_\gamma : \mathbb{T}^2 \to \mathbb{T}^2$, defined by

$$T_\gamma : (q_1, q_2) \mapsto (q_1 + \gamma_1, q_2 + \gamma_2),$$

for some $\gamma_1, \gamma_2 \in \mathbb{R}$. Recall that the numbers $\gamma_1, \gamma_2, \ldots, \gamma_n$ are called "rationally independent" if

$$\sum_{i=1}^{n} k_i\gamma_i = 0 \quad \Rightarrow \quad k_i = 0,\ i = 1, \ldots, n$$

for all n-tuples $(k_1, \ldots, k_n) \in \mathbb{Z}^{n+1}$.

Proposition 3 *If $\gamma_1, \gamma_2, 1$ are rationally independent, then the translation T_γ is ergodic with respect to Lebesgue measure. [KH, §4.2]*

Proof. Let $f : \mathbb{T}^2 \to \mathbb{R}$ be a bounded, measurable function that is T_γ-invariant. By the periodicity of our domain, we can write f as a Fourier expansion

$$f(q_1, q_2) = \sum_{(k_1,k_2)\in\mathbb{Z}^2} f_{k_1,k_2} \exp\left(2\pi i \sum_{j=1}^{2} k_j q_j\right) \tag{3.24}$$

for suitable $f_{k_1,k_2} \in \mathbb{R}$. Now consider

$$\begin{aligned} f(T(q_1, q_2)) &= \sum_{(k_1,k_2)\in\mathbb{Z}^2} f_{k_1,k_2} \exp\left(2\pi i \sum_{j=1}^{2} k_j (q_j + \gamma_j)\right) \\ &= \sum_{(k_1,k_2)\in\mathbb{Z}^2} f_{k_1,k_2} \exp\left(2\pi i \sum_{j=1}^{2} k_j q_j\right) \exp\left(2\pi i \sum_{j=1}^{2} k_j \gamma_j\right) \end{aligned}$$

But $f(T(q_1, q_2)) = f(q_1, q_2)$. By uniqueness of the Fourier expansion, we thus have

$$f_{k_1,k_2} \cdot \left(1 - \exp\ 2\pi i \sum_{j=1}^{2} k_j \gamma_j\right) = 0, \quad \forall k_1, k_2 \in \mathbb{Z}^2.$$

Unless $k_1 = k_2 = 0$, rational independence implies $k_1\gamma_1 + k_2\gamma_2 \notin \mathbb{Z}$, and hence $f_{k_1,k_2} = 0$. The f is therefore constant almost everywhere, and T_γ is ergodic. □

Will will now adapt this discrete result for a continuous map, using an approach given in [KH, §1.5] for their Proposition 1.5.1. The results are related as follows: for a given flow Φ on a manifold M, every Φ-invariant function will by definition be invariant under the discrete transformations $\Phi^t : M \to M$, although each of these discrete transformations may admit other invariant functions, which are potentially non-constant. On the other hand, if any discrete transformation $\Phi^t : M \to M$ is ergodic, the flow Φ must also be ergodic, since its invariant functions are a subset of the former.

For simplicity, we will again use co-ordinates $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$; there are then γ_1, γ_2 such that

$$T_\gamma^t : (q_1, q_2) \mapsto (q_1 + t\gamma_1, q_2 + t\gamma_2)$$

corresponds to the solution (3.23) in local co-ordinates.

Proposition 4 *If γ_1, γ_2 are rationally independent, then the translation T_γ^t is*

ergodic with respect to Lebesgue measure.

Proof. Let $f : \mathbb{T}^2 \to \mathbb{R}$ be a bounded, measurable function that is T_γ^t-invariant. Suppose $t \in \mathbb{R}$ is fixed. Again, we consider the Fourier expansion of f, given by (3.24), and use invariance to show

$$f_{k_1,k_2} \cdot \left(1 - \exp\ 2\pi i \sum_{j=1}^{2} k_j t \gamma_j\right) = 0, \quad \forall k_1, k_2 \in \mathbb{Z}^2.$$

We wish to find a value of t for which

$$t \cdot (k_1\gamma_1 + k_2\gamma_2) + k_3 = 0 \quad \Rightarrow \quad k_i = 0,\ i = 1, \ldots, 3; \tag{3.25}$$

then, we will have f constant almost everywhere, and T_γ^t ergodic as before. Since γ_1, γ_2 are rationally independent, for every triple $(k_1, k_2, k_3) \in \mathbb{Z}^3$, there exists precise one value of t for which $t\,(k_1\gamma_1 + k_2\gamma_2) + k_3 = 0$:

$$\tilde{t} = \frac{k_3}{k_1\gamma_1 + k_2\gamma_2}.$$

Finally, since there are only countably many such values, but uncountably many choices of t, we can choose a t such that (3.25) holds. □

For each γ_1 there are only countable many γ_2 for which $k_1\gamma_1 + k_2\gamma_2 = 0$; indeed, there is just one for each pair k_1, k_2. The set of all such γ_2 thus has Lebesgue measure zero over the line. Integrating over all choices of γ_1, we thus find that the set of all pairs of rationally dependent real numbers—i.e., those for which $k_1\gamma_1 + k_2\gamma_2 = 0$ for some $k_1, k_2 \in \mathbb{Z}$—has Lebesgue measure zero in the plane. Finally, since the correspondence between (γ_1, γ_2) and $(p_1(0), p_2(0))$ is linear and therefore measure preserving, we can say that flow arising from (3.20) is ergodic for almost all initial conditions.

This, however, does not solve the problem of ergodicity for the system (3.10). Suppose, for instance, that the initial conditions giving rise to ergodic behaviour in (3.23) also give rise to a flow Φ in the non-linear component which preserves some set A of partial measure (i.e., $0 < \mu(A) < 1$). Then the set $B := \mathbb{T}^2 \times A$ would have measure $\mu(B) = \mu(A)$ in the product space. Since B is an invariant set of the total flow, and $0 < \mu(B) < 1$, the product of these systems would not be ergodic. It follows that, to determine the ergodicity of (3.10), we will also need to determine the ergodicity of the non-linear component.

3.4 The Non-linear Component

As mentioned in the previous section, the ergodicity of the original Hamiltonian system (3.10) now depends largely on the non-linear system with Hamiltonian (3.21). Moreover, since we have dispensed with its linear counterpart (3.20), and the remainder of this paper deals with the latter system, the subscripts on P,Q and H will be dropped. Finally, by a linear rescaling of both $T\mathbb{T}^2$ and $U : \mathbb{R}^+ \to \mathbb{R}$, we can rewrite (3.21) as follows

$$H = \frac{1}{2}\|P\|^2 + U(d(0, Q)). \tag{3.26}$$

Now, suppose the torus has be endowed with its Cartesian representation as the periodic domain $\Omega = [-\alpha, \alpha] \times [-\beta, \beta]$, and let $Q = (x, y)$ and $P = (p_x, p_y)$. The Hamiltonian then gives rise to the differential equations

$$\frac{dQ}{dt} = \frac{\partial H}{\partial P} = P, \quad \frac{dP}{dt} = -\frac{\partial H}{\partial Q} = -\nabla_Q U(d(0, Q)) \tag{3.27}$$

which in turn correspond to the vector field $v_H : \mathbb{T}^2 \times \mathbb{R}^2 \to \mathbb{R}^4$ defined as follows:

$$v_H : \begin{bmatrix} x \\ y \\ p_x \\ p_y \end{bmatrix} \mapsto \begin{bmatrix} p_x \\ p_y \\ -\frac{\partial}{\partial x}U(d(0, Q)) \\ -\frac{\partial}{\partial y}U(d(0, Q)) \end{bmatrix}, \tag{3.28}$$

where the partial derivatives are taken with respect to x and y as variables, and then evaluated at the point in question. Simply knowing the vector field of the reduced system, one can make concrete assertions about the existence of symmetric trajectories. In particular, we can now show the existence of a continuum of trajectories that are each symmetric about a co-ordinate axis—each of which, if satisfying certain elementary conditions, will be periodic for a special choice of the geometry.

3.4.1 Existence of Periodic Orbits

The results in this section follow from certain discrete symmetries of the vector field (3.28). Let $\rho_y : (x, y, p_x, p_y) \mapsto (x, -y, -p_x, p_y)$ denote the component-wise reflection of the system in y and p_x, and note that the function $\frac{\partial}{\partial y}U(d(0, Q))$ is

odd in its dependence on y:

$$\frac{\partial}{\partial y}U(d(0,Q))\bigg|_{-y} = -\frac{\partial}{\partial y}U(d(0,Q))\bigg|_{y}.$$

Then we clearly have the reversibility relation

$$v_H \circ \rho_y = -\rho_y \circ v_H$$

everywhere, and the flow Φ_H^t arising from v_H satisfies

$$\rho_y \circ \Phi_H^t = \Phi_H^{-t} \circ \rho_y, \quad \forall t \in \mathbb{R}.$$

This reversibility immediately gives rise to a set of trajectories that are symmetric about $y = 0$ in configuration space:

Proposition 5 *If a trajectory of the system passes through*[1] either *the point* $(\tilde{x}, 0, 0, \tilde{p}_y)$ or *the point* $(\tilde{x}, \pm\beta, 0, \tilde{p}_y)$*, for any pair* $\tilde{x} \in \mathbb{S}^1$ *and* $\tilde{p}_y \in \mathbb{R}$*, then that trajectory, projected into the configuration space, will be symmetric about* $y = 0$*.*

Proof. Both sets of points are invariant under ρ_y. Consider any trajectory passing through one such point, and assume without loss of generality that this point corresponds to the initial condition (Q_0, P_0). For any such point,

$$\Phi_H^{-t}(Q_0, P_0) = \rho_y \circ \Phi_H^t(Q_0, P_0)$$

—i.e., reversibility of the flow amounts to symmetry of the orbit about $t = 0$. Thus, if $(x, y, p_x, p_y) = \Phi_H^t(Q_0, P_0)$ for some $t \in \mathbb{R}$, then $\Phi_H^{-t}(Q_0, P_0) = (x, -y, -p_x, p_y)$; in particular, the trajectory, extended in both temporal directions, is symmetric about $y = 0$ when projected onto the configuration space. □

Note that the quotient structure of the torus ensures that any curve symmetric about $y = 0$ is also symmetric about $y = \pm\beta$ and vice versa. Moreover, the vector field is symmetric with respect to the interchanging of (x, p_x) with (y, p_y); in particular, we have the reflection $\rho_x : (x, y, p_x, p_y) \mapsto (-x, y, p_x, -p_y)$, which

[1] In using the phrase "passes through," we rule out those trajectories for which $p_x = p_y = 0$ for all $t \in \mathbb{R}$.

means that the flow is also reversible under ρ_x with

$$\rho_x \circ \Phi_H^t = \Phi_H^{-t} \circ \rho_x, \quad \forall t \in \mathbb{R}.$$

Now, we could thus repeat the above proposition (as well as those derived below) using reversibility in x and the corresponding invariant points and axes—although such amendments would be merely notational. Of greater interest are the assertions this additional reversibility allows us to make concerning the existence of periodic orbits. Before stating this result, we note that the point $(\pm\alpha, \pm\beta)$ in configuration space—which corresponds to any of the corners of Ω, all of which have been identified in the construction of the torus—is invariant under both reflections.

