Sapienza Università di Roma

Scuola di Dottorato "Vito Volterra"
Dottorato di Ricerca in Matematica – XXI ciclo

Problems of Celestial Mechanics

I. On the chaotic motions and the integrability of the planar 3-centre problem of Celestial Mechanics.

II. The use of the Kepler integrals for Orbit Determination.

Thesis submitted to obtain the degree of
Doctor of Philosophy ("Dottore di Ricerca") in Mathematics
18th January 2010

by

Linda Dimare

Program Coordinator
Prof. Tesei Alberto

Thesis Advisors
Prof. Negrini Piero
Dr. Gronchi Giovanni Federico

"Ours, according to Leibniz, is the best of all possible worlds, and the laws of nature can therefore be described in terms of extremal principles."

C. L. Siegel, J. K. Moser,
Lectures on Celestial Mechanics, Springer 1971.

"When, however, one attempts to formulate a precise definition of integrability, many possibilities appear, each with a certain intrinsic theoretic interest. Let us consider briefly the concept of integrability, not forgetting the dictum of Poincaré, that a system of differential equations is only more or less integrable."

G. D. Birkhoff,
Dynamical Systems, AMS 1966.

Abstract

This thesis is divided into two main parts.

In the first part we investigate the question of integrability for the planar restricted 3-centre problem on the small negative energy level sets. The case of positive energy has been studied by Bolotin (see [4]), who showed that no independent real-analytic integral exists on positive energy shells. For negative energies we know that chaotic motions exist, if we make the assumption that one of the centres is far away from the other two (see [11]). This result has been obtained by the use of the Poincaré-Melnikov theory. Here we change the assumption on the third centre: we do not make any hypothesis on its position, and we obtain a perturbation of the 2-centre problem by assuming its intensity to be very small. Then we prove the existence of uniformly hyperbolic invariant sets of periodic and chaotic almost collision orbits by the use of a general result of Bolotin and Mackay (see [8], [9]). To apply it, we must preliminarily construct *chain of collision arcs* in a proper way. We succeed in doing that by the classical regularisation of the 2-centre problem and the use of the periodic orbits of the regularised problem passing through the third centre.

In the second part we study a problem of orbit determination in the context of the current and next generation observational techniques. We investigate a method to compute a finite set of preliminary orbits for a solar system body, using the first integrals of the Kepler problem. This method is thought for the applications to the modern sets of astrometric observations, where often the available information allows only to compute, by interpolation, two angular positions of the observed body and their time derivatives at a given epoch; we call this set of data *attributable*. Given two attributables of the same body at two different epochs we can use the energy and angular momentum integrals of the two-body problem to write a system of polynomial equations for the topocentric distance and the radial velocity at the two epochs. We define two different algorithms for the computation of the solutions, based on different ways to perform elimination of variables and obtain a univariate polynomial. Moreover we can use the redundancy of the data to test the hypothesis that the two attributables belong to the same body (*linkage problem*). It is also possible to compute a covariance matrix, describing the uncertainty of the preliminary orbit which results from the observation error statistics. The performance of this method has been investigated by using a large set of simulated observations given to us by astronomers of the Pan-STARRS project. Finally, the method has been tested also for the space debris, by processing one year data from ESA Optical Ground Station telescope, provided by the University of Bern.

Acknowledgements

*I wish to thank my advisors Prof. P. Negrini (*University of Roma I 'La Sapienza'*) and Dr. G. F. Gronchi (*University of Pisa, Celestial Mechanics Group*) for having introduced me to these beautiful subjects, for their precious suggestions and help during my Ph.D. studies.*

Contents

Preface

During the years of my Ph.D. I have devoted myself to the study of two different subjects of the same major area of Celestial Mechanics.

As a consequence, this thesis is divided into two well distinguished parts. In the first one, the unsolved question about the analytic integrability of the n-centre problem is investigated, and a result of existence of chaotic and periodic quasi collision motions is obtained. The second part is dedicated to the subject of orbit determination, which is of more practical interest for the astronomical applications in the identification of new detected celestial bodies. In this part, new algorithms to find the preliminary orbits of an observed body are defined.

In the first chapter a general introduction to the question of integrability for the n-body and n-centre problem is carried out. The only known integrable cases are the 2-body problem (which is the same as the 1-centre problem), and the 2-centre problem. The word integrability can have different meanings: in these particular cases integrability by quadratures is obtained. Some very important classical results of non-integrability (in accordance to a certain definition of it) are also recalled, such as Brun's and Poincaré theorems, which are valid for the general n-body problem and the restricted 3-body problem respectively. Then the question of analytic integrability for the n-centre problem is introduced by referring to some modern results by Kozlov, Bolotin, Negrini, Knauf, and Taimanov ([28], [4], [5], [10], [11], [27]). In the final section of the chapter a general shadowing result of Bolotin and Mackay is described, together with some applications of it to the restricted 3-body problem (see [8], [9]).

The latter outcomes are the basis for the subsequent study on the integrability of the planar 3-centre problem, which is carried on in Chapter 2. This chapter contains an original result about the existence of periodic and chaotic trajectories which undergo close encounters with the third centre, that is assumed of infinitesimal mass. The result is obtained by the application of the main theorem of [8], which states the existence of chaotic motions shadowing chains of orbits which start and end at collision with a centre, for a certain class of Lagrangian systems with Newtonian singularities. We are able to construct chains of orbits satisfying the assumptions of this theorem, so that it can be applied and it allows us to get a hyperbolic invariant set formed by the shadowing orbits. Moreover, on this set we have a symbolic dynamics.

In Chapter 3 the proofs of the theorem of [8] and of the hyperbolicity of the shadowing orbits obtained from it are both recalled to conclude the description of the result attained.

In the second part of the thesis we start, in Chapter 4, with a rapid glance at the problem of orbit determination, that is the problem to find an orbit of a solar system body from observations. After describing the classical methods by Gauss [22] and Laplace [33], we put the attention on the more recent procedures conceived to deal with the huge amount of data of the current and next generation surveys ([31], [38], [41]).

In Chapter 5 we describe a method to determine the preliminary orbits by the use of the two-body integrals of angular momentum and energy. The data from which the procedure starts are two sets of observations, which are used to compute two *attributables*. Each of these sets is typically formed by observations of the same object made in the same night, which, due to different possible reasons, are not enough to compute an orbit. Nevertheless, they allow to compute by interpolation two angular positions and the corresponding angular velocities at an average time: these four quantities form an attributable. We describe the algebraic problem which arises by imposing the equality of the integrals at the two attributables. It is a polynomial system of total degree 48, whose unknowns are the radial distance and velocity of the solar system body with respect to the observer positions. Two different algorithms are developed to find all the solutions of this problem, from which it is possible to compute the preliminary orbits.

In Chapter 6 some numerical experiments are presented. After a first test with a known object, we describe a simulation performed on a large set of data, that have been prepared thinking to the expected performance of the next generation surveys: the astronomers of the Pan-STARRS project have provided us a data-set with the features required. Another large scale test has been made for the case of the space debris using one year data provided by the University of Bern.

Part I

On the chaotic motions and the integrability of the planar 3-centre problem of Celestial Mechanics

Chapter 1

Introduction

In this chapter we want to introduce our first problem, inserting it in a proper mathematical background. First of all, we make a short review of the more important classical results about the integrability of both the n-body and n-centre problems. Then, we make the point for the question of analytic integrability of the n-centre problem, giving a survey of the existing literature on the subject. We also recall a shadowing result of Bolotin and Mackay, which is the starting point of our investigation and, at the end of the chapter, we state our theorem on the existence of chaotic motions of the planar 3-centre problem shadowing collision chains.

1.1 Classical results on the integrability of the n-body problem

The n-centre problem enters in the more general and difficult study of the n-body problem as a rough simplification: the hope is that any information about it could give some insight even for the other. Anyway, even if it is simpler than the general n-body problem, it is very difficult too and it continue to be a source for much mathematical work.

The n-body problem is described as follows: n point masses move in space under their mutual gravitational attraction and the goal is to determine their orbits. There are no constraints, the masses are arbitrary and the bodies are initially moving in any given manner. Fixed a Cartesian reference system in the space, the equations of motion are

$$m_i \ddot{\mathbf{r}}_i = -\sum_{j \neq i} m_i m_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \,, \tag{1.1}$$

where $\mathbf{r}_1, \dots, \mathbf{r}_n$ are the position vectors of the bodies and $m_1, \dots, m_n$ their masses.

The only known integrable case is the 2-body problem. It can be reduced to a problem with a central force field by considering the relative motion in the reference of the centre of mass. Then, the solution by quadratures is obtained by the use of the angular momentum and energy integrals.

Already for $n = 3$ the situation becomes very complicate. Equations (1.1) form a system of the $6n$-th order: as it is classically known (see [55]), the order of this system can be reduced to $(6n - 12)$ by the use of the six integral of the motion of the centre of mass, the three integrals of angular momentum, the integral of energy, the *elimination of the time* and the *elimination of the nodes*.

The ten integrals that we have just cited are known as the *classical integrals* of the n-body problem and are the only known algebraic integrals of it. In 1887 Bruns showed (see [12]) that *the classical integrals are the only independent algebraic integrals of the problem of three bodies*, where independent means algebraically independent. A proof of this theorem can be found in [55, Chapter XIV]. Actually, the same result is true for the n-body problem, with n arbitrary. More precisely, any integral of the n-body problem which is an algebraic function of the time, the coordinates and the velocities must be a combination of the classical integrals. A further generalisation of Bruns' theorem is due to Painlevé (see [45]), who showed that any integral of the n-body problem which is an algebraic function of the velocities and is analytic in the coordinates is a combination of the classical integrals.

If the initial positions and velocities of the bodies lie on a plane, then the motion of the particles will take place on it: this is the planar problem. For the planar 3-body problem the order of the system is 12 and it can be reduced to 4, by a procedure very similar to the one used in the spatial case.

However, all these results do not say much about the complexity of the system: in particular, the question of analytic integrability remains open. After the outcomes just recalled, due to the difficulty of the general problem, the research has concentrated on some important simplifications of it. In the *restricted problem* of three bodies, the 3rd point mass moves under the gravitational attraction of the other two, but it is supposed not to influence their motion. The problem is to determine the motion of this third body, which is generally called *Planetoid*. In the *circular* restricted problem, the first two point masses, called *primaries*, are assumed to be moving on circular orbits about their centre of mass, subject only to their mutual gravitational attraction. In the *planar* case, the Planetoid is supposed to move on the same plane of motion of the primaries.

The assumption that the 3rd body does not influence the motion of the other ones is a good approximation only if its mass is sufficiently small: indeed, only when this mass is zero we can correctly affirm that this assumption is satisfied. Often the third body is said to have zero mass, meaning simply that it does not influence the motion of the other bodies.

The formulation of the circular restricted problem was originally suggested by the approximately circular motion of the planets around the Sun and by the small masses of the asteroids with respect to the planets' ones. Even the system Sun-planet-satellite can be modelled by the restricted problem. Indeed, the importance of the restricted problem lies on its applicability: it's not at all surprising that it is more often a suitable model than the general 3-body problem, especially in space dynamics and in solar system dynamics.

Anyway, even if the restricted problem is simpler than the general one, it is still very difficult to deal with from the mathematical viewpoint: in particular, the question of analytic integrability is an open problem. This is true for any version of the restricted problem.

To describe the planar circular restricted problem, it is suitable to pass to the reference system with origin at the centre of mass of the primaries and rotating with them. Let (x, y) be the position of the third mass in the rotating frame. If we use suitable dimensionless coordinates (see [51, Section 1.5]), the angular velocity of the primaries and their distance are normalised to one, and their masses a_1, a_2 are such that $a_1 + a_2 = 1$. We can introduce a single parameter $a = a_2$ in place of a_1, a_2, so that the positions of the primaries are $(a, 0)$ and $(-(1-a), 0)$. The equations of motion are

$$\begin{aligned} \frac{d^2x}{dt^2} - 2\frac{dy}{dt} &= -\left(\frac{(1-a)(x-a)}{r_1^3} + \frac{a(x+1-a)}{r_2^3}\right) + x\,, \\ \frac{d^2y}{dt^2} + 2\frac{dx}{dt} &= -\left(\frac{(1-a)y}{r_1^3} + \frac{ay}{r_2^3}\right) + y\,, \end{aligned} \tag{1.2}$$

where r_1, r_2 are the distances of (x, y) from the primaries. The only known algebraic integral of the restricted problem is the Jacobi integral and it can be easily obtained by multiplying the equations (1.2) respectively by $\frac{dx}{dt}$ and $\frac{dy}{dt}$, adding and integrating with respect to the dimensionless time t. The Jacobi integral is given by the expression

$$\frac{1}{2}\left(\frac{dx}{dt}^2 + \frac{dy}{dt}^2\right) - \frac{1}{2}\left(x^2 + y^2\right) - \frac{(1-a)}{r_1} - \frac{a}{r_2} = -\frac{C}{2}\,. \tag{1.3}$$

The constant C is usually called the Jacobi constant.

By the use of this integral and the elimination of the time, the problem can be reduced to one of the second order. A theorem of Siegel (see [50, 1936]) states that no other independent algebraic integral exists besides the Jacobian integral.

Another result on the nonexistence of a certain type of integrals is due to Poincaré (see [47, 1890]). The restricted problem is an Hamiltonian system. The change to the rotating frame is a canonical or contact transformation, so that the new system (1.2) is also Hamiltonian and the new Hamiltonian function coincides with the Jacobian integral given by (1.3). After another suitable contact transformation the new Hamiltonian is an analytic function $H(\mathbf{q}, \mathbf{p}, a)$ of the new coordinates $\mathbf{q}$ and the new conjugate momenta $\mathbf{p}$, which is periodic in the coordinates $\mathbf{q}$ with period 2π. Moreover, it is analytic in the parameter a, for sufficiently small values of it. Then H can be expanded with respect to a in a convergent power series, for sufficiently small values of a, of the form

$$H = H_0 + aH_1 + a^2H_2 + \dots\,,$$

where $H_0 = H_0(\mathbf{p})$, $H_i = H_i(\mathbf{p}, \mathbf{q})$ are analytic functions of the coordinates and the momenta, periodic in the coordinates $\mathbf{q}$ with period 2π. The transformations used by Poincaré to obtain an Hamiltonian of this form can be found in [55, Ch. XIII §162]. Let $\phi(\mathbf{q}, \mathbf{p}, a)$ be an analytic function for small enough values of a and for $\mathbf{p}$ in an arbitrarily small domain and suppose that ϕ is periodic in $\mathbf{q}$ with period 2π. Then ϕ can be expanded in a convergent power series

$$\phi = \phi_0 + a\phi_1 + a^2\phi_2 + \dots\,,$$

where the ϕ_i are analytic functions of $\mathbf{q}, \mathbf{p}$, periodic in $\mathbf{q}$ with period 2π. Poincaré's theorem says that no integral ϕ of this kind exists for the restricted problem, except the Jacobian integral of energy and integrals equivalent to it.

We don't go further into the analysis of these kind of results. A complete survey of the history of the n-body problem would require an entire book. In this thesis we are interested on integrability, then we have recalled only theorems on this question. Even with this restriction we had to make a choice and we have finally mentioned only those results which seem to us the very fundamental steps of the two-hundred years of research on the problem. Actually, we are going to study the 3-centre problem, then it is useful at this point to give the description of the n-centre problem and of the known results about its integrability.

In the n-centre problem the positions of n point masses, called centres, are fixed in space and a particle moves under the action of their gravitational attraction. Note that the planar 2-centre problem can be seen as a simplified version of the restricted planar circular 3-body problem, in which the centrifugal and Coriolis forces have been neglected. This can be a good approximation for example in the case in which the two primaries move very slowly on their circular orbits.

The central problem faced in this thesis is the 3-centre problem on a plane. In the same way, it can be considered as an approximation of a special case of the restricted 4-body problem, in which the first three masses are fixed on a plane that rotates with constant angular velocity orthogonal to the plane itself, and the fourth body is supposed to move on this plane, subject to the gravitational attraction of the others, but not influencing their motion. This problem can be referred to as the *restricted circular planar* 4*-body problem*. Even in this case the problem of three centres is a good approximation when the centrifugal and Coriolis forces are negligible.

To conclude properly our brief review of the classical outcomes, we must remind that the only known integrable cases for the n-centre problem are $n = 1, 2$. For $n = 1$ we have a motion in a central force field. For $n = 2$, integrability by quadratures was obtained by Euler (1760) in the planar case, using *elliptic coordinates* and a suitable reparametrisation of time. Indeed, in this way we obtain the regularisation of singularities and the separation of the problem: we will use this strategy in Section 2.1.1, where it is recalled in detail. The 2-centre problem in $\mathbb{R}^3$ can be solved in an analogous manner.

In a few words, the only known integrable problems are the one of two bodies and of two centres, which are for this reason the starting point of any work on Celestial Mechanics. The passage from two to three bodies or centres complicates the matter in such a measure that after more than two centuries the question of integrability has not been completely solved. Nevertheless, there are many research articles on the subject, which give important qualitative and numerical results. In particular, there are non-integrability results valid for the problem of n centres, with $n \geq 3$, reduced on energy levels $E > E_{th}$, where E_{th} is a suitable threshold.

1.2 The question of analytic integrability for the n-centre problem

We are going to make the point on the question of analytic integrability for the n-centre problem.

1.2.1 Non-integrability for positive energies

The first result that we recall is due to Bolotin (see [4]) and concerns analytic non-integrability of the n-centre problem for $n \geq 3$. He considers the n-centre problem on the plane $\mathbb{R}^2$, which has Lagrangian

$$L = \frac{|\dot{\mathbf{r}}|^2}{2} + \sum_{i=1}^{n} \frac{a_i}{|\mathbf{r} - \mathbf{r}_i|},$$

where $\mathbf{r} \in \mathbb{R}^2$ denote the position of the particle, $\mathbf{r}_1, \ldots, \mathbf{r}_n$ the fixed positions of the centres on the plane and $a_1, \ldots, a_n$ their densities, which are assumed to be positive. The problem is well defined on the domain $M = \mathbb{R}^2 \setminus \{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$, where the potential energy

$$V = -\sum_{i=1}^{n} \frac{a_i}{|\mathbf{r} - \mathbf{r}_i|}$$

is smooth.

Bolotin succeeds in showing and applying an extension of a result due to Kozlov (see [28]), which assures analytic non-integrability of natural mechanical systems on the basis of geometric properties of the configuration manifold.

Theorem 1.2.1 (Kozlov, 1980) *Let M be a connected compact orientable analytic surface. If M has genus $g > 1$, i.e. M is not homeomorphic to the sphere S^2 or to the torus $\mathbb{T}^2$, then any natural system with real analytic Lagrangian $L = T - V$ on the tangent bundle TM does not admit analytic first integrals on TM independent of the energy integral.*

Theorem 1.2.1 follows from a stronger result of non-integrability, valid for fixed sufficiently large values of the total energy. Given a natural system on a surface M, denote the total energy function by $H = T + V$ and let $M_E = H^{-1}(E) \subset TM$ be the level set of energy E. The following holds:

Theorem 1.2.2 *Let M be a connected compact orientable analytic surface and consider a natural system with real analytic Lagrangian $L = T - V$ on TM. If M has genus $g > 1$, then for all $E > E_{th} = \sup_M V$, the reduced problem on M_E does not have a first integral analytic on M_E.*

Bolotin obtains an analogous result valid for non-compact surfaces. Let N be a closed submanifold with boundary on an analytic surface M, which is not assumed to be compact. Fixed an energy value E, let N_E denote the set of all points of the level set M_E, which are taken by the projection $\pi : TM \to M$ into points of N. The manifold N is *geodesically convex* if for any two close points of the boundary ∂N, the minimal geodesic of the Maupertuis metric joining these points is entirely contained in N.

Let now V be a fixed analytic potential energy on M and let $E_{th} = \sup_M V$.

Theorem 1.2.3 (Bolotin, 1984) *If on the connected analytic surface M there is a compact two-dimensional geodesically convex submanifold N of negative Euler characteristic, then for all $E > E_{th}$ the reduced system on M_E does not have an analytic first integral. Moreover, an analytic first integral does not even exist in a neighbourhood of the set N_E.*

After a global regularisation of singularities and by the use of this result, Bolotin shows that if $n > 2$, for any strictly positive value $E > 0$ of the energy, the planar n-centre problem has no analytic integrals, which are non-constant on the energy shell $H^{-1}(E)$.

In [5] the same author extends his theorem to a wider class of Lagrangian systems defined on any 2-dimensional configuration manifold M, with n Newtonian singularities on M. The Lagrangian function is of the form

$$L = T + \Lambda - V \,,$$

where, denoting by $\mathbf{q}$ the generalised coordinates on M, $T = \frac{1}{2} < \dot{\mathbf{q}}, \dot{\mathbf{q}} >$ is the kinetic energy defined on TM by the Riemannian metric $<,>$, V is the potential energy and $\Lambda =< \omega(\mathbf{q}), \dot{\mathbf{q}} >$ is a linear function of the velocity defined by the vector field ω on M. The functions T, Λ are assumed of class at least C^2 on M, while V is a function of class C^2 everywhere on M, except a finite set of n singular points of Newtonian type. A point $P \in M$ is a singular point of Newtonian type for V, if in local coordinates $\mathbf{q}$ with origin in P,

$$V = -\frac{f(\mathbf{q})}{|\mathbf{q}|} \,,$$

where f is a function of class C^2 on a neighbourhood of P, $f(\mathbf{0}) > 0$, and the symbol $|\cdot|$ denotes the norm defined by the Riemannian metric $<,>$ on M. Bolotin shows non-integrability when n is greater than two times the Euler characteristic of the manifold, $n > 2\chi(M)$, and the energy is over a proper threshold, $E > E_{th}$,

$$E_{th} = \sup_{\mathbf{q}\in M} \left(\frac{1}{2} < \omega(\mathbf{q}), \omega(\mathbf{q}) > + V(\mathbf{q}) \right) \,.$$

Clearly this result comprises the previous case of the planar n-centre problem. In particular, we have analytic non-integrability for the restricted circular many-body problem, in which a particle moves in a rotating plane, under the action of the gravitational attraction of n centres fixed on this plane, when $n > 2$.

Let us now look at the spatial n-centre problem. In [10] Bolotin and Negrini show that if $n \geq 3$ and $E \geq 0$, then the topological entropy is positive. It suggests that the system should not be analytically integrable. In fact this is true, as proved by Knauf and Taimanov in [27]: no analytic independent integral exists for the n-centre problem in the space, if $n \geq 3$ and the energy is greater than some threshold $E > E_{th}$, which depends on the data of the problem, that is on the positions and intensities of the centres. Moreover, for both the planar and the spatial problem, smooth independent integrals are exhibited on the energy levels with $E > E_{th}$.

1.2.2 The case of negative energy

The case of negative energy $E < 0$ has been investigated in [11]. In this work Bolotin and Negrini study the restricted 3-centre problem on the plane, when the third centre is very far from the other two, and consider small negative energies E, in the limit $E \to 0$. Then they have a two-parameter perturbation of the 2-centre problem on the zero-energy level. They succeed in applying the Poincaré-Melnikov theory, thus proving the existence of a hyperbolic invariant set with chaotic dynamics.

We too study the case of small negative energies, but our point of view is different, in that the position of the third centre does not go to infinity. We will see that chaotic motions exist for a dense set of possible positions of the third centre in $\mathbb{R}^2$. Moreover, the obtained trajectories shadow chains of collision orbits through the third centre.

1.3 Chaotic motions for the 3-centre problem

In this section we define the three centre problem on the plane with the assumptions from which our study begins. The starting point of our investigation is a general result about shadowing chains of collision orbits in the n-centre problem, due to Bolotin and Mackay (see [8]). We describe briefly this outcome and some applications of it to the restricted problem of three bodies and finally we introduce our problem and state our main result on the existence of chaotic motions.

1.3.1 The theorems of Bolotin and Mackay

Let M be a smooth Riemannian manifold of dimension $d = 2, 3$ and $\mathcal{C} = \{C_1, \ldots, C_n\}$ a finite set of points on M. Denote by $<, >$ the fixed Riemannian metric on M. Consider a Lagrangian system (L_ε) with configuration space $M \setminus \mathcal{C}$ and Lagrangian function

$$L_\varepsilon(\mathbf{q}, \dot{\mathbf{q}}) = L_0(\mathbf{q}, \dot{\mathbf{q}}) - \varepsilon V(\mathbf{q}) \ . \tag{1.4}$$

The function L_0 is assumed to be at least C^4 on the tangent bundle TM and of the form

$$L_0(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) + < \omega(\mathbf{q}), \dot{\mathbf{q}} > - W(\mathbf{q}) \, , \tag{1.5}$$

where $T(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} < A(\mathbf{q})\dot{\mathbf{q}}, \dot{\mathbf{q}} >$ is a positive definite quadratic form on TM, the term ω is a vector field on M, and the function V is a C^4 function on $M \setminus \mathcal{C}$ with Newtonian singularities at the *centres* $C_i \in \mathcal{C}$. As reminded in Subsection 1.2.1, this means that in a neighbourhood U_i of any point $C_i \in \mathcal{C}$,

$$V(\mathbf{q}) = -\frac{f_i(\mathbf{q})}{\mathrm{dist}(q, C_i)} \, ,$$

where f_i is a C^4 function on U_i, with $f_i(C_i) > 0$, and the distance $\mathrm{dist}(\mathbf{q}, C_i)$ is defined by means of the Riemannian metric T. The energy integral is

$$H_\varepsilon = H_0 + \varepsilon V \, , \qquad \text{with} \qquad H_0(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) + W(\mathbf{q}) \ . \tag{1.6}$$

If the parameter $\varepsilon > 0$ is small, then we can consider the system (L_ε) as a perturbation of the system with Lagrangian L_0, which has no singularities on M.

Consider the unperturbed system (L_0).

Definition 1.3.1 *A solution* $\gamma : [0, T] \to M$ *of fixed energy* E *for the unperturbed system* (L_0) *is called a* collision arc *if* $\gamma(0), \gamma(T) \in \mathcal{C}$ *and* $\gamma(t) \notin \mathcal{C}$ *for any* $t \in (0, T)$.

In particular, the latter condition means that there are no *early* collisions.

Fix an energy value $E \in \mathbb{R}$, such that the set $\mathcal{C}$ is contained in the region $D = \{\mathbf{q} \in M | \ W(\mathbf{q}) < E\}$. Given $\alpha, \beta \in \{1, \ldots, n\}$, denote by $\Omega_{\alpha,\beta}$ the space of

$W^{1,2}$ curves in D, starting and ending respectively at C_α, C_β. A collision arc γ of energy E starting from C_α and ending at C_β is a critical point of the Maupertuis-Jacobi functional J_E on $\Omega_{\alpha,\beta}$. If $\gamma : [0, \tau_f] \to D$ is a curve in $\Omega_{\alpha,\beta}$, then the Maupertuis-Jacobi functional of γ is given by

$$J_E(\gamma) = \int_0^{\tau_f} g_E(\gamma(t), \dot\gamma(t))dt\,,$$

where g_E is the Jacobi metric

$$g_E(\mathbf{q}, \dot{\mathbf{q}}) = 2\sqrt{(E - W(\mathbf{q}))T(\mathbf{q}, \dot{\mathbf{q}})} + < \omega(\mathbf{q}), \dot{\mathbf{q}} >\,,$$

which is a pseudo-Riemannian metric on D. Note that it is positive definite if

$$E > \sup_{\mathbf{q}\in M} \left(\frac{1}{2} < \omega(\mathbf{q}), A^{-1}(\mathbf{q})\omega(\mathbf{q}) > + W(\mathbf{q}) \right)\,.$$

The collision arcs of energy E are the critical points of the functional J_E, and a collision arc γ is said to be *nondegenerate* if it is a nondegenerate critical point of J_E on $\Omega_{\alpha,\beta}$.

Equivalently, we can consider the space $\Omega'_{\alpha,\beta}$ of $W^{1,2}$ curves $u : [0, 1] \to M$, such that $u(0) = C_\alpha$, $u(1) = C_\beta$. For any $(u, \tau_f) \in \Omega'_{\alpha,\beta} \times \mathbb{R}^+$, we have a curve $\gamma : [0, \tau_f] \to M$, defined by $\gamma(t) := u(t/\tau_f)$. Fixed the energy value E, the action of γ is

$$F(\gamma) = \int_0^{\tau_f} \left(L_0(\gamma(t), \dot\gamma(t)) + E\right) dt\,.$$

The collision arcs of energy E starting from C_α and ending at C_β are critical points of this functional on $\Omega'_{\alpha,\beta} \times \mathbb{R}^+$, and nondegeneracy for this functional is equivalent to nondegeneracy for the Maupertuis-Jacobi functional.

These definitions of nondegeneracy are the most natural, but it is quite complicated to use them for verifications in concrete examples. In Subsection 2.3.1 we will describe a sufficient condition for nondegeneracy, to be used in our proofs.

Suppose that the system (L_0) has a finite set of non-degenerate collision arcs $\gamma_k : [0, \tau_k] \to D$, $k \in K$, with the same energy E, where K denote a finite set of labels.

Definition 1.3.2 *A sequence of collision arcs $(\gamma_{k_i})_{i\in\mathbb{Z}}$, $k_i \in K$, with the same energy E, is called a* collision chain *if $\gamma_{k_i}(\tau_{k_i}) = \gamma_{k_{i+1}}(0)$ and the condition of* direction change *is satisfied:*

$$\dot\gamma_{k_i}(\tau_{k_i}) \neq \pm\dot\gamma_{k_{i+1}}(0)\,, \text{ for any } i \in \mathbb{Z}\,. \tag{1.7}$$

Collision chains correspond to paths in the graph Γ with the set of vertices K and the set of edges

$$\Gamma = \{(k, k') \in K^2 |\ \gamma_k(\tau_k) = \gamma_{k'}(0)\,,\ \dot\gamma_k(\tau_k) \neq \pm\dot\gamma_{k'}(0)\}\,. \tag{1.8}$$

Any edge corresponds to an ordered couple of collision arcs which meet transversely at collision, so that the passage from the first arc to the second one through a collision is allowed to form a piece of a collision chain. The paths are formed by sequences of vertices in K such that any two subsequent vertices form an edge: then a path corresponds to a sequence of collision arcs joined transversely at collisions, and a sequence of this type is exactly a collision chain as previously defined.