Proposition 6 *Let $x \in \Omega$ and $p_y \in \mathbb{R}$. If a trajectory passing through either $(x, 0, 0, p_y)$ or $(x, \pm\beta, 0, p_y)$ in phase space also passes through any point corresponding to $(\pm\alpha, \pm\beta)$ in configuration space, then the trajectory will be periodic.*

Proof. Again, we assume without loss of generality that (Q_0, P_0) is one of the invariant points specified. Suppose further that this trajectory eventually passes through $(\pm\alpha, \pm\beta)$, taking on those values for some $t_0 \in \mathbb{R}$; i.e., $\Phi_H^{t_0}(Q_0, P_0) = (\pm\alpha, \pm\beta, \tilde{p}_x, \tilde{p}_y)$, for some $\tilde{p}_x, \tilde{p}_y \in \mathbb{R}$. Then, by the symmetry in y, we also have $\Phi_H^{-t_0}(P_0, Q_0) = (\pm\alpha, \pm\beta, -\tilde{p}_x, \tilde{p}_y)$. Together, these imply

$$\Phi_H^{2t_0}(\pm\alpha, \pm\beta, -\tilde{p}_x, \tilde{p}_y) = (\pm\alpha, \pm\beta, \tilde{p}_x, \tilde{p}_y), \tag{3.29}$$

and applying reversibility in ρ_x to this equation gives

$$\Phi_H^{-2t_0}(\pm\alpha, \pm\beta, -\tilde{p}_x, -\tilde{p}_y) = (\pm\alpha, \pm\beta, \tilde{p}_x, -\tilde{p}_y).$$

But Φ_H is a Hamiltonian flow, and is therefore reversibile under the mapping $\rho : (x, y, p_x, p_y) \mapsto (x, y, -p_x, -p_y)$ [HLW, §V1], which gives $\rho \circ \Phi_H^t = \Phi_H^{-t} \circ \rho$ everywhere. In particular, when applied to the previous equation, this implies $\Phi_H^{2t_0}(\pm\alpha, \pm\beta, \tilde{p}_x, \tilde{p}_y) = (\pm\alpha, \pm\beta, -\tilde{p}_x, \tilde{p}_y)$. Taken in tandem with equation (3.29), we thus obtain

$$\Phi_H^{4t_0}(\pm\alpha, \pm\beta, \tilde{p}_x, \tilde{p}_y) = \Phi_H^{2t_0}(\pm\alpha, \pm\beta, -\tilde{p}_x, \tilde{p}_y) = (\pm\alpha, \pm\beta, \tilde{p}_x, \tilde{p}_y),$$

and the trajectory is therefore periodic. □

This may not seem like a particularly useful result in itself; however, by fixing an initial condition $(\tilde{x}, 0, 0, \tilde{p}_y)$, and then altering the geometry accordingly, it is possible to adjust the "corners" of Ω so that a given trajectory will pass through them. In particular, we fix $\beta \in \mathbb{R}$ and consider the differential equation acting on the domain $\overline{\Omega} := \mathbb{R} \times [-\beta, \beta]$. Until the trajectory reaches $x = \pm\alpha$, its behaviour on Ω is indistinguishable from its behaviour on $\overline{\Omega}$; thus, we could determine when a trajectory intersects $y = \pm\beta$ in $\overline{\Omega}$ and choose an α as follows whenever it exists:

$$\alpha = \inf \left\{ |x(t)| \;\middle|\; y(t) = \pm\beta \text{ and } |x(t)| = \sup_{|s|<t} |x(s)| \right\},$$

where $x(t)$ and $y(t)$ are solutions in $\overline{\Omega}$ corresponding to the given initial condition. Note that the condition $|x(t)| = \sup_{|s|<t} |x(s)|$ is necessary to ensure that the trajectory on $\overline{\Omega}$ will indeed be identical with that on Ω until time t, at which point it will reach the prospective corner of some Ω. Note also that the infimum merely serves to make α unique; any element of the set on which it acts will produce a suitable geometry. Unfortunately, we cannot properly treat these matters without first computing solutions of the system. We give a more complete account of the existence and cardinality of such geometries in §5.3, by which point we will have presented the necessary machinery.

3.4.2 Local Quadratures

We will now attempt a local description of the trajectories arising from system (3.27). To do so, we will restrict ourselves to a set of local co-ordinates, where the global structure of the torus is lost. In particular, our co-ordinate system will be precisely that of Ω—although restricted to its interior, $\tilde{\Omega}$, to avoid any confusion between the previously identified edges. The system will, of course, leave this region in finite time; however, since the system is linear near the boundary of Ω, we can readily extend these trajectories some small distance on the torus, and then return to the local co-ordinate system.

Recall that, within these local co-ordinates, $d(0, Q) = \|Q\|$, the usual Euclidean 2-norm, and so U depends only on the distance between $(x, y) =: Q$ and $(0, 0)$. In this context, the system defined by H describes a point-mass with unit mass moving in a central potential field—that is, a field whose vectors are invariant to the group of motions on the plane that fix the origin [Arn2, §2.6 ff]. For

any such system, the angular momentum is a first integral. This can be made readily apparent in our particular case (and all others) by rewriting the planar system as a new Hamiltonian in polar co-ordinates. For position co-ordinates, we choose $q = (r, \theta)$ such that

$$\left.\begin{array}{rcl} x &=& r\cos\theta \\ y &=& r\sin\theta \end{array}\right\} \Rightarrow \left\{\begin{array}{l} p_x = \dot{x} = \dot{r}\cos\theta - r\dot{\theta}\sin\theta \\ p_y = \dot{y} = \dot{r}\sin\theta + r\dot{\theta}\cos\theta \end{array}\right. . \tag{3.30}$$

Now, the potential and kinetic energies of the system are independent of the co-ordinates chosen for the configuration space; thus, we have

$$T = \frac{1}{2}\|P\|^2 = \frac{1}{2}\left[(\dot{x})^2 + (\dot{y})^2\right] = \frac{1}{2}\left[(\dot{r})^2 + (r\dot{\theta})^2\right]$$

in terms of q and $\dot{q}$. This allows us to define the Lagrangian $L = L(q, \dot{q})$ of the system, given by $L = T - U$; the generalized momenta can then be recovered by the usual relation—

$$p = \frac{\partial L}{\partial \dot{q}}$$

—which gives $p = (\dot{r}, r^2\dot{\theta})$ for $q = (r, \theta)$. The Hamiltonian may now be written in terms of the components of p and q:

$$H = T + U = \frac{1}{2}\left[p_1^2 + \frac{p_2^2}{r^2}\right] + U(r). \tag{3.31}$$

Note that the total energy is independent of q_2. Finally, the dynamics of the system are given by Hamilton's equations as follows:

$$\begin{array}{rclcrcl} \dot{r} &=& p_1, & & \dot{p_1} &=& \frac{p_2^2}{r^3} - U'(r); \\ \dot{\theta} &=& \frac{p_2}{r^2}, & & \dot{p_2} &=& 0. \end{array} \tag{3.32}$$

The momentum p_2—which is traditionally denoted M, and known as the angular momentum—is clearly an invariant of the system. Moreover, r_1 depends only on p_1, and vice versa, since $p_2 =: M$ is constant. Thus, $q_1 = r$ is defined entirely by the equation

$$\ddot{r} = \frac{M^2}{r^3} - U'(r) =: f(r) \tag{3.33}$$

and appropriate initial conditions. In general, this differential equation cannot be solved explicitly for r as a function of t, but we can construct an implicit solution via quadratures. To see this, return to the Hamiltonian (3.31), keeping in mind

that $p_2 =: M$ is constant. Now, H is also constant along any trajectory; let us denote this value by $H =: E \in \mathbb{R}$. Independently, the equation $H = E$ determines a curve in $(r, \dot{r})$-space; since the Hamiltonian is quadratic in $p_1 = \dot{r}$, this curve is also symmetric about $\dot{r} = 0$. Provided these curves contain no stationary points, each will correspond to a trajectory of (3.32): for fixed $r \in \Omega$, equation (3.31) can then be solved for $\dot{r}$ as follows

$$\Rightarrow \dot{r} = \pm\sqrt{2\,(E - V(r))}. \tag{3.34}$$

where $V(r) = \frac{M^2}{2r^2} + U(r)$ can be seen as a potential energy, now that M is fixed. The shape of each level-curve, and the nature of the corresponding trajectory, will be determined (in part) by the set of r values for which $V(r) = E$.

If U is a potential with repelling core—that is, it satisfies equations (3.5–3.7)—then V will also have a repelling core, since $r \mapsto \frac{M^2}{2r^2}$ is (strictly) decreasing. In particular, V is also strictly decreasing on I, and one-to-one between I and $V(I) =: J \subset \mathbb{R}$. We now restrict ourselves to values of $E \in J$, noting that J is in fact an interval of positive measure. For any such E, there exists a unique $\tilde{r} \in \mathbb{R}$ such that $V(\tilde{r}) = E$, and this $\tilde{r}$ must be an element of I. We then have $V(r) > E$ for all $r \in [0, \tilde{r})$, since V is decreasing on some $[0, \xi)$ containing $\tilde{r}$; we also have $V(r) < E$ for all $r \in (\tilde{r}, \infty)$, since V is decreasing on $[0, \xi)$ and

$$\inf_{r\in[0,\xi)} V(r) \geq \sup_{r\in[\xi,\infty)} V(r).$$

Finally, we can guarantee that, for such H, the trajectory along the level-curve does not contain an equalibrium point, since

$$\dot{r} = 0 \;\Rightarrow\; V(r) = E \;\Rightarrow\; r = \tilde{r} \;\Rightarrow\; f(r) = V'(\tilde{r}) < 0,$$

since V is strictly decreasing on I. The trajectory can then be extended indefinitely in time [Arn1, §12.5], during which it will traverse the entire level-curve, with the half for which $\dot{r} < 0$ being traversed first with respect to time.