We now recall the definition of shadowing orbit for a given collision chain.

Definition 1.3.3 *Let $(\gamma_{k_i})_{i\in\mathbb{Z}}$ be a collision chain of energy E. For each $k \in K$ let W_k be a neighbourhood of $\gamma_k([0,\tau_k])$. A trajectory $\gamma : \mathbb{R} \to M \setminus \mathcal{C}$ of energy E for the system (L_ε) is said to* shadow *the chain $(\gamma_{k_i})_{i\in\mathbb{Z}}$ if there exists an increasing sequence of times $(t_i)_{i\in\mathbb{Z}}$ such that $\gamma([t_i, t_{i+1}]) \subset W_{k_i}$.*

In [8] Bolotin and Mackay have proved the following general theorem, valid for systems (L_ε) of the type described so far.

Theorem 1.3.4 (Bolotin-Mackay, 2000) *Given a finite set of nondegenerate collision arcs $\{\gamma_k |\ k \in K\}$, with the same energy E, there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$ and any collision chain $(\gamma_{k_i})_{i\in\mathbb{Z}}$, $k_i \in K$, there exists a unique (up to a time shift) trajectory $\gamma : \mathbb{R} \to D\backslash\mathcal{C}$ of energy E of system (L_ε), which shadows the chain $(\gamma_{k_i})_{i\in\mathbb{Z}}$ within order ε, and at the same time avoids collisions by a distance of order ε. More precisely, there exist constants $B, B' > 0$, independent of ε and the collision chain, and a sequence of times $(t_i)_{i\in\mathbb{Z}}$, such that $|t_{i+1} - t_i - \tau_{k_i}| \leq B\varepsilon$, $dist(\gamma(t), \gamma_{k_i}([0, \tau_{k_i}])) \leq B\varepsilon$ for $t_i \leq t \leq t_{i+1}$, and $dist(\gamma(t), \mathcal{C}) \geq B'\varepsilon$.*

From Theorem 1.3.4 it follows that there is an invariant subset Λ_ε on the energy shell $\{H_\varepsilon = E\}$ on which the system (L_ε) is a suspension of a subshift of finite type. The subshift is given by the shift on the set of paths in the graph Γ, that is the set of all the possible collision chains, and the set Λ_ε is formed by the orbits which shadow them. The important fact about the invariant set Λ_ε is that it is uniformly hyperbolic.

Theorem 1.3.5 (Bolotin-Mackay, 2006) *There exists a cross-section $N \subset \{H_\varepsilon = E\}$, such that the corresponding invariant set $M_\varepsilon = \Lambda_\varepsilon \cap N$ of the Poincaré map is uniformly hyperbolic with Lyapunov exponents of order $\log \varepsilon^{-1}$.*

In particular, the set Λ_ε is uniformly hyperbolic as a suspension of a hyperbolic invariant set with bounded transition times.

We will recall the explicit construction of the cross-section N and the definition of the Poincaré map in Chapter 3: we omit it for the moment, because it would require much machinery, which is not necessary for the sequel. In particular, it is fundamental for this construction a result of Aubry, Mackay and Baesens which gives a characterisation of uniform hyperbolicity through the phonon gap for Frenkel-Kontorova models (see [1]).

In the development of part I, we will apply the Theorems 1.3.4 and 1.3.5 to obtain periodic and chaotic quasi collision orbits for the problem of three centres on the plane. Before starting with our application, we would like to recall analogous results concerning the restricted problem of three bodies, which are still due to Bolotin and Mackay.

1.3.2 Applications to the restricted 3-body problem

In chapter XXXII of the volume III, pp. 362-371, of the famous *Les Méthodes Nouvelle de la Mécanique Céleste* of Poincaré ([48, 1899]), he conjectured the existence of periodic solutions of a particular kind for the three body problem of Celestial Mechanics, in the hypothesis that the second and third mass μ, m are small compared to the primary mass M. These periodic orbits are characterised by the fact that, as μ, m tend to zero, each one of the smaller bodies describes a sequence formed by segments of Kepler ellipses joined at collisions between the two. Poincaré called these kind of trajectories periodic orbits of *second species* , to distinguish them from the *first species* orbits, which do not involve passages close to singularities.

At least in the restricted circular case, Poincaré's conjecture has been proved by Bolotin and Mackay by the use of their Theorem 1.3.4: actually they have proved much more and we are going to briefly illustrate their results. Our work, which is itself an application of Theorem 1.3.4, follows the same ideas and presents similar difficulties. An independent result on the existence of periodic orbits of the second species has been obtained by Font, Nunes and Simó in [20], by a completely different method: we will briefly describe their outcomes at the end of this subsection.

We have defined the planar restricted circular 3-body problem in Section 1.1: we recalled there the equations of motion in the rotating reference system with origin at the centre of mass of the primaries and with normalised dimensionless coordinates (equations (1.2)).

Suppose now that the mass $(1-a)$ of the first primary is much greater than the one of the second primary a. Then $\varepsilon = a$ is our perturbation parameter. With this assumption we can think to the problem as a model of the system formed by the Sun, Jupiter and an Asteroid, with Jupiter supposed to move on a circular orbit about the Sun. It is convenient in this situation to take the rotating system with origin at the position of the Sun, instead of the centre of mass of Sun and Jupiter, so that the position of Jupiter is fixed at $J = (1, 0)$.

With a little abuse of notation we still use the symbols $\mathbf{q} = (x, y)$ to denote the position of the Asteroid with respect to this reference system, even if we have already used them for the rotating frame centred at the centre of mass of the primaries. The Lagrangian function L_ε and the integral of energy H_ε are

$$\begin{aligned} L_\varepsilon(\mathbf{q}, \dot{\mathbf{q}}) &= \frac{1}{2}|\dot{\mathbf{q}}|^2 + x\dot{y} - y\dot{x} - W(\mathbf{q}) - \varepsilon V(\mathbf{q}), \\ H_\varepsilon(\mathbf{q}, \dot{\mathbf{q}}) &= \frac{1}{2}|\dot{\mathbf{q}}|^2 + W(\mathbf{q}) + \varepsilon V(\mathbf{q}), \\ W(\mathbf{q}) &= -\frac{1}{2}|\mathbf{q}|^2 - \frac{1}{|\mathbf{q}|}, \\ V(\mathbf{q}) &= \frac{1}{|\mathbf{q}|} - \frac{1}{|\mathbf{q} - J|} + x \; . \end{aligned} \tag{1.9}$$

The configuration manifold M of L_0 is $\mathbb{R}^2 \setminus \{0\}$. It is clear that this system has the form desired for the application of Theorem 1.3.4 to be possible. In particular, the set $\mathcal{C}$ of singularities contains only one point, the position J of Jupiter.

In [8] Bolotin and Mackay perform an application of their theorem to this situation. What they have to do is to search for a finite number of nondegenerate collision arcs, that is nondegenerate orbits of system (L_0) which start and end at collision with Jupiter, and to construct collision chains with them.

To study the system (L_0), it is better to return to the fixed reference frame with origin at the Sun, in which this system is the well known Kepler problem and Jupiter describes a unit circle around the Sun with frequency 1. It is sufficient to consider the elliptic orbits for the Asteroid and, among them, only the ones which intersect the unit circle are needed. Defining the Jacobi constant $C = -2E$, where E is the fixed energy value for H_0, the ellipses of energy E are completely characterised by C and their mean motion. Indeed, if h is the norm of the angular momentum about O and $\mathcal{E}$ the energy in the fixed coordinate frame, the following relations hold

$$h = \sqrt{\Omega^{-\frac{2}{3}}(1 - e^2)}\,, \quad C = 2(\pm h - \mathcal{E})\,, \tag{1.10}$$

where e is the eccentricity of the ellipse, Ω the frequency or mean motion, and the choices of sign "+" or "−" correspond respectively to the Asteroid moving in the same or opposite direction of Jupiter.

Given $C \in (-\sqrt{8}, 3)$ there is an open interval A_C of frequencies for which any corresponding ellipse crosses the unit circle exactly at two distinct points. To obtain a collision orbit of given C it's enough to choose a rational frequency $\Omega \in A_C$, with $\Omega = \frac{m}{k}$ in lowest terms, then let Jupiter and the Asteroid start at either of the two intersection points. After k revolutions of Jupiter, the Asteroid will have made m revolutions and they will collide again. The delicate point here is to avoid early collision of the Asteroid with Jupiter, in a way that successive collisions happens exactly at the same point in the not-rotating system.

Suppose this is the case. There are exactly two transverse ellipses with the same values of C, Ω which are orbits of the Kepler problem and have a common intersection with the unit circle. This means that there is no problem in constructing a collision chain if we take for any allowed rational frequency Ω both the ellipses.

For any given $C \in (-\sqrt{8}, 3)$, it is proved the existence of a dense subset S_C of rational frequencies in A_C, such that early collisions are excluded. Finally, the following holds:

Theorem 1.3.6 (Bolotin-Mackay, 2000) *In the planar circular restricted 3-body problem with masses $1-\varepsilon, \varepsilon, 0$, for all values of the Jacobi constant $C \in (-\sqrt{8}, 3)$ there exists a dense subset S_C of rationals in the set A_C of allowed frequencies for Kepler ellipses crossing the unit circle, such that for all finite subsets $T \subset S_C$ there exist $\varepsilon_0 > 0$ such that for any sequence $\sigma = (\Omega_i = m_i/k_i)_{i\in\mathbb{Z}}$ in T and $0 < \varepsilon < \varepsilon_0$ there is a unique trajectory of Jacobi constant C near a chain of collision trajectories formed by transforming ellipses of frequencies Ω_i traversed m_i times to the rotating frame, and it converges to the chain as $\varepsilon \to 0$.*

It is worth remarking that the choice to have a whole number of revolutions between two successive collisions with Jupiter is not obligatory, but it's simpler and more reasonable to examine only one type of orbit at a time and this seems at a first sight the more easy case. There is also another question to arise: the Asteroid is assumed to move on the plane of motion of Jupiter, but this constraint can be suppressed.

As was remarked by Poincaré in [48, 1899], segments of Kepler orbits about the Sun between two intersections with a given one, in this case the unit circle, fall into four classes:

1. a whole number of revolutions of a coplanar orbit;

2. a segment of coplanar orbit between distinct intersection points;
3. a whole number of revolutions of a non-coplanar orbit;
4. a segment of a non-coplanar orbit between points at opposite ends of a straight line through the Sun.

As observed in [8], the third type of collision orbits cannot be taken in consideration, because they do not satisfy the nondegenerate condition: indeed they are determined by five conditions on the six orbital elements. In [9] the authors study the orbits of the fourth type instead, and obtain the desired collision arcs for the spatial restricted circular 3-body problem, where the Asteroid is not constrained on the plane of Jupiter's circular motion about the Sun.

They take normalised dimensionless coordinates as in the planar case and consider the frame $Oxyz$, with origin at the Sun, the z-axis orthogonal to the plane of Jupiter's orbit, and rotating anti-clockwise about the z-axis at angular frequency 1 together with Jupiter. The Lagrangian and energy functions have still the form (1.9), putting $\mathbf{q} = (x, y, z)$, $J = (1, 0, 0)$ and $W(\mathbf{q}) = -\frac{1}{2}(x^2 + y^2) - \frac{1}{|\mathbf{q}|}$.

In the application of Theorem 1.3.4 to this case, the more difficult point is to find the algebraic conditions on the parameters to have an arc which starts and ends at collision with Jupiter. First of all, the ellipse must intersect the unit circle at two opposite points and this defines the open interval A_C of allowed frequencies Ω for any given value of the Jacobi constant $C \in (-2, 3)$. The condition of starting and ending at collision is satisfied for a dense subset of A_C. Moreover, collision arcs starting and ending at the same two points of the unit circle and corresponding to the same values of C, Ω are symmetric with respect to the plane of Jupiter's orbit, and then they intersect transversely at the collisions. The final result is

Theorem 1.3.7 (Bolotin-Mackay, 2006) *For any $C \in (-2, 3)$ there exists a dense subset S of the set A_C of allowed frequencies, such that for any finite set $\Lambda \subset S$ there exists $\varepsilon_0 > 0$ such that for any sequence $(\Omega_i)_{i\in\mathbb{Z}}$ in Λ and $\varepsilon \in (0, \varepsilon_0)$ there is a trajectory of the spatial circular restricted three-body problem with Jacobi constant C, which avoids collisions by order ε and is within order ε a concatenation of collision orbits formed from arcs of Kepler ellipses of frequencies Ω_i and with inclination ι_i satisfying $cos(\iota_i) = C/2 - \Omega_i^{2/3}$.*

In [7] Bolotin puts newly the attention on the planar problem, but without assuming Jupiter's orbit to be circular: it is assumed to be an ellipse with focus at the centre of mass of Jupiter and the Sun, and eccentricity $\epsilon \in (0, 1)$. The Hamiltonian of this system is time periodic and cannot be put in the form considered in the Theorem 1.3.4 simply by a change of the reference frame. It is necessary a new result, analogous to Theorem 1.3.4, and valid for a suitable class of time-periodic Hamiltonian functions. Bolotin succeeds in proving such a general shadowing result. Then, in a subsequent work he applies his new theorem to the elliptic problem, proving for small mass ratio and small eccentricity the existence of periodic orbits shadowing collision chains (see [6]). We don't enter further in the details of this problem.

In general, for any problem to which we desire to apply the Theorem 1.3.4, after a finite set of collision arcs has been found, the points to be verified are:

- nondegeneration,
- absence of early collision,
- presence of pairs of arcs transverse at collision,

so that we have the possibility to construct collision chains as defined in Subsection 1.3.1. The characterisation of the collision arcs through suitable parameters and any of these verifications can be difficult and need investigation, even for well known systems (L_0) as the Kepler problem.

For the restricted circular problem of three bodies the assumptions of the Theorem 1.3.4 are satisfied and it is applied to find orbits that shadow collision chains. The Theorem 1.3.5 holds too, so that we can assert that the invariant set formed by the shadowing orbits is uniformly hyperbolic with symbolic dynamics, and we can properly talk of chaotic motion.

As mentioned before, a result of existence of periodic orbits of the second species for the restricted circular three body problem can be obtained by completely different methods, as made in [20]. A direct study of the orbits which have close encounters with the small primary is carried out and the proof in this case is constructive. Indeed, an approximation of the *first return map*, defined on a region of the phase space whose projection is a small circle around the second primary, is explicitly computed. It is proved that the first return map is *horseshoe* like and this allows to conclude about the existence of orbits with consecutive infinite close approaches with the small primary. A complete numerical study of these orbits is carried out in [21]: in particular it is given a proof of the symbolic dynamics for the whole set of planar orbits of second species. This is a very interesting point because it involves *all* second species orbits, while in the results of Bolotin and Mackay the symbolic dynamics is defined only for a subset of periodic orbits of second species, the ones that shadow the collision chains constructed with the fixed finite set of collision arcs.

1.3.3 Chaotic motions for the planar 3-centre problem

We are now going to describe the central problem of the first part of the thesis.

Consider the motion of a particle in the plane, under the gravitational action of three point masses at fixed positions (the planar restricted 3-centre problem). We fix a Cartesian reference system Oxy on the plane and choose suitable dimensionless coordinates such that the two centres with greater masses occupy the positions $C_1 = (1, 0), C_2 = (-1, 0)$. Following the common terminology, we refer to these as the primaries. We suppose for simplicity that the primaries have equal intensities $a_1 = a_2 = a > 0$ (symmetric problem), and we assume all the three centres having positive intensities.

Let $C = (x_0, y_0) \in \mathbb{R}^2 \setminus \{C_1, C_2\}$ be the position of the third centre and $\varepsilon > 0$ be its intensity. We assume ε to be very small and consider the limit $\varepsilon \to 0$: this means that ε is a perturbation parameter. In other words, we make the hypothesis that the conditions are such that we can consider the problem as a one-parameter perturbation of an integrable one: the 2-centre problem.

Let M be the smooth Riemannian manifold $M = \mathbb{R}^2 \setminus \{C_1, C_2\}$, with the induced Euclidean metric on the tangent bundle TM. The configuration space for the motion of a particle in the gravitational field generated by the centres C_1, C_2, C is $M \setminus \{C\}$. Let (x, y) denote the position of the particle on $M \setminus \{C\}$. Then the system has smooth Lagrangian function on $T(M \setminus \{C\})$ given by

$$L_\varepsilon = L_0 + \frac{\varepsilon}{\sqrt{(x-x_0)^2 + (y-y_0)^2}}\,, \tag{1.11}$$

with L_0 the Lagrangian of the symmetric 2-centre problem

$$L_0 = \frac{\dot{x}^2 + \dot{y}^2}{2} + \frac{a}{\sqrt{(x+1)^2 + y^2}} + \frac{a}{\sqrt{(x-1)^2 + y^2}}\ . \tag{1.12}$$

Here we use the Newtonian notation for the derivatives with respect to the time t: $\dot{x} = dx/dt, \dot{y} = dy/dt$. Note that L_0 is a smooth function on M, while L_ε has a Newtonian singularity at the point C. To simplify the notation, let us denote by $W(x, y)$ and $\varepsilon V(x, y)$ the potential energies due respectively to the primaries and the third centre C, so that

$$W(x,y) = -\frac{a}{\sqrt{(x+1)^2 + y^2}} - \frac{a}{\sqrt{(x-1)^2 + y^2}}\,, \tag{1.13}$$

$$V(x,y) = -\frac{1}{\sqrt{(x-x_0)^2 + (y-y_0)^2}}\ . \tag{1.14}$$

The Hamiltonian of the problem has the form

$$H_\varepsilon = H_0 + \varepsilon V(x,y)\,, \tag{1.15}$$

with H_0 the Hamiltonian of the 2-centre problem

$$H_0 = \frac{p_x^2 + p_y^2}{2} + W(x,y)\ . \tag{1.16}$$

We will investigate the question of integrability for negative values of the energy $E < 0$, with $E \to 0$. As recalled in Section 1.2, the problem of three centres has no analytic integrals on positive energy levels, while the case of negative energy is still an open question. To our knowledge, only Bolotin and Negrini have studied the negative energy case in a joined work (see [11] and Subsection 1.2.2), proving the existence of chaotic motions.

We too show existence of chaotic motions but our assumptions are different: while Bolotin and Negrini obtain a perturbation of the 2-centre problem by assuming the third centre to be very far from the other two, we make no hypothesis on the position of the third centre, which can be in principle everywhere on the plane minus the positions of the primaries.

We will see that chaotic motions exist for a dense subset of possible positions of the third centre in $\mathbb{R}^2$. Moreover, the obtained trajectories shadow chains of collision orbits through the third centre. To prove it we will apply the general result of Bolotin and Mackay, Theorem 1.3.4, reminded in Subsection 1.3.1, to show the existence of orbits shadowing collision chains, and then Theorem 1.3.5 to show that these motions define a hyperbolic invariant set.

Our task is to find a finite number of collision arcs, satisfying the assumptions required by Theorem 1.3.4. In our case the only possible collisions are those with the third centre, then the collision arcs are orbits of the 2-centre problem (L_0) which start and end at the third centre C, without intermediate passage through the point C itself.

After the classical regularisation of the 2-centre problem, we find periodic trajectories through C and we show that they do not pass through the primaries if C belongs to a suitably defined dense subset of $\mathbb{R}^2$ (Chapter 2). Then, these are periodic orbits of the 2-centre problem and we can use them to construct the collision arcs. We can verify that the assumption of nondegeneracy is satisfied: this is a fundamental point for the application of Theorem 1.3.4, and we will spent some time on this subject in Section 2.3. We face also the problem of early collision and of transversality at C, and conclude by applying Theorems 1.3.4 and 1.3.5, thus getting our result

Theorem 1.3.8 *Let $I \subset \mathbb{Q}^+$ be a finite set of positive rationals. There exists a dense open subset $X_I \subset M$ of possible positions for the third centre C, such that, fixed $C \in X_I$, the following is true. There is a small value $E_0 > 0$, depending on C and I, such that, fixed an energy value $E \in (-E_0, 0)$, we have that there is $\varepsilon_0 > 0$, such that for any $\varepsilon \in (0, \varepsilon_0)$ and any sequence $(q_k)_{k\in\mathbb{Z}}, q_k \in I$, there exists a trajectory of the planar restricted 3-centre problem (L_ε) on the energy shell $\{H_\varepsilon = E\}$, which avoids collision with the third centre C by order ε and is within order ε a concatenation of pieces of periodic orbits for the planar restricted 2-centre problem (L_0), passing through C and of classes $q_k, k \in \mathbb{Z}$ (see Subsection 2.1.3 for notation).*

The resulting invariant set formed by these orbits is uniformly hyperbolic.

In particular, as it follows from Theorem 1.3.5, fixed a small enough energy value $E < 0$ and $\varepsilon > 0$, there is a cross section in the energy shell $\{H_\varepsilon = E\}$, such that the associated Poincaré map has a hyperbolic invariant set with Lyapunov exponents of order $\log \varepsilon^{-1}$. This invariant set contains infinitely many periodic orbits, corresponding to periodic collision chains. This topic will be treated in Chapter 3.

Theorem 1.3.8 is still true if we substitute in the statement the set X_I with a set X, which is dense in M and is independent of the set of rationals I. In this case, we do not know if the set X is open or not, the only thing that we can say about it is that it is dense (see Remark 2.2.12).

Chapter 2

Shadowing chains of collision orbits in the 3-centre problem

We are going to prove our main result Theorem 1.3.8. We study the problem of three centres on the negative energy level sets, with the assumptions that make it possible to consider the system as a perturbation of the 2-centre problem. In particular, the intensity of the third centre is supposed to be very small. We want to show the existence of a special kind of motion, which follows chains of arcs starting and ending at collision with the third centre and which tends to the collision chain when the mass of the third centre approaches zero. Such orbits are said to shadow the collision chain.

2.1 Periodic orbits of the regularised 2-centre problem

As outlined in the previous chapter, our starting point is the Theorem 1.3.4. In order to apply it, we must find collision arcs with fixed energy $E < 0$ for the unperturbed system (L_0). A natural choice is to look for periodic orbits through the third centre C. As a first step, we recall the classical regularisation of singularities of Euler, for which a brief account can be found in [55, Section 53]. A deep analysis of the 2-centre problem was made by Charlier in [15]: following his approach, we get the existence of infinite classes of periodic orbits for the separated problem obtained after regularisation[1].

2.1.1 The classical regularisation of the planar 2-centre problem

By Bonnet's theorem (see [55]), elliptic and hyperbolic trajectories with foci at the two centres $C_1 = (1, 0), C_2 = (-1, 0)$ are admissible for the planar 2-centre problem, because they are possible trajectories for the central motions with centre of force at one of C_1, C_2. Actually, this result is due to Legendre.

Theorem 2.1.1 (Legendre, 1817) *If a given orbit can be described in each of n given fields of force, taken separately, the velocities at any point P of the orbit being*

[1] A classification of the periodic orbits of the 2-centre problem based on the variation of the energy parameters can be found in [18] and [19]. We will not need this classification: according to it our orbits are all of the same kind, because of the constraints that we impose on the parameters.

$v_1, v_2, \dots, v_n$, respectively, then the same orbit can be described in the field of force which is obtained by superposing all these fields, the velocity at the point P being $(v_1^2 + v_2^2 + \cdots + v_n^2)^{\frac{1}{2}}$.

It is then natural to replace the rectangular coordinates (x, y) by *elliptic coordinates* (ξ, ϕ), defined by the map $x+iy = \cosh(\xi+i\phi)$ from the cylinder $\mathbb{R}\times S^1$ to $\mathbb{R}^2$. This transformation has two ramification points at the two primaries, which in elliptic coordinates are $C_1 = (0, 0), C_2 = (0, \pi)$. The transformation can be written

$$\begin{cases} x = \cosh\xi\cos\phi \\ y = \sinh\xi\sin\phi \end{cases} .$$

Then we have

$$\frac{x^2}{\cosh^2\xi} + \frac{y^2}{\sinh^2\xi} = 1\,, \qquad \frac{x^2}{\cos^2\phi} - \frac{y^2}{\sin^2\phi} = 1\ .$$

From this expressions it is evident that the equations ξ =constant and ϕ =constant define respectively an ellipse and a hyperbola with foci at the centres C_1, C_2 and by Bonnet's theorem they are a particular family of possible trajectories for the system (L_0).

With elliptic coordinates the Lagrangian L_ε becomes

$$L_\varepsilon = \frac{\dot\xi^2 + \dot\phi^2}{2}(\cosh^2\xi - \cos^2\phi) - W(\xi, \phi) - \varepsilon V(\xi, \phi)\,,$$

where the potentials W, V have the form

$$\begin{aligned} W(\xi, \phi) &= -\frac{2a\cosh\xi}{\cosh^2\xi - \cos^2\phi}\,, \\ V(\xi, \phi) &= -\frac{1}{\sqrt{\cosh^2\xi - \sin^2\phi + (x_0^2 + y_0^2) - 2(x_0\cosh\xi\cos\phi + y_0\sinh\xi\sin\phi)}}\ . \end{aligned}$$

Consider the problem on the energy level set $\{H_\varepsilon = E\}$. The Hamiltonian in elliptic coordinates is

$$H_\varepsilon - E = \frac{1}{\cosh^2\xi - \cos^2\phi}\mathcal{H}_\varepsilon\,,$$

with

$$\mathcal{H}_\varepsilon = \frac{{p_\xi}^2 + {p_\phi}^2}{2} - 2a\cosh\xi - [E - \varepsilon V(\xi, \phi)]\,(\cosh^2\xi - \cos^2\phi)\,,$$

where the symbols p_ξ, p_ϕ denote the conjugate momenta. If we denote by $Z = (p_\xi, p_\phi, \xi, \phi)$, the system on the energy level $\{H_\varepsilon = E\}$ is

$$\begin{cases} \dot Z = \dfrac{1}{\cosh^2\xi - \cos^2\phi}J\nabla\mathcal{H}_\varepsilon(Z) \\ \mathcal{H}_\varepsilon(Z) = 0 \end{cases} .$$

The orbits of the problem with Hamiltonian (H_ε) on the energy level $\{H_\varepsilon = E\}$ are, up to time parametrisation, orbits of the system with Hamiltonian $\mathcal{H}_\varepsilon$ on the energy level $\{\mathcal{H}_\varepsilon = 0\}$. The regularised Hamiltonian $\mathcal{H}_\varepsilon$ has no singularities at C_1, C_2: this means that an orbit for $\mathcal{H}_\varepsilon$ is an orbit for H_ε only if it does not pass through the primaries. The new time parameter τ is given by

$$\tau = \int_0^t \frac{1}{\cosh^2 \xi(s) - \cos^2 \phi(s)} ds \,, \tag{2.1}$$

and this formula allows us to pass from a solution for the system ($\mathcal{H}_\varepsilon$), on the zero energy level and not passing through the primaries, to a solution for (H_ε) with energy E.

The Lagrangian corresponding to $\mathcal{H}_\varepsilon$ is

$$\mathcal{L}_\varepsilon = \frac{(\xi')^2 + (\phi')^2}{2} + 2a \cosh \xi + [E - \varepsilon V(\xi, \phi)] (\cosh^2 \xi - \cos^2 \phi) \,,$$

where the prime sign denote derivation with respect to the new time parameter: $\xi' = d\xi/d\tau, \; \phi' = d\phi/d\tau$.

$\mathcal{L}_0$ and $\mathcal{H}_0$ are the Lagrangian and the Hamiltonian of the regularised 2-centre problem, obtained after the passage to elliptic coordinates and the reparametrisation of time: this is the regularisation of Euler. The singularities at the centres C_1, C_2 have disappeared and the solutions of the original problem correspond to the ones of the regularised system by a reparametrisation of the time. After only the change of coordinates the system (L_0) becomes of Liouville type (see [55]) and it is known that a system of this type must separate after a suitable choice of the time parameter, as given by equation (2.1). We are going to remind some properties of the obtained separated problem in the next section. Moreover, we will put the ground for the investigations of the subsequent sections by a proper choice of the parameters and of the intervals of values allowed for them.

2.1.2 The separated problem.

We have to find orbits of the 2-centre problem (L_0) with fixed energy $E < 0$, then we put $\varepsilon = 0$ and study the regularised system on the energy level $\{\mathcal{H}_0 = 0\}$. The Lagrangian is

$$\mathcal{L}_0 = \frac{(\xi')^2 + (\phi')^2}{2} + 2a \cosh \xi + E(\cosh^2 \xi - \cos^2 \phi) \;.$$

The system separates and we have the two one-dimensional problems

$$\begin{cases} \dfrac{(\xi')^2}{2} - 2a \cosh \xi - E \cosh^2 \xi = -E_1 \\ \dfrac{(\phi')^2}{2} + E \cos^2 \phi = E_1 \end{cases} . \tag{2.2}$$

If $E \geq 0$, then every motion in ξ is unbounded except for the equilibrium point $\xi = 0$. The potential in ξ has a maximum at $\xi = 0$ and goes to minus infinity for $\xi \to \pm\infty$ (see Figure 2.1).

But we are interested in the case of negative energies, $E < 0$, then we can write the system (2.2) as

$$\begin{cases} \dfrac{(\xi')^2}{2} - 2a \cosh \xi + |E| \cosh^2 \xi = -E_1 \\ \dfrac{(\phi')^2}{2} - |E| \cos^2 \phi = E_1 \end{cases} . \tag{2.3}$$

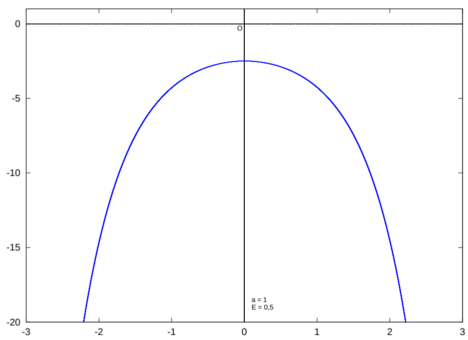

Figure 2.1. Graph of the potential energy of ξ for $E \geq 0$: in this picture $a = 1$ and $E = 0.5$.

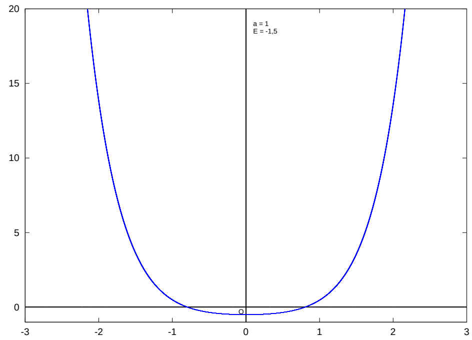

Figure 2.2. Graph of the potential energy of ξ when $|E| \geq a$: in this picture $a = 1$ and $E = -1.5$.

If $E < 0$ and $|E| \geq a$, then the potential of ξ has a minimum at $\xi = 0$ and it goes to infinity for $\xi \to \pm\infty$ (see Figure 2.2). Any motion in ξ is bounded and periodic.

We want to consider values of the energy E approaching zero, then we assume $|E| < a$. In this case the potential in the variable ξ has a local maximum when $\xi = 0$, with negative maximal value $|E| - 2a$, and two minima $\pm\xi_m$, defined by $\cosh(\pm\xi_m) = a/|E|$, where the potential takes the value $-a^2/|E|$. Furthermore, it goes to infinity for $\xi \to \pm\infty$. Finally, if we choose $-E_1 > |E| - 2a$, i.e. we consider a level over the separatrix, the motion in ξ is periodic with two inversion points at ξ_- and ξ_+, defined by the relation

$$\cosh(\xi_\pm) = -\frac{a}{E}\left(1 + \sqrt{1 + \frac{EE_1}{a^2}}\right) .$$

The motion in ϕ is simply the motion of the standard pendulum. When $E_1 > 0$ we have rotational closed orbits for ϕ and there exist action-angle variables. The graphs of the potential energies for ξ and ϕ separately are given in Figures 2.3 and 2.4 below, when $a = 1$ and $E = -0.5$.

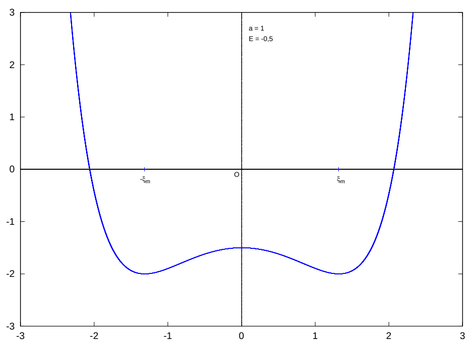

Figure 2.3. Potential energy for the variable ξ when $a = 1$ and $E = -0.5$.

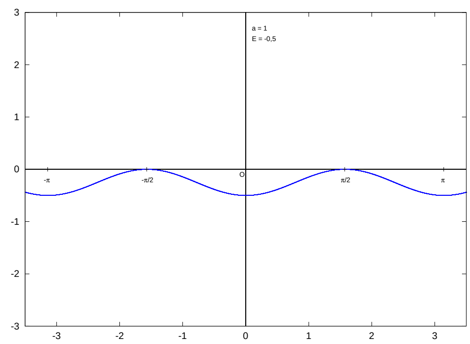

Figure 2.4. Potential energy for the variable ϕ when $a = 1$ and $E = -0.5$.

After these considerations we make the following assumptions:

$$|E| < a\,, \quad E_1 > 0\,, \quad |E| + E_1 < 2a\ . \tag{2.4}$$

For future convenience, we scale the energy parameter E_1 and define

$$A_1 := E_1/2a\,, \quad \beta := |E|/E_1\ .$$

In this manner the conditions (2.4) become

$$A_1, \beta > 0\,, \quad 2\beta A_1 < 1\,, \quad A_1 < \frac{1}{1+\beta}\,, \tag{2.5}$$

and the regularised system (2.3) takes the form

$$\begin{cases} \dfrac{(\xi')^2}{4a} = \cosh\xi - \beta A_1 \cosh^2\xi - A_1 \\ \dfrac{(\phi')^2}{4a} = \beta A_1 \cos^2\phi + A_1 \end{cases} . \tag{2.6}$$

With the assumptions (2.5) on β, A_1, both the one-dimensional motions are periodic, with periods T_1, T_2 respectively for ξ, ϕ, which depends only on the values of β, A_1. If the periods T_1, T_2 have rational ratio $T_1/T_2 \in \mathbb{Q}$, then the corresponding orbit on the cylinder $\mathbb{R} \times S^1$ is periodic. To find a periodic orbit for the system $(\mathcal{L}_0)$, corresponding to a fixed value of the energy parameter β, we have to show that there is at least a value of A_1 for which T_1, T_2 have rational ratio. Before doing that we must compute the analytical expressions of the periods.

Lemma 2.1.2 *Consider the system (2.6), with the assumptions (2.5). Then the one-dimensional motions are both periodic with periods T_1, T_2, for the coordinates ξ, ϕ respectively, given by*

$$T_1 = \frac{2\sqrt{2a^{-1}}}{\sqrt[4]{1-4\beta A_1^2}}\mathrm{K}(\kappa_1)\,, \qquad T_2 = \frac{2\sqrt{a^{-1}}}{\sqrt{A_1(1+\beta)}}\mathrm{K}(\kappa_2)\,,$$

where

$$\kappa_1^2 = \frac{A_1(1-\beta) + \sqrt{1-4\beta A_1^2}}{2\sqrt{1-4\beta A_1^2}}\,, \quad \kappa_2^2 = \frac{\beta}{1+\beta}\,, \quad \kappa_1, \kappa_2 > 0\,,$$

and $\mathrm{K}(\kappa)$ *is the elliptic integral of the first type:*

$$\mathrm{K}(\kappa) = \mathrm{cn}^{-1}(0,\kappa) = \int_0^1 \frac{dv}{\sqrt{(1-v^2)(1-\kappa^2 v^2)}}\ .$$

Proof. It is a straightforward computation, starting from the integral expressions of $\tau(\xi), \tau(\phi)$.
As functions of the parameters β, A_1 the inversion points $\xi_\pm$ are given by

$$\cosh(\xi_\pm) = \frac{1 \pm \sqrt{1-4\beta A_1^2}}{2\beta A_1}\ . \tag{2.7}$$

The progressive motion of ξ is given by

$$\tau(\xi) = \frac{1}{2\sqrt{a}\sqrt{\beta A_1}} \int_{\xi_0}^{\xi} \frac{ds}{\sqrt{(\cosh\xi_+ - \cosh s)\left(\cosh s - \frac{2A_1+1-\sqrt{1-4\beta A_1^2}}{1+2\beta A_1+\sqrt{1-4\beta A_1^2}}\right)}}\ .$$

We substitute $u = \tanh(\frac{\xi}{2})$ and obtain

$$\tau = \frac{1}{\sqrt{a}\sqrt{1 + A_1(\beta+1)}} \int_{u_0}^{u} \frac{ds}{\sqrt{(u_+^2 - s^2)\,(s^2 + d^2)}} \,,$$

where $u_0 = \tanh(\frac{\xi_0}{2})$ and

$$u_+^2 = \frac{A_1(1-\beta) + \sqrt{1 - a\beta A_1^2}}{1 + A_1(1+\beta)} \,, \qquad d^2 = \frac{1 - A_1(1+\beta)}{A_1(1-\beta) + \sqrt{1 - 4\beta A_1^2}} \,.$$

Clearly we have $u_+ = \tanh(\frac{\xi_+}{2})$. Define

$$I_1 = \int_0^{u_+} \frac{ds}{(u_+^2 - s^2)(s^2 + d^2)} \,,$$

so that the period for ξ is $T_1 = \frac{4}{\sqrt{a}\sqrt{1+A_1(\beta+1)}} I_1$. With the substitution $s = vu_+$ we obtain

$$I_1 = \frac{\kappa_1}{u_+} \int_0^1 \frac{dv}{\sqrt{(1-v^2)(\kappa_1^2 v^2 + 1 - \kappa_1^2)}} = \frac{\kappa_1}{u_+} \mathrm{cn}^{-1}(0, \kappa_1) \,,$$

and, after the change of variable $z = \sqrt{1 - v^2}$, this gives the expression of T_1 that we desired.

Now we pass to the computation of T_2. The progressive motion for ϕ is given by

$$\tau(\phi) = \frac{1}{2\sqrt{a}} \int_{\phi_0}^{\phi} \frac{ds}{\sqrt{A_1(\beta \cos^2 s + 1)}} \,,$$

then the period is

$$T_2 = \frac{1}{\sqrt{a}\sqrt{A_1(1+\beta)}} \int_0^{\pi} \frac{ds}{\sqrt{1 - \kappa_2^2 \sin^2 s}} \,,$$

with $\kappa_2 = \frac{1}{1+\beta}$. We substitute $\nu = \frac{\pi}{2} - s$ for $s \in (0, \frac{\pi}{2})$ and $\nu = s - \frac{\pi}{2}$ for $s \in (\frac{\pi}{2}, \pi)$, thus obtaining

$$T_2 = \frac{2}{\sqrt{a}\sqrt{A_1(1+\beta)}} \int_0^{\frac{\pi}{2}} \frac{d\nu}{\sqrt{1 - \kappa_2^2 \cos^2 \nu}} \,.$$

Putting $v = \cos\nu$ we obtain the desired expression for T_2. ∎

Remark 2.1.3 *Observe that* $\lim_{\kappa \to 1^-} \mathrm{K}(\kappa) = +\infty$ *and*

$$\kappa_1 = 1 \iff \begin{cases} \beta & < & 1 \\ A_1 & = & \frac{1}{1+\beta} \end{cases} \,.$$

In particular, the second condition means that the motion of ξ takes place on the separatrix energy level. Then, if $\beta < 1$ the period T_1 goes to infinity as $A_1 \to \frac{1}{1+\beta}$:

$$\lim_{A_1 \to \frac{1}{1+\beta}} T_1 = +\infty \,.$$

Correspondingly, for T_2 we have:

$$\lim_{A_1 \to 0} T_2 = +\infty \ .$$

We are interested to the limit $E \to 0$, then we can suppose that the parameter E_1 is greater than $|E|$. It corresponds to make the hypothesis that $\beta < 1$. Our definitive assumptions are:

$$\beta \in (0,1)\,, \quad 0 < A_1 < \frac{1}{1+\beta} \ .$$

2.1.3 Periodic orbits of the regularised 2-centre problem

The expressions of the periods T_1, T_2 given in Lemma 2.1.2 and the Remark 2.1.3 give the limits:

$$\lim_{A_1 \to \frac{1}{1+\beta}} T_1 = +\infty\,, \qquad \lim_{A_1 \to 0} T_2 = +\infty\,,$$

$$\lim_{A_1 \to 0} T_1 = \frac{4}{\sqrt{2a}}\mathrm{K}(\frac{1}{\sqrt{2}})\,, \qquad \lim_{A_1 \to \frac{1}{1+\beta}} T_2 = \frac{2}{\sqrt{a}}\mathrm{K}(\sqrt{\frac{\beta}{1+\beta}})\ . \tag{2.8}$$

Moreover, from the definition of $\mathrm{K}(\kappa)$, we easily conclude that, fixed $\beta \in (0,1)$, T_2 is a strictly decreasing function of $A_1 \in (0, \frac{1}{1+\beta})$, while T_1 is strictly increasing. Indeed, $K(\kappa)$ is a strictly increasing function of κ^2, κ_2 does not depend on A_1 and κ_1^2 has positive derivative with respect to A_1 given by

$$\frac{\partial \kappa_1^2}{\partial A_1} = \frac{1-\beta}{2}(1 - 4\beta A_1^2)^{-\frac{3}{2}} > 0\ .$$

By these observations, we have shown the following

Proposition 2.1.4 *Let $\beta \in (0,1)$ be fixed. For any positive rational $q \in \mathbb{Q}^+$, there exists a unique value $\hat{A}_1(\beta, q) \in (0, \frac{1}{1+\beta})$ for A_1, such that*

$$qT_1(\beta, \hat{A}_1) = T_2(\beta, \hat{A}_1)\ . \tag{2.9}$$

In particular, the system (2.6) has a periodic solution in correspondence of the value $\hat{A}_1$, with energy $E = -2a\beta\hat{A}_1$. ■

This situation is drawn in the Figures 2.5 and 2.6 below, where the value of β is taken to be $\frac{1}{7}$.

We have seen that for any $\beta \in (0,1)$ and any positive rational $q \in \mathbb{Q}^+$, there is a periodic orbit for the regularised system $(\mathcal{L}_0)$ with energy given by the relation $E = -2a\beta\hat{A}_1$. We can classify these periodic orbits, identifying each *class* with the rational number q. Then for any fixed value of the parameter $\beta \in (0,1)$, we have exactly one value of $\hat{A}_1$ for each class $q \in \mathbb{Q}^+$. Orbits of different classes do not have the same energy E; more precisely we have

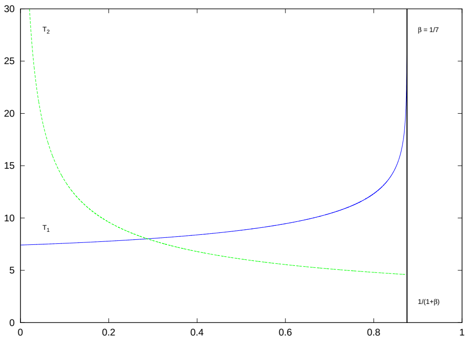

Figure 2.5. Graphs of the periods T_1 and T_2 when $\beta = \frac{1}{7}$.

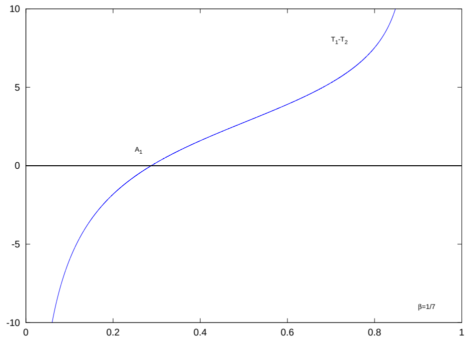

Figure 2.6. Graph of the difference $T_1 - T_2$ when $\beta = \frac{1}{7}$.

Proposition 2.1.5 *Let $\beta \in (0,1)$ fixed. The function $\hat{A}_1(\beta, \cdot) : \mathbb{Q}^+ \to (0, \frac{1}{1+\beta})$, defined by the equality (2.9), is strictly decreasing. Moreover, there exist the limits*

$$\lim_{q\to 0^+} \hat{A}_1(\beta, q) = \frac{1}{1+\beta}\,, \qquad \lim_{q\to +\infty} \hat{A}_1(\beta, q) = 0\ .$$

Proof. If $q > q'$ then $qT_1(\beta, A_1) > q'T_1(\beta, A_1)$ and we have

$$T_2(\beta, \hat{A}_1(\beta, q)) = qT_1(\beta, \hat{A}_1(\beta, q)) > q'T_1(\beta, \hat{A}_1(\beta, q)))\ .$$

T_2 is strictly decreasing with respect to A_1, while T_1 is strictly increasing, then $\hat{A}_1(\beta, q) < \hat{A}_1(\beta, q')$.

The monotony of $\hat{A}_1(\beta, \cdot)$ assures the existence of the limits

$$L_1 = \lim_{q\to 0^+} \hat{A}_1(\beta, q)\,, \quad L_2 = \lim_{q\to\infty} \hat{A}_1(\beta, q)\ .$$

Clearly $0 \leq L_2 < L_1 \leq \frac{1}{1+\beta}$ and from the knowledge of the limits (2.8), we easily obtain the desired value for L_1, L_2. Indeed, we have by definition $qT_1(\hat{A}_1) = T_2(\hat{A}_1)$, for any $q \in \mathbb{Q}^+$, and then

$$\lim_{q\to 0^+} qT_1(\hat{A}_1) = \lim_{q\to 0^+} T_2(\hat{A}_1)\,, \qquad \lim_{q\to +\infty} qT_1(\hat{A}_1) = \lim_{q\to +\infty} T_2(\hat{A}_1)\ .$$

Since $\lim_{q\to 0^+} T_2(\hat{A}_1) \gneqq 0$ then $\lim_{q\to 0^+} qT_1(\hat{A}_1) \gneqq 0$, which is possible only when $\lim_{q\to 0^+} T_1(\hat{A}_1) = +\infty$, that is when $L_1 = \frac{1}{1+\beta}$. On the other hand, $\lim_{q\to +\infty} qT_1(\hat{A}_1) = +\infty$ then $\lim_{q\to +\infty} T_2(\hat{A}_1) = +\infty$, which implies that $L_2 = 0$. ■

It follows that periodic orbits with many "loops" in ξ and few in the variable ϕ tend to the separatrix level for ϕ; viceversa, orbits with many loops in ϕ tend to the separatrix level for ξ. In other words, if we increase only for a single variable the number of loops before the orbit closes, we will obtain a limit orbit which does not close anymore in finite time. For example, take $q = m/n$, with $m, n \in \mathbb{N}$. Increasing the number of loops in ξ corresponds to make m larger and consequently $\hat{A}_1$ smaller. The periodic orbit increases the number of "oscillations" in ξ, while making the same number of revolutions in the variable ϕ. As m goes to infinity we have that the corresponding orbit makes an infinite number of times the same trajectory in ξ, tending to close, but without being able to reach the limit values $\phi = \pm\pi/2$ in a finite time interval, in the future and in the past respectively: in particular, it does not complete even a single revolution for ϕ. Furthermore, the energy $E = -2a\beta\hat{A}_1$ tends to zero.

We conclude that to form easily a finite set of collision arcs with the same energy, we should fix $q \in \mathbb{Q}^+$ and look for collision arcs only in the set of periodic orbits of the same class q.

2.2 Construction of collision arcs

In the previous section we have found infinite classes of periodic orbits for the regularised problem $\mathcal{L}_0$. We would like to use them for constructing collision arcs.

The next step is then to show that, among the periodic orbits of the regularised system ($\mathcal{L}_0$), there is at least one which passes through the third centre C. Actually this is true if the parameter β is sufficiently small. We must also verify that the obtained orbits are solutions of the not regularised problem (L_0), that is they do not pass through the primaries: this is the most delicate point of the proof of Theorem 1.3.8.

2.2.1 Periodic orbits through the third centre

Let $(\xi_0, \phi_0) \in \mathbb{R} \times S^1 \setminus \{(0,0),(0,\pi)\}$ be fixed elliptic coordinates for the position of the third centre C. Then, among the orbits corresponding to the value $\hat{A}_1(\beta, q)$, surely there is one which pass through the centre C, if $\xi_0 \in (\xi_-(\beta,\hat{A}_1), \xi_+(\beta,\hat{A}_1))$, where $\xi_\pm$ are the inversion points. Note that in Cartesian coordinates this corresponds to say that the centre C lies in the region internal to the ellipse defined by the equation $\xi = \xi_+$.

By construction, for each $\beta \in (0,1)$, we have $\hat{A}_1 \in (0, \frac{1}{1+\beta})$, then $\lim_{\beta\to 0} \beta\hat{A}_1 = 0$ and

$$\lim_{\beta\to 0} \cosh(\xi_\pm) = \lim_{\beta\to 0} \frac{1+\sqrt{1-4\beta\hat{A}_1^2}}{2\beta\hat{A}_1} = +\infty \ .$$

This means that, as β tends to zero, the ellipse of equation $\xi = \xi_+$, which encloses the orbits associated to $\hat{A}_1$, becomes larger and larger, tending to cover all the plane. Then for β sufficiently small the point C falls into the region internal to the ellipse. We conclude that

Proposition 2.2.1 *Fixed a class $q \in \mathbb{Q}^+$, there exists $\beta_0 > 0$, such that for any $\beta \in (0, \beta_0)$ there is a periodic orbit of system ($\mathcal{L}_0$), associated with the value $\hat{A}_1(\beta, q)$, which passes through C, and the coordinate ξ_0 is not an inversion point of the corresponding one-dimensional motion in ξ.* ■

Remark 2.2.2 *At this point it is worth to give an estimate of the smallness of β. A sufficient condition to have $|\xi_0| < \xi_+(\beta,\hat{A}_1))$ is that $\beta \le \beta_0$ with $\beta_0 < \frac{1}{\cosh\xi_0}$.*

If $\xi_0 = 0$ *we have nothing to show. Then suppose* $\xi_0 \neq 0$. *We must have* $\cosh(\xi_0) < \cosh(\xi_+)$. *From (2.7) we have*

$$2\beta\hat{A}_1\cosh(\xi_0) - 1 < \sqrt{1 - 4\beta\hat{A}_1^2}\ .$$

Squaring we obtain

$$\hat{A}_1(\beta\cosh^2(\xi_0) + 1) < \cosh(\xi_0)\ .$$

Since $\hat{A}_1(\beta, q) < \frac{1}{1+\beta}$, *a sufficient condition is*

$$\beta\cosh^2(\xi_0) + 1 < (1+\beta)\cosh(\xi_0)$$

and then if $\xi_0 \neq 0$ *we obtain* $\beta < \frac{1}{\cosh(\xi_0)}$.

Then it seems to us that the parameter β *should be really small, because the hyperbolic function grows very fast. For example* $\operatorname{acosh}(10^4) \simeq 9.9$ *and* $\operatorname{acosh}(10^5) \simeq 12.2$, *then if* ξ_0 *is of order* 10 *we must choose* β *of order less then* 10^{-4}. *Nevertheless, we must bear in mind that this is a very rough estimate, made with the limit value* $\frac{1}{1+\beta}$ *for* $\hat{A}_1$*: in general, the value of* $\hat{A}_1$ *can be much smaller and it implies an enlargement of the ellipse of admissible positions* $\xi = \xi_+$.

Note that the periodic orbits associated with the same value of $\hat{A}_1$ differ only for the sign of the velocities ξ_0', ϕ_0', at the centre C. Thus we have exactly two orbits on the configuration space $\mathbb{R} \times S^1$: if one has velocity (ξ_0', ϕ_0') at C, the other has velocity $(-\xi_0', \phi_0')$. The remaining two possibilities give the same orbits, but with the opposite direction of motion. Moreover, $\xi_0' \neq 0$, because ξ_0 is not an inversion point. Then the two trajectories corresponding to $\hat{A}_1$ meet transversely at C on the cylinder $\mathbb{R} \times S^1$.

We remark that there is the possibility that the two orbits coincide: it can happen when the trajectory has an autointersection at (ξ_0, ϕ_0) before closing. This is a case of *early collision* and it will be treated in Proposition 2.2.13.

The obtained solutions are not yet the collision arcs that we desire. In fact, they are orbits for the regularised system $(\mathcal{L}_0)$: for being orbits of the 2-centre problem with Lagrangian L_0, it's enough they do not pass through the primaries C_1, C_2. This is a delicate problem and to face it we will need a general result about the regularity of $\hat{A}_1$ as function of β, in a neighbourhood of $\beta = 0$.

2.2.2 Avoiding collision with the primaries: first exceptions

In this subsection we describe some properties of the periodic orbits of the regularised problem $(\mathcal{L}_0)$, which pass at least through one of the primaries. Note that periodic orbits through the primaries exist for any value of $\beta \in (0,1)$, because a primary has $\xi = 0$. Then we consider the periodic orbits through the third centre C and show how it is possible to exclude the passage through the primaries for some particular positions of C.

Proposition 2.2.3 *Let* $\beta \in (0,1)$ *arbitrarily fixed. Given* $q \in \mathbb{Q}^+$, *let* $m, n \in \mathbb{N}$ *such that* $q = m/n$ *and* $(m,n) = 1$. *Let* γ *be a periodic orbit of the system* $(\mathcal{L}_0)$ *associated with* $\hat{A}_1(\beta, q)$ *and suppose that it passes through one of the centres* C_1, C_2, *which have elliptic coordinates* $(0,0), (0,\pi)$ *respectively.*

If n is odd, then the orbit goes through both the primaries in a period, and the collisions happens at a time distance of one-half the period from each other.

If n is even, then the orbit passes through only one of the primaries and it happens two times in a period, at a distance of one-half the period. In the configuration space $\mathbb{R} \times S^1$ the orbit has a transverse self-intersection at the position of the centre.

Proof. The system $(\mathcal{L}_0)$ has the form (2.6). Without loss of generality we can suppose that the orbit $\gamma(\tau) = (\xi(\tau), \phi(\tau))$ passes through one of the centres C_1, C_2 at time $\tau = 0$. The orbit γ collides with a primary at time $\tau \neq 0$ if and only if $\xi(\tau) = 0$ and $\phi(\tau) \in \{0, \pi\}$. Then we must have $\tau = k\frac{T_1}{2} = j\frac{T_2}{2}$, with $k, j \in \mathbb{Z} \setminus \{0\}$. Then $\frac{k}{j} = \frac{m}{n} = q$ and this implies that there is $i \in \mathbb{Z}$ such that $k = im$ and $j = in$, because $(m, n) = 1$. It follows that $\gamma(\tau)$ is a primary if and only if $\tau = im\frac{T_1}{2} = in\frac{T_2}{2} = i\frac{T}{2}$, where $T = mT_1 = nT_2$ is the period of the orbit. This concludes the proof. ∎

In the Figures 2.7, 2.8, 2.9 and 2.10 we have drawn in the Cartesian reference frame some periodic orbits which pass through the primaries. Note that while in

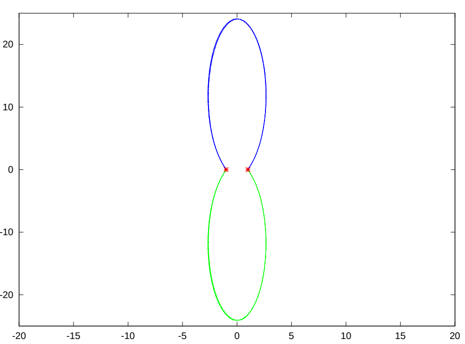

Figure 2.7. Periodic orbits colliding with the primaries corresponding to the values: $a = 1$, $\beta = \frac{1}{7}$, $q = 1$.

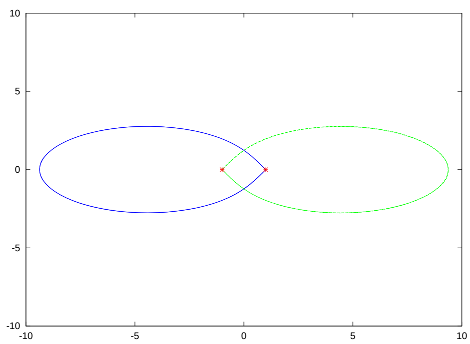

Figure 2.8. Periodic orbits colliding with the primaries corresponding to the values: $a = 1$, $\beta = \frac{1}{7}$, $q = \frac{1}{2}$.

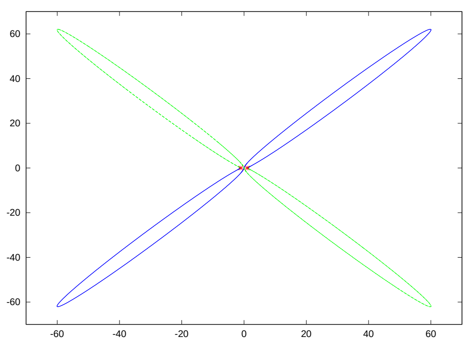

Figure 2.9. Periodic orbits colliding with the primaries corresponding to the values: $a = 1$, $\beta = \frac{1}{7}$, $q = 2$.

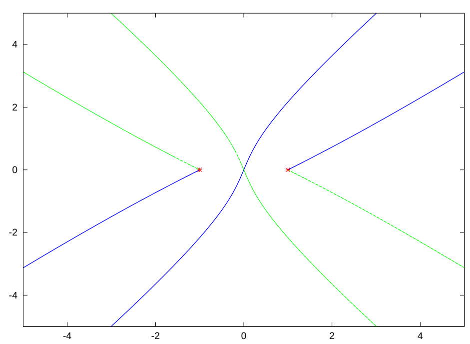

Figure 2.10. An enlargement of Figure 2.9.

the configuration space $\mathbb{R} \times S^1$ the orbits passes really *through* the primaries with finite velocities, when we turn to the Cartesian coordinates instead this passage transforms in a *collision* and a solution which arrives at a primary reverses the direction of motion at the instant of collision. In other words, the passage to the Cartesian coordinates implies that the velocities at the primaries become zero and these positions become inversion points. This is not true when we consider the original time parameter t instead of τ. In this case, when a collision with a primary occurs, the module of the velocity tends to infinity and there is not an inversion of the motion: the motion cannot continue through a collision. According to Proposition 2.2.3, after passing from elliptic to Cartesian coordinates and maintaining the time parameter τ, we have two collisions in a period and the same trajectory between two collisions is followed twice, one time in a direction and the other time in the reversed one.

If we assume $q = 1$, then, using the above Proposition 2.2.3, we can exclude at once the collision with the primaries for some simple cases. To simplify notation we place $\hat{A}_1(\beta) = \hat{A}_1(\beta, 1)$.

Proposition 2.2.4 *Let γ be a periodic orbit through the third centre C, corresponding to the value $\hat{A}_1(\beta)$, with $\beta < \beta_0$, as in Proposition 2.2.1. Let (ξ_0, ϕ_0) be fixed elliptic coordinates for the centre C. If $\phi_0 = k\frac{\pi}{2}, k \in \mathbb{Z}$, or $\xi_0 = 0$, then the orbit γ cannot pass through the primaries.*

Remark 2.2.5 *Note that the positions of C with $\phi_0 = k\frac{\pi}{2}, k \in \mathbb{Z}$, and the ones with $\xi_0 = 0$ correspond in Cartesian coordinates (x, y) to points on the coordinate axes. In particular, if $\xi_0 = 0$ then the three centres are collinear on the x-axis, and C lies between the primaries. If $\phi_0 = k\pi$ then the three centres are still aligned on the x-axis, but C is external. Finally, if $\phi_0 = (2k+1)\frac{\pi}{2}$, then the centre C lies on the y-axis and the configuration of the centres is symmetric with respect to this axis.*

In particular, Proposition 2.2.4 says that, for $q = 1$, the periodic orbits through the primaries intersect the coordinate axes only at the primaries and for $\xi = \xi_\pm$ as in Figure 2.7.