Equation (3.34) is, of course, separable, and can be integrated to yield another quadrature:

$$t = \int_0^t d\zeta = \pm \int_{r_0}^{r} \frac{d\xi}{\sqrt{2\,(E - V(\xi))}}.$$

Assuming $t \in \mathbb{R}^+$, only one of these solutions is admissible; if r is initially decreasing—which will (or, rather, *can*) be the case whenever we have an ini-

tial condition within U's compact support—then we take the negative root. This quadrature then only makes sense only until r takes on the critical value $r = \tilde{r}$, after which the integrand is undefined, and this will happen at some finite time $t = \tilde{t}$, since the trajectory does not contain an equilibrium; until that point, it defines t explicitly in terms of r. Moreover, since the integrand is always positive, the relation between t and r is one-to-one, and so r is also implicitly defined on $(-\infty, \tilde{t})$ as a function of t. Finally, we can extend our solution beyond this point using the symmetry relation

$$r(\tilde{t} + \tau) = r(\tilde{t} - \tau), \quad \forall \tau > 0,$$

which arises from the reversibility of the system (3.32) under $\rho : (r, \dot{r}) \mapsto (r, -\dot{r})$, and the ρ-invariance of $(r, \dot{r})$ at $t = \tilde{t}$.

Having determined r implicitly for all t, we can now solve for the other spatial co-ordinate in our polar equations, $q_2 = \theta$. Noting that $\dot{\theta} = M/r^2$, and by the chain rule

$$\frac{d\theta}{dt} = \frac{d\theta}{dr} \cdot \frac{dr}{dt},$$

we have for $t < \tilde{t}$

$$\theta - \theta_0 = -\int_{r_0}^{r} \frac{M}{\xi^2 \sqrt{2(E - V(\xi))}}\, d\xi, \tag{3.35}$$

which defined θ explicitly as a function of r (and thus t) that is again one-to-one. If we define $\tilde{\theta} = \theta(\tilde{r}) = \theta(r(\tilde{t}))$, then we can also, much as before, extend this result via a symmetry relation:

$$r(\tilde{\theta} + \phi) = r(\tilde{\theta} - \phi), \quad \forall \phi > 0.$$

Thus we have solved the Hamiltonian system arising from H in the local co-ordinates.

3.4.3 Integrability and the Global Geometry

Unfortunately, the local solutions from the previous section tell us little about the global dynamics of the system. As mentioned earlier, the original system is not integrable: when a trajectory moves between two charts, it generally experiences a discrete jump between distinct angular momenta, and as such the various invariants utilized in the previous section do not exist for a system properly defined on the torus.

To see this, take $M = r^2\dot{\theta}$ as determined in the previous section, and rewrite that expression in Cartesian co-ordinates via (3.30), yielding

$$M = xp_y - yp_x. \tag{3.36}$$

Now, consider a trajectory that exits the local co-ordinate system Ω through the line $y = \beta$, and "reappears" on the line $y = -\beta$ in an identical co-ordinate system. All other Cartesian co-ordinates remain unchanged in this scenario; thus, if M is the angular momentum before the change of co-ordinates and M' is the momentum after, we simultaneously have

$$\begin{cases} M &= xp_y - \beta p_x \\ M' &= xp_y + \beta p_x \end{cases}$$

Subtracting the former from the latter then gives the following relation:

$$M' = M + 2\beta p_x. \tag{3.37}$$

Thus, the two momenta are equal if and only if $p_x = 0$, and there are certainly many trajectories for which the latter does not hold at the boundary. In a more general system, this would only show that our representation of the angular momentum is dependent on the choice of local co-ordinates—and that is hardly a surprising result. But our representation of M is valid almost everywhere on $\mathbb{T}^2$. If M is to be a global integral on the 2-torus, there must be a way of identifying the different local values of M that correspond to the same global trajectory. And yet this correspondence is given by equation (3.37), and there is no obvious change of co-ordinates in which this relation is made obsolete.

Example. For the sake of comparison, let us briefly describe a manifold for which the angular momentum of this system is globally preserved. Consider the 2-sphere embedded in Euclidean space as $\mathbb{S}^2 = \{(x, y, z) \mid x^2+y^2+z^2 = 1\}$, and suppose the scatterer comprises $\mathbb{S}^2 \cap \{(x, y, z) \mid z < R\}$ for some $R \in [-1, 0]$. Suppose further that the potential energy U depends only on $\|(x, y, z)\|$, and define $H = T + U$ as before; the resulting vector field $v_H : T\mathbb{S}^2 \to T(T\mathbb{S}^2)$ is invariant under rotations about the z-axis. By Noether's Theorem, these rotations correspond to an invariant whose Hamiltonian vector field generates them. But, defined as a function on

$T\mathbb{S}^2$ so embedded, $M = xp_y - yp_x$ gives rise to the vector field

$$\dot{q} = \begin{bmatrix} -y \\ x \\ 0 \end{bmatrix}, \qquad \dot{p} = \begin{bmatrix} -p_y \\ p_x \\ 0 \end{bmatrix}$$

which generates precisely these rotations. It follows that M is an integral of the system; moreover, since the system has only four dimensions, the first integral M and the Hamiltonian function (which is itself a first integral) are sufficient to solve the system by quadratures. The Hamiltonian thus gives rise to an integrable system on the sphere, and such a system is trivially non-ergodic. $\diamond$

Returning to $\mathbb{T}^2$, one may intuit that the geometry of the torus is incompatible with the symmetry about the scatterer. However, without choosing a particular embedding of the torus, and then a particular representation of M, this is not quite obvious. For our Cartesian representation, it suffices to consider a trajectory tangent to the scatterer with $p_x = 0$, passing through the point $(R, 0)$ in configuration space. Such a trajectory is contained entirely in a set $\{x\} \times \mathbb{S}^1 \subset \mathbb{T}^2$, where $x = R$ in local co-ordinates, and never enters the non-linear region. But rotating that point about the scatter, within a given chart, produces a trajectory which will eventually enter the non-linear region, either in the next set of local co-ordinates or, when extended backwards in time, the previous. Moreover, in order to prevent such an occurrence, one would need to introduce a non-linearity outside the scatterer. Thus, the underlying vector field can have no such symmetries, and the angular momentum is not a global invariant of the system. Finally, in relation to equation (3.37), we note that under those rotations valid in a neighbourhood of the scatter, the value of M corresponding to the "vertical" trajectory above would remain constant within the chart, but give rise to a continuum of M' values. Hence, we could also dispense with the idea that M might be preserved under a different representation.

Chapter 4

The Collision Map

In order to ascertain the global behaviour of system (3.27), it is necessary to examine the way the local solutions from §3.4.2 link-together. This task, at its most general, might be accomplished via a Poincaré map. If we identify the torus with the rectangular domain Ω that gave rise to it, and let S denote the set of initial conditions with $q \in \partial\Omega$ and p inward-pointing, then it would be enough to consider a mapping between the points of S and the exit coordinates of their corresponding trajectories. Then, by identifying opposing edges as before, we obtain a self-map $\phi : S \to S$, whose iterates $\{\phi^n x\}_{n\in\mathbb{Z}}$ define an orbit for each $x \in S$. This map is a particular case of the more general construction known as a Poincaré map, which allows the dynamics of a continuous system to be modelled using a lower-dimensional discrete one.

Definition 11 *Let Φ^t be a flow acting on a phase space $X \subset \mathbb{R}^n$, and let $\Sigma \subset X$ be a submanifold of co-dimension one such that each trajectory of Φ^t intersects Σ laterally.*[1] *Then the map $\phi : \Sigma \to \Sigma$ determined by*

$$\phi(x) = \min_{t>0}\{t \mid \Phi^t x \in \Sigma\}, \quad \forall x \in \Sigma$$

is know as the Poincaré map *or* first-return map *defined by Σ, arising from the flow Φ^t.*

In our example, the surface in question was $\Sigma = S$. Note that each trajectory $\{\Phi^t x_0\}_{t\in\mathbb{R}}$ of the original flow Φ^t gives rise to an orbit of ϕ; moreover, since the equation is autonomous and reversible, each orbit of ϕ corresponds to a specific

[1] That is, for all $x_0 \in X$, $\{\Phi^t x_0\}_{t\in\mathbb{R}}$ is never tangent to Σ. Since our flow is at least once differentiable, this ensures that $\{\Phi^t x_0\}_{t\in\mathbb{R}} \cap \Sigma$ is countable for each x_0.

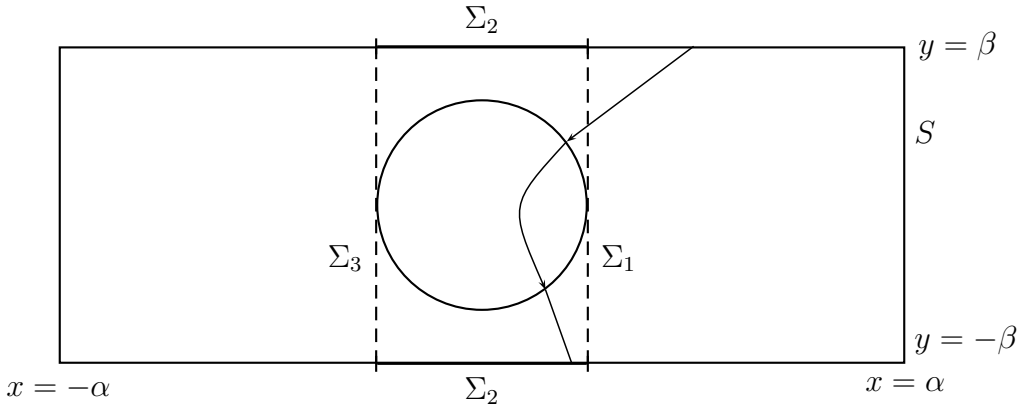

Figure 4.1: The Poincaré sections and a trajectory, as they appear in local coordinates.

trajectory of Φ^2—obtained by integrating, in both temporal directions, any one of its elements. It is easy to see that every periodic trajectory of Φ thus gives rise to a periodic orbit of ϕ and vice versa. It can also be shown that, for a smooth dynamical system, stability in the discrete map implies stability in the original flow; and thus *if* we are able to find stable periodic orbits of ϕ, we will have also found stable periodic trajectories of system (3.27).