Proof. We are in the case $q = 1$, then from Proposition 2.2.3 we know that the orbit γ passes through a primary if and only if for any time τ such that $\xi(\tau) = 0$, we have $\phi(\tau) = k\pi, k \in \mathbb{Z}$, and viceversa. The centre C does not coincide with a primary, then it cannot happen that $\phi_0 = k\pi, k \in \mathbb{Z}$, and $\xi_0 = 0$ at the same time and in these cases the statement is obvious. Now suppose $\phi_0 = (2k+1)\frac{\pi}{2}$. The shortest time to pass from the centre C to a position with $\phi = i\pi, i \in \mathbb{Z}$, is $\frac{T}{4}$, where $T = T_1 = T_2$ is the period of γ. Look at the variable ξ: the only positions which have a time distance of $\frac{T}{4}$ from $\xi = 0$ are the inversion points $\xi_\pm$, but $\xi_0 \neq \xi_\pm$. ■

Remark 2.2.6 *Note that the proof of Proposition 2.2.4 cannot be generalised to arbitrary fixed values of $q \in \mathbb{Q}^+$. For example, if $q = 1/2$, then, if a periodic orbit passes through C_1, it certainly passes through a point with elliptic coordinates $(\xi, \pi/2)$, with $\xi \in (0, \xi_+)$: in fact the time passed from the last passage through C_1 to this point is $T_2/4 = T_1/8$. Nevertheless, it's easy to see that even for this case the passage through the primaries is excluded when $\xi_0 = 0$ or $\phi_0 = k\pi$: indeed, as we* *can see in Figure 2.8, an orbit which collides with a primary intersects the x-axis only at the primary and when $\xi = \xi_\pm$.*

2.2.3 Avoiding collision with the primaries: the general theorem

In this subsection we will obtain the central result of the chapter: we will show that for almost all the possible positions of the third centre in $\mathbb{R} \times S^1$, the periodic orbits through C corresponding to $\hat{A}_1$ don't collide with the primaries for any sufficiently small value of β. It means that they are solutions of the not-regularised system (L_0) and allows us to proceed with the final verifications, in order to apply Theorem 1.3.4.

Our first step is to study the regularity of the function $\hat{A}_1(\beta, q)$ with respect to the real parameter β: in particular, we are interested in the behaviour near $\beta = 0$.

Lemma 2.2.7 *Let $q \in \mathbb{Q}^+$ be fixed. Then $\hat{A}_1$ is a smooth function of $\beta \in (0,1)$ and it can be smoothly extended to $\beta = 0$: in particular, there exists the limit*

$$\hat{A}_1(0,q) := \lim_{\beta \to 0} \hat{A}_1(\beta, q) \in (0,1) \ ,$$

the periods T_1 and T_2 are smooth for $\beta = 0$ and

$$qT_1(0, \hat{A}_1(0,q)) = T_2(0, \hat{A}_1(0,q)) \ .$$

Proof. Denote by F the function

$$F(\beta, A_1) := [qT_1 - T_2](\beta, A_1) \, ,$$

defined on the domain $\mathcal{D} = \Big\{(\beta, A_1) \in \mathbb{R}^2 | \quad \beta \in (0,1) \, , \ 0 < A_1 < \frac{1}{1+\beta}\Big\}$.

We observe that κ_1, κ_2 are C^∞ functions of $(\beta, A_1) \in \mathcal{D}$, and $\kappa_1, \kappa_2 \in (0,1)$; then, by derivation under the integral sign, we conclude that F is C^∞ on the same domain. Furthermore $\frac{\partial F}{\partial A_1} \gneqq 0$, then from the implicit function theorem we can assert the regularity of $\hat{A}_1$ on $(0,1)$.

We want to extend the definition of the function $\hat{A}_1$ to $\beta = 0$. First of all we prove that the function F can be smoothly extended to $\beta = 0$. We study T_1 and T_2 separately.

We start from T_1. If $\beta \in [0,1)$ and $A_1 < \frac{1}{1+\beta}$, then $(1 - 4\beta A_1^2) \gneqq 0$. It follows that k_1^2 is well defined and C^∞ for $\beta = 0$ and that $k_1^2 \in (\frac{1}{2}, 1)$. Then T_1 is C^∞ for $\beta = 0$. Now see T_2. For $\beta \in [0,1)$, we have $\kappa_2^2 \in [0, \frac{1}{2})$ and it is a C^∞ function of β. To assert the regularity of T_2, it was enough to show $\kappa_2^2 < 1$, then we surely have T_2 defined and smooth for $\beta = 0$.

The next step is to verify that for $\beta = 0$ there is a unique value $\hat{A}_1(0) \in (0,1)$, such that $F(0, \hat{A}_1(0)) = 0$ and that for this value $\frac{\partial F}{\partial A_1}(0, \hat{A}_1(0)) \gneqq 0$. From the definition, we easily see that $\lim_{\kappa \to 0} \mathrm{K}(\kappa) = \frac{\pi}{2}$, while $\lim_{\kappa \to 1^-} \mathrm{K}(\kappa) = +\infty$. Moreover

$$qT_1(0, A_1) = \frac{4q}{\sqrt{2a}} \mathrm{K}(\sqrt{\frac{A_1 + 1}{2}}) \, , \quad T_2(0, A_1) = \frac{\pi}{\sqrt{aA_1}} \ .$$

$T_1(0, A_1)$ is a strictly increasing function of A_1, while $T_2(0, A_1)$ is strictly decreasing, and

$$\lim_{A_1 \to 0} F(0, A_1) = -\infty \, ,$$
$$\lim_{A_1 \to 1^-} F(0, A_1) = \frac{1}{\sqrt{2a}} \left[\lim_{A_1 \to 1^-} 4q\mathrm{K}(\sqrt{\tfrac{A_1+1}{2}}) - \sqrt{2}\pi \right] = +\infty \ .$$

We conclude that $\hat{A}_1(0,q)$ is uniquely determined from the equality $F(0,\hat{A}_1(0,q)) = 0$, and $\hat{A}_1(0,q) \in (0,1)$. Moreover $\frac{\partial F}{\partial A_1}(0,\hat{A}_1(0,q)) \gneq 0$, then the regularity of the function $\hat{A}_1$ in $\beta = 0$ follows from the implicit function theorem. ∎

Now we are ready to state and show the main result.

Theorem 2.2.8 *Let $q \in \mathbb{Q}^+$ a fixed positive rational number. There is a dense open subset $X'_q \subset \mathbb{R} \times S^1$, such that for each $(\xi_0, \phi_0) \in X'_q$, there is $\beta_0 > 0$, such that for each $\beta \in (0, \beta_0)$, the periodic orbits through (ξ_0, ϕ_0), associated with $\hat{A}_1(\beta, q)$, do not pass through the primaries C_1, C_2. In particular, after scaling time, they are orbits of the not-regularised 2-centre problem with Lagrangian L_0, with energy $E = -2a\beta\hat{A}_1$.*

Proof. Let $(\xi_0, \phi_0) \in \mathbb{R} \times S^1$ be the position of the centre C in elliptic coordinates and let $\gamma(\tau) = (\xi(\tau), \phi(\tau))$ be a periodic orbit associated with $\hat{A}_1(\beta, q)$, which pass through (ξ_0, ϕ_0) with velocity (ξ'_0, ϕ'_0). Without loss of generality we can assume $\phi'_0 > 0$ (see the end of Subsection 2.2.1). Suppose that the orbit γ passes through a primary C_1 or C_2: at that instant we have $\xi = 0$ and $\phi \in \{0, \pi\}$. Let $q = \frac{m}{n}$, with $m, n \in \mathbb{Z}$ positive integers and $(m, n) = 1$. Let $\Delta\tau$ be the minimal time interval to go from a primary to the centre C along the orbit γ. Thanks to Proposition 2.2.3, we must have $\Delta\tau < \frac{T}{2}$, where $T = mT_1 = nT_2$ is the period of γ.

We have two possibilities, corresponding to start from C_1 or from C_2. The orbit solves the separated system (2.6). Then, in the first case

$$\Delta\tau = \frac{1}{2\sqrt{a}} \int_0^{\phi_0} \frac{1}{\sqrt{\beta\hat{A}_1 \cos^2\phi + \hat{A}_1}} d\phi + iT_2\,,$$

while in the second case

$$\Delta\tau = \frac{1}{2\sqrt{a}} \int_0^{\phi_0} \frac{1}{\sqrt{\beta\hat{A}_1 \cos^2\phi + \hat{A}_1}} d\phi \pm \frac{T_2}{2} + i'T_2\,,$$

with $i, i' \in \mathbb{Z}$, $i, i' \geq 0$, and in the second case with sign $\pm$ respectively when $\phi_0 < \pi$, $\phi_0 \geq \pi$.

The condition $\Delta\tau < \frac{T}{2} = \frac{nT_2}{2}$ implies that $i < \frac{n}{2}$ and $i' < \frac{n+1}{2}$. Look now at the variable ξ. In both cases we must have

$$\Delta\tau = \pm\frac{1}{2\sqrt{a}} \int_0^{\xi_0} \frac{1}{\sqrt{\cosh\xi - \beta\hat{A}_1\cosh^2\xi - \hat{A}_1}} d\xi + j\frac{T_1}{2}\,.$$

with $j \in \mathbb{Z}$, $j \geq 0$, and the condition $\Delta\tau < \frac{T}{2} = \frac{mT_1}{2}$ implies that $j \leq m$.

We denote by $P(\phi_0, \beta)$ and $Q(\xi_0, \beta)$ the functions

$$\begin{aligned} P(\phi_0,\beta) &= \frac{1}{2\sqrt{a}} \int_0^{\phi_0} \frac{1}{\sqrt{\beta\hat{A}_1 \cos^2\phi + \hat{A}_1}} d\phi\,, \\ Q(\xi_0,\beta) &= \frac{1}{2\sqrt{a}} \int_0^{\xi_0} \frac{1}{\sqrt{\cosh\xi - \beta\hat{A}_1\cosh^2\xi - \hat{A}_1}} d\xi\,. \end{aligned}$$

From Lemma 2.2.7 and the fact that ξ_0 is not an inversion point, we deduce that these functions are smooth in a neighbourhood of $\beta = 0$. The condition to pass through a primary for the first case is

$$\Delta\tau = P(\phi_0, \beta) + iT_2 = \pm Q(\xi_0, \beta) + j\frac{T_1}{2} ,$$

while for the second case it is

$$\Delta\tau = P(\phi_0, \beta) \pm \frac{T_2}{2} + i'T_2 = \pm Q(\xi_0, \beta) + j\frac{T_1}{2} .$$

We have a finite set of possible values for the integers i, i', j and the equality $qT_1 = \frac{m}{n}T_1 = T_2$ holds, then we can summarise all the conditions with the following one: if the orbit γ passes through a primary, then $G^+(\xi_0, \phi_0, \beta) \in S$ or $G^-(\xi_0, \phi_0, \beta) \in S$, where $S \subset \mathbb{Q}$ is a finite subset of rationals, depending only on the fixed parameter $q \in \mathbb{Q}^+$; here the functions G^+, G^- are defined by

$$G^{\pm}(\xi_0, \phi_0, \beta) = \frac{P(\phi_0, \beta) \pm Q(\xi_0, \beta)}{T_1(\beta, \hat{A}_1(\beta, q))} .$$

The functions $G^{\pm}$ are smooth in all the variables: in particular, if $G^{\pm}(\xi_0, \phi_0, 0) \notin S$, there is $\beta_0 > 0$, such that $G^{\pm}(\xi_0, \phi_0, \beta) \notin S$ for any $\beta \in (0, \beta_0)$. Then, to exclude the collision with the primaries for small enough values of β, it is enough that the centre C belongs to the set

$$X'_q = \{(\xi_0, \phi_0) \in \mathbb{R} \times S^1 | \quad G^{\pm}(\xi_0, \phi_0, 0) \notin S\} .$$

The set S is finite, then it is clear that X'_q is open and dense in $\mathbb{R} \times S^1$. Indeed, we easily see from the definitions of the functions $G^{\pm}$ that their partial derivatives with respect to ϕ_0 are different from zero; then X'_q is a finite intersection of open dense subset of $\mathbb{R} \times S^1$. Moreover, the complement of X'_q has zero Lebesgue measure.
■

Remark 2.2.9 *When $q = 1$, it follows easily from Proposition 2.2.3 that at least one of the two periodic orbits through C corresponding to the same $\hat{A}_1(\beta)$ does not collide with the primaries. Indeed, without loss of generality we can assume $\phi'_0 > 0$ (see the end of Subsection 2.2.1). Suppose $\phi_0 \in (0, \frac{\pi}{2})$: then the time to pass from the centre C_1 to the centre C must be less than $\frac{T}{4}$, where $T = T_1 = T_2$ is the period. If $\xi_0 > 0$ the only possibility to collide with the primaries is that $\xi'_0 > 0$, while if $\xi_0 < 0$ we must have $\xi'_0 < 0$. For the other possible intervals of values of ϕ_0 a similar reasoning works.*

For our purposes, the fact of having at least one orbit through C that does not collide with the primaries is not enough: indeed, after we have obtained the collision arcs, we want to construct collision chains *with them, and to do it we need at least two orbits of the system (L_0) with the same energy E, which pass through C with not-parallel tangent fields. The latter condition will be investigated in the next section.*

Remark 2.2.10 *Theorem 2.2.8 would be improved if we showed one of the following:*

i) the partial derivatives $\frac{\partial G^{\pm}}{\partial \beta}(\xi_0, \phi_0, 0)$ *are zero only for isolated values of* (ξ_0, ϕ_0)*;*

ii) the partial derivatives $\frac{\partial G^{\pm}}{\partial \beta}(\xi_0, \phi_0, 0)$ *do not vanish for every* $\xi_0 \in \mathbb{R}$*,* $\phi_0 \in S^1$*, with* (ξ_0, ϕ_0) *not a primary.*

For case (i) we would have that the collision with the primaries is possible only for isolated positions $(\xi_0, \phi_0) \in \mathbb{R} \times S^1$*, while for case (ii) the collision would be excluded for any position of the centre* C*.*

At present, we don't have any of these improvements.

Corollary 2.2.11 *Let* $I \subset \mathbb{Q}^+$ *be a finite set of positive rationals. There is a dense open subset* $X'_I \subset \mathbb{R} \times S^1$*, such that for each* $(\xi_0, \phi_0) \in X'_I$*, there is* $\beta_0 > 0$*, such that for any* $\beta \in (0, \beta_0)$*, the periodic orbits through* (ξ_0, ϕ_0) *corresponding to* $\hat{A}_1(\beta, q)$*, with* $q \in I$*, do not pass through the primaries. In particular, these are periodic orbits for the system* (L_0)*.*

Proof. It is an immediate consequence of the construction of the subset $X'_q \subset \mathbb{R} \times S^1$ in the proof of Theorem 2.2.8: we can apply this theorem for any $q \in I$, then take the intersection of the sets X'_q thus obtained, and the resulting set maintains the same properties of the sets X'_q. ∎

Remark 2.2.12 *Denote by* $\psi : \mathbb{R} \times S^1 \to \mathbb{R}^2$ *the map to pass from elliptic to Cartesian coordinates,* $\psi(\xi, \phi) = \cosh(\xi + i\phi) = x + iy$*. The map* ψ *is an open map, then, fixed a finite subset* $I \subset \mathbb{Q}^+$*, the image* $X_I = \psi(X'_I)$ *is an open dense subset of* $\mathbb{R}^2$*. Then we can say that the thesis of Corollary 2.2.11 holds for any position of the centre* C *in an open dense subset* $X_I \subset \mathbb{R}^2$*.*

We also note that it doesn't matter which elliptic coordinates we choose for a point $C \in X_I$*. Indeed, if one possibility is* $(\xi_0, \phi_0) \in X'_I$*, the other is* $(-\xi_0, -\phi_0)$*. By the symmetries of the problem (2.6) we see that the two orbits through* $(-\xi_0, -\phi_0)$ *in* $\mathbb{R} \times S^1$ *are obtained from the orbits through* (ξ_0, ϕ_0) *by the symmetry through the origin and then they correspond to the same trajectories in* $\mathbb{R}^2$*. We can simply enlarge our set* X'_I *using this symmetry.*

Consider now the set $X' = \cap_{q \in \mathbb{Q}^+} X'_q$*: by the classical Baire's Category Theorem (see for example [26, Appendix A]), the set* X' *is dense in* $\mathbb{R} \times S^1$*. Then, the set* $X = \psi(X')$ *is dense in* $\mathbb{R}^2$ *and for any finite subset* $I \subset \mathbb{Q}^+$*, it is contained in* X_I*. Then we can say that the thesis of Corollary 2.2.11 holds for any position* $C \in X \subset \mathbb{R}^2$*. In this manner the only information that we have lost is that we don't know if the set* X *is open, while we have gained the independence of the dense set* X *of positions of* C *from the set of rationals* I*. Note that the choice of a sufficiently small* β *cannot be independent of* I*, instead.*

2.2.4 Early collisions

We defined a collision arc in Section 1.3.1 to be a critical point of the Maupertuis-Jacobi functional which starts and ends at collision with the centre C and does not meet this centre at intermediate times. If we take one of the periodic orbits through C found in the preceding section, and we pass to Cartesian coordinates, then we cannot be sure that the orbit does not pass newly through C before a period is passed. This can happen in two ways: when the orbit, considered in elliptic coordinates, meets the point (ξ_0, ϕ_0) and when it meets $(-\xi_0, -\phi_0)$.

When a periodic orbit starting from the centre C passes newly through C in a time shorter than its period, then we talk of *early collision*. To understand when an early collision occurs we need to study the behaviour of periodic orbits a little deeper.

An easy example is given by the case $q = 1$: in this case any periodic orbit crosses the y-axis, is symmetric with respect to this axis and, if it does not collide with the primaries, it has a transverse autointersection at the point of crossing. The symmetry comes from the fact that after half the orbit's period we have the passage from a point (ξ, ϕ) to a point $(-\xi, \phi + \pi)$, that is from a point $(x, y) = (\cosh(\xi)\cos(\phi), \sinh(\xi)\sin(\phi))$, to its symmetric $(-x, y)$. The intersections with the y-axis occur when $\phi = \pm\pi/2$ and from one intersection to another there is a time interval of one half the period. In elliptic coordinates the orbit passes through a point $(\xi, \pi/2)$ and after half a period it arrives at $(-\xi, -\pi/2)$, but these two points coincide in Cartesian coordinates and are on the y-axis. Then each orbit autointersects at a point on the y-axis. The transformation of the velocities when we pass from elliptic to Cartesian coordinates is given by:

$$v = V(\xi, \phi) \cdot \begin{pmatrix} \xi' \\ \phi' \end{pmatrix}, \quad V(\xi, \phi) = \begin{pmatrix} \sinh\xi\cos\phi & -\cosh\xi\sin\phi \\ \cosh\xi\sin\phi & \sinh\xi\cos\phi \end{pmatrix}, \tag{2.10}$$

where $V(\xi, \phi)$ is an invertible matrix, except when $(\xi, \phi) \in \{(0, 0), (0, \pi)\}$. We observe that $V(-\xi, -\phi) = -V(\xi, \phi)$. If at the point of intersection with the y-axis $\xi = \pm\xi_+$, then $\xi' = 0$ and the velocity has zero y-component: then the two crossings are not transverse. But we see that this is the case in which the orbit collides with the primaries, reversing its direction at the collisions. If $\xi \neq \pm\xi_+$, instead, the velocities at the point of intersection with the y-axis have both components different from zero, then they are transverse, because by symmetry they differ only in the sign of their x-component.

An illustration of this situation is given in Figure 2.11, where we have drawn 18 periodic orbits, included the two orbits that collide with the primaries: these latter orbits are in green, while the primaries are marked with red asterisks. We have also drawn the periodic orbit $\xi = \xi_+$: we cannot use it to construct collision arcs, because it has different energy, but this orbit is interesting in itself because it enclose all the trajectories with the same parameters (and then the same energy), and it is tangent to all. The behaviour of a couple of periodic orbits through the same point can be seen in Figure 2.12, where the starting point is marked with a black asterisk.

Then, if $q = 1$, we can say that there is an early collision at C if and only if C lies on the y-axis: in this case there is a unique, up to reversing the direction of motion, periodic orbit through C.

In particular, we see that the period does not change in the passage from elliptic to Cartesian coordinates. Actually, this holds for any value of the parameter $q \in \mathbb{Q}^+$. In fact, if we take a periodic orbit that starts from a point (ξ_0, ϕ_0), which is not a primary, we see that the matrix $V(\xi_0, \phi_0)$ given in (2.10) is invertible and $V(-\xi_0, -\phi_0) = -V(\xi_0, \phi_0)$. Then the only possibility to have a shorter period when we pass to Cartesian coordinates is that the orbit passes through $(-\xi_0, -\phi_0)$, with

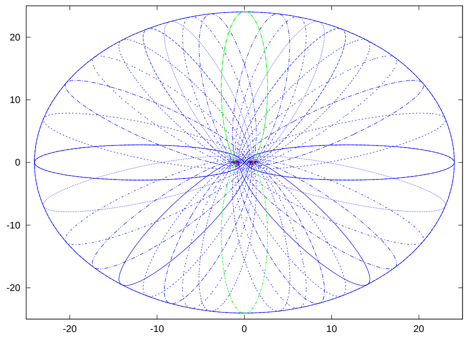

Figure 2.11. Periodic orbits in Cartesian coordinates when $a = 1$, $q = 1$, $\beta = 1/7$ and the orbit through the primaries.

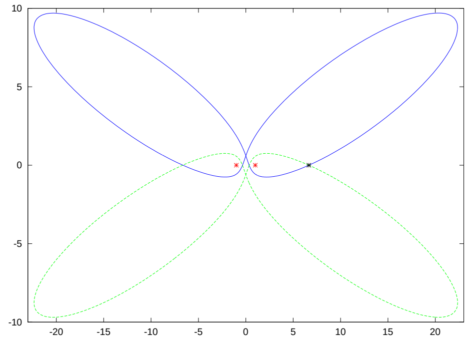

Figure 2.12. Periodic orbits through $(\xi_0, \phi_0) = (\frac{2}{3}\xi_+, 0)$ in Cartesian coordinates, for $a = 1$, $q = 1$, $\beta = 1/7$.

velocity $(-\xi_0', -\phi_0')$. This is clearly impossible because the sign of ϕ' never changes along the orbit.

Anyway, the situation complicates if $q \neq 1$ and we cannot get a global view easily: as an example, in Figure 2.13 we have drawn the orbits through the point $(\xi_0, \phi_0) = (\frac{2}{3}\xi_+, 0)$, when $q = 2$. To better see the autointersections we have put an enlargement in Figure 2.14.

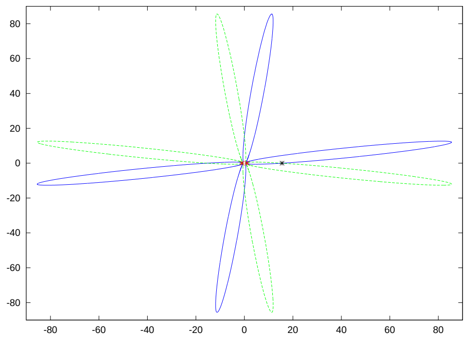

Figure 2.13. Periodic orbits through $(\xi_0, \phi_0) = (\frac{2}{3}\xi_+, 0)$ in Cartesian coordinates, when $a = 1$, $q = 2$, $\beta = 1/7$.

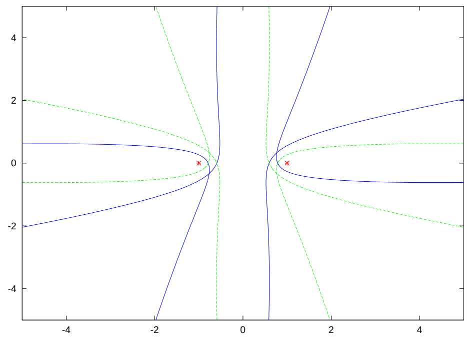

Figure 2.14. An enlargement of Figure 2.13, where the autointersections of the two orbits are visible.

We note the following remarkable fact:

Proposition 2.2.13 *Let* $(\xi_0, \phi_0) \in \mathbb{R} \times S^1$ *be fixed and consider a periodic orbit* $(\xi(\tau), \phi(\tau))$ *through this point, corresponding to the values* $\beta, \hat{A}_1(\beta, q)$*. If* $(\xi(\tau), \phi(\tau))$ *passes newly through* (ξ_0, ϕ_0) *before closing, then it is the unique periodic orbit through* (ξ_0, ϕ_0)*, corresponding to the same values of* β, q*, up to reverse the direction of motion. Moreover, if* ξ_0 *is not an inversion point, the orbit has a transverse autointersection at* (ξ_0, ϕ_0)*.*

Proof. If the initial velocity in (ξ_0, ϕ_0) is (ξ_0', ϕ_0'), then the only other possible velocity through the same point along the same orbit is $(-\xi_0', \phi_0')$. If the orbit arrives at (ξ_0, ϕ_0) before that a period is passed, then it must happen with velocity

$(-\xi_0', \phi_0')$, and this implies that the two possible periodic orbits through (ξ_0, ϕ_0) coincide. ■

At these stage, we know completely early collisions only for the case $q = 1$. In general, we can only say that early collisions cannot be excluded: we don't have any knowledge about the conditions that determine them. Anyway, it is not a problem for the proof of Theorem 1.3.8. When an early collision occurs, we will take as collision arc the partial arc of the periodic orbit through C, which starts from C and ends at the first next passage through C. The final time T of the collision arc in this case will not be the period, but it will be the time of the first return to C.

Finally, we want to stress that a collision arc must end at C with one of the four possible velocities determined by the choice of the parameters, which in elliptic coordinates are $\{\pm(\xi_0', \phi_0'), \pm(-\xi_0', \phi_0')\}$, where (ξ_0', ϕ_0') is the initial velocity of the arc.

2.3 Nondegeneracy

In this section we will verify that the collision arcs obtained from Corollary 2.2.11 satisfy the nondegeneracy condition. Before doing that we recall the definition of nondegeneracy and describe a particular characterisation of it which will be used for the verification.

2.3.1 A sufficient condition for nondegeneracy

Let us denote by (M, L) our Lagrangian system: M is a smooth Riemannian manifold, the configuration manifold, and $L : TM \times \mathbb{R} \to M$ is the Lagrangian function, which has the general form

$$L(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) + < \omega(\mathbf{q}), \dot{\mathbf{q}} > - W(\mathbf{q}) \,,$$

where $T = \frac{1}{2} < A(\mathbf{q})\dot{\mathbf{q}}, \dot{\mathbf{q}} >$ is the kinetic energy and ω is a vector field on M. The kinetic energy gives a positive definite quadratic form on TM. The term $< \omega(\mathbf{q}), \dot{\mathbf{q}} >$ can appear for example in the presence of a magnetic field or when there are constraints depending on the time.

Fixed a value E for the generalised energy $H = T + W$, define the *region of possible motion* $D = D_E = \{\mathbf{q} \in M |\ W(\mathbf{q}) < E\}$. Fixed two points $\mathbf{q}_1, \mathbf{q}_2 \in D$, let $\Omega = \Omega(D; \mathbf{q}_1, \mathbf{q}_2) = \{u : [0,1] \to D |\ u \in W^{1,2}([0,1]; D),\ u(0) = \mathbf{q}_1,\ u(1) = \mathbf{q}_2\}$ be the space of $W^{1,2}$ curves in D starting from $\mathbf{q}_1$ and ending at $\mathbf{q}_2$.

As reminded in Subsection 1.3.1, the Jacobi metric g_E on the domain D is defined by

$$g_E(\mathbf{q}, \dot{\mathbf{q}}) = 2\sqrt{(E - W(\mathbf{q}))T(\mathbf{q}, \dot{\mathbf{q}})} + < \omega(\mathbf{q}), \dot{\mathbf{q}} > ,$$

and the Maupertuis-Jacobi functional $J_E : \Omega \to \mathbb{R}$ is given by

$$J_E(u) = \int_0^1 g_E(u(s), \dot{u}(s)) ds \ .$$

Note that the functional J_E is independent of the parametrisation. The orbits of the system (M, L) with energy E are, up to a reparametrisation of time, the critical points of the Maupertuis action functional.