If the analysis of the above Poincaré map could be carried to fulfillment, it would allow for a complete classification of the system's dynamics; however, by necessity, complicated behaviour in our flow gives rise to complicated behaviour in the Poincaré map. Instead, it is necessary to restrict our attention to a particular class of trajectories as they interface with a subsection of the original Poincaré surface. For instance, one set of interest comprises those trajectories for which $p_y < 0$, and for which the ratio $|\frac{p_x}{p_y}|$ is sufficiently small. For such trajectories, a initial series of collisions with the scatterer will result in a kind of "deflection"—after which p_x has changed sign; p_y remains negative; and $|\frac{p_x}{p_y}|$ is still small, although potentially larger than before. Ignoring the mechanics of the collisions themselves, such trajectories could be approached as a sequence of distinct momentum-states, so long as the ratio $|\frac{p_x}{p_y}|$ remains small enough to ensure further deflections (see figure 4.2).

To a similar end, we introduce three "partial" Poincaré sections, and define maps between them which, when properly composed, will produce a map corre-

[2]That is, specific up to translations in time.

sponding to an actual first-return map of the flow for certain trajectories.[3] The sections, as they occur in position-space, are defined as follows (see figure 4.1):

$$\begin{aligned}\Sigma_1 &:= \{(p,q) \mid q_x = r\};\\ \Sigma_2 &:= \{(p,q) \mid q_y = \pm\beta, -r < q_x < r\};\\ \Sigma_3 &:= \{(p,q) \mid q_x = -r\}.\end{aligned}$$

The first such surface, Σ_1, is tangent to the scatterer, and intersects trajectories that approach the scatterer from the right or are repelled to the right. The second, Σ_2, creates a boundary between successive collisions with the scatterer, and also operates as a boundary between successive solutions in local co-ordinates. The third surface operates to the same end as Σ_1, but intersects those trajectories which approach from the left or are repelled to the left. The region between Σ_1 and Σ_3, including the scatterer but excluding Σ_2, will be referred to as the collision region—

$$R_c := \{(p,q) \mid |q_x| < r,\ q_y \neq \pm\beta\}.$$

The remainder of the position space, excluding the surfaces themselves, will be referred to as the linear region—

$$R_l := \{(p,q) \,|\, |q_x| > r\}$$

—because here the point's motion is always linear in the local co-ordinates.

In this context, a *collision* is said to occur when a trajectory enters the collision region through any surface and subsequently exits through any other; the term *deflection* is reserved for trajectories which exit the linear region through Σ_1 or Σ_3 and re-enter the linear region through the same surface. (Note that the trajectory may intersect with Σ_2 multiple times in the interim.) Deflections to the right are thus described by a mapping $\phi_r : \Sigma_1 \to \Sigma_1$, while deflections to the left are described by some $\phi_l : \Sigma_3 \to \Sigma_3$. Collisions, on the other hand, may begin and end on any of the three surfaces. In this way, deflections comprise a simpler case than collisions—although each deflection comprises a certain combinations of collisions, and thus it is necessary to study at least a restriction of the possible collision maps before discussing possible deflections. The remainder of this chapter enacts such a study. We return to the "actual" first-return map in Chapter 5.

[3] In particular, the map corresponding to the surface $\{(p,q) \mid p_x > 0, q_x = r\}$

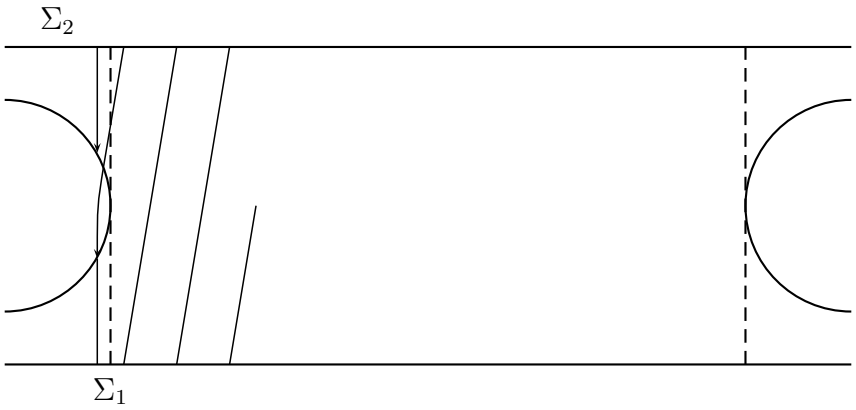

Figure 4.2: A trajectory leading to a "deflection"—as it appears in local coordinates.

4.1 Constructing the Map

Each of the Poincaré sections introduced in the previous section is a one dimensional subset of the configuration space, taken such that one spatial co-ordinate is fixed while the other varies along its length. The domain and range of the collision map, which always correspond to one of these surfaces, are thus of one dimension less than the total phase space. This, however, leaves three degrees of freedom in both sets, which makes the analysis rather unwieldy. To proceed, we limit the flow to a particular energy surface, which (as we have seen) reduces by one spatial dimension the set of states which the system may attain.

With that in mind, consider the quadrature from the previous section, given by equation (3.35)—

$$\theta - \theta_0 = -\int_{r_0}^{r} \frac{M}{\xi^2\sqrt{2(E - V(\xi))}}\, d\xi$$

—which gives the trajectory locally in position space as a curve in r and θ, for fixed values of E and M. Since the trajectory is linear outside the scatter itself, the curve will intersect the boundary ∂R_c at two points; moreover, the direction of the flow along the curve is obvious from the initial conditions. We thus have "in hand" the Poincaré-like map associating an initial position $q = (r, \theta) \in \partial R_c$ and inward-pointing momentum $p = (\dot{r}, M)$ with the corresponding trajectory's exit co-ordinate $q' \in \partial R_c$ and the momentum on exit p'. But we have only one degree of freedom in q, and $\dot{r}$ is determined by q, E and M; thus, for fixed E, every trajectory passing through a given section can be identified entirely by M and one co-ordinate of q.

The quadrature thus allows us to find exit values corresponding to each initial

condition, but within the transformed state space from the previous section: both the Poincaré sections and Ω itself are more naturally formulated in Cartesian co-ordinates; moreover, when a trajectory passes through $\partial\Omega$—and into a new local-coordinate system—both momenta and one q co-ordinate are unchanged when written in Cartesian form, while the remaining q co-ordinate merely changes sign; in polar form, the angular momentum undergoes the change described by equation (3.37), which itself requires an expression for p_x. In short, it is convenient to switch between co-ordinate systems near $\partial\Omega$. Here, in addition to the usual formulae, we have $U(q) = 0$, and thus $E = \frac{1}{2}\left(p_x^2 + p_y^2\right)$. We also have $M = xp_y - yp_x$ in Cartesian co-ordinates. This gives us, in total, the following transformations at the boundary—

$$\left.\begin{array}{rcl} r^2 &=& x^2 + y^2 \\ \theta &=& \arctan\left(\frac{y}{x}\right) \end{array}\right\} \quad \Leftrightarrow \quad \left\{\begin{array}{rcl} x &=& r\cos\theta \\ y &=& r\sin\theta \end{array}\right. ;$$

$$E = \frac{1}{2}\left(p_x^2 + p_y^2\right) \quad \Leftrightarrow \quad p_y^2 = 2E - p_x^2,$$

$$M = xp_y - yp_x.$$

The leftmost set of equations is a well-defined transformation from (x, y, p_x, p_y) to (r, θ, M, E), which is one-to-one when restricted to the configuration space. Is is not, however, one-to-one in the momenta: each pair (M, E) corresponds to two, potentially different values of (p_x, p_y).[4] However, for our particular system, $(p_x, p_y) = (\dot{x}, \dot{y})$, and so the following procedure will suffice: whenever we have a solution curve in (r, θ), we can first transform the curve into (x, y)-space; then, knowing the total kinetic energy, and the direction of the tangent vector to the curve in configuration space, we can determine the unique, corresponding momentum vector.

Armed with these transformations, we can return to the problem in Cartesian co-ordinates: Suppose we are given an initial condition (p, q) on the boundary of the collision region, $\partial R_c = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$. Suppose further that p is inward-pointing, and $p_y \neq 0$. Now, as in §3.4.1, we fix β and employ the extension $\overline{\Omega} = \mathbb{R} \times [-\beta, \beta]$. If q is contained in Σ_2, then q is contained in its extension, $\overline{\Sigma}_2 := \{q \mid q_y = \pm\beta\}$; if $q \in \Sigma_1 \cup \Sigma_3$, then the trajectory may be extended backward in time—into the linear region—until it intersects $\overline{\Sigma}_2$. Hence, by assuming $p_y \neq 0$, we have without

[4] Noting that x and y can be determined independent of M, equation (3.36) defines a line in the space of points (p_x, p_y), while $2E = \frac{1}{2}\left(p_x^2 + p_y^2\right)$ defines a circle. They are guaranteed to have at least one intersection, but almost all pairs (M, E) correspond to two distinct points in the space of Cartesian momenta. In §A.1, we derive these points explicitly.

further loss of generality $q \in \overline{\Sigma}_2$, at least initially. Moreover, since the problem is symmetric about both local co-ordinate axes, it is enough to consider the case where $q_x \geq 0$ and $p_x, p_y \leq 0$; i.e., when the trajectory begin in the first quadrant. The domain of our map, in Cartesian co-ordinates, is thus the following—

$$\overline{\Sigma}_2' := \{(p,q) \mid q_y = \beta,\ q_x > 0;\ p_x, p_y \leq 0\} \tag{4.1}$$

—and the range comprises, for the moment, all (p,q) for which $q \in \partial\Omega$ and p is outward pointing. Beyond these simplifications, the set of trajectories we will consider is governed by one further assumption:

Assumption. Every trajectory satisfies $p_y < 0$ on both its entry into the local chart structure and its exit therefrom; that is, the trajectories begin on $\overline{\Sigma}_2'$ with $p_y < 0$, and subsequently cross $\overline{\Sigma}_2'$ in the same direction.

This assumption holds for the "deflecting" trajectories mentioned earlier, but allows for non-negative p_y values within the non-linear region. Unfortunately, we cannot say *a priori* which points in our domain satisfy this condition; in addition, a trajectory which satisfies the assumption within a given chart—and is thus aptly described by our map—will not necessarily satisfy the assumption in a subsequent chart. Fortunately, we will readily ascertain when the map is valid once its construction is completed.

In constructing the map, it is convenient to replace the Cartesian momenta (p_x, p_y) with the invariants (M, E) from the polar representation, while continuing to use position co-ordinates (x, y). Since E is fixed in advance, and y is fixed by the surface $\overline{\Sigma}_2$, this gives us (effectively) a map $\varphi : (x, M) \mapsto (x', M')$. Each pair (x, M) uniquely specifies a trajectory in phase space, which can be followed until it leaves the current chart; we use M (as opposed to p_x) because it is constant within a given chart, and because it specifies certain geometric properties of the system, as will be examined below. In these terms the domain of φ, as given by (4.1), is as follows:

$$\overline{\Sigma}_2' := \{(x, M) \mid x > 0, M \leq 0\}. \tag{4.2}$$

Finally, we are ready to derive the components of this map. This proceeds in two stages: first, since M is constant within a given chart, we determine the component $\varphi_1 : (x, M) \mapsto x'$; then, since the change in M depends on p_x—which in turn depends the solution curve in some neighbourhood of x'—we determine $\varphi_2 : (x, M) \mapsto M'$.

4.1.1 The x-Component

To construct the first component, we begin by treating the case $U \equiv 0$, for which the system is linear. Assume further that we are solving the system within a given chart, and so M is considered constant. Here, $V(r) = \frac{M^2}{2r^2}$, and the quadrature (3.35) has the form

$$\theta - \theta_0 = -\int_{r_0}^{r} \frac{M}{\xi^2\sqrt{2E - \frac{M^2}{\xi^2}}}\, d\xi.$$

This can be readily solved by assuming each $\xi = \frac{M}{\sqrt{2E}} \sec\vartheta$, for some $\vartheta \in \mathbb{R}$, which is true provided

$$r \geq V^{-1}(E) = \frac{|M|}{\sqrt{2E}} =: \hat{r}.$$

But, as we saw in the previous chapter, $E - V(r) \geq 0$.[5] The solution then has the form

$$\theta - \theta_0 = \operatorname{arcsec}\left(\frac{\xi}{\hat{r}}\right)\Bigg|_{r_0}^{r},$$

which is valid whenever $r > \hat{r}$. Now, if we choose $r_0 = \hat{r}$, we find

$$r = \hat{r}\sec(\theta - \theta_0), \tag{4.3}$$

where θ_0 is achieved when $r = \hat{r}$. Of course this is merely a line, but each such line is of distance $\hat{r}$ from the origin, and thus solutions comprise the family of lines tangent to the circle $\{(r, \theta) \mid r = \hat{r}\}$. What explicitly differentiates solutions is θ_0; that is, how far each line is rotated from the vertical line given by (4.3) when $\theta_0 = 0$.

Each value of θ_0 corresponds to a particular trajectory, and each such trajectory is determined by x. Assuming a trajectory passes through the Cartesian point (x, y), we have through obvious trigonometric manipulations

$$\hat{r} = x\cos\theta_0 + y\sin\theta_0. \tag{4.4}$$

Substituting $y = \beta$ and solving for θ_0 (as detailed in §A.2) gives the following

$$\theta_0 = \arccos\left[\frac{x\hat{r} + \beta\sqrt{x^2 + \beta^2 - \hat{r}^2}}{x^2 + \beta^2}\right]. \tag{4.5}$$

[5] One can also show (algebraically) that $M^2 \leq 2Er^2$ for any trajectory outside the scatterer. [See §A.1.]

Then, using equation (4.4) with $x = x'$ and $y = -\beta$, we can easily find an expression for x' in terms of θ_0:

$$x' = \frac{\hat{r} + \beta \sin\theta_0}{\cos\theta_0}. \tag{4.6}$$

Finally, substituting equation (4.5) into the latter yields, after manipulations presented in §A.2,

$$x' = x + \frac{2\beta(\hat{r}^2 - x^2)}{\beta x + \hat{r}\sqrt{x^2 + \beta^2 - \hat{r}^2}}$$

—which, as we would expect, has a fixed point if and only if $x = \hat{r}$ (which also gives $\theta_0 = 0$).

Now consider the non-linear case. Here we do not assume that U is everywhere zero, but we still have $U \equiv 0$ outside the scatter—that is, on $\{(r, \theta) \mid r \geq R\}$. Since the non-linear region has compact support and is also convex, and since the linear-region comprises the remainder of each chart, any trajectory that leaves the scatterer must exit the current chart before entering the scatterer again. This, taken together with the reversibility of the Hamiltonian, ensures that each trajectory corresponding to some $x \in \overline{\Sigma}_2$, when restricted to a given chart, is such that: (i) it is both initially and finally linear; (ii) it is non-linear only within $\{(r, \theta) \mid r < R\}$; and (iii) those non-linearities occur within a single continuous interval in time. We thus divide the local trajectory $\Phi^t x$ into three parts $\ell_1 \cup c_s \cup \ell_2 = \Phi^t x$ where: ℓ_1 comprises the initial linear region; ℓ_2 comprises the final linear region; and $c_s = \Phi^t x \cap \{(r, \theta) \mid r \geq R\}$ covers the trajectory within the scatter. In all cases, we are principally concerned with these objects as subsets of the configuration space, and bracket their dependence on t, short of the ordering it entails.

Since M is constant within each chart, both ℓ_1 and ℓ_2 are described by equation (4.3), where $\hat{r} = \frac{M}{\sqrt{2E}}$ as before, but with potentially different values of θ_0; for further reference, these values will be denotes θ_1 and θ_2 respectively. The former can be found just as in the linear case:

$$\theta_1 = \arccos\left[\frac{x\hat{r} + \beta\sqrt{x^2 + \beta^2 - \hat{r}^2}}{x^2 + \beta^2}\right]. \tag{4.7}$$

The latter, of course, requires some additional work. Let $\overline{\ell}_1$ and $\overline{\ell}_2$ denote the lines obtained by extending the segments ℓ_1 and ℓ_2; let C_L denote a circle centred at the origin with radius $L \in \mathbb{R}$; and let D denote the origin. Note that each extension is

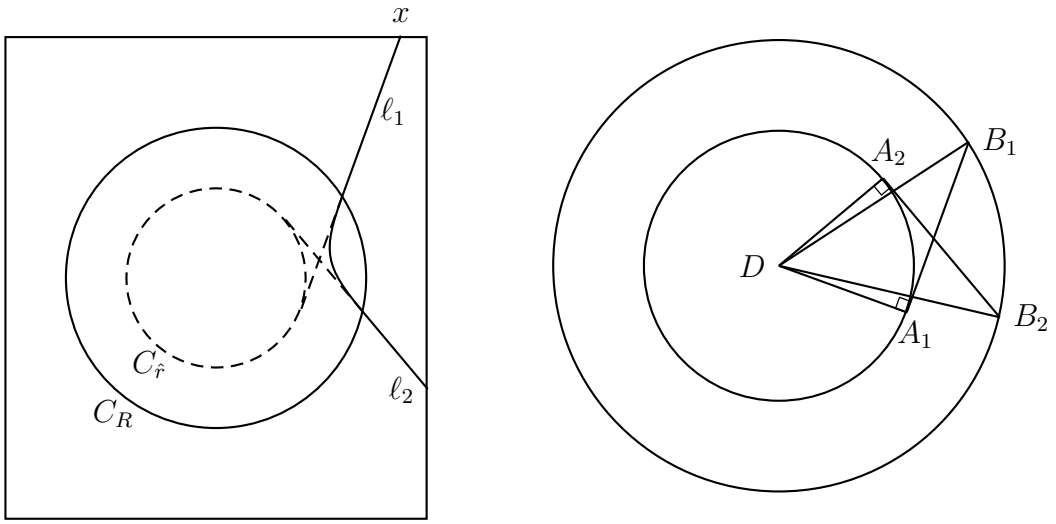

Figure 4.3: The geometry of $\Phi^t x$; its components, ℓ_1, ℓ_2 and c_s; and the extensions $\overline{\ell}_1$ and $\overline{\ell}_2$—both within a given chart (left), and inside the scatterer (right).

secant to the circle C_R, and tangent to the circle $C_{\hat{r}}$. We now consider the points of intersection denoted as follows:

$$A_1 := \overline{\ell}_1 \cap C_{\hat{r}} \qquad B_1 := \ell_1 \cap C_R$$
$$A_2 := \overline{\ell}_2 \cap C_{\hat{r}} \qquad B_2 := \ell_2 \cap C_R$$

—noting that, to define each B_i, we use ℓ_i and not $\overline{\ell}_i$ to prevent ambiguity. (See figure 4.3.) Letting $\theta(\cdot)$ denotes the projection $(r, \theta) \mapsto \theta$, we see that $\theta_1 = \theta(A_1)$ and $\theta_2 = \theta(A_2)$. Noting that $\triangle DA_1B_1$ and $\triangle DA_2B_2$ are congruent right triangles, we have

$$\angle A_1DB_1 = \angle A_2DB_2 = \arccos\left(\frac{\hat{r}}{R}\right),$$

The same geometric construction relates each $\theta(A_i)$ to the corresponding $\theta(B_i)$, as depicted in figure 4.3, although both the construction and the resulting relations depend on the sign of M :

$$\begin{aligned}\theta(A_1) &= \theta(B_1) + \operatorname{sgn}(M) \cdot \arccos\left(\frac{\hat{r}}{R}\right); \\ \theta(A_2) &= \theta(B_2) - \operatorname{sgn}(M) \cdot \arccos\left(\frac{\hat{r}}{R}\right)\end{aligned}$$

We note in passing that these equations could also be obtained using the solution (4.5) with appropriate initial conditions. Now, $\theta(B_1)$ and $\theta(B_2)$ are related by the

results from §3.4.2: their difference is simply the net change in the trajectory's θ co-ordinate during the period of non-linearity, which can be calculated using the quadrature (3.35) and the symmetry about $\theta = \tilde{\theta}$. We thus have

$$\theta(B_1) - \theta(B_2) = 2[\theta(B_1) - \tilde{\theta}] = 2\int_{\tilde{r}}^{R} \frac{M}{\xi^2\sqrt{2(E - V(\xi))}}\, d\xi,$$

When all these relations are combined, we are left with the following expression for θ_2, in terms of θ_1 and M

$$\theta_2 = \theta_1 - 2\,\mathrm{sgn}(M)\arccos\left(\frac{\hat{r}}{R}\right) - 2\int_{\tilde{r}}^{R} \frac{M}{\xi^2\sqrt{2(E - V(\xi))}}\, d\xi, \tag{4.8}$$

where everything on the right-hand side is known. Finally, to make this expression consistent when the point-mass doesn't interact with the scatterer, we adopt the following conventions: $\arccos(z) = 0$ when $z > 1$, and the above integral is zero whenever $\hat{r} > R$.