By Morse theory, a path $u \in \Omega$ is a critical point of the Maupertuis functional J_E if and only if it is a geodesic with respect to the Jacobi metric g_E and a characterisation of degeneracy is the following (see [43]):

Theorem 2.3.1 *A path $u \in \Omega$, which is a critical point of the Maupertuis functional J_E, is degenerate if and only if there exists a variation of it through geodesics for the metric g_E, with fixed endpoints $\mathbf{q}_1, \mathbf{q}_2$.*

If we denote the general solution of system (M, L) with energy E by $\mathbf{q}(t) = \mathbf{f}(\mathbf{q}_0, \boldsymbol{v}_0, t)$, where $\mathbf{q}_0, \boldsymbol{v}_0$ are the initial position and velocity, then the orbits from $\mathbf{q}_1$ to $\mathbf{q}_2$ with energy E correspond to solutions of the system

$$\mathbf{f}(\mathbf{q}_1, \boldsymbol{v}_0, t_f) = \mathbf{q}_2\,, \quad H_0(\mathbf{q}_1, \boldsymbol{v}_0) = E\ . \tag{2.11}$$

From Theorem 2.3.1 we can derive the following

Proposition 2.3.2 *A sufficient condition for $u \in \Omega$ to be a nondegenerate solution of system (M, L) with energy E is that the Jacobian of system (2.11) is not zero at the corresponding solution $\mathbf{q}(t) = \mathbf{f}(\mathbf{q}_0, \boldsymbol{v}_0, t)$.*

Proof. Suppose $u \in \Omega$ is a degenerate critical point. Then there is a variation α of u through geodesics for the metric g_E. But the geodesics are, up to time parametrisation, solutions of the system (2.11). Then, given a variation $\alpha(\eta, s)$ of u through geodesics, if we leave η fixed, we can define the new parameter t by

$$t(\eta, s) = \int_0^s \sqrt{\frac{T(\alpha(\eta, z), \dot{\alpha}(\eta, z))}{E - W(\alpha(\eta, z))}}dz\ .$$

With the parameter t the curves defined by the variation α have energy E. Moreover, the parameter t is a smooth function of (η, s). Then $\boldsymbol{v}_0 = \boldsymbol{v}(\eta, 0) = \frac{\partial\alpha}{\partial t}(\eta, 0)$ and $t_f = t(\eta, 1)$ are smooth functions of η. This implies that there is a continuum of solutions of system (2.11) with the values $\mathbf{q}_2$ and E fixed and then the Jacobian at u is zero. ■

2.3.2 Nondegeneracy of the collision arcs

Let $q \in \mathbb{Q}^+$ be fixed and $m, n \in \mathbb{Z}$ be positive coprime integers such that $q = m/n$. Denote by $(\xi_0, \phi_0) \in \mathbb{R} \times S^1$ the position of the centre C in elliptic coordinates. Suppose that $(\xi_0, \phi_0) \in X'_q$, where the set $X'_q \subset \mathbb{R} \times S^1$ is the one given by Theorem 2.2.8. Take β small enough, so that the periodic orbits of system $(\mathcal{L}_0)$ corresponding to $\hat{A}_1(\beta, q)$ and passing through C do not pass through the primaries. Let $\gamma(t)$ be one of the resulting collision arcs for the system (L_0), which starts from the centre C with velocity v_0 and whose energy is $E = -2a\beta\hat{A}_1(\beta, q)$. Denote the general solution of the system (L_0) by $\mathbf{q}(t) = \mathbf{f}(\mathbf{q}_0, \boldsymbol{v}_0, t)$, where $\mathbf{q}_0 \in M, \boldsymbol{v}_0 \in TM_{q_0}$ are the initial position and velocity. Then $\gamma(t)$ solves the system

$$\mathbf{f}(C, \boldsymbol{v}_0, T) = C\,, \quad H_0(C, \boldsymbol{v}_0) = E\,, \tag{2.12}$$

with respect to the variables $(\boldsymbol{v}_0, T)$. To show the nondegeneracy of γ it is sufficient to verify that the Jacobian of system (2.12) is nonzero. Actually we will verify a slight variant of this condition.

It is convenient to consider as variables the parameters (β, A_1) instead of the coordinates of the initial velocity $\boldsymbol{v}_0$. This procedure is right only if there is a local diffeomorphism which allows to pass from $\boldsymbol{v}_0$ to (β, A_1). The transformation of the velocities in the passage from elliptic to Cartesian coordinates is given by the invertible matrix $V(\xi_0, \phi_0)$ defined in (2.10). We know that the orbit γ corresponds to a solution of the separated system (2.6), and in elliptic coordinates we have $(\dot{\xi}_0, \dot{\phi}_0) = \frac{d\tau}{dt}(0)(\xi_0', \phi_0')$, with $\frac{d\tau}{dt}(0) = (\cosh^2(\xi_0) - \cos^2(\phi_0))^{-1}$. Then $\boldsymbol{v}_0$ is a C^∞ function of (ξ_0', ϕ_0'). From the system (2.6) we see that (ξ_0', ϕ_0') is a C^∞ function of (β, A_1) and then $\boldsymbol{v}_0$ is. By the same reasoning we see that (β, A_1) is locally a C^∞ function of the velocity (ξ_0', ϕ_0') and then of $\boldsymbol{v}_0$.

For all nearby trajectories from the same initial point C, we evaluate the m-th positive instant of time at which they meet the ellipse $\xi = \xi_0$, with the velocity $\dot{\xi}$ equal to the initial velocity $\dot{\xi}_0 \neq 0$; then we consider the time distance from the n-th passage through ϕ_0, which is given by $nF(\beta, A_1)$, where

$$F(\beta, A_1) = [qT_1 - T_2](\beta, A_1) \ .$$

This is equivalent to consider the value of the angle coordinate ϕ at the instant at which we have the m-th *oriented* crossing of the line $\xi = \xi_0$.

Note that in this manner the time variable T is fixed as function of (β, A_1), $T = mT_1(\beta, A_1)$, so that we have reduced the order of the system (2.12) of a unit. The orbit γ satisfies

$$F(\beta, A_1) = 0\,, \quad -2a\beta A_1 = E\,,$$

and it is nondegenerate if the Jacobian determinant of this system at the solution $(\beta, \hat{A}_1(\beta, q))$, corresponding to the orbit γ, is different from zero.

The Jacobian matrix is

$$J = \begin{pmatrix} \frac{\partial F}{\partial \beta} & \frac{\partial F}{\partial A_1} \\ -2aA_1 & -2a\beta \end{pmatrix} \ .$$

As observed in Subsection 2.1.3, we have $\partial F/\partial A_1 \gtrneqqless 0$. Furthermore, we have that $\hat{A}_1(\beta, q) \in (0, \frac{1}{1+\beta})$. These properties remain true for $\beta = 0$, as proved in Lemma 2.2.7, and moreover T_1, T_2 are C^∞ functions of (β, A_1). It follows that when $\beta = 0$ and $A_1 = \hat{A}_1(0, q)$, the determinant is well defined and different from zero, and by regularity the same is true also for values $\beta, \hat{A}_1(\beta, q)$, with $\beta > 0$ sufficiently small.

We conclude that the collision arc γ is nondegenerate for β small enough.

2.4 Proof of the shadowing theorem for the planar 3-centre problem

In this section we will verify that the collision arcs constructed so far meet transversely at the centre C: this will allow us to apply Theorem 1.3.4 and derive our central result, Theorem 1.3.8.

2.4.1 Direction change

Let $q \in \mathbb{Q}^+$ be fixed and $m, n \in \mathbb{Z}$ be positive coprime integers such that $q = m/n$. In order to construct *collision chains* to which apply Theorem 1.3.4, we should have at least two collision arcs which start and arrive at the centre C with transverse tangent fields.

From the results of Section 2.2, if in elliptic coordinates our centre (ξ_0, ϕ_0) belongs to the dense open subset $X'_q \subset \mathbb{R} \times S^1$, then for small enough values of β, after reparametrisation of time and change of coordinates, the orbits associated to $\hat{A}_1(\beta, q)$ do not collide with the primaries and they are periodic orbits for the not-regularised two-centre problem (L_0). As observed in Subsection 2.2.1, there are exactly two orbits corresponding to $\hat{A}_1(\beta, q)$, which pass through the centre (ξ_0, ϕ_0) with velocities (ξ'_0, ϕ'_0), $(-\xi'_0, \phi'_0)$ respectively. Changing sign to ϕ'_0, we obtain simply the same orbits, with reversed direction of motion. Moreover, the two transverse orbits coincide in the case of an autointersection at (ξ_0, ϕ_0), as we have seen in Proposition 2.2.13.

Now consider the two orbits associated with $\hat{A}_1(\beta, q)$: suppose they pass through the centre (ξ_0, ϕ_0) at time $\tau = 0$ and that their velocities are (ξ'_0, ϕ'_0), $(-\xi'_0, \phi'_0)$. They are obviously transverse at the point (ξ_0, ϕ_0) in the cylinder $\mathbb{R} \times S^1$, because we have chosen β small enough to have $\xi'_0 \neq 0$ (see Proposition 2.2.1). We must verify that this transversality is conserved after time reparametrisation and changing from elliptic to Cartesian coordinates. The reparametrisation of time is given by formula (2.1): it maintains the directions, because the centre (ξ_0, ϕ_0) is not a primary. As seen in subsection 2.2.4, the passage to Cartesian coordinates is given by an invertible matrix V(see (2.10)), then the transversality is conserved.

Look now at the other possible elliptic coordinates for C: they are $(-\xi_0, -\phi_0)$, then the possible velocities at this point are the same as the ones at (ξ_0, ϕ_0). It follows that there are no more orbits through C that we can consider, for the same values of β, q.

We conclude that there are two transverse directions for the collision arcs starting from C. In correspondence of the same values of the parameters (β, q) we obtain four collision arcs, divided in pairs of arcs with transverse initial velocities.

We summarise the results of this and the preceding subsection in the following

Proposition 2.4.1 *Let $I \subset \mathbb{Q}^+$ be a finite set of positive rationals. Suppose that the centre C has elliptic coordinates $(\xi_0, \phi_0) \in X'_I$, where $X'_I \subset \mathbb{R} \times S^1$ is the dense subset given by Corollary 2.2.11. Then, there is $\beta_0 > 0$ such that for each $\beta \in (0, \beta_0)$ and for each $q \in I$, there are exactly four collision arcs at C for the system (L_0), associated with the value $\hat{A}_1(\beta, q)$, which are nondegenerate. Moreover, the four collision arcs divide in two pairs, according to their initial velocities at C: any couple is formed by two arcs with opposite initial velocities at C and each arc in one pair has transverse initial velocity to each arc in the other.* ∎

2.4.2 Proof of the theorem

We're going to conclude the proof of Theorem 1.3.8.

Let $I \subset \mathbb{Q}^+$ be a finite set of positive rationals and $X_I = \psi(X'_I) \subset M$ as in Remark 2.2.12. The energy of the collision arcs at C is given as function of the parameters (β, q), $q \in I$, by the relation

$$E = -2a\beta\hat{A}_1(\beta, q) \ .$$

Thanks to the regularity of the function $\hat{A}_1(\beta, q)$ for $\beta \in [0, 1)$ and to the fact that $\hat{A}_1(0, q) \in (0, 1)$ (Lemma 2.2.7), to require β small enough is equivalent to ask for the absolute value of the energy E to be small enough. In particular, if $\beta < \beta_0$, with β_0 sufficiently small, then the energy $E(\beta, q)$ is a strictly decreasing function of β:

$$E(\cdot, q) : (0, \beta_0) \to (-E_0, 0) \, ,$$

with $E(0, q) = 0$, $E(\beta_0, q) = -E_0 < 0$. It follows that all the results can be stated using the energy parameter E instead of β.

If we choose the energy $E < 0$ sufficiently close to zero, then for each $q \in I$, there are four nondegenerate collision arcs of energy E through the centre C: they are pieces of periodic trajectories on the configuration space $M = \mathbb{R}^2 \setminus \{C_1, C_2\}$.

For β fixed, the energy increase with the class q (see Proposition 2.1.5). This means that we cannot state a general result valid for a fixed energy E and any $q \in \mathbb{Q}^+$. On the other hand, we don't need such a result, because to apply Theorem 1.3.4 we only want a finite number of collision arcs. What is certainly true is that for any finite set of classes $I \subset \mathbb{Q}^+$ of cardinality i, we can choose a small enough energy value $E < 0$ to form a set of $4i$ collision arcs of energy E, by taking for each $q \in I$ the four arcs obtained by our procedure.

By the monotony of the function $\hat{A}_1(q, \beta)$ with respect to q and its relation with the energy E, we are sure that the arcs with the same energy E, but different classes, cannot have the same value for $\hat{A}_1$: then they cannot coincide. In general, we can't state that collision arcs of different classes will determine a different set of directions at the point C: what we can certainly assure is that if we fix an arc of class $q \in I$, then for each class $q' \in I$, not necessarily different from q, we can always choose a pair of arcs of class q', which start at C with a direction transverse to the velocity with which the arc of class q has arrived at C.

Thus, for any fixed sequence $\{q_k\}_{k\in\mathbb{Z}}$, $q_k \in I$, we can construct infinite collision chains with the following property: if $\{\gamma_k\}_{k\in\mathbb{Z}}$ is one of them, any γ_k is a piece of a periodic orbit of the 2-centre problem (L_0) of class q_k.

The assumptions of Theorem 1.3.4 are all satisfied and we can apply this theorem and then Theorem 1.3.5 to get our central result Theorem 1.3.8, which is now definitely proved.

2.5 Some final considerations

In this section we put the attention on the possible extensions of our main result Theorem 1.3.8. First of all we must remark that the existence of chaotic motions does not necessarily imply non-integrability. However, it would be interesting to investigate more deeply the properties of the motions found.

We are quite sure that Theorem 1.3.8 is valid in the non-symmetric case in which the intensities of the primaries are different. We can also try to show an analogous result for an arbitrary number $N \geq 2$ of perturbing centres: in this case the possibility to choose collision arcs with different centres as end-points complicates the matter and a result of this kind could be quite difficult.

An important question to analyse is the generalisation to higher dimension, that is to the 3-centre problem in $\mathbb{R}^3$. We think it should be possible to repeat our construction step by step, since we have also in $\mathbb{R}^3$ elliptic coordinates by which regularise the problem at the singularities corresponding to the primaries.

Finally, we want to remark that the 2-centre problem is integrable on the sphere S^2 and that we know some of its topological properties: see [29] and [54]. It would be interesting to try to prove a result about the existence of chaotic quasi-collision motions for the 3-centre problem on S^2.

Chapter 3

Hyperbolicity of the shadowing orbits

In this chapter we sketch the proof of Theorem 1.3.5 (see [9]), thus giving the definition of the Poincaré map to which the theorem refers. Before that, we must preliminarily recall a characterisation of uniform hyperbolicity through the use of Frenkel-Kontorova models (see [1]).

3.1 Equivalence of uniform hyperbolicity for symplectic twist maps and phonon gap for Frenkel-Kontorova models

First of all let us remind the definition of uniformly hyperbolic invariant set for a discrete time dynamical system.

Definition 3.1.1 *Let F be a C^1 diffeomorphism of a Riemannian manifold M with metric $|\cdot|$. A closed invariant set Λ for F is* uniformly hyperbolic *if there is a continuous splitting of the tangent space TM_z at each point $z \in \Lambda$ into the direct sum of spaces E^s_z and E^u_z, such that $E^s_{F(z)} = DF(E^s_z)$, $E^u_{F(z)} = DF(E^u_z)$ and there are constants $C > 0, \lambda < 1$, such that*

$$|DF^n_z s| \le C\lambda^n |s|\,, \qquad \text{for all } s \in E^s_z\,,\ z \in \Lambda\,,\ n \ge 0\,,$$

$$|DF^{-n}_z u| \le C\lambda^n |u|\,, \qquad \text{for all } u \in E^u_z\,,\ z \in \Lambda\,,\ n \ge 0\,;$$

finally there exists $\alpha \in \mathbb{R}$ such that for all tangent vectors ξ on Λ,

$$|s|, |u| \le \alpha |\xi|\,,$$

where s, u are respectively the stable and unstable components of ξ.

The objects of study of this section are the twist maps and the Frenkel-Kontorova models, which we now define.

Definition 3.1.2 *A* symplectic twist map *is a C^1 diffeomorphism*

$$\begin{array}{rcl} F : \mathbb{R}^d \times \mathbb{R}^d & \to & \mathbb{R}^d \times \mathbb{R}^d \\ (x, p) & \mapsto & (x', p') \end{array}$$

such that

(i) F *preserves the symplectic form* $\omega = \sum_{\nu=1}^{d} dp^{\nu} \times dx^{\nu}$,

(ii) *the map* $p \mapsto x'(x,p)$ *is a diffeomorphism of* $\mathbb{R}^d$ *for each fixed* $x \in \mathbb{R}^d$ *(*twist condition*).*

The orbits of F *are the sequences* $(x_n, p_n)_{n\in\mathbb{Z}}$*, such that* $(x_{n+1}, p_{n+1}) = F(x_n, p_n)$*.*

Definition 3.1.3 *Given a* C^2 *function* $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$*, a* Frenkel-Kontorova model *assigns an* energy *or* action functional

$$\Phi_{MN}(x) = \sum_{n=M}^{N-1} L(x_n, x_{n+1})$$

to any finite segment $M \le n \le N$ *of a configuration* $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}}$*, with* $x_n \in \mathbb{R}^d$*. The* equilibrium states *of the model are the configurations* $\mathbf{x}$*, such that, for each* $M < N$ *the energy* Φ_{MN} *is stationary with respect to all variations of* $x_n, M < n < N$*.*

Denoting by Φ the formal sum Φ_{MN} from $M = -\infty$ to $N = +\infty$, the last statement of this definition means that

$$\frac{\partial \Phi}{\partial x_n^{\nu}} = \frac{\partial L}{\partial x_n^{\nu}}(x_n, x_{n+1}) + \frac{\partial L}{\partial x_n^{\nu}}(x_{n-1}, x_n) = 0\,, \tag{3.1}$$

for $\nu = 1 \dots d$, $n \in \mathbb{Z}$.

There is a one-to-one correspondence between twist maps and Frenkel-Kontorova models, in which the orbits of a twist map correspond to the equilibrium states of the associated Frenkel-Kontorova model. It is described as follows.

Let Φ be the action of a Frenkel-Kontorova model with generating function $L(x, x')$ and assume that the map from $\mathbb{R}^d$ to itself defined by

$$p'(x) = \frac{\partial L(x, x')}{\partial x'} \tag{3.2}$$

is a diffeomorphism for each fixed $x' \in \mathbb{R}^d$. Then the equation (3.1) defines a symplectic twist map

$$(x_{n+1}, p_{n+1}) = F(x_n, p_n)\,,$$

by the relation

$$p_n^{\nu} = \frac{\partial L}{\partial x_n^{\nu}}(x_{n-1}, x_n)\,,$$

which defines the conjugate variables $p_n = \{p_n^{\nu}\}_{\nu=1\dots d}$.

Conversely, given a symplectic twist map $(x,p) \mapsto F(x,p) = (x', p')$, the differential form $p'dx' - pdx$ is exact and hence there is a function $K(x,p)$, such that $dK = p'dx' - pdx$. The twist condition allows to define p as a function of (x, x'), then we can define $L(x,x') = K(x, p(x,x'))$.

In the way described above, we obtain the desired one-to-one correspondence between twist maps and Frenkel-Kontorova models for which (3.2) defines a diffeomorphism. In this correspondence, the uniformly hyperbolic invariant sets correspond to *phonon gaps*, which we now define.

Let Φ be the action of a Frenkel-Kontorova model. For each equilibrium state $\mathbf{x}$, denote by $D^2\Phi(\mathbf{x})$ the infinite matrix of block triangular form with components

$$\frac{\partial^2 \Phi}{\partial x_i^{\nu} \partial x_j^{\nu'}} \ .$$

The *phonon spectrum* $\mathcal{E}$ of $\mathbf{x}$ is the spectrum of the second derivative $D^2\Phi(\mathbf{x})$ so defined. The equilibrium $\mathbf{x}$ has *phonon gap* if 0 is not in the spectrum. Since the spectrum is closed, this is equivalent to

$$E_{min} = \inf\{|E| : \ E \in \mathcal{E}\} > 0 \ .$$

Defining $E_{max} = \sup\{|E| : \ E \in \mathcal{E}\} \in \mathbb{R}^+ \cup \{\infty\}$, the *gap parameter* $G = E_{min}/E_{max}$ gives a measure of the phonon gap. Finally, a set Λ' of equilibrium states has phonon gap if

$$\inf\{E_{min}(\mathbf{x}) : \ \mathbf{x} \in \Lambda'\} > 0 \ .$$

Let $l_2 = \{\xi = \{\xi_n\}_{n\in\mathbb{Z}} | \ \sum_{n\in\mathbb{Z}} |\xi_n|^2 < \infty\}$ with the norm

$$\|\xi\|_2 = \sqrt{\sum_{n\in\mathbb{Z}} |\xi_n|^2} \ .$$

Being $D^2\Phi(\mathbf{x})$ a symmetric linear operator on l_2, then $\|D^2\Phi(\mathbf{x})\|_2$ is the supremum of its spectrum in absolute value, where as usual

$$\|D^2\Phi(\mathbf{x})\|_2 = \sup_{\xi\in l_2}\{\|D^2\Phi(\mathbf{x})\xi\|_2/\|\xi\|_2 : \ \|\xi\|_2 > 0\} \ .$$

Then we have a phonon gap at an equilibrium state $\mathbf{x}$, if

$$\Delta(\mathbf{x}) = \|D^2\Phi(\mathbf{x})^{-1}\|_2^{-1} > 0 \ .$$

The gap parameter at $\mathbf{x}$ is

$$G = (\|D^2\Phi\|_2 \|D^2\Phi^{-1}\|_2)^{-1} \ .$$

Finally, a set Λ' of equilibrium states has phonon gap if

$$\Delta = \inf_{\mathbf{x}\in\Lambda'} \Delta(\mathbf{x}) > 0 \ . \tag{3.3}$$

With the notation just introduced, the following holds

Theorem 3.1.4 (Aubry et al. , 1992) *Let $M = \mathbb{R}^d$. Let Λ be a compact invariant set for a C^1 symplectic twist map F on M^2 and $\Lambda' \subset M^{\mathbb{Z}}$ be the associated set of equilibrium states for the corresponding Frenkel-Kontorova model. Then Λ is uniformly hyperbolic if and only if Λ' has phonon gap.*

If the invariant set Λ is not compact, then the notion of hyperbolicity depends on the choice of the Riemannian metric on M^2. Let $|\cdot|$ be a fixed Riemannian metric on M and denote by the pairs (ξ, π), with $\xi, \pi \in M$, the tangent vectors to M^2. Given a twist map F on M^2, choose on M^2 the metric defined by

$$|(\xi, \pi)| = \sqrt{|\xi|^2 + |\xi'|^2} \, ,$$

where $(\xi', \pi') = DF(\xi, \pi)$. By the twist condition the metric so defined is a Riemannian metric. Denoting the second partial derivatives by $D_{ij}L$, with $i, j \in \{1, 2\}$, we have

Theorem 3.1.5 (Aubry et al. , 1992) *Let Λ be a closed invariant set for a symplectic twist map F on M^2, and let $\Lambda' \subset M^{\mathbb{Z}}$ be the associated set of equilibrium states for the corresponding Frenkel-Kontorova model. Suppose that the matrix norms $|D_{12}L(x_n, x_{n+1})|$, $|D_{12}L(x_n, x_{n+1})^{-1}|$, $|D_{11}L(x_n, x_{n+1})|$, $|D_{22}L(x_n, x_{n+1})|$ are bounded for $\mathbf{x} \in \Lambda'$, $n \in \mathbb{Z}$. Then Λ is uniformly hyperbolic if and only if Λ' has phonon gap.*

In the next section we will recall how to use Theorem 3.1.4 to show the hyperbolicity of the shadowing orbits coming from the Theorem 1.3.4.

3.2 Hyperbolicity of the shadowing orbits and construction of the Poincaré map

We are going to illustrate the ideas of the proofs of Theorems 1.3.4 and 1.3.5, which, together with the results of the previous section, allow to show the hyperbolicity of the orbits which shadow the collision chains. We will use the notation of Subsection 1.3.1.

3.2.1 Scheme of the proof of the existence of quasi-collision orbits

The proof of Theorem 1.3.4 is based on the construction of a discrete action functional, whose properties allow also to prove the hyperbolicity of the orbits which shadow collision chains.

Regularisation

The first step is the regularisation of the Newtonian singularities. Let K be the finite set of labels for the collision arcs $\gamma_k : [0, \tau_k] \to D$, where $D = \{\mathbf{q} \in M | \; W(\mathbf{q}) < E\}$. We have chosen the energy value E such that the set of singularities $\mathcal{C}$ is contained in D. For any $C_i \in \mathcal{C}$, let U_i be a small ball centred at C_i in the metric defined by the kinetic energy T. For any point $a \in U_i$, there exist a unique trajectory $\gamma_a^+ : [0, \tau^+(a)] \to U_i$ of energy E for the system (L_0) which goes from a to C_i. Similarly, for any $b \in U_i$ there is a unique trajectory $\gamma_b^- : [-\tau^-(b), 0] \to U_i$ with energy E for the system (L_0), which connects C_i to b (see the Figure 3.1).

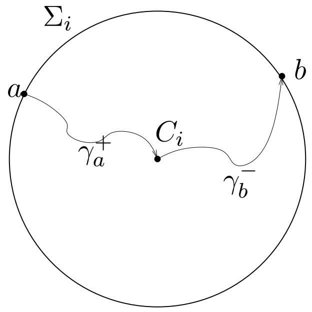

Figure 3.1. The collision trajectories γ_a^+, γ_b^- in the neighbourhood U_i.

Denote the actions of these trajectories by

$$S_i^+(a) = \int_0^{\tau^+(a)} \Big(L_0(\gamma_a^+(t), \dot\gamma_a^+(t)) + E \Big)\, dt\,, \tag{3.4}$$

$$S_i^-(b) = \int_{-\tau^-(b)}^{0} \Big(L_0(\gamma_b^-(t), \dot\gamma_b^-(t)) + E \Big)\, dt\ . \tag{3.5}$$

The functions $S_i^\pm$ are continuous on U_i and of class C^3 on $U_i \setminus \{C_i\}$. Denote with

$$u^+(a) = \dot\gamma_a^+(\tau^+(a))\,, \qquad u^-(b) = \dot\gamma_b^-(-\tau^-(b))\,,$$

the tangent vectors at the point C_i. Let $\Sigma_i = \partial U_i$ be the spherical boundary of the ball and $\delta > 0$ a real parameter, and define

$$X_i = \{(a,b) \in \Sigma_i^2 |\ \|u^+(a) - u^-(b)\| \geq \delta\}\,, \tag{3.6}$$

$$Y_i = \{(a,b) \in X_i |\ \|u^+(a) + u^-(b)\| \geq \delta\}\,, \tag{3.7}$$

where the norm $\|\cdot\|$ is given by the Riemannian metric of M. A pair $(a,b) \in \Sigma_i^2$ belongs to X_i if the solution of the system (L_0) connecting a to b with energy E does not pass too close to the centre C_i.

By the use of Levi-Civita regularisation (see [35]) in the 2-dimensional case and of KS-regularisation (see [32]) in the 3-dimensional one, together with the λ-lemma (see [26], [17]) it can be proved the following

Lemma 3.2.1 (Bolotin-Mackay, 2000) *There exists $\varepsilon_0 > 0$ such that:*

- *For any $\varepsilon \in (0, \varepsilon_0]$ and $(a,b) \in X_i$, there is a unique trajectory $\gamma = \gamma_{a,b}^\varepsilon : [0,\tau] \to U_i$ of energy E for the system (L_ε), with $\gamma_{a,b}^\varepsilon(0) = a$ and $\gamma_{a,b}^\varepsilon(\tau) = b$.*

- *$\tau = \tau(a,b,\varepsilon)$ is a C^2 function on $X_i \times (0,\varepsilon_0]$ and $\tau(a,b,\varepsilon) \to \tau^+(a) + \tau^-(b)$ uniformly as $\varepsilon \to 0$.*

- *$\gamma_{a,b}^\varepsilon\Big|_{[0,\tau^+(a)]} \to \gamma_a^+$ and $\gamma_{a,b}^\varepsilon(\cdot + \tau)\Big|_{[-\tau^-(b),0]} \to \gamma_b^-$ uniformly as $\varepsilon \to 0$.*

- *The action of the trajectory γ*

  $$S_i(a,b,\varepsilon) = \int_0^\tau (L_\varepsilon(\gamma(t), \dot\gamma(t)) + E)\, dt$$

  is a C^2 function on $X_i \times (0,\varepsilon_0]$ and $S_i(a,b,\varepsilon) = S_i^+(a) + S_i^-(b) + \varepsilon s_i(a,b,\varepsilon)$, where s_i is uniformly C^2 bounded on X_i as $\varepsilon \to 0$.

- *If $(a,b) \in Y_i$, then the trajectory $\gamma_{a,b}^\varepsilon$ does not pass too close to C_i:*

  $$\min_{0 \leq t \leq \tau} \mathrm{dist}(\gamma_{a,b}^\varepsilon(t), C_i) \geq c\varepsilon\,, \qquad c > 0\ .$$

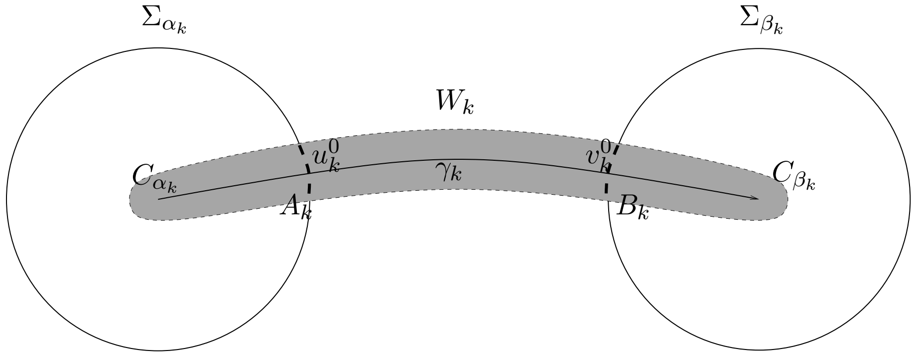

Figure 3.2. The intersections of the collision arc γ_k with the neighbourhoods of the centres C_{α_k} and C_{β_k}.

Construction of a discrete action functional

The second step of the proof of Theorem 1.3.4 is the construction of a discrete action functional, so that the orbits of (L_ε) correspond to critical points of this functional. For any $k \in K$, let $\gamma_k : [0, \tau_k] \to D$ be the collision arc from C_{α_k} to C_{β_k}, with $C_{\alpha_k}, C_{\beta_k} \in \mathcal{C} = \{C_1, \ldots, C_n\}$. Let $u_k^0 \in \Sigma_{\alpha_k}$ and $v_k^0 \in \Sigma_{\beta_k}$ be the intersection points of γ_k with Σ_{α_k} and Σ_{β_k}, respectively, as in Figure 3.2.

Then we must have

$$\begin{aligned}\gamma_k(t) &= \gamma^-_{u_k^0}(t - \tau^-(u_k^0))\,, \qquad 0 \le t \le \tau^-(u_k^0)\,,\\ \gamma_k(t) &= \gamma^+_{v_k^0}(t - (\tau_k - \tau^+(v_k^0)))\,, \qquad \tau_k - \tau^+(v_k^0)) \le t \le \tau_k\ .\end{aligned}$$

Without loss of generality we can assume that the points u_k^0 and v_k^0 are not conjugate along γ_k: indeed, there are only finitely many points which are conjugate to u_k^0 along γ_k (see [43, §15]). If u_k^0 and v_k^0 are conjugate, we change the radius of one of the balls $U_{\alpha_k}, U_{\beta_k}$ a little, so that the assumption is satisfied.