To determine x', all that remains is to find the intersection of ℓ_2 with the line $y = -\beta$. Again, this follows as in the linear case, with

$$x' = \frac{\hat{r} + \beta\sin\theta_2}{\cos\theta_2}. \tag{4.9}$$

Since we cannot simplify the integral in equation (4.8), there is little point in expressing x' explicitly in terms of x and M—although equations (4.7), (4.8) and (4.9) in tandem would allow us to do so.

4.1.2 The M-Component

Most of the work required for the calculation of M' has been carried out in preceeding sections. As given by equation (3.37),

$$M' = M + 2\beta p_x.$$

Moreover, as noted earlier, p_x can be found for fixed E using the solution curve in configuration space that contains x'; in particular, this information is given by ℓ_2. Since the curve satisfies

$$\hat{r} = x\cos\theta_2 + y\sin\theta_2$$

as it approaches the boundary, we must also have

$$p_x \cos\theta_2 + p_y \sin\theta_2 = 0, \tag{4.10}$$

since $p_x = \dot{x}$ and $p_y = \dot{y}$. In addition, by considering equation (3.36) and taking the limit as the trajectory approaches the boundary, we find

$$x' p_y + \beta p_x = M.$$

Solving for p_x then gives

$$p_x = \frac{M \sin\theta_2}{\beta \sin\theta_2 - x' \cos\theta_2} = -\mathrm{sgn}(M) \cdot \sqrt{2E} \sin\theta_2, \tag{4.11}$$

where the last expression follows from equation (4.9) and the definition of $\hat{r}$. This leaves us with

$$M' = M - \mathrm{sgn}(M) \cdot \sqrt{2E} \sin\theta_2, . \tag{4.12}$$

Insofar as this expression depends on θ_2, it is only marginally more endearing than the series of expressions that generate x'. Fortunately, as its form would indicate, the expression for M' readily lends itself to a fixed point analysis.

4.2 Fixed Points of the Components

Rather than analyse the collision map $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ in its entirety, it is helpful to treat each component individually with one of its variables fixed. Thus, we instead consider the one-dimensional maps $\varphi_1(\,\cdot\,, M) : x \mapsto x'$ and $\varphi_2(x, \,\cdot\,) : M \mapsto M'$, which better facilitate analysis. In particular, we consider the fixed points of both maps, which are directly linked to the existence of symmetric orbits as hypothesized in §3.4.1. However, before we can bring the components of the map to bear on the underlying system, we must determine when the assumption made in the previous section is valid.

We had already assumed, without loss of generality, that $p_y > 0$ in the initial condition. To determine when $p_y > 0$ on exit, we again rely on the information implicit in ℓ_2; namely, equation (4.10). Using the solution for p_x given in equation (4.11), we have

$$p_y = \mathrm{sgn}(M) \cdot \sqrt{2E} \cos\theta_2, \tag{4.13}$$

although this expression is not valid when $M = \hat{r} = 0$. We thus have $p_y < 0$

whenever one of the following conditions holds:

$$\theta_2 \bmod 2\pi \in (-\frac{\pi}{2}, \frac{\pi}{2}) \quad \text{and} \quad M < 0; \tag{4.14}$$

$$\text{or} \qquad \theta_2 \bmod 2\pi \in (\frac{\pi}{2}, \frac{3\pi}{2}) \quad \text{and} \quad M > 0. \tag{4.15}$$

For trajectories beginning in the first quadrant of a given chart, the first of these is of principle interest when the potential is essentially repelling. Thus, while either condition is sufficient for validity of the map, the former will be used more frequently in the subsequent analysis.

We can now return to the question of symmetric orbits:

Proposition 7 *Both the fixed points of $\varphi_1(\,\cdot\,, M)$, and those of $\varphi_2(x, \,\cdot\,)$, give rise to trajectories symmetric about $y = 0$ in the configuration space of system (3.27), provided the corresponding pair (x, M) satisfies one of the conditions (4.14) and (4.15).*

Proof. Both conditions, in addition to validating the maps, imply the corresponding trajectory passes through $(x, -\beta)$ in the local co-ordinate system. In the first case, fix M and suppose $\varphi_1(x, M) = x$. Consider the linear components of the trajectory ℓ_1 and ℓ_2 as defined above: by virtue of the fixed point, these contain the points $(x, \beta) =: P_1$ and $(x, -\beta) := P_2$ respectively. Assume that the trajectory enters the non-linear region; otherwise, the trajectory is simply P_1P_2 and we are done. Recall from §3.4.2 that the trajectory is symmetric about some time $t = \tilde{t}$, and thus each region of linearity is a reflection of the other (when properly extended). Moreover, the map $t \mapsto r$ is strictly monotone—and thus injective—in those regions. Noting that $\|P_1\| = \|P_2\|$, it immediately follows that these points are reflections of one another about the critical point in time. Now, by the corresponding symmetry in θ, we have $\theta(P_1) = -\theta(P_2)$; in particular, integrating away from the critical time in both directions gives

$$\tilde{\theta} = \frac{1}{2}\left(\theta(P_1) + \theta(P_2)\right) = 0.$$

Recalling that $\theta = \tilde{\theta}$ is the line of symmetry in configuration space, we have proven the trajectory is symmetric about $y = 0$.

For the second case, $\varphi_2(x, M) = M$ immediately implies $p_x = 0$, as evinced in equation (3.37). If we let $x' = \varphi_1(x, M)$ as above, then the trajectory in question clearly passes through the point $(x', \pm\beta, 0, p_y)$, and symmetry follows from

Proposition 5. □

Thus, to find symmetric trajectories of the system, it is enough to find the fixed points of either component and verify that one of the conditions (4.14–4.15) hold. We begin with $\varphi_1(\,\cdot\,, M)$: If $x' = x$, then we immediately have $\theta_2 = -\theta_1$, since both linear portions of the trajectories are tangent to the same circle, beginning at the same x co-ordinate, and with the same orientation, although reflected across the x-axis. Substituting equation (4.8) into this condition then gives

$$\theta_1 = \int_{\tilde{r}}^{R} \frac{M}{\xi^2\sqrt{2(E - V(\xi))}}\, d\xi \;+\; \operatorname{sgn}(M) \cdot \arccos\left(\frac{\hat{r}}{R}\right). \tag{4.16}$$

Note that the quantity on the right-hand side depends only on M, and let the function $g(M)$ denote this expression. To find a pair (x, M) for which the condition holds, we thus fix an M value and find an x such that

$$\theta_1(x, M) = g(M) \bmod 2\pi. \tag{4.17}$$

Now, $g(M)$ depends on various parameters of the underlying system: the energy surface E; the radius R of the potential field $U : \mathbb{R}^+ \to \mathbb{R}$; and the potential field itself. However, it is clear based on our conventions that: $g(M) = 0$ whenever $|M| \geq R\sqrt{2E}$; g is only discontinuous at $M = 0$; and g is odd. Deferring additional assumptions, to be made in the following chapter, we instead ask which values in $\mathbb{R}/2\pi\mathbb{Z}$ can be obtained by $\theta_1(\,\cdot\,, M)$?

This can be ascertained readily from the map's construction. If we let x range over all of $\mathbb{R}$ and assume $M < 0$, then each θ_1 is determined by the line through x that is clock-wise tangent to the circle with radius $\hat{r} = \frac{|M|}{\sqrt{2E}}$. Note that for $x = \hat{r}$ we have $\theta_1 = 0$. Then as x is moved continuously toward negative infinity, and then toward positive infinity, θ_1 achieves every value in $(-\frac{\pi}{2}, \frac{\pi}{2})$ and only these; hence,

$$\theta_1(\mathbb{R}, M) = (-\frac{\pi}{2}, \frac{\pi}{2}) \quad \text{for} \quad M < 0.$$

Now, note that g is continuous at $M = -R\sqrt{2E}$ and that $g(-R\sqrt{2E}) = 0$. We can guarantee the following:

Lemma 3 *There exists a $\delta > 0 \in \mathbb{R}$ such that, for all $M \in [-R\sqrt{2E}, -R\sqrt{2E} + \delta) =: K$, there exists a unique $x \in \mathbb{R}$ such that equation (4.17) holds.*

Proof. Since g is continuous at $M = -R\sqrt{2E}$, there exists a $\delta > 0$ such that

$g(K) \subset (-\frac{\pi}{2}, \frac{\pi}{2})$. The result follows since $\theta_1(\,\cdot\,, M)$, considered as a map into the latter interval, is a bijection for any $M < 0$. □

Proposition 8 *For every energy level E, there is a (generally finite) curve in (x, M)-space which generates symmetric trajectories.*

The proof follows directly from the Lemma and Proposition 7. It differs from our earlier result in that these points are located on the line $y = \pm\beta$ in physical space, and their x-momenta are not necessarily zero; the importance of this feature will be highlighted in the following sections. Note that while, in the name of completeness, we could also assume $M > 0$ and examine the clockwise tangents to obtain

$$\theta_1(\mathbb{R}, M) = [-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi) \quad \text{for} \quad M > 0,$$

it is possible that $g^{-1}\left([-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi)\right) = \emptyset$. Thus, we cannot guarantee periodic trajectories for $M > 0$ while assuming $p_y < 0$. In particular, one could show that there are no such trajectories if $U : \mathbb{R}^+ \to \mathbb{R}$ is non-negative. However, we can always reflect the trajectories from Proposition 8 to find trajectories for which (at least locally) we would have $M, p_y > 0$.