Let W_k be a small neighbourhood of $\gamma_k([0, \tau_k])$ and let $A_k = W_k \cap \Sigma_{\alpha_k}$, $B_k = W_k \cap \Sigma_{\beta_k}$ be the corresponding neighbourhoods of the points u_k^0, v_k^0. If these neighbourhoods are small enough, then by the non-conjugacy of u_k^0 and v_k^0 along γ_k and the implicit function theorem, for any $u \in A_k, v \in B_k$, there is a unique solution $\sigma = \sigma^\varepsilon_{uv} : [0, \tau] \to W_k$, $\tau = \tau^\varepsilon_{uv}$ of energy E for the system (L_ε), such that $\sigma(0) = u, \sigma(\tau) = v$, and σ is close to γ_k. This solution is a C^3 function of u, v. Denote the action of the trajectory σ^ε_{uv} by

$$f_k(u, v, \varepsilon) = \int_0^\tau (L_\varepsilon(\sigma(t), \dot\sigma(t)) + E)\, dt\ .$$

Then $f_k(\cdot, \cdot, \varepsilon)$ is a C^3 function on $A_k \times B_k$. By the nondegeneracy of γ_k as a critical point of the action functional it follows that

Lemma 3.2.2 (Bolotin-Mackay, 2000) *The function $g_k(u, v, \varepsilon) = f_k(u, v, \varepsilon) + S^-_{\alpha_k}(u) + S^+_{\beta_k}(v)$ defined on $A_k \times B_k$, has a nondegenerate critical point for $\varepsilon = 0$ at (u_k^0, v_k^0).*

At this point the strategy to prove Theorem 1.3.4 is to study the formal functional defined as follows. Given a sequence $(k_i)_{i\in\mathbb{Z}}$ defining a collision chain, take the set

$$Y = \prod_{i\in\mathbb{Z}} (A_{k_i} \times B_{k_i}) \ .$$

Note that if the neighbourhoods A_k and B_k are small enough, than by the direction change condition on the collision chain, we have automatically $B_{k_i} \times A_{k_{i+1}} \subset Y_{\beta_{k_i}}$, for any $i \in \mathbb{Z}$.

As showed in [8], the trajectories of system L_ε of energy E near the chain $(\gamma_{k_i})_{i\in\mathbb{Z}}$ correspond to critical points of the formal functional defined on Y by

$$F_\varepsilon(u,v) = \sum_{i\in\mathbb{Z}} f_{k_i,k_{i+1}}(u_i, v_i, u_{i+1}, \varepsilon) \,,$$

with

$$f_{k,k'}(u,v,u,\varepsilon) = g_k(u,v,\varepsilon) + \varepsilon s_\alpha(v,u',\varepsilon) \,, \qquad \alpha = \alpha_{k'} = \beta_k \,,$$

for any couple of consecutive indexes $(k,k') \in \Gamma$, where $\Gamma = \{(k,k') \in K^2 |\ \beta_k = \alpha_{k'},\ \dot\gamma_k(\tau_k) \neq \pm\dot\gamma_{k'}(0)\}$. The functions g_k and s_α are the ones defined in Lemma 3.2.2 and Lemma 3.2.1.

From these Lemmas and the implicit function theorem, it can be shown that for small $\varepsilon \in (0, \varepsilon_0]$ the functional F_ε has a nondegenerate critical point near $(u^0, v^0) = (u^0_{k_i}, v^0_{k_i})_{i\in\mathbb{Z}}$, which gives the shadowing orbit.

The formal functional F_ε allows also to use Frenkel-Kontorova models to find hyperbolicity, as we see in the following subsection.

3.2.2 Hyperbolicity and Poincaré map

The first step to find hyperbolicity is to reduce the functional F_ε to the form

$$\Phi_\varepsilon(u) = \sum_{i\in\mathbb{Z}} S_{k_i,k_{i+1}}(u_i, u_{i+1}, \varepsilon) \,, \qquad \Phi_0(u) = \sum_{i\in\mathbb{Z}} \phi_{k_i}(u_i) \,,$$

by the elimination of the variables v.

Up to change the radius of Σ_{α_k} a little, we can assume that the points u^0_k and C_{β_k} are not conjugate along γ_k, so that

$$\det D^2_v g_k(u^0_k, v^0_k, 0) \neq 0 \ .$$

Then, for ε small enough we can locally solve the equality

$$D_v f_{kk'}(u,v,u',\varepsilon) = 0 \,,\ u \in A_k \,,\ v \in B_k \,,\ u' \in A_{k'} \,,$$

thus gaining

$$v = w_k(u) + \mathcal{O}(\varepsilon) \ .$$

Then

$$f_{kk'}(u,v,u',\varepsilon) = S_{kk'}(u,u',\varepsilon) = \phi_k(u) + \varepsilon\psi_{kk'}(u,u') + \mathcal{O}(\varepsilon^2) \,,$$

with

$$\phi_k(u) = g_k(u, w_k(u), 0) \,, \qquad \psi_{kk'}(u,u') = s_\alpha(w_k(u), u', 0) + D_\varepsilon g_k(u, w_k(u), 0) \ .$$

In this last formulae we have used the fact that $D_v f_{kk'}(u, w_k(u) + \mathcal{O}(\varepsilon), u', \varepsilon)|_{\varepsilon=0} = D_v g_k(u, w_k(u), 0) \equiv 0$.

We have obtained, for any couple $(k, k') \in \Gamma$, a function $S_{kk'}$ defined in a small neighbourhood $A_k \times A_{k'}$ of $(u^0_k, u^0_{k'})$, such that stationary sequences (u, v) of F_ε correspond to stationary sequences u of the action

$$\Phi_\varepsilon(u) = \sum_{i\in\mathbb{Z}} S_{k_i,k_{i+1}}(u_i, u_{i+1}, \varepsilon) \ .$$

This action is not exactly of the form given in the previous section for the energy of a Frenkel-Kontorova model, because the $S_{kk'}$ are not equal for different couples of indexes (k, k'). But there is only a finite number of these functions, then we can replace the $S_{kk'}$ with a unique map defined on a disjoint union of the $A_k \times A_{k'}$, and taking the same values of $S_{kk'}$.

We know that for small ε, Φ_ε has a nondegenerate critical point u near u^0. If the action Φ_ε satisfy the condition (3.2), then the $S_{kk'}$, which are defined on the sets $A_k \times A_{k'}$, are the generating functions of symplectic maps $T_{kk'} : N_k \to N_{k'}$, where $N_k \subset T^*A_k$ is an open subset of the cotangent bundle on A_k. It can be proved (see [9]) that

$$\det D^2_{vu'} s_\alpha(v, u', 0) \neq 0 \,, \qquad (v, u') \in Y_\alpha \ .$$

It follows that the condition (3.2) is uniformly satisfied in $A_k \times A_{k'}$, more precisely that

$$\|(D^2_{uu'} S_{kk'}(u, u', \varepsilon))^{-1}\| \leq C\varepsilon^{-1} \,, \qquad (u, u') \in A_k \times A_{k'} \ . \tag{3.8}$$

Then the symplectic maps $T_{kk'} : (u, p) \mapsto (u', p')$ are well defined and by (3.8) we have $\|DT_{kk'}\| \leq c\varepsilon^{-1}$ uniformly on N_k. This means in particular that the cross section N_k can be identified with an open set in $T_{A_k} M \cap \{H_\varepsilon = E\}$ via the Legendre transform.

By the results of [1], recalled in Section 3.1, the nondegenerate stationary points of the Frenkel-Kontorova model given by Φ_ε, that is points with phonon gap, correspond to uniform hyperbolic orbits for the finite sequence of symplectic twist maps $T_{kk'}$. This proves uniform hyperbolicity.

Consider now the Poincaré map defined by the $T_{kk'}$. The inequality $\|DT_{kk'}\| \leq c\varepsilon^{-1}$ gives an upper bound for the Lyapunov exponents of order $\log \varepsilon^{-1}$. A lower bound of the same order comes from the proof of Proposition 1 of [1]: here we have a precise expression of the constants C and λ that appear in the Definition 3.1.1 of hyperbolic set, as functions of the parameter Δ defined by (3.3). From the estimate of Δ we can thus gain the desired bound for the Lyapunov exponents.

Part II

The use of the Kepler integrals for Orbit Determination: a method to find preliminary orbits and applications to the Orbit Determination of Asteroids and Space Debris.

Chapter 4

Introduction

In this first chapter we give a general introduction to the problem of orbit determination. After recalling the classical methods by Gauss and Laplace, we put the attention on some questions related to the current and next generation surveys. In particular, in Section 4.2 we introduce the problem of *linkage*, which is a particular kind of identification problem, and describe the program of solution proposed by Milani et al. in [38] and [41].

In the final subsection, we shall talk about the possibility of computing a finite set of preliminary orbits and solving the linkage problem by the use of the first integrals of the Kepler problem. This is the central subject of the next chapters, where the study of a method for preliminary orbit determination based on this idea is carried on, together with some numerical experiments to test the efficiency of one of the developed algorithms.

4.1 The classical methods of Gauss and Laplace

The words *Laplace's and Gauss' methods* usually refer to two classes of methods following the original ideas of the two authors respectively. We are going to expose these ideas following a modern approach, as we can find them in the more recent manuals (see [37]).

4.1.1 The orbit determination problem

When we consider the motion of a solar system body like a comet, an asteroid or even an artificial body, like an interplanetary spacecraft, the two-body model around the Sun is a very good approximation for short time intervals, provided that we can assume that close encounters with the planets are excluded. Then, an orbit is completely defined by the values of the six Keplerian elements at a reference time. If we know the orbital elements, we can compute the ephemeris of the object, giving the position of the body on the celestial sphere at any given time. The problem of orbit determination is exactly the reversed process, which consists of finding the elements of an orbit from observations. This is a very complicated work: usually, there is not a direct way to obtain the solution and we have to proceed by successive approximations.

A significant event in the history of orbit determination is the discovery of Ceres by Piazzi in 1801. Ceres is the first of the minor planets to have been discovered. When Piazzi detected this new object, he was able to follow it only for a few nights. It was important to predict when and where it could be observed again to be able to newly detect it. The problem to find an orbit on the basis of the knowledge of a small arc of one revolution arose. It was brilliantly solved by Gauss [22] and the asteroid was newly observed thanks to his predictions. Clearly the orbit found by this method was not very accurate, but it could be the starting point for successive improvements after the recovery of the object, when more observations were available.

As just seen, the first step of orbit determination is to find a *preliminary orbit* of a newly discovered comet or minor planet, on the basis of a minimum of observations. The classical methods of preliminary orbit determination by Laplace [33] and Gauss [22] are based on the knowledge of at least three observations of a solar system body in three different nights. An observation provides two angles at a given time, which identify a direction and then a position on the celestial sphere. Then, at least three observations are necessary to compute the six elements of an orbit.

Laplace's and Gauss' methods have been often revisited in the last two centuries, as we can see in [49], [34], [36], and, in their actual form, they are still used in the current orbit determination procedures. Both methods may produce more than one preliminary orbit for the same object: a detailed analysis of the occurrence of multiple solutions is in [24].

The determination of a preliminary orbit is followed by the *differential corrections* [2], an iterative method to obtain the minimum of a target function, that improves the orbit in the sense of the least squares fit of the residuals: this sequence of operations was already proposed in [22].

4.1.2 Laplace's method

An observation defines a point on the celestial sphere centered at the observer position. Let $\boldsymbol{\rho}$ be the topocentric position vector of the observed body and $\rho = |\boldsymbol{\rho}|$ the topocentric distance. Using spherical coordinates with respect to an equatorial reference system (e.g. J2000), let α denote the right ascension and δ the declination, so that the observation is represented by the unitary vector

$$\hat{\boldsymbol{\rho}} = (\cos\alpha\cos\delta, \sin\alpha\cos\delta, \sin\delta)\,.$$

Denote the heliocentric position of the observed body with $\mathbf{r}$ and the position of the observer on the Earth by $\mathbf{q}$, then

$$\mathbf{r} = \mathbf{q} + \rho\hat{\boldsymbol{\rho}}\ .$$

Consider the curve described by the topocentric vector $\boldsymbol{\rho} = \rho\hat{\boldsymbol{\rho}}$. Denote the time parameter by t and let s be the arc length parameter of the corresponding path observed on the celestial sphere. Then

$$\dot{\boldsymbol{\rho}} = \frac{d\boldsymbol{\rho}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\eta\hat{\mathbf{v}}\,,$$

where η is the *proper motion*, $\eta = \frac{ds}{dt} = |\frac{d\hat{\boldsymbol{\rho}}}{dt}|$, and $\hat{\mathbf{v}} = \frac{d\hat{\boldsymbol{\rho}}}{ds}$ the tangent unit vector. We shall use the moving orthonormal frame $\{\hat{\boldsymbol{\rho}}, \hat{\mathbf{v}}, \hat{\mathbf{n}}\}$, with $\hat{\mathbf{n}} = \hat{\boldsymbol{\rho}} \times \hat{\mathbf{v}}$. By standard computations we obtain $\frac{d\hat{\mathbf{v}}}{ds} = -\hat{\boldsymbol{\rho}} + \kappa\hat{\mathbf{n}}$, which defines the *geodesic curvature* κ. The acceleration of $\boldsymbol{\rho}$ is

$$\ddot{\boldsymbol{\rho}} = (\ddot{\rho} - \rho\eta^2)\hat{\boldsymbol{\rho}} + (2\dot{\rho}\eta + \rho\dot{\eta})\hat{\mathbf{v}} + (\rho\eta^2\kappa)\hat{\mathbf{n}} \ .$$

Both the Earth and the body observed, which could be for example a comet or a minor planet, have negligible mass with respect to the mass of the Sun. The lunar and planetary perturbations can be neglected, so that the only force operating is the gravitational attraction by the Sun. By using the *geocentric approximation* in which the observer position coincides with the centre of mass of the Earth, $\mathbf{q} = \mathbf{q}_\oplus$, we have for the accelerations

$$\ddot{\mathbf{r}} = -\frac{\mu}{r^3}\mathbf{r} \,, \qquad \ddot{\mathbf{q}} = -\frac{\mu}{q^3}\mathbf{q} \,,$$

with $r = |\mathbf{r}|$, $q = |\mathbf{q}|$, and μ the mass of the Sun multiplied by the gravitational constant. Then

$$\ddot{\boldsymbol{\rho}} = \ddot{\mathbf{r}} - \ddot{\mathbf{q}} = -\mu\left(\frac{\mathbf{q}+\boldsymbol{\rho}}{r^3} - \frac{\mathbf{q}}{q^3}\right) \ .$$

Considering the component in the direction $\hat{\mathbf{n}}$, we obtain

$$-\mu\left(\frac{1}{r^3} - \frac{1}{q^3}\right)(\mathbf{q}\cdot\hat{\mathbf{n}}) = \rho\eta^2\kappa \,,$$

and finally we have the *dynamical equation*

$$C\frac{\rho}{q} = 1 - \frac{q^3}{r^3} \,, \qquad \text{where} \qquad C = \frac{\eta^2\kappa q^3}{\mu(\hat{\mathbf{q}}\cdot\hat{\mathbf{n}})} \ . \tag{4.1}$$

Now, using the *geometric equation*

$$r^2 = q^2 + \rho^2 + 2q\rho\cos\epsilon \,, \tag{4.2}$$

where $\cos\epsilon = \hat{\mathbf{q}}\cdot\hat{\boldsymbol{\rho}}$, we can write a polynomial equation of degree eight for r

$$C^2r^8 - q^2(C^2 + 2C\cos\epsilon + 1)r^6 + 2q^5(C\cos\epsilon + 1)r^3 - q^8 = 0 \ . \tag{4.3}$$

The component of the acceleration $\ddot{\boldsymbol{\rho}}$ along $\hat{\mathbf{v}}$ gives

$$-\mu\left(\frac{1}{r^3} - \frac{1}{q^3}\right)(\mathbf{q}\cdot\hat{\mathbf{v}}) = 2\dot{\rho}\eta + \rho\dot{\eta} \ . \tag{4.4}$$

Equations (4.1), (4.2), (4.4) are the body of the method of Laplace. Suppose to have three observations at times t_i, $i = 1, 2, 3$, represented by the vectors

$$\hat{\boldsymbol{\rho}}_i = (\cos\alpha_i\cos\delta_i, \sin\alpha_i\cos\delta_i, \sin\delta_i) \,, \qquad i = 1, 2, 3 \ .$$

Then we can make a quadratic interpolation of the observed path, by using for example a degree 2 model to approximate the right ascension α and declination δ. This allows to compute approximate values for the quantities C, $\cos\epsilon$, $\dot{\eta}$ and the tangent unit vector $\hat{\mathbf{v}}$, at an intermediate time $\bar{t}$, usually the arithmetic mean of the times of the observations.

By the use of the equation (4.1), we can compute for any solution r of the polynomial equation (4.3), the corresponding value of ρ at the mean time $\bar{t}$. By equation (4.4), we obtain $\dot{\rho}$. From the values of $\rho, \dot{\rho}$ thus given, we can compute $\mathbf{r}$, $\dot{\mathbf{r}}$ and then the elements of the orbit at the corresponding epoch $\bar{t} - c/\rho$, which must be corrected by aberration, due to the finiteness of the velocity of light c.

We have described the classical Laplace's method which uses a geocentric approximation for the position of the observer. It is possible to consider the topocentric position by writing $\mathbf{q} = \mathbf{q}_{\oplus} + \boldsymbol{P}$, where $\boldsymbol{P}$ is the geocentric observer position: in this way we obtain the *topocentric Laplace's method*, described in [39], which is in fact preferable, especially when processing large data-sets, containing different classes of celestial objects that span a wide range of distances. In Laplace's classical geocentric method the geocentric acceleration $\ddot{\boldsymbol{P}}$ of the observer is not taken into account and the errors introduced in this way can be important. However, when observations from different nights are obtained from the same station at the same sidereal time, the observer's acceleration cancels out and the geocentric classical Laplace's method is a good approximation.

The use of the topocentric formulae was already proposed by Poincaré in 1906: he suggested (see [49, pp. 177–178]) to make a quadratic interpolation of the three geocentric positions of the observer at the times of the observations, thus obtaining approximate values for the geocentric position, velocity and acceleration $\boldsymbol{P}, \dot{\boldsymbol{P}}, \ddot{\boldsymbol{P}}$, to be inserted in the topocentric equations.

Curvature

If we have at least three observations, we can make a quadratic interpolation of the observed angles α_i and δ_i, thus obtaining approximate values for $(\alpha, \dot{\alpha}, \ddot{\alpha}, \delta, \dot{\delta}, \ddot{\delta})$ at a central time $\bar{t}$. This make it possible to compute the geodesic curvature of the apparent path described by the observed body on the celestial sphere.

The geodesic curvature κ is defined by the relation

$$\hat{\mathbf{v}}' = \frac{d\hat{\mathbf{v}}}{ds} = -\hat{\boldsymbol{\rho}} + \kappa\hat{\mathbf{n}} \ .$$

Then $\kappa = \hat{\mathbf{v}}' \cdot \hat{\mathbf{n}}$. By using the expression of $\hat{\boldsymbol{\rho}}$ in terms of the angles α and δ, we obtain

$$\hat{\mathbf{v}} = \frac{d\hat{\boldsymbol{\rho}}}{ds} = \alpha'\hat{\boldsymbol{\rho}}_\alpha + \delta'\hat{\boldsymbol{\rho}}_\delta \ ,$$
$$\hat{\mathbf{n}} = -\frac{\delta'}{\cos\delta}\hat{\boldsymbol{\rho}}_\alpha + \alpha'\cos\delta\hat{\boldsymbol{\rho}}_\delta \ ,$$

where $\hat{\boldsymbol{\rho}}_\alpha, \hat{\boldsymbol{\rho}}_\delta$ are the partial derivatives of $\hat{\boldsymbol{\rho}}$ with respect to α, δ respectively. Denoting by $\hat{\boldsymbol{\rho}}_{\alpha\alpha}, \hat{\boldsymbol{\rho}}_{\alpha\delta}, \hat{\boldsymbol{\rho}}_{\delta\delta}$ the second partial derivatives, we have

$$\hat{\mathbf{v}}' = (\alpha''\hat{\boldsymbol{\rho}}_\alpha + \delta''\hat{\boldsymbol{\rho}}_\delta) + (\alpha'^2\hat{\boldsymbol{\rho}}_{\alpha\alpha} + 2\alpha'\delta'\hat{\boldsymbol{\rho}}_{\alpha\delta} + \delta'^2\hat{\boldsymbol{\rho}}_{\delta\delta}) \ .$$

By the explicit computation of the second derivatives in terms of α, δ we arrive at

$$\kappa = \hat{\mathbf{v}}' \cdot \hat{\mathbf{n}} = (\delta''\alpha' - \alpha''\delta')\cos\delta + \alpha'(1 + \delta'^2)\sin\delta \ .$$

Finally, using the proper motion η, we obtain the desired expression of the geodesic curvature

$$\kappa = \frac{1}{\eta^3}\left[(\ddot{\delta}\dot{\alpha} - \ddot{\alpha}\dot{\delta})\cos\delta + \dot{\alpha}(\eta^2 + \dot{\delta}^2)\sin\delta\right] \ .$$

Moreover, from the expression

$$\eta = \sqrt{\dot{\alpha}^2 \cos^2 \delta + \dot{\delta}^2}$$

we have

$$\dot{\eta} = \frac{\ddot{\alpha}\dot{\alpha}\cos^2\delta + \ddot{\delta}\dot{\delta} - \dot{\alpha}^2\dot{\delta}\cos\delta\sin\delta}{\eta} \ .$$

These formulae allow to compute the approximate values of the coefficients appearing in the equations of Laplace's method, as we have previously stated.

4.1.3 Gauss' method

Let $\mathbf{r}_i$, $\boldsymbol{\rho}_i$, $i = 1, 2, 3$, be the heliocentric and topocentric positions respectively of the observed body at times t_i, with $t_1 < t_2 < t_3$. Let $\mathbf{q}_i$ be the corresponding heliocentric positions of the observer, so that

$$\mathbf{r}_i = \mathbf{q}_i + \boldsymbol{\rho}_i \,, \qquad i = 1, 2, 3 \ . \tag{4.5}$$

Assume that the interval between any two of these observations is much smaller than a period. With a two body approximation, the orbit of the observed object lies on a plane and the condition for coplanarity can be written

$$\lambda_1 \mathbf{r}_1 + \lambda_3 \mathbf{r}_3 = \mathbf{r}_2 \,, \tag{4.6}$$

with $\lambda_1, \lambda_3 \in \mathbb{R}$. For the same reason the vectors $\mathbf{r}_i \times \mathbf{r}_j$, with $i < j$, have the same orientation as the angular momentum per unit mass $\mathbf{c} = \mathbf{r}_h \times \dot{\mathbf{r}}_h$, for each $h = 1, 2, 3$. Then, by the vector product with $\mathbf{r}_i$ for $i = 1, 3$, we find

$$\lambda_1 = \frac{\mathbf{r}_2 \times \mathbf{r}_3 \cdot \hat{\mathbf{c}}}{\mathbf{r}_1 \times \mathbf{r}_3 \cdot \hat{\mathbf{c}}} \,, \qquad \lambda_3 = \frac{\mathbf{r}_1 \times \mathbf{r}_2 \cdot \hat{\mathbf{c}}}{\mathbf{r}_1 \times \mathbf{r}_3 \cdot \hat{\mathbf{c}}} \ .$$

These quantities are traditionally called *triangle area ratios*. From the scalar product of $\hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3$ with (4.6), and using (4.5) we obtain

$$\rho_2(\hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3 \cdot \hat{\boldsymbol{\rho}}_2) = \hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3 \cdot (\lambda_1 \mathbf{q}_1 - \mathbf{q}_2 + \lambda_3 \mathbf{q}_3) \ . \tag{4.7}$$

Consider now the f, g series formalism for the orbit at hand. Since the orbit is planar, functions f, g must exist such that

$$\mathbf{r}(t) = f(t_0, t)\mathbf{r}_0 + g(t_0, t)\dot{\mathbf{r}}_0 \,,$$

where t_0 is a fixed reference time and $\mathbf{r}_0 = \mathbf{r}(t_0)$, $\dot{\mathbf{r}}_0 = \dot{\mathbf{r}}(t_0)$ are the heliocentric position and velocity of the body at the time t_0. Consider the Taylor series

$$\mathbf{r}(t) = \mathbf{r}_0 + (t - t_0)\dot{\mathbf{r}}_0 + \frac{1}{2}(t - t_0)^2 \frac{d^2\mathbf{r}}{dt^2}(t_0) + \frac{1}{6}(t - t_0)^3 \frac{d^3\mathbf{r}}{dt^3}(t_0) + \mathcal{O}\left((t - t_0)^4\right) \ . \tag{4.8}$$

By the two body approximation,

$$\frac{d^2\mathbf{r}}{dt^2}(t_0) = -\frac{\mu \mathbf{r}_0}{r_0^3} = -\sigma \mathbf{r}_0 \,, \qquad \frac{d^3\mathbf{r}}{dt^3}(t_0) = -\frac{\mu \dot{\mathbf{r}}_0}{r_0^3} + \frac{3\mu \dot{r}_0 \mathbf{r}_0}{r_0^4} = -\sigma \dot{\mathbf{r}}_0 + 3\sigma\tau \mathbf{r}_0 \,,$$

where $r_0 = |\mathbf{r}_0|$, $\sigma = \frac{\mu}{r_0^3}$, and $\tau = \frac{\dot{r}_0}{r_0} = \frac{\mathbf{r}_0 \cdot \dot{\mathbf{r}}_0}{r_0^2}$. Thus, substituting into (4.8), we gain the first terms of the f and g series

$$\begin{cases} f(t_0, t) = 1 - \frac{1}{2}\sigma(t - t_0)^2 + \frac{1}{2}\sigma\tau(t - t_0)^3 + \mathcal{O}((t - t_0)^4)\,, \\ g(t_0, t) = (t - t_0) - \frac{1}{6}\sigma(t - t_0)^3 + \mathcal{O}((t - t_0)^4)\ . \end{cases}$$

Taking t_2 as the central time t_0, we have $\mathbf{r}_i = f_i \mathbf{r}_2 + g_i \dot{\mathbf{r}}_2$, with

$$f_i = 1 - \frac{\mu}{2}\frac{\tau_{i2}^2}{r_2^3} + \mathcal{O}(\Delta t^3)\,, \quad g_i = \tau_{i2}\left(1 - \frac{\mu}{6}\frac{\tau_{i2}^2}{r_2^3}\right) + \mathcal{O}(\Delta t^4)\,, \tag{4.9}$$

where we have written $\mathcal{O}(\Delta t)$ for the order of magnitude of the time differences. Then $\mathbf{r}_i \times \mathbf{r}_2 = -g_i \mathbf{c}$, $\mathbf{r}_1 \times \mathbf{r}_3 = (f_1 g_3 - f_3 g_1)\mathbf{c}$ and

$$\lambda_1 = \frac{g_3}{f_1 g_3 - f_3 g_1}\,, \qquad \lambda_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1}\,, \tag{4.10}$$

$$f_1 g_3 - f_3 g_1 = \tau_{31}\left(1 - \frac{\mu}{6}\frac{\tau_{31}^2}{r_2^3}\right) + \mathcal{O}(\Delta t^4)\ . \tag{4.11}$$

Substituting (4.9) and (4.11) in (4.10), we obtain λ_1 and λ_3 in terms of the time intervals and the radial distance r_2:

$$\begin{aligned} \lambda_1 &= \frac{\tau_{32}}{\tau_{31}}\left(1 + \frac{\mu}{6r_2^3}\left(\tau_{31}^2 - \tau_{32}^2\right)\right) + \mathcal{O}(\Delta t^3)\,, \\ \lambda_3 &= \frac{\tau_{21}}{\tau_{31}}\left(1 + \frac{\mu}{6r_2^3}\left(\tau_{31}^2 - \tau_{21}^2\right)\right) + \mathcal{O}(\Delta t^3)\ . \end{aligned} \tag{4.12}$$

Let $V = \hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_2 \cdot \hat{\boldsymbol{\rho}}_3$. By entering the expressions of λ_1, λ_3 in (4.7) and using the relations $\tau_{31}^2 - \tau_{32}^2 = \tau_{21}(\tau_{31} + \tau_{32})$ and $\tau_{31}^2 - \tau_{21}^2 = \tau_{32}(\tau_{31} + \tau_{21})$, we find

$$\begin{aligned} &- V\rho_2\tau_{31} = \hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3 \cdot (\tau_{32}\mathbf{q}_1 - \tau_{31}\mathbf{q}_2 + \tau_{21}\mathbf{q}_3) + \\ &+ \hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3 \cdot \left(\frac{\mu}{6r_2^3}\tau_{32}\tau_{21}\left((\tau_{31} + \tau_{32})\mathbf{q}_1 + (\tau_{31} + \tau_{21})\mathbf{q}_3\right)\right) + \mathcal{O}(\Delta t^4)\ . \end{aligned} \tag{4.13}$$

Neglecting the terms of order $\mathcal{O}(\Delta t^4)$, the coefficient of $\frac{1}{r_2^3}$ is

$$B(\mathbf{q}_1, \mathbf{q}_3) = \frac{\mu}{6}\tau_{32}\tau_{21}\hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3 \cdot \left((\tau_{31} + \tau_{32})\mathbf{q}_1 + (\tau_{31} + \tau_{21})\mathbf{q}_3\right)\ .$$

Multiplying (4.13) by $\frac{q_2^3}{B(\mathbf{q}_1, \mathbf{q}_3)}$, we obtain

$$-\frac{V\rho_2\tau_{31}}{B(\mathbf{q}_1, \mathbf{q}_3)}q_2^3 = \frac{q_2^3}{r_2^3} + \frac{A(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)}{B(\mathbf{q}_1, \mathbf{q}_3)}\,,$$

where

$$A(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = q_2^3\hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_3 \cdot (\tau_{32}\mathbf{q}_1 - \tau_{31}\mathbf{q}_2 + \tau_{21}\mathbf{q}_3)\ .$$

Letting

$$C = \frac{V\tau_{31}q_2^4}{B(\mathbf{q}_1, \mathbf{q}_3)}\,, \qquad \gamma = -\frac{A(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)}{B(\mathbf{q}_1, \mathbf{q}_3)}\,,$$

we finally obtain the dynamical equation of Gauss' method

$$C\frac{\rho_2}{q_2} = \gamma - \frac{q_2^3}{r_2^3}\ .$$

Putting together this equation with the geometric equation $r_2^2 = \rho_2^2 + q_2^2 + 2\rho_2 q_2 \cos\epsilon_2$, we can compute the possible values of r_2. Then we can use for example the Gibbs' formulae to obtain the velocity vector $\dot{\mathbf{r}}_2$ (see [25, Ch. 8]).

We are going to briefly describe this formulae. Given the values of λ_1, λ_3, from the scalar product of (4.6) with $\hat{\boldsymbol{\rho}}_1 \times \hat{\boldsymbol{\rho}}_2$ and with $\hat{\boldsymbol{\rho}}_2 \times \hat{\boldsymbol{\rho}}_3$, we obtain linear equations for ρ_3, ρ_1 respectively. Then we can compute $\mathbf{r}_1, \mathbf{r}_3$ by (4.5).

By expanding $\mathbf{r}(t)$ in a Taylor series about the time t_2, we have

$$\begin{aligned}
\mathbf{r}_1 &= \mathbf{r}_2 + \tau_{12}\frac{d\mathbf{r}_2}{dt} + \frac{\tau_{12}^2}{2}\frac{d^2\mathbf{r}_2}{dt^2} + \frac{\tau_{12}^3}{6}\frac{d^3\mathbf{r}_2}{dt^3} + \frac{\tau_{12}^4}{24}\frac{d^4\mathbf{r}_2}{dt^4} + \mathcal{O}(\Delta t^5)\,,\\
\mathbf{r}_3 &= \mathbf{r}_2 + \tau_{32}\frac{d\mathbf{r}_2}{dt} + \frac{\tau_{32}^2}{2}\frac{d^2\mathbf{r}_2}{dt^2} + \frac{\tau_{32}^3}{6}\frac{d^3\mathbf{r}_2}{dt^3} + \frac{\tau_{32}^4}{24}\frac{d^4\mathbf{r}_2}{dt^4} + \mathcal{O}(\Delta t^5)\,,
\end{aligned}$$

where we have denoted by $\frac{d^k\mathbf{r}_2}{dt^k}$ the derivatives of $\mathbf{r}(t)$ at time t_2. We can eliminate the term $\frac{d^2\mathbf{r}_2}{dt^2}$ by multiplying $\mathbf{r}_1$ for τ_{32}^2, $\mathbf{r}_3$ for τ_{12}^2 and subtracting, thus obtaining the equality

$$\begin{aligned}
&\tau_{32}^2\mathbf{r}_1 + (\tau_{12}^2 - \tau_{32}^2)\mathbf{r}_2 - \tau_{12}^2\mathbf{r}_3 = \\
&= \tau_{12}\tau_{13}\tau_{32}\left(-\frac{d\mathbf{r}_2}{dt} + \frac{\tau_{12}\tau_{32}}{6}\frac{d^3\mathbf{r}_2}{dt^3} + \frac{\tau_{12}\tau_{32}(\tau_{12}+\tau_{32})}{24}\frac{d^4\mathbf{r}_2}{dt^4} + \mathcal{O}(\Delta t^5)\right)\ .
\end{aligned} \tag{4.14}$$

Now we consider the Taylor expansion of the second derivatives

$$\begin{aligned}
\frac{d^2\mathbf{r}_1}{dt^2} &= \frac{d^2\mathbf{r}_2}{dt^2} + \tau_{12}\frac{d^3\mathbf{r}_2}{dt^3} + \frac{\tau_{12}^2}{2}\frac{d^4\mathbf{r}_2}{dt^4} + \mathcal{O}(\Delta t^3)\,,\\
\frac{d^2\mathbf{r}_3}{dt^2} &= \frac{d^2\mathbf{r}_2}{dt^2} + \tau_{32}\frac{d^3\mathbf{r}_2}{dt^3} + \frac{\tau_{32}^2}{2}\frac{d^4\mathbf{r}_2}{dt^4} + \mathcal{O}(\Delta t^3)\ .
\end{aligned}$$

By a procedure analogous to the previous one, we can eliminate the terms $\frac{d^3\mathbf{r}_2}{dt^3}$ and $\frac{d^4\mathbf{r}_2}{dt^4}$, thus obtaining

$$\tau_{32}\frac{d^2\mathbf{r}_1}{dt^2} + \tau_{13}\frac{d^2\mathbf{r}_2}{dt^2} - \tau_{12}\frac{d^2\mathbf{r}_3}{dt^2} = \frac{\tau_{12}\tau_{13}\tau_{32}}{2}\frac{d^4\mathbf{r}_2}{dt^4} + \mathcal{O}(\Delta t^4)\,, \tag{4.15}$$

$$\tau_{32}^2\frac{d^2\mathbf{r}_1}{dt^2} + (\tau_{12}^2 - \tau_{32}^2)\frac{d^2\mathbf{r}_2}{dt^2} - \tau_{12}^2\frac{d^2\mathbf{r}_3}{dt^2} = -\tau_{12}\tau_{13}\tau_{32}\frac{d^3\mathbf{r}_2}{dt^3} + \mathcal{O}(\Delta t^4)\ . \tag{4.16}$$

Keeping only terms up to the order $\mathcal{O}(\Delta t^4)$, we can insert (4.15) and (4.16) in (4.14), thus having

$$\begin{aligned}
&\tau_{32}^2\mathbf{r}_1 + (\tau_{12}^2 - \tau_{32}^2)\mathbf{r}_2 - \tau_{12}^2\mathbf{r}_3 + \tau_{12}\tau_{13}\tau_{32}\frac{d\mathbf{r}_2}{dt} = \\
&= \frac{\tau_{12}\tau_{13}\tau_{32}}{12}\left(\tau_{32}\frac{d^2\mathbf{r}_1}{dt^2} - (\tau_{12}+\tau_{32})\frac{d^2\mathbf{r}_2}{dt^2} + \tau_{12}\frac{d^2\mathbf{r}_3}{dt^2}\right)\ .
\end{aligned} \tag{4.17}$$

Using now the two-body motion $\frac{d^2\mathbf{r}_i}{dt^2} = -\mu\frac{\mathbf{r}_i}{r_i^3}$, $i = 1, 2, 3$, we arrive at

$$\dot{\mathbf{r}}_2 = -\omega_1\mathbf{r}_1 + \omega_2\mathbf{r}_2 + \omega_3\mathbf{r}_3 \, ,$$

where

$$\omega_i = G_i + \frac{H_i}{r_i^3} \, , \qquad i = 1, 2, 3 \, ,$$

$$G_1 = \frac{\tau_{32}^2}{\tau_{12}\tau_{13}\tau_{32}} \, , \; G_3 = \frac{\tau_{12}^2}{\tau_{12}\tau_{13}\tau_{32}} \, , \; G_2 = G_1 - G_3 \, ,$$

$$H_1 = \frac{\mu\tau_{32}}{12} \, , \; H_3 = -\frac{\mu\tau_{12}}{12} \, , \; H_2 = H_1 - H_3 \; .$$

When $\mathbf{r}_2$ and $\dot{\mathbf{r}}_2$ are available, they provide a preliminary two-body orbit, from which new solutions $\mathbf{r}_1, \mathbf{r}_3$ can be computed. Note that the times of the observations t_i, $i = 1, 2, 3$, are not the times of the celestial body, because of the finiteness of the velocity of light c. Then we have to consider the times $t_i - \rho_i/c$, for $i = 1, 2, 3$, as the times of the positions $\mathbf{r}_i$, i.e. the times corrected by aberration. After $\mathbf{r}_1$, $\mathbf{r}_3$ have been obtained by a propagation along the orbit at the corrected times, we can compute new values of λ_1, λ_3 by (4.6). Then (4.7) gives an improved value of ρ_2, from which a new iteration could be started. If the procedure converge, at the end we will have an optimal orbit with respect to the three given observations. We can not exclude the possibility of divergence. A recent analysis of the convergence of this iterative procedure, named *Gauss' map*, can be seen in [13] and [14], where it is shown that effectively each step of the iteration improve the solution of the two-body problem, but despite this the Gauss' map can diverge when the solution r_2 of the degree 8 polynomial equation is outside the convergence domain. Thus the Gauss map should be used with some caution, for example with a recovery procedure in case of divergence.

There is a critical difference between the methods of Gauss and Laplace. In the method of Gauss a truncation to order $\mathcal{O}(\Delta t^2)$ is used only for the heliocentric motion of the body $\mathbf{r}(t)$, but the observer positions $\mathbf{q}_i$ are used with their exact values, being them coincident with the centre of the Earth or not. Laplace instead uses a truncation to the same order for the relative motion $\boldsymbol{\rho}(t)$, thus introducing implicitly an approximation for the motion of the observer. A rigorous comparison between the two methods can be found in [39], where an equivalence of the two techniques to the order zero in Δt is found when the averaged time is $\bar{t} = t_2$, and in Gauss' method the geocentric approximation and the truncation to the second order of the observer motion is introduced, with the use of the f, g series even for the Earth's motion. In the same paper it is also proved an equivalence to order $\mathcal{O}(\Delta t)$ between Gauss' and topocentric Laplace's methods, which is valid if in Gauss' method a truncation to order $\mathcal{O}(\Delta t^2)$ is used to compute the observer positions $\mathbf{q}_1, \mathbf{q}_3$ and the small terms $\mathcal{O}\left(\frac{P}{q}\right)$ are neglected, where P is the distance of the observer from the centre of the Earth. The conclusion is that Gauss' method seems superior, because it naturally accounts for topocentric observations.

4.2 Orbit determination with very short arcs

In this section we describe the problem of linkage, which is related to the modern techniques used for the astrometric observations. We talk about some recent methods for solving it, which are efficient also under the conditions expected for the next generation surveys.

4.2.1 The linkage problem

At the times of Piazzi, Olbers, and Gauss, asteroids were detected by comparison of the observations with a star catalogue. The amount of work for any individual observation was clearly high and multiple observations in the same night were rare. On the contrary, the data of the current surveys generally do not provide a single observation for an object in an observing night: in fact, the moving objects are distinguished from fixed stars by detecting them a few times in the same night.

By the use of the CCD, the detecting process is completely automated: a number N, with $3 \leq N \leq 5$, of digital images of the same area of the celestial sphere is taken within a short time span, typically within a single night, and an object is detected if, with the analysis of the images by a computer program, it is found to move on a straight line with uniform velocity. When such an occurrence happens, the sequence of N observations of the detection is reported to the Minor Planet Center (MPC): such a sequence is called a *Very Short Arc*. The possibility of developing algorithms to detect a moving object between different frames is guaranteed by the short time span from one image to another, so that the curvature of the arc is necessarily small. But, due to the very low curvature of the arc, even if the sequence of observations is enough for a detection, in most cases it is not for computing a full orbit: in this case we speak of a *Too Short Arc* (see [40]).

The information contained in a short arc of observations can be used to define an *attributable* [42], consisting of a reference time, which is just the mean of the observing times, two average angular coordinates and two corresponding angular rates at the reference epoch. The topocentric distance and the radial velocity at that time are unknown. Therefore two short arcs of observations belonging to the same object provide us 8 scalar data, from which we can try to compute an orbit.

In the current surveys, the number of detected objects per night is very large, of the order of many thousands for every night of operation. With the new observational techniques of the next generation surveys, like Pan-STARRS and LSST [1], the number of moving objects detected in each night of observations is expected to increase by two orders of magnitude with respect to the current surveys. We can expect this number to be between 100,000 and 1 million of detections per night.

The large number of detections per night makes it difficult to decide whether two sequences of observations made in different nights belong to the same object: this gives rise to the problem of *linkage* of two short arcs. The problem can be described as follows. Given a finite number of nights of observations, usually covering a time interval of about one month, we have to decide which couples of attributables, one from a night, the other from a different night, can belong to the same object. We also have to compute some preliminary orbit which roughly satisfies the observations of both arcs of an identified object. The computed preliminary orbits have to be good enough to start a differential correction procedure, which can converge to give an orbit fitting well both arcs. Moreover, it is necessary an overall control of the computational complexity.

[1] see the web pages `http://pan-starrs.ifa.hawaii.edu`, and `http://www.lsst.org`

Note that while this description is suitable for a Main Belt asteroid, it is not so for a Trans-Neptunian or for a body very close to the Earth at the time of the detection. In the first case, to form an attributable we need observations covering many days, up to one month, while in the second case we may have more than one arc in the same night.

A solution to the linkage problem is proposed in [38] and [41]. We have used the algorithms developed in these works as a preliminary step in our numerical experiments, as described in Section 6.2. Nevertheless, before exposing the solution proposed in these papers, we need to establish more precisely the basic concept of attributable, by describing how it is computed.

4.2.2 Attributables

Suppose that an asteroid is observed from the Earth at a time t. Let $\mathbf{r}$ be its heliocentric position and $\mathbf{q}$ be the heliocentric position of the observer on the Earth. Let $(\rho, \alpha, \delta) \in \mathbb{R}^+ \times [-\pi, \pi) \times (-\pi/2, \pi/2)$ be spherical coordinates for the topocentric position of the asteroid, which we denote by $\boldsymbol{\rho} = \rho\hat{\boldsymbol{\rho}}$. The reference system can be selected as necessary: usually we use an equatorial system, e.g. J2000, so that α is the right ascension and δ the declination with respect to this system.

An astrometric observation measures the quantities $(t, \alpha, \delta, h) \in \mathbb{R} \times [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}$, where the angles α, δ define the position observed on the celestial sphere at time t and h is the apparent magnitude, which is an optional parameter. A sequence of $m \geq 2$ astrometric observations is called a *Very Short Arc* if they are fitted well by a polynomial curve of low degree, so that it can be assumed that they belong to the same object without the computation of an orbit.

Given a very short arc of observations $(t_i, \alpha_i, \delta_i)$, $i = 1, \dots m$, with $m \geq 2$, it is often possible to compute an attributable[2] with its uncertainty. An *attributable* is a vector

$$\mathcal{A} = (\alpha, \delta, \dot{\alpha}, \dot{\delta}) \in [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}^2 \,,$$

representing the angular position and velocity of the body at an average time $\bar{t}$ in the selected coordinates (see [42]). The topocentric distance ρ and radial velocity $\dot{\rho}$ are completely unknown from the attributable.

The time $\bar{t}$ is the mean of the times t_i. Since the observations are given with their uncertainty, then $\bar{t}$ must be a weighted mean. Usually, the observations from the same station at the same date have the same weight, thus in most cases $\bar{t}$ is just the arithmetic mean of the observing times t_i.

The computation of the attributable $\mathcal{A}$ corresponding to the given sequence of observations is made by a Linear Least Squares fit (see [37, Ch. 5]), using a polynomial model for the angles α and δ. If $m = 2$ then a linear model must be used, while for $m \geq 3$ a quadratic model can always be considered. Moreover, the fit must take into account the weights of the individual observations.

[2]The name refers to the possibility of attributing the observations of the short arc to an already known orbit.

Assuming $m \geq 3$, the approximating functions are

$$\begin{aligned} \alpha(t) &= \alpha + \dot{\alpha}(t-\bar{t}) + \frac{1}{2}\ddot{\alpha}(t-\bar{t})^2 \,, \\ \delta(t) &= \delta + \dot{\delta}(t-\bar{t}) + \frac{1}{2}\ddot{\delta}(t-\bar{t})^2 \;. \end{aligned} \tag{4.18}$$

The target function is given by

$$Q = \frac{1}{2m}\boldsymbol{\xi} \cdot W\boldsymbol{\xi} \,,$$

where W is the weight matrix and $\boldsymbol{\xi}$ is the vector of the residuals

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_\alpha \\ \boldsymbol{\xi}_\delta \end{pmatrix}, \quad \boldsymbol{\xi}_\alpha = \begin{pmatrix} \alpha_1 - \alpha(t_1) \\ \vdots \\ \alpha_m - \alpha(t_m) \end{pmatrix}, \quad \boldsymbol{\xi}_\delta = \begin{pmatrix} \delta_1 - \delta(t_1) \\ \vdots \\ \delta_m - \delta(t_m) \end{pmatrix} .$$

The matrix W is related to the error distribution of the observations and it is symmetric with positive eigenvalues. Usually, we assume that the α and δ error components are not correlated, so that W is of the form

$$W = \begin{pmatrix} W_\alpha & 0 \\ 0 & W_\delta \end{pmatrix},$$

with W_α and W_δ symmetric positive definite $m \times m$ matrices. We define

$$\mathbf{x} = (\alpha, \dot{\alpha}, \ddot{\alpha}, \delta, \dot{\delta}, \ddot{\delta})^T \,, \qquad \boldsymbol{\lambda} = (\alpha_1, \ldots, \alpha_m, \delta_1, \ldots, \delta_m)^T \;.$$

The value $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{x})$ that minimizes the target function is obtained from the solution of the normal equation

$$C\mathbf{x} = -B^T W \boldsymbol{\lambda} \,, \qquad \text{with } B = \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} \,, \; C = B^T W B \;.$$

The normal matrix C is symmetric and it is positive definite. Its inverse $\Gamma = C^{-1}$ is the covariance matrix. It has the form

$$\Gamma = \begin{pmatrix} \Gamma_\alpha & 0 \\ 0 & \Gamma_\delta \end{pmatrix},$$

where $\Gamma_\alpha, \Gamma_\delta$ are the two 3×3 marginal covariance matrices associated with the solutions $(\alpha, \dot{\alpha}, \ddot{\alpha})$ and $(\delta, \dot{\delta}, \ddot{\delta})$ respectively. Actually, this is true because we have assumed that the α and δ error components are not correlated. Otherwise the matrix Γ could be full. The marginal covariance matrix $\Gamma_{\mathcal{A}}$ associated with the attributable is obtained simply by extracting the relevant 4×4 submatrix, and the normal matrix is $C_{\mathcal{A}} = \Gamma_{\mathcal{A}}^{-1}$. The covariance matrix $\Gamma_{\mathcal{A}}$ is symmetric and positive definite.

If there are only two observations, a linear model must be used in place of (4.18). Assuming that α and δ are not correlated, the fit gives the values $(\alpha, \dot{\alpha}, \delta, \dot{\delta})$ with the 2×2 covariance matrices $\Gamma_\alpha, \Gamma_\delta$, which form the covariance matrix $\Gamma_{\mathcal{A}}$ of the attributable. If the two observations have equal weight $1/\sigma^2$ and the time difference is Δt, then the correlations $Corr(\alpha, \dot{\alpha})$ and $Corr(\delta, \dot{\delta})$ are zero and $\Gamma_{\mathcal{A}}$ is diagonal: the standard deviation of both angles is $\sigma/\sqrt{2}$ and the standard deviation of both angular rates is $\sqrt{2}\sigma/\Delta t$.

In conclusion, the output of the polynomial fits are: the attributable four coordinates $(\alpha, \delta, \dot{\alpha}, \dot{\delta})$, the central time $\bar{t}$, the estimated second derivatives $\ddot{\alpha}, \ddot{\delta}$ (only if the fit is quadratic), and the covariance matrices $\Gamma_\alpha, \Gamma_\delta$. If the observations provide also the apparent magnitudes h_i, then we can associate to the attributable an apparent magnitude $\bar{h}$, which is the mean of the h_i.

4.2.3 Linkage by the triangulation of the admissible region

Given an attributable $\mathcal{A}$, the topocentric distance and radial velocity are not available. However, if we can assume that the observed body belongs to the solar system and not to the Earth-Moon system, then the values $(\rho, \dot{\rho})$ are constrained to a compact subset of $\mathbb{R}^2$, which is called *Admissible Region* (see [38]).

The hypothesis which define the admissible region can be put into analytical form by considering the expression of the following quantities in terms of $\rho, \dot{\rho}$:

- heliocentric two-body energy

$$\mathcal{E}_\odot(\rho, \dot{\rho}) = \frac{1}{2}|\dot{\mathbf{r}}|^2 - k^2 \frac{1}{|\mathbf{r}|} ,$$

  with k denoting the Gauss constant;

- geocentric two-body energy

$$\mathcal{E}_\oplus(\rho, \dot{\rho}) = \frac{1}{2}|\dot{\boldsymbol{\rho}}|^2 - k^2 \mu_\oplus \frac{1}{|\boldsymbol{\rho}|} ,$$

  with $\mu_\oplus$ the ratio between the mass of the Earth and that of the Sun.

Then we must take into account the physical radius $R_\oplus$ of the Earth and the radius of the sphere of influence of the Earth

$$R_{SI} = a_\oplus \left(\frac{\mu_\oplus}{3}\right)^{\frac{1}{3}} ,$$

which is the distance from the Earth to the collinear Lagrangian point L_2, apart from terms of order $\mu_\oplus^{2/3}$. Here $a_\oplus$ denotes the semimajor axis of the orbit of the Earth.

The admissible region is the set

$$\mathcal{D} = \{\mathcal{D}_1 \cup \mathcal{D}_2\} \cap \mathcal{D}_3 \cap \mathcal{D}_4 ,$$

where the subsets $\mathcal{D}_i \subset \mathbb{R}^2$, for $i = 1, \ldots, 4$, are defined as follows:

$$\begin{aligned} \mathcal{D}_1 &= \{(\rho, \dot{\rho}) : \ \mathcal{E}_\oplus \geq 0\} , \qquad & \mathcal{D}_2 &= \{(\rho, \dot{\rho}) : \ \rho \geq R_{SI}\} , \\ \mathcal{D}_3 &= \{(\rho, \dot{\rho}) : \ \mathcal{E}_\odot \leq 0\} , \qquad & \mathcal{D}_4 &= \{(\rho, \dot{\rho}) : \ \rho \geq R_\oplus\} . \end{aligned}$$

If the vector $(\rho, \dot{\rho})$ belongs to $\mathcal{D}_1 \cup \mathcal{D}_2$ then the Asteroid is not a satellite of the Earth, or it is not in its sphere of influence. The set $\mathcal{D}_3$ corresponds to the condition of belonging to the solar system, while the set $\mathcal{D}_4$ is simply the condition to be outside the Earth. The admissible region is compact and it consists of at most two connected components.

Any point of the region together with the attributable $\mathcal{A}$ is a set of six initial conditions and then it defines uniquely an orbit: the set $X = [\alpha, \delta, \dot{\alpha}, \dot{\delta}, \rho, \dot{\rho}]$ is called a set of *attributable orbital elements*.

Anyway, the points of the admissible region are infinitely many and we cannot proceed with computations for each point of the region. There is the need to sample it with a finite and not too large subset. In [38], after giving the analytical definition and geometrical properties of the admissible region, the authors define an algorithm to construct a *Delaunay triangulation* of it.

A Delaunay triangulation is defined by the following properties:

(i) it maximizes the minimum angle;

(ii) it minimizes the maximum circumcircle;

(iii) for any triangle T_i of the triangulation, the interior part of its circumcircle does not contain any node of the triangulation.

The conditions (i), (ii), (iii) are equivalent if the domain is convex.

A first *constrained* triangulation is constructed by sampling the boundary of $\mathcal{D}$, taking the selected points as nodes, and imposing the boundary edges to be edges of the triangulation. A constrained triangulation with properties (i), (ii) always exists: it is called *constrained Delaunay triangulation* and a procedure can be defined to construct it, by the iteration of the *edge-flipping technique*, which is now explained. For a convex quadrangle, there are only two triangulations with its vertices as nodes and the boundary edges belonging to the triangulation. One of them is certainly a Delaunay triangulation and it can be obtained from the other one by substituting the diagonal of the quadrangle which is an edge of the triangulation with the other diagonal: this is the edge-flipping technique, which must be iterated over the couples of adjacent triangles.

The initial triangulation can then be refined by adding new points internal to the domain, maintaining at each insertion the Delaunay properties.

As described in [41], the admissible region can be used to solve the linkage problem on a very large data-set containing many short arcs of observations. We summarize here the ideas of this procedure.

To each attributable $\mathcal{A}$ it is associated an admissible region $\mathcal{D}(\mathcal{A})$, and for each point $B = (\rho, \dot{\rho}) \in \mathcal{D}(\mathcal{A})$ we can define an orbit with attributable elements $X = [\mathcal{A}, B]$. Their uncertainty can be described by defining formally the covariance and normal matrices Γ_X, C_X as

$$\Gamma_X = \begin{pmatrix} \Gamma_{\mathcal{A}} & 0 \\ 0 & 0 \end{pmatrix}, \qquad C_X = \begin{pmatrix} C_{\mathcal{A}} & 0 \\ 0 & 0 \end{pmatrix}, \tag{4.19}$$

where $\Gamma_{\mathcal{A}}, C_{\mathcal{A}}$ are the covariance and normal matrices of the attributable $\mathcal{A}$. Note that Γ_X and C_X are not positive-definite and are not inverse of each other, but pseudo-inverse. Anyway, we can use them to compute the uncertainty of the predictions and this justifies the use of the names *covariance matrix* and *normal matrix* for Γ_X and C_X. In particular, the lower right 2×2 zero submatrix expresses the fact that the values of B have no uncertainty: indeed, they have been assumed, they are not measured or computed.

Given an attributable $\mathcal{A}_0$, with time $\bar{t}_0$, and chosen $B_0 \in \mathcal{D}(\mathcal{A}_0)$, the *confidence region* of the attributable orbital elements $X = [\mathcal{A}_0, B_0]$ can be approximated with the product

$$Z_X^0(\sigma) = \{\mathcal{A} : (\mathcal{A} - \mathcal{A}_0)^T C_{\mathcal{A}_0} (\mathcal{A} - \mathcal{A}_0) \leq \sigma^2\} \times \mathcal{D}(\mathcal{A}_0)\,,$$

where σ is a parameter.

A natural way to sample this set is to use a Delaunay triangulation of $\mathcal{D}(\mathcal{A}_0)$. Let $\{B^i = (\rho_i, \dot{\rho}_i)\}_{i=1,\ldots,k} \subset \mathcal{D}(\mathcal{A}_0)$ be the nodes of the triangulation. Then the Virtual Asteroids sampling $Z_X^0(\sigma)$ are given by the attributable elements $\{X^i = [\mathcal{A}_0, B^i]\}_{i=1,\ldots,k}$, with epoch times $t_0^i = \bar{t}_0 - \rho_i/c$, being $\bar{t}_0$ the time of $\mathcal{A}_0$.

Taken two attributables $\mathcal{A}_0, \mathcal{A}_1$, with times $\bar{t}_0, \bar{t}_1$ respectively, we want to test if they belong to the same object.

After sampling the confidence region of $\mathcal{A}_0$, the following step is to propagate each orbit X^i, for $i = 1, \ldots, k$, from time t_0^i to time $\bar{t}_1$, thus obtaining the orbital elements Y^i, with covariance matrix

$$\Gamma_{Y^i} = \left.\frac{\partial Y}{\partial X}\right|_{X=X^i} \Gamma_{X^i} \left.\frac{\partial Y}{\partial X}\right|_{X=X^i}^T .$$

Given the set of orbital elements Y^i, let $\mathcal{A}^i$ be the attributable computed from them. Denote by $\mathcal{A}'(X)$ the global function to pass from the orbital elements at the observing time $\bar{t}_0$ to the corresponding attributable at time $\bar{t}_1$. By the covariance propagation rule, the covariance and normal matrices of $\mathcal{A}^i$ are respectively

$$\Gamma_{\mathcal{A}^i} = \left.\frac{\partial \mathcal{A}'}{\partial \mathcal{A}}\right|_{X^i} \Gamma_{A_0} \left.\frac{\partial \mathcal{A}'}{\partial \mathcal{A}}\right|_{X^i}^T , \qquad C_{\mathcal{A}^i} = M^T C_{\mathcal{A}_0} M\,, \quad \text{where } M = \left(\left.\frac{\partial \mathcal{A}'}{\partial \mathcal{A}}\right|_{X^i}\right)^{-1} .$$

Then the confidence ellipsoid for the prediction $\mathcal{A}^i$ is

$$Z_{\mathcal{A}^i}(\sigma) = \{\mathcal{A}' :\ (\mathcal{A}' - \mathcal{A}^i)^T C_{\mathcal{A}^i} (\mathcal{A}' - \mathcal{A}^i) \leq \sigma^2\}\ .$$

To test the hypothesis that the two attributables $\mathcal{A}_0, \mathcal{A}_1$ belong to the same object, we need to find a minimum for the joint target function, obtained from the weighted sum of squares of the discrepancies:

$$Q = \frac{1}{2}\Big((\mathcal{A} - \mathcal{A}_0)^T C_{\mathcal{A}_0} (\mathcal{A} - \mathcal{A}_0) + (\mathcal{A}' - \mathcal{A}_1)^T C_{\mathcal{A}_1} (\mathcal{A}' - \mathcal{A}_1)\Big)\ .$$

By using an approximation to the first order for the propagation of the residuals

$$\mathcal{A}' - \mathcal{A}^i = \left.\frac{\partial \mathcal{A}'}{\partial \mathcal{A}}\right|_{X^i} (\mathcal{A} - \mathcal{A}_0)\,,$$

we obtain $\mathcal{A} - \mathcal{A}_0 = M(\mathcal{A}' - \mathcal{A}^i)$ and then

$$2Q(\mathcal{A}') = (\mathcal{A}' - \mathcal{A}_1)^T C_{\mathcal{A}_1} (\mathcal{A}' - \mathcal{A}_1) + (\mathcal{A}' - \mathcal{A}^i)^T C_{\mathcal{A}^i} (\mathcal{A}' - \mathcal{A}^i)\ .$$

Let $\mathcal{A}_1^i$ be the minimising value and define the minimum identification penalty associated to the node X^i of the triangulation by $K^i = 2Q(\mathcal{A}_1^i)$.