As for the fixed points of $\varphi_2(x, \,\cdot\,)$, it is clear from equation (4.12) that these occur only when $\sin\theta_2 = 0$. Moreover, assuming $p_y > 0$ on exit rules out the roots $\theta_2 = \pi \mod 2\pi$, and so we are left with $\theta_2 = 0 \mod 2\pi$. Substituting equation (4.8) as before, and using $g(M)$ as defined above, gives

$$\theta_1(x, M) = \frac{1}{2} g(M) \mod 2\pi, \tag{4.18}$$

which is essentially a re-scaling of the earlier condition. Just as before, we can use the continuity of g at $M = -R\sqrt{2E}$ to obtain the following result:

Lemma 4 *There exists a $\delta > 0 \in \mathbb{R}$ such that, for all $M \in [-R\sqrt{2E}, -R\sqrt{2E} + \delta) =: K$, there exists a unique $x \in \mathbb{R}$ such that equation (4.18) holds.*

In fact, we could have used this lemma to prove Proposition 8, since it too guarantees a (generally finite) curve in x and M that generates symmetric orbits. Note that these curves coincide if and only if $M = 0$, and that the points of their correspondence generate vertical, periodic trajectories in the configuration space. Finally, we would expect more such trajectories, in the sense that

$$(\frac{1}{2}g)^{-1}\left((-\frac{\pi}{2}, \frac{\pi}{2})\right) = g^{-1}\left((-\pi, \pi)\right) \supset g^{-1}\left((-\frac{\pi}{2}, \frac{\pi}{2})\right).$$

Chapter 5

Geometry and Periodic Orbits

5.1 Deflections

In the previous section, we showed that, by studying the collision map, we could ascertain the existence of a continuum of trajectories, each symmetric about the x-axis. While we had a similar result in §3.4.1, the trajectories were there described either by the curve

$$S_1 := \{(x, 0, 0, p_y) \mid x \in \mathbb{S}^1, p_y \in \mathbb{R}\},$$

or by the curve

$$S_2 := \{(x, \pm\beta, 0, p_y) \mid x \in \mathbb{S}^1, p_y \in \mathbb{R}\}.$$

However, we in general know very little about the trajectories arising from these points. On the other hand, Lemmas 3 and 4 give the curves

$$C_1 := \{(x, M) \mid M \in K_1 \text{ and } x = \theta_M^{-1}(g(M))\} \tag{5.1}$$

$$\text{and} \qquad C_2 := \{(x, M) \mid M \in K_2 \text{ and } x = \theta_M^{-1}(\frac{1}{2}g(M))\} \tag{5.2}$$

respectively, where K_1 and K_2 are both intervals of the form $[-R\sqrt{2E}, -R\sqrt{2E} + \delta)$, for some $\delta > 0$, and $\theta_M^{-1}(y)$ is defined such that $\theta_1(\theta_M^{-1}, M) = y$. Each of their trajectories is contained in either S_1 or S_2, but we now know their momenta at the boundary of the local co-ordinate system. We will now use this information to examine the possibility of deflections—in particular, deflections to the right, in the sense of the previous chapter. (That is, those that enter the collision region through Σ_1 and (eventually) re-enter the linear region through Σ_1 without

crossing Σ_3.) Note that, since the trajectories are symmetric, we need verify only that they encounter Σ_1, and subsequently reach the point where they begin to trace a symmetric pattern in space, to prove they are deflected—provided, of course, they do not reach Σ_3 in the interim.

Throughout this section, in the interest of brevity, we treat only those trajectories that arise from the points of C_1. Similar results exist for C_2 in each case, and the arguments require only minor (in effect, notational) modifications. In the first case, we make repeated use of two facts that

1. If $(x, M) \in C_1$ corresponds to an initial condition $(x, \pm\beta, p_x, p_y)$, then the corresponding trajectory exits the local chart at $(x, \pm\beta, -p_x, p_y)$.

2. Additionally, these points are symmetric when considered *in time*—that is, there is point $t = \hat{t} \in \mathbb{R}$ depending on the initial condition (p_0, q_0) such that $\Phi^{\hat{t}+\delta}(p_0, q_0) = \rho_x \circ \Phi^{\hat{t}-\delta}(p_0, q_0)$.

While the first cannot be said of $(x, M) \in C_2$, there is an analogous result: the trajectory arising from $(x, M) = (x, \pm\beta, p_x, p_y)$ is symmetric in time about the point $(x', \pm\beta, 0, p_y)$, where it exits the initial chart. Thus, if we were to integrate to the boundary of the following chart, we would find ourselves at $(x, \pm\beta, -p_x, p_y)$. In both cases, the pair of points is symmetric in time and distinct only insofar as their x momenta have opposite signs. These facts forms the basis for many of our arguments; they are also the starting point for showing that certain such trajectories deflect.

Any trajectory that begins outside the scatter will eventually leave the scatter: this follows from certain arguments in §3.4.2. It is unclear, however, if our symmetric trajectories will enter the linear region.[1] In fact, if the potential function U is non-positive in a neighbourhood of R, then we would expect that no trajectories arising from C_1 would enter the linear region without additional collisions with the scatterer. Thus, in order to ensure simple deflections for the trajectories in question, we will need to make additional assumptions about the potential function. Bracketing this for the moment, we note that for an arbitrary potential there may be pairs $(x, M) \in C_1$ such that $(x, \pm\beta)$ is already in the linear region, and thus the trajectory is guaranteed to re-enter after local integration. Unfortunately, we have only indirect information concerning x: the values of M are given explicitly

[1] Although it is obvious that such trajectories exist in the space of all initial conditions. In general, if given a trajectory that moves from Σ_2 to Σ_2 within a collision, one can simply rotate the trajectory 90° about the origin for a trajectory that is deflected.

in the definition, while those for x depend on the functions θ_M^{-1} and g, neither of which has been explicitly defined.

Focusing on what we have, note then that the local trajectory is guaranteed to be linear whenever $|M| > R\sqrt{2E}$. While this result does not hold for any $M \in K_1$, it may hold for the corresponding trajectory on a subsequent chart, whereupon the value of M will have changed. If $M' = \varphi_2(x, M) < -R\sqrt{2E}$, then the trajectory will avoid the scatterer and enter the linear region. Now M' is given by the now-familiar relation

$$M' = M - \operatorname{sgn}(M)\sqrt{2E}\sin\theta_2.$$

But $\theta_2 = -\theta_1$ and $\theta_1 = g(M)$. Thus $M' < -R\sqrt{2E}$ if and only if

$$\sin g(M) > \frac{M + R\sqrt{2E}}{\sqrt{2E}}$$

for a given $M \in K_1$. Writing $M = -R\sqrt{2E} + \epsilon$ for some $\epsilon \in [0, \delta)$ gives the condition

$$\sin g(-R\sqrt{2E} + \epsilon) > \frac{\epsilon}{\sqrt{2E}} \quad \iff \quad g(-R\sqrt{2E} + \epsilon) > \arcsin\frac{\epsilon}{\sqrt{2E}}. \tag{5.3}$$

If this holds for all $\epsilon \in [0, \delta')$, for some $\delta' \leq \delta$, then each of the trajectories, given by a corresponding (x, M) pair, will deflect regardless of the associated x co-ordinate.

Under what assumptions will this condition hold? In §4.1.1, we noted in passing that $\arccos\left(\frac{\hat{r}}{R}\right)$ could be found using the linear solution—more explicitly,

$$\operatorname{sgn}(M)\arccos\left(\frac{\hat{r}}{R}\right) = \theta(B_1) - \theta_1 = -\int_{\tilde{r}}^{R} \frac{M}{\xi^2\sqrt{2(E - V(\xi))}}\,d\xi,$$

when $U \equiv 0$. Thus we can write g as follows

$$g(M) = \int_{\tilde{r}}^{R} \frac{M}{\xi^2\sqrt{2E - \frac{M^2}{\xi^2} - 2U(\xi)}}\,d\xi - \int_{\hat{r}}^{R} \frac{M}{\xi^2\sqrt{2E - \frac{M^2}{\xi^2}}}\,d\xi, \tag{5.4}$$

and hence g can be seen as the difference (in the θ co-ordinate) between the non-linear trajectory and its linear counterpart. Unfortunately, it is unclear from this expression whether the condition (5.3) holds. Instead we assume that U

can be chosen such that g is negative on the some subinterval of K_1 containing $M_0 = -R\sqrt{2E}$ for which the remaining points are deflected. Henceforth, we let $K_1' \subset K_1$ denote the largest such subinterval. We then define

$$C_1' := \{(x, M) \mid M \in K_1' \text{ and } x = \theta_M^{-1}(g(M))\},$$

which gives the set of pairs (x, M) corresponding to those trajectories. We note in passing that there *should* be a geometric argument to this end when $U > 0$, although we are not able to present such a result at this time.

5.2 Periodicity for a Special Case

In this section, we consider the case in which x is constant for all $(x, M) \in C_1'$, which implies $x = R$. While this is obviously a very particular case, it provides an avenue to explore the effects of the geometry on these trajectories, and to develop the machinery for our final result. Moreover, it is obvious in this case that deflections do indeed occur, since such trajectories always pass through the point (R, β) in configuration space.

Example. Suppose that x was constant along the curve $(x, M) \in C_1'$. Then we would have a set of symmetric trajectories passing through the point $(R, \pm\beta)$ in phase space that achieved a subinterval of the possible momentum-states, including the state $p_x = 0$. Moreover, if we then truncated the geometry so that we were left with the torus $\Omega = [-R, R] \times [-\beta, \beta]$, then all of these trajectories would be periodic by Proposition 6. ◇

Now we consider the effect of a more general geometry on the same system.

Proposition 9 *Suppose x is constant in C_1' (and therefore $x = R$ for all M). For* any *choice of $\alpha > R$, the system arising from the geometry $\Omega = [-\alpha, \alpha] \times [-\beta, \beta]$ will have infinitely many periodic trajectories.*

Before beginning, we define $a = 2(\alpha - R)$, and note that the linear region corresponding to the given geometry has horizontal length a. For convenience, we will use co-ordinates $R_l := [0, a] \times [-\beta, \beta]$ within the linear region, noting that the lines $x = 0$ and $x = a$ in these co-ordinates correspond to the lines $x = R$ and $x = -R$ in Ω. Thus, we begin with trajectories that enter the linear region through $(0, 0)$ with varying momenta. Note also that, because of various discrete

symmetries,[2] and the fact that the Carestian momenta are constant in the linear region, it will be enough to show that an infinite subset of those trajectories pass through the point $(a, \pm\beta)$.

Proof. Consider the ratio $\frac{p_x}{p_y}$ for each pair $(x, M) \in C_1'$. In fact, this quantity is independent of x for the trajectories in question: using equations (4.11) and (4.13) we clearly have

$$\frac{p_x}{p_y} = -\tan\theta_2 = \tan\theta_1 = \tan g(M) =: \rho(M).$$

Now, $g(M)$ is assumed non-positive on K_1' and $g(-R\sqrt{2E}) = 0$. By continuity of g at $M_0 := -R\sqrt{2E}$, it follows that there must be some subinterval of K_1' containing M_0 for which g is decreasing. Let $\tilde{I}$ denote this interval, and define $I = \rho(\tilde{I})$, noting that $I = (-\delta, 0]$ for some δ. Now suppose we are working in the extension of R_l defined as $\overline{R}_l = [0, \infty) \times [-\beta, \beta]$. Each trajectory begins at the point $(0, \beta) \in R_l$ and moves along the line with slope ρ until it intersects with the line $y = -\beta$ at some point $(x', -\beta)$; as the flow continues, it traces another line with slope ρ from $(x,'\beta)$ to another intersection, and so on. Let $\psi(\,\cdot\,, \rho) : [0, \infty) \to [0, \infty)$ denote this map, for which we have the relation

$$\psi(x, \rho) = (x - 2\beta\rho, \rho).$$

Since ρ depends only on the Cartesian momentum vector, which is constant in R_l, ρ too must be constant in R_l; hence ρ is preserved in the map, which is thus another Poincaré section whose orbits describe the paths of our symmetric trajectories within the linear region.