Now the idea of the procedure is quite evident. Given the attributable $\mathcal{A}_0$ and the triangulation $\{B^i\}_{i=1,\ldots,k}$ of $\mathcal{D}(\mathcal{A}_0)$, we scan the list of attributables of the second night of observations at time $\bar{t}_1$. For each of these attributables, say $\mathcal{A}_1$, we compute $K^i, i = 1, \ldots, k$. Fixed a suitable value K_{max}, to be chosen on the basis of experience, if $K^i > K_{max}$, for each $i = 1, \ldots, k$, then the couple $(\mathcal{A}_0, \mathcal{A}_1)$ must be discarded. On the contrary, if there exist some nodes $B^i, i \in I \subset \{1, \ldots, k\}$, such that $K^i \leq K_{max}$, for each $i \in I$, then we apply the differential corrections to any preliminary orbit $X^i = [\mathcal{A}_0, B^i]$, $i \in I$: we will keep only those orbits for which the differential corrections converge. Note that for a single couple of attributables we can obtain more preliminary orbits.

Subsequently, it is possible to try to link to any couple a third attributable, then to any *linked* triplet a fourth attributable, and so on. These attributions will be made by least squares fits.

To assess the efficiency of the method, in [41] a large scale test is performed by the use of a simulation reproducing the conditions of the next generation surveys. Two iterations of the described procedure are runned, with the use of two different kinds of sample. The first is chosen to have a low average number of virtual asteroids per attributable, close to one, so as to make a first rough selection of the couples, with a lower computational complexity. In the second iteration, a triangulation is made with an average of 50 virtual asteroids per attributable.

The results of the simulation are very good and suggest that the method is efficient enough to be used as a primary identification method.

We have taken into account these results while developing a new method for the linkage, which uses the first integrals of the Kepler problem. The idea is to add the new procedure to the existing ones, in order to improve the overall process and find as many as possible reliable orbits from a given set of many observations. Then the new method should be successful where the other ones fail. In particular, we should test the new algorithms on a data-set on which the previous procedures have failed to link pairs of attributables.

4.2.4 The use of the Kepler integrals

As we previously outlined, two attributables of the same solar system body at two different epochs provide us with 8 scalar data, so that we can try to compute the 6 elements of an orbit from them. A natural idea is to use directly the first integrals of the Kepler problem. After writing the angular momentum and the energy in terms of the attributables and equating them, we arrive at a system of polynomial equations of total degree 48 for the topocentric distance and radial velocity at the two epochs of the given attributables. Then we have to solve the not simple problem of finding the solutions of this system to get preliminary orbits.

In 1977 Taff and Hall had already proposed to use the angular momentum and the energy integrals to perform orbit determination starting from a data set that corresponds to two attributables of the same observed body (see [52] and [53]). They noticed that the problem can be written in an algebraic form but, since the total degree is high, they suggested to use a Newton-Raphson method to solve the problem. This approach deals with the solutions only locally, and there are alternative possible solutions that can be lost.

The use of only the angular momentum integral to determine an orbit of a solar system body is already present in [44]. More recently, in [30], the author proposed a method to compute a preliminary orbit from four observations divided in two pairs close in time: the basic idea is to equate the angular momentum vectors at the two mean epochs of the couples of observations.

In the next chapters we will investigate a method to compute a finite set of preliminary orbits for a solar system body, starting from two attributables. Like Taff and Hall, we shall start from both the first integrals of the Kepler problem, but we shall exploit the algebraic character of the problem, keeping in this way a global control on the solutions. In particular, we shall present two different methods to solve by elimination the polynomial system corresponding to this problem, and to compute all the preliminary orbits defined by the two attributables.

Since an orbit is defined by 6 scalar data, the available information is redundant, and we can use this to set compatibility conditions for the solutions (see 5.16), that should be fulfilled if the two attributables belong to the same solar system object. The unavoidable errors in the observations affect also the computation of the attributables. Given a covariance matrix for the two attributables, expressing their uncertainty, we can use this to compute the value of an *identification norm*, based on the compatibility conditions, to decide if the attributables may be related to the same body (i.e. if the linkage is successful) and to choose among possible alternative solutions.

Our method is thought for the applications to the modern sets of astrometric observations, in particular in the conditions that we expect for the next generation surveys. Some numerical experiments are presented in Chapter 6.

The method is thought for asteroids, but it can be applied also to the case of the space debris. The result of a numerical test for this case is also described in Chapter 6.

Chapter 5

Orbit Determination with the two-body Integrals

We are going to describe in detail our method for linking two attributables by the use of the Kepler integrals. In particular, in Section 5.3 we shall present two algorithms to be used for solving the algebraic system arising from this method, thus computing all the possible preliminary orbits from the two attributables. In Section 5.4 we deal with the uncertainty of the data, introduce the identification norm and provide the covariance matrices of the preliminary orbits.

5.1 Notation

In this section we recall for clarity some of the notations defined in the previous chapter, adding a few remarks about the computation of the observer's position and velocity.

Let $(\rho, \alpha, \delta) \in \mathbb{R}^+ \times [-\pi, \pi) \times (-\pi/2, \pi/2)$ be spherical coordinates for the topocentric position of a solar system body. The angular coordinates (α, δ) are defined by a topocentric coordinate system that can be arbitrarily selected. Usually, in the applications, α is the right ascension and δ the declination with respect to an equatorial coordinate system (e.g., J2000).

Given a short arc of observations of a celestial body $(t_i, \alpha_i, \delta_i)$, for $i = 1, .., m$ with $m \geq 2$, it is often possible to compute an attributable, that is a vector

$$\mathcal{A} = (\alpha, \delta, \dot{\alpha}, \dot{\delta}) \in [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}^2 \,,$$

representing the angular position and velocity of the body at a mean time $\bar{t}$ in the selected coordinates. Usually we choose $\bar{t}$ as the mean $(\sum_i t_i)/m$. As recalled in Subsection 4.2.2, the attributable is computed by a polynomial fit, typically linear or quadratic, and the observations used in the computation need to be made by the same observatory. If the observations are enough, i.e. $m \geq 2$ for a linear fit, $m \geq 3$ for a quadratic one, then we can compute also a covariance matrix $\Gamma_{\mathcal{A}}$, representing the uncertainty of the attributable. Note that the topocentric distances ρ_i at times t_i are completely unknown.

We introduce the heliocentric position and velocity of the body at time $\bar{t}$

$$\mathbf{r} = \mathbf{q} + \rho\hat{\boldsymbol{\rho}} \,, \qquad \dot{\mathbf{r}} = \dot{\mathbf{q}} + \dot{\rho}\hat{\boldsymbol{\rho}} + \rho(\hat{\boldsymbol{\rho}}_\alpha\dot{\alpha} + \hat{\boldsymbol{\rho}}_\delta\dot{\delta}) \,, \tag{5.1}$$

with ρ, $\dot{\rho}$ the topocentric distance and the radial velocity, and with $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\rho}}_\alpha$, $\hat{\boldsymbol{\rho}}_\delta$ the observation direction and its partial derivatives with respect to α and δ.

The vectors $\mathbf{q}, \dot{\mathbf{q}}$ represent the heliocentric position and velocity of the observer on the Earth. The observer position is know as a function of time, but for consistency, if the attributable is computed by a fit to polynomials with low degree, the values $\mathbf{q}(\bar{t}), \dot{\mathbf{q}}(\bar{t})$ need to be computed by the same interpolation. Therefore we make a quadratic fit with the actual geocentric positions $\mathbf{q}(t_i) - \mathbf{q}_\oplus(t_i)$ at the times of the individual observations ($\mathbf{q}_\oplus$ is the heliocentric position of the Earth centre) to obtain the interpolating function $\mathbf{q}_{obs}(t)$; then we take $\mathbf{q}(\bar{t}) = \mathbf{q}_\oplus(\bar{t}) + \mathbf{q}_{obs}(\bar{t})$ and $\dot{\mathbf{q}}(\bar{t}) = \dot{\mathbf{q}}_\oplus(\bar{t}) + \dot{\mathbf{q}}_{obs}(\bar{t})$. This method was suggested by Poincaré in [49], and it is important to obtain preliminary orbits of better quality, see [39].

In rectangular coordinates we have

$$\begin{aligned}
\hat{\boldsymbol{\rho}} &= (\cos\alpha\cos\delta, \sin\alpha\cos\delta, \sin\delta)\,, \\
\hat{\boldsymbol{\rho}}_\alpha &= (-\sin\alpha\cos\delta, \cos\alpha\cos\delta, 0)\,, \\
\hat{\boldsymbol{\rho}}_\delta &= (-\cos\alpha\sin\delta, -\sin\alpha\sin\delta, \cos\delta)\ .
\end{aligned}$$

These vectors form an orthogonal system, in particular

$$|\hat{\boldsymbol{\rho}}| = |\hat{\boldsymbol{\rho}}_\delta| = 1\,, \qquad |\hat{\boldsymbol{\rho}}_\alpha| = \cos\delta\,, \qquad \hat{\boldsymbol{\rho}} \cdot \hat{\boldsymbol{\rho}}_\alpha = \hat{\boldsymbol{\rho}} \cdot \hat{\boldsymbol{\rho}}_\delta = \hat{\boldsymbol{\rho}}_\alpha \cdot \hat{\boldsymbol{\rho}}_\delta = 0\,,$$

where the dot indicates the Euclidean scalar product and $|\cdot|$ the corresponding norm.

We shall use the orthonormal basis $\{\hat{\boldsymbol{\rho}}, \hat{\mathbf{v}}, \hat{\mathbf{n}}\}$ adapted to the apparent path $\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\rho}}(t)$ of the observed body on the celestial sphere: the unit vector $\hat{\mathbf{v}}$ is defined by the relation

$$\frac{d}{dt}\hat{\boldsymbol{\rho}}(\bar{t}) = \eta\,\hat{\mathbf{v}}\,,$$

where $\eta = \sqrt{\dot{\alpha}^2\cos^2\delta + \dot{\delta}^2}$ is the *proper motion*, and $\hat{\mathbf{n}} = \hat{\boldsymbol{\rho}} \times \hat{\mathbf{v}}$.

5.2 Linkage by the two-body integrals

Given two attributables $\mathcal{A}_1, \mathcal{A}_2$ at different epochs $\bar{t}_1, \bar{t}_2$, in the hypothesis that they belong to the same observed body, we write down polynomial equations for the topocentric distance and radial velocity of the body at the two epochs by using the angular momentum and the energy integrals.

5.2.1 Angular momentum and Energy

For a given attributable $\mathcal{A}$ the angular momentum vector (per unit mass) can be written as a polynomial function of the radial distance and velocity $\rho, \dot{\rho}$:

$$\mathbf{c}(\rho, \dot{\rho}) = \mathbf{r} \times \dot{\mathbf{r}} = \mathbf{D}\dot{\rho} + \mathbf{E}\rho^2 + \mathbf{F}\rho + \mathbf{G}\,,$$

where

$$\begin{aligned}
\mathbf{D} &= \mathbf{q} \times \hat{\boldsymbol{\rho}}\,, \\
\mathbf{E} &= \dot{\alpha}\hat{\boldsymbol{\rho}} \times \hat{\boldsymbol{\rho}}_\alpha + \dot{\delta}\hat{\boldsymbol{\rho}} \times \hat{\boldsymbol{\rho}}_\delta = \eta\hat{\mathbf{n}}\,, \\
\mathbf{F} &= \dot{\alpha}\mathbf{q} \times \hat{\boldsymbol{\rho}}_\alpha + \dot{\delta}\mathbf{q} \times \hat{\boldsymbol{\rho}}_\delta + \hat{\boldsymbol{\rho}} \times \dot{\mathbf{q}}\,, \\
\mathbf{G} &= \mathbf{q} \times \dot{\mathbf{q}}\,,
\end{aligned}$$

depend only on the attributable $\mathcal{A}$ and on the motion of the observer $\mathbf{q}, \dot{\mathbf{q}}$ at the time $\bar{t}$ of the attributable. For the given $\mathcal{A}$ we can also write the two-body energy as a function of $\rho, \dot{\rho}$, as in [38]

$$2\mathcal{E}(\rho, \dot{\rho}) = \dot{\rho}^2 + c_1\dot{\rho} + c_2\rho^2 + c_3\rho + c_4 - \frac{2k^2}{\sqrt{\rho^2 + c_5\rho + c_0}} ,$$

where k is Gauss' constant and

$$\begin{array}{lll} c_0 = |\mathbf{q}|^2 , & c_1 = 2\, \dot{\mathbf{q}} \cdot \hat{\boldsymbol{\rho}} , & c_2 = \eta^2 , \\ c_3 = 2(\dot{\alpha}\, \dot{\mathbf{q}} \cdot \hat{\boldsymbol{\rho}}_\alpha + \dot{\delta}\, \dot{\mathbf{q}} \cdot \hat{\boldsymbol{\rho}}_\delta) , & c_4 = |\dot{\mathbf{q}}|^2 , & c_5 = 2\, \mathbf{q} \cdot \hat{\boldsymbol{\rho}} , \end{array}$$

depend only on $\mathcal{A}, \mathbf{q}, \dot{\mathbf{q}}$.

5.2.2 Equating the integrals

Now we take two attributables $\mathcal{A}_1 = (\alpha_1, \delta_1, \dot{\alpha}_1, \dot{\delta}_1)$, $\mathcal{A}_2 = (\alpha_2, \delta_2, \dot{\alpha}_2, \dot{\delta}_2)$ at epochs $\bar{t}_1, \bar{t}_2$; we shall use the notation of Section 5.2.1, with index 1 or 2 referring to the epoch. If $\mathcal{A}_1$, $\mathcal{A}_2$ correspond to the same physical object, then the angular momentum vectors at the two epochs must coincide:

$$\mathbf{D}_1\dot{\rho}_1 - \mathbf{D}_2\dot{\rho}_2 = \mathbf{J}(\rho_1, \rho_2) , \tag{5.2}$$

where

$$\mathbf{J}(\rho_1, \rho_2) = \mathbf{E}_2\rho_2^2 - \mathbf{E}_1\rho_1^2 + \mathbf{F}_2\rho_2 - \mathbf{F}_1\rho_1 + \mathbf{G}_2 - \mathbf{G}_1 \ .$$

Relation (5.2) is a system of three equations in the four unknowns $\rho_1, \dot{\rho}_1, \rho_2, \dot{\rho}_2$, with constraints

$$\rho_1 > 0 , \ \ \rho_2 > 0 \ .$$

By scalar multiplication of (5.2) with $\mathbf{D}_1 \times \mathbf{D}_2$ we eliminate the variables $\dot{\rho}_1, \dot{\rho}_2$ and obtain the equation

$$\mathbf{D}_1 \times \mathbf{D}_2 \cdot \mathbf{J}(\rho_1, \rho_2) = 0 \ . \tag{5.3}$$

The left hand side in (5.3) is a quadratic form in the variables ρ_1, ρ_2; we write it as

$$q(\rho_1, \rho_2) \stackrel{\text{def}}{=} q_{20}\rho_1^2 + q_{10}\rho_1 + q_{02}\rho_2^2 + q_{01}\rho_2 + q_{00} , \tag{5.4}$$

with

$$\begin{array}{ll} q_{20} = -\mathbf{E}_1 \cdot \mathbf{D}_1 \times \mathbf{D}_2 , & q_{02} = \mathbf{E}_2 \cdot \mathbf{D}_1 \times \mathbf{D}_2 , \\ q_{10} = -\mathbf{F}_1 \cdot \mathbf{D}_1 \times \mathbf{D}_2 , & q_{01} = \mathbf{F}_2 \cdot \mathbf{D}_1 \times \mathbf{D}_2 , \end{array}$$

$$q_{00} = (\mathbf{G}_2 - \mathbf{G}_1) \cdot \mathbf{D}_1 \times \mathbf{D}_2 \ .$$

Equation (5.4) defines a conic section in the (ρ_1, ρ_2) plane, with symmetry axes parallel to the coordinate axes. Since the directions of $\mathbf{E}_1, \mathbf{E}_2$ correspond to $\hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2$, for a time span $\bar{t}_2 - \bar{t}_1$ small enough the angle between these two directions is small and the coefficients q_{20}, q_{02} have opposite signs, thus in this case (5.4) defines a hyperbola.

We can compute the radial velocities $\dot\rho_1, \dot\rho_2$ by vector multiplication of (5.2) with $\mathbf{D}_1$ and $\mathbf{D}_2$, projecting on the direction of $\mathbf{D}_1 \times \mathbf{D}_2$:

$$\dot\rho_1(\rho_1,\rho_2) = \frac{(\mathbf{J}\times\mathbf{D}_2)\cdot(\mathbf{D}_1\times\mathbf{D}_2)}{|\mathbf{D}_1\times\mathbf{D}_2|^2}\,, \quad \dot\rho_2(\rho_1,\rho_2) = \frac{(\mathbf{J}\times\mathbf{D}_1)\cdot(\mathbf{D}_1\times\mathbf{D}_2)}{|\mathbf{D}_1\times\mathbf{D}_2|^2}\,. \tag{5.5}$$

For the given $\mathcal{A}_1, \mathcal{A}_2$ we can also equate the corresponding two-body energies $\mathcal{E}_1, \mathcal{E}_2$. We use the expressions of $\dot\rho_1(\rho_1,\rho_2), \dot\rho_2(\rho_1,\rho_2)$ above and substitute them into $\mathcal{E}_1 = \mathcal{E}_2$, thus we obtain

$$\mathcal{F}_1(\rho_1,\rho_2) - \frac{2k^2}{\sqrt{\mathcal{G}_1(\rho_1)}} = \mathcal{F}_2(\rho_1,\rho_2) - \frac{2k^2}{\sqrt{\mathcal{G}_2(\rho_2)}}\,, \tag{5.6}$$

for some polynomial functions $\mathcal{F}_1(\rho_1,\rho_2)$, $\mathcal{F}_2(\rho_1,\rho_2)$, $\mathcal{G}_1(\rho_1)$, $\mathcal{G}_2(\rho_2)$ with degrees $\deg(\mathcal{F}_1) = \deg(\mathcal{F}_2) = 4$ and $\deg(\mathcal{G}_1) = \deg(\mathcal{G}_2) = 2$. By squaring we have

$$(\mathcal{F}_1 - \mathcal{F}_2)^2\mathcal{G}_1\mathcal{G}_2 - 4k^4(\mathcal{G}_1 + \mathcal{G}_2) = -8k^4\sqrt{\mathcal{G}_1\mathcal{G}_2}\,. \tag{5.7}$$

Squaring again we obtain the polynomial equation

$$p(\rho_1,\rho_2) \stackrel{\text{def}}{=} \Big[(\mathcal{F}_1 - \mathcal{F}_2)^2\mathcal{G}_1\mathcal{G}_2 - 4k^4(\mathcal{G}_1 + \mathcal{G}_2)\Big]^2 - 64k^8\mathcal{G}_1\mathcal{G}_2 = 0\,, \tag{5.8}$$

with total degree 24. Some spurious solutions may have been added as a result of squaring expressions with unknown sign.

Note that, if the observations were made from the center of the Earth, $\mathbf{G}_i$ ($i = 1,2$) would be the angular momentum of the Earth at epochs $\bar t_1$, $\bar t_2$, thus $\mathbf{G}_1 = \mathbf{G}_2$ and $q_{00} = 0$. With this simplifying assumption $\rho_1 = \rho_2 = 0$ is a solution of the system $q(\rho_1,\rho_2) = p(\rho_1,\rho_2) = 0$ that corresponds to the Earth center and therefore is not acceptable. This solution also appears in the geocentric version of the method of Laplace for a preliminary orbit from 3 observations. Actually we use topocentric observations, for which the zero solution is replaced by one with both ρ_1 and ρ_2 very small.

5.2.3 Degenerate cases

The quadratic form (5.4) degenerates into a linear function when

$$\mathbf{E}_1\cdot\mathbf{D}_1\times\mathbf{D}_2 = \mathbf{E}_2\cdot\mathbf{D}_1\times\mathbf{D}_2 = 0\,.$$

A simple computation shows that

$$\begin{aligned}\mathbf{E}_1\cdot\mathbf{D}_1\times\mathbf{D}_2 &= \eta_1(\hat{\mathbf{n}}_1\cdot\mathbf{q}_1)(\hat{\boldsymbol\rho}_1\times\hat{\boldsymbol\rho}_2\cdot\mathbf{q}_2)\,,\\ \mathbf{E}_2\cdot\mathbf{D}_1\times\mathbf{D}_2 &= \eta_2(\hat{\mathbf{n}}_2\cdot\mathbf{q}_2)(\hat{\boldsymbol\rho}_1\times\hat{\boldsymbol\rho}_2\cdot\mathbf{q}_1)\,,\end{aligned}$$

thus, assuming that the proper motions η_1, η_2 do not vanish and setting $\mathbf{n_{12}} = \hat{\boldsymbol\rho}_1\times\hat{\boldsymbol\rho}_2$, the degeneration occurs when either $\mathbf{n_{12}}$ vanishes (C0) or at least one the following relations holds:

$$\hat{\mathbf{n}}_1\cdot\mathbf{q}_1 = \hat{\mathbf{n}}_2\cdot\mathbf{q}_2 = 0\,, \tag{C1}$$
$$\mathbf{n_{12}}\cdot\mathbf{q}_1 = \mathbf{n_{12}}\cdot\mathbf{q}_2 = 0\,, \tag{C2}$$
$$\hat{\mathbf{n}}_1\cdot\mathbf{q}_1 = \mathbf{n_{12}}\cdot\mathbf{q}_1 = 0\,, \tag{C3}$$
$$\hat{\mathbf{n}}_2\cdot\mathbf{q}_2 = \mathbf{n_{12}}\cdot\mathbf{q}_2 = 0\,. \tag{C4}$$

The interpretation of these conditions is the following: (C0) means that $\hat{\boldsymbol{\rho}}_1, \hat{\boldsymbol{\rho}}_2$ point to either exactly the same or exactly the opposite direction in the sky; (C1) means that both sets $\{\hat{\boldsymbol{\rho}}_1, \hat{\mathbf{v}}_1, \mathbf{q}_1\}$ and $\{\hat{\boldsymbol{\rho}}_2, \hat{\mathbf{v}}_2, \mathbf{q}_2\}$ are constituted of coplanar vectors; (C2) says that $\hat{\boldsymbol{\rho}}_1, \hat{\boldsymbol{\rho}}_2, \mathbf{q}_1, \mathbf{q}_2$ are coplanar. Let us discuss condition (C3): $\hat{\mathbf{n}}_1 \cdot \mathbf{q}_1 = 0$ means that $\mathbf{q}_1, \hat{\boldsymbol{\rho}}_1, \hat{\mathbf{v}}_1$ are coplanar and $\mathbf{n_{12}} \cdot \mathbf{q}_1 = 0$ means that $\hat{\boldsymbol{\rho}}_1, \hat{\boldsymbol{\rho}}_2, \mathbf{q}_1$ are coplanar as well. If $\mathbf{D}_1 \neq 0$ we obtain that the four vectors $\mathbf{q}_1, \hat{\boldsymbol{\rho}}_1, \hat{\mathbf{v}}_1, \hat{\boldsymbol{\rho}}_2$ all lie in the same plane. In particular (C3) implies that $\hat{\boldsymbol{\rho}}_2$ belongs to the great circle defined by the intersection of the plane generated by $\hat{\boldsymbol{\rho}}_1, \hat{\mathbf{v}}_1$ with the celestial sphere. This degeneration condition can be compared with the failure condition of the classical orbit determination methods with three observations by Gauss and Laplace [46], due to vanishing of the curvature in the apparent path of the observed body on the celestial sphere. The discussion of condition (C4) is similar to the previous one and corresponds to the coplanarity of $\mathbf{q}_2, \hat{\boldsymbol{\rho}}_2, \hat{\mathbf{v}}_2, \hat{\boldsymbol{\rho}}_1$.

5.3 Computation of the solutions

In this section we introduce two different methods to search for the solutions of the semi-algebraic problem

$$\left\{ \begin{array}{l} p(\rho_1, \rho_2) = 0 \\ q(\rho_1, \rho_2) = 0 \end{array} \right. , \qquad \rho_1, \rho_2 > 0 \tag{5.9}$$

for the polynomials p, q introduced in (5.8), (5.4) respectively. Moreover we explain the full procedure for the computation of the preliminary orbits and introduce compatibility conditions to decide whether the attributables used to define the problem are related to the same solar system body.

5.3.1 Computation of the resultant via DFT

The first method consists in writing the resultant (see [16]) of p and q with respect to one variable, say ρ_1. In this way we find a univariate polynomial in the ρ_2 variable whose real positive roots are the only possible ρ_2-components of a solution of (5.9). By grouping the monomials with the same power of ρ_1 we can write

$$p(\rho_1, \rho_2) = \sum_{j=0}^{20} a_j(\rho_2)\, \rho_1^j\,, \qquad \text{where} \tag{5.10}$$

$$\deg(a_j) = \left\{ \begin{array}{lll} 20 & \text{for } j = 0 \ldots 4 & \\ 24 - (j+1) & \text{for } j = 2k-1 & \text{with } k \geq 3 \\ 24 - j & \text{for } j = 2k & \text{with } k \geq 3 \end{array} \right.$$

and

$$q(\rho_1, \rho_2) = b_2\, \rho_1^2 + b_1\, \rho_1 + b_0(\rho_2) \tag{5.11}$$

for some univariate polynomial coefficients a_i, b_j, depending on ρ_2 (actually b_1, b_2 are constant). We consider the resultant $Res(\rho_2)$ of p, q with respect to ρ_1: it is generically a degree 48 polynomial defined as the determinant of the Sylvester matrix

$$\mathtt{S}(\rho_2) = \begin{pmatrix} a_{20} & 0 & b_2 & 0 & \dots & \dots & 0 \\ a_{19} & a_{20} & b_1 & b_2 & 0 & \dots & 0 \\ \vdots & \vdots & b_0 & b_1 & b_2 & \dots & \vdots \\ \vdots & \vdots & 0 & b_0 & b_1 & \dots & \vdots \\ a_0 & a_1 & \vdots & \vdots & \vdots & b_0 & b_1 \\ 0 & a_0 & 0 & 0 & 0 & 0 & b_0 \end{pmatrix} . \tag{5.12}$$

The positive real roots of $Res(\rho_2)$ are the only possible values of ρ_2 for a solution (ρ_1, ρ_2) of (5.9). We could use the resultant method to eliminate the variable ρ_2 by a different grouping of the terms of p, q:

$$p(\rho_1, \rho_2) = \sum_{j=0}^{20} a'_j(\rho_1)\ \rho_2^j\ , \qquad q(\rho_1, \rho_2) = b'_2\ \rho_2^2 + b'_1\ \rho_2 + b'_0(\rho_1)\ ,$$

where the degrees of a'_j are described by the same rule as for a_j.

Apart from non-real and non-positive solutions, we shall see that there are additional different reasons to discard some pairs of solutions of (5.9), thus we expect that the number of acceptable ones is not large. We use a scheme similar to [23] to compute the coefficients of the resultant $Res(\rho_2)$:

1) evaluate $a_i(\rho_2), b_j(\rho_2)$ at the 64-th roots of unit $\omega_k = e^{2\pi i \frac{k}{64}}$, $k = 0, .., 63$ by a DFT (Discrete Fourier Transform) algorithm;

2) compute the determinant of the 64 Sylvester matrices; by relation

$$\det\left(\mathtt{S}(\rho_2)|_{\rho_2=\omega_k}\right) = \left(\det \mathtt{S}(\rho_2)\right)|_{\rho_2=\omega_k}$$

we have the values of $Res(\rho_2)$ at the 64-th roots of unit;

3) apply an IDFT (Inverse Discrete Fourier Transform) algorithm to obtain the coefficients of $Res(\rho_2)$ from its evaluations.

The use of the DFT and IDFT allows us to interpolate the resultant $Res(\rho_2)$ in an efficient way. The use of numerical evaluations, e.g. at the roots of unit, avoids the difficulty of writing a very long symbolic expression for the resultant, that could be cumbersome to be managed by a programming language compiler.

The complete set of complex roots of $Res(\rho_2)$, with an error bound for each of them, are computed using the algorithm described in [3], which is based on simultaneous iterations. Let $\rho_2(k)$, $k = 1, .., n \leq 48$, be the subset of the real and positive roots of $Res(\rho_2)$. Then for each k we perform the sequence of operations below:

4) solve the equation $q(\rho_1, \rho_2(k)) = 0$ and compute the two possible values $\rho_1(k, 1)$, $\rho_1(k, 2)$ for ρ_1, discarding negative solutions. Then define $\rho_1(k)$ equal to either $\rho_1(k, 1)$ or $\rho_1(k, 2)$, selecting the one that gives the smaller value of $|p(\rho_1, \rho_2(k))|$;

5) discard spurious solutions, resulting from the squaring used to reduce the energy equality to the polynomial equation (5.8). The spurious solutions are the solutions of (5.9) that do not satisfy either (5.7) or (5.6);

6) compute the corresponding values of $\dot\rho_1(k), \dot\rho_2(k)$ by (5.5) and obtain a pair of orbits defined by the sets $(\alpha_i, \delta_i, \dot\alpha_i, \dot\delta_i, \rho_i, \dot\rho_i)$ of *attributable elements*,[1] for $i = 1, 2$;

7) change from attributable elements to Cartesian heliocentric coordinates by relation $\mathbf{r}_i = \rho_i(k)\hat{\boldsymbol{\rho}}_i + \mathbf{q}_i$ for $i = 1, 2$, and the corresponding formula for $\dot{\mathbf{r}}_i$. Note that the observer position $\mathbf{q}_i$ is not the actual $\mathbf{q}(t_i)$, but is obtained by interpolation as proposed by Poincaré (see Section 5.1). Then a standard coordinate change allows us to obtain the related pairs of orbital elements: we shall use Keplerian elements $(a, e, I, \Omega, \omega, \ell)$, where ℓ is the mean anomaly[2]. The epochs of the orbits are $\tilde t_1(k), \tilde