We can now apply the iterates of this map to the interval I_ρ which describes the location and momentum of each trajectory upon entry to the region. Note first that $\psi^n(x, \rho) = (x - 2n\beta\rho, \rho)$. Now,

$$I = \{(0, \rho) \mid \rho \in (-\delta, 0]\}.$$

and hence

$$\psi^n(I) = \{(-2n\beta\rho, \rho) \mid \rho \in (-\delta, 0]\}, \quad \forall n \in \mathbb{N}.$$

Projecting this into configuration space gives intervals of the form $I^n = [0, 2n - \beta\rho)$ for all $n \in \mathbb{N}$, which satisfy $I_n \subset I_m$ for all $n < m$. Note that for each choice

[2] In particular, the symmetry relating φ_l and φ_r, the left and right collision maps respectively.

of $a \in \mathbb{R}^+$ there is an $N \in \mathbb{N}$ such that $a \in I_N$–that is, $a = -2n\beta\rho$ for some $\rho \in (-\delta, 0]$—and thus a is contained in all I_n for $n > N$. Moreover, for each such interval, the ρ that generates a clearly satisfies $\rho = -\frac{a}{2n\beta}$ for each $n > N$, and arises from the point $(0, -\frac{a}{2n\beta}) \in I$. Thus, there are a countably infinite number of points in I that pass through $(a, \pm\beta)$. □

Note that, while the premise of this result is singularly restrictive, one could potentially obtain such a curve of initial conditions by simultaneously varying the energy level E in an otherwise rather general system. The result would then give countably many *energies* for which a periodic solution exists (given the geometry)—although such a result may require further assumptions on the potential energy function $U : \mathbb{R}^+ \to \mathbb{R}$.

5.3 A Criterion for Periodic Orbits

This brings us to our final result:

Proposition 10 *Suppose we have a curve of initial conditions* $C := \{(x_\epsilon, \rho_\epsilon)\}_{\epsilon \in [0,\delta)}$ *corresponding to symmetric trajectories where, for some* $\delta > 0$*, we have the following: (i)* $x_0 = R$ *and* $\rho_0 = 0$*; (ii) the maps* $\epsilon \mapsto x_\epsilon$ *and* $x_\epsilon \mapsto -\rho_\epsilon$ *are increasing, continuous functions. Then, for any choice of geometry with* $a \geq 0$*, the system will have a countably infinite number of periodic orbits.*

Proof. The result follows essentially as above, except we now have

$$\psi^n(I) = \{(x\epsilon - 2n\beta\rho_\epsilon, \rho_\epsilon) \mid \epsilon \in [0, \delta)\}, \quad \forall n \in \mathbb{N}.$$

but since $\epsilon \mapsto x_\epsilon$ and $x_\epsilon \mapsto -\rho_\epsilon$ are increasing, continuous functions, they still give rise to a series of nested intervals I_n in the configuration space, the span of which increases by $\sup_\epsilon \rho_\epsilon$ with each iteration. The monotonicities also ensure the points in C that give rise to a in subsequent intervals are in fact distinct. □

As in the previous section, we note that each of these trajectories $(x_\epsilon, \rho_\epsilon)$ will be initially deflected, since $x_\epsilon > x_0 = R$.

5.4 Endnote: Periodic Solutions and Ergodicity

Ergodicity tells us a great deal about the behaviour of a flow. For example, let U be any set with $\mu(U) > 0$. Then $\Phi^t U$ is an invariant set with $\mu(\Phi^t U) \geq \mu(U) > 0$,

and thus $\mu(\Phi^t U) = 1$. Thus, any set of positive measure will fill up the phase space under the action of the flow.

Now suppose that one of the trajectories described in the previous section is stable in the following sense: within every neighbourhood of the trajectories there are a set of trajectories of positive measure which stay sufficiently near it for all time. Such a set could be chosen such that it never enters a particular region of positive measure, and thus we would have shown that the system cannot be ergodic. Donnay presents such a result in [Don2], which follows straightforwardly from a more difficult result in [Don1]. Essentially, Donnay fixes the triple (U, R, E) and then constructs a particular geometry which admits a periodic trajectory. He then introduces a Poincaré map of the system and proves the corresponding orbit is stable. Interestingly, Donnay's trajectories seem to be a subset of those presented here; moreover, while he hypothesizes that such trajectories are stable, he is unable to show stability for strictly positive U. It is possible that further study of the Poincaré maps presented here, as well as the continuum of symmetric orbits they describe, would facilitate such a result.

Appendix A

Derivations

A.1 The Transformation $(M, E; r, \theta) \mapsto (p_x, p_y)$

Recall the formula for M in Cartesian co-ordinates, given by equation (3.36):

$$M = xp_y - yp_x.$$

By isolating xp_y and squaring, we have

$$M^2 + 2Myp_x + y^2p_x^2 = x^2p_y^2 = x^2(2E - p_x^2).$$

Collecting terms in p_x and writing $x^2 + y^2 = r^2$ gives

$$r^2p_x^2 + 2Myp_x + (M^2 - 2x^2E) = 0,$$

which has discriminant

$$4x^2\left(2Er^2 - M^2\right).$$

Note that this expression is never negative:

$$\begin{aligned} M^2 &= x^2p_y^2 - 2xyp_xp_y + y^2p_x^2 \\ &\leq x^2p_y^2 + x^2p_x^2 + y^2p_y^2 + y^2p_x^2 = 2Er^2, \end{aligned}$$

since $0 \leq (xp_x + yp_y)^2$. The solution is thus defined everywhere, with

$$p_x = \frac{-My \pm x\sqrt{2Er^2 - M^2}}{r^2}.$$

As mentioned earlier, neither solution can be excluded. To each solution, we

can associate a (respective) value of p_y by substitution into equation (3.36):

$$p_y = \frac{Mx \pm y\sqrt{2Er^2 - M^2}}{r^2}.$$

A.2 Equation (4.5): $(x, M) \mapsto \theta_0$

Recall equation (4.4), which gives

$$\hat{r} = x\cos\theta_0 + y\sin\theta_0.$$

for all Cartesian co-ordinates (x, y) on the line whose polar form is $r = \hat{r}\sec(\theta - \theta_0)$. Now let $y = \beta$ and assume x is given. We then have

$$\begin{aligned} & \hat{r} - x\cos\theta_0 = \beta\sin\theta_0 \qquad (A.1) \\ \Rightarrow\ & (\hat{r} - x\cos\theta_0)^2 = \beta^2\sin^2\theta_0 \\ \Rightarrow\ & (x^2 + \beta^2)\cos^2\theta_0 - 2x\hat{r}\cos\theta_0 + (\hat{r}^2 - \beta^2) = 0. \end{aligned}$$

Thus, $\cos\theta_0$ arises as the root of a quadratic polynomial with discriminant

$$4\beta^2\left(x^2 + \beta^2 - \hat{r}^2\right)$$

—which is positive since $\beta \geq R \geq \hat{r}$. Solving the quadratic for $\cos\theta_0$ then gives

$$\cos\theta_0 = \frac{x\hat{r} \pm \beta\sqrt{x^2 + \beta^2 - \hat{r}^2}}{x^2 + \beta^2}.$$

By considering the case $\theta_0 = 0$, which corresponds to $x = \hat{r}$, we can exclude the negative root.

Note that, having found $\cos\theta_0$, we can also determine $\sin\theta_0$ from equation (A.1):

$$\sin\theta_0 = \frac{\beta\hat{r} - x\sqrt{x^2 + \beta^2 - \hat{r}^2}}{x^2 + \beta^2}.$$

Substituting both these values into equation (4.6) produces

$$\begin{aligned} x' &= \frac{\hat{r}x^2 + 2\beta^2\hat{r} - \beta x\sqrt{x^2 + \beta^2 - \hat{r}^2}}{\hat{r}x + \beta\sqrt{x^2 + \beta^2 - \hat{r}^2}} \\ &= x + 2\beta\,\frac{\beta\hat{r} - x\sqrt{x^2 + \beta^2 - \hat{r}^2}}{\hat{r}x + \beta\sqrt{x^2 + \beta^2 - \hat{r}^2}}. \end{aligned}$$

Rationalizing the numerator in the latter term, and then cancelling a factor of $x^2 + \beta^2$, gives the more useful form found in §4.1.1.

Bibliography

[Arn1] V. I. Arnol′d. *Ordinary differential equations*. MIT Press, Cambridge, Mass., 1978. Translated from the Russian and edited by Richard A. Silverman.

[Arn2] V. I. Arnol′d. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 199? Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[Bol] L. Boltzmann. *Theoretical Physics and Philosophical Problems*. D. Reidel, Dordrecht, Holland, 1974. Edited by Brian McGuinness. Translated from the German by Paul Foulkes.

[CFS] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.

[DL] Victor Donnay and Carlangelo Liverani. Potentials on the two-torus for which the Hamiltonian flow is ergodic. *Comm. Math. Phys.*, 135(2):267–302, 1991.

[DLT] Andrei Draganescu, Rich Lehoucq, and Paul Tupper. Hamiltonian molectular dynamics for computational mechanicians and numerical analysts. Technical report, Sandia National Laboratories, 2006.

[Don1] Victor J. Donnay. Elliptic islands in generalized Sinai billiards. *Ergodic Theory Dynam. Systems*, 16(5):975–1010, 1996.

[Don2] Victor J. Donnay. Non-ergodicity of two particles interacting via a smooth potential. *J. Statist. Phys.*, 96(5-6):1021–1048, 1999.

[HLW] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. Structure-preserving algorithms for ordinary differential equations.

[KH] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[Olv] Peter J. Olver. *Applications of Lie groups to differential equations*, volume 107 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1993.

[Sin] Ja. G. Sinaĭ. Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat. Nauk*, 25(2 (152)):141–192, 1970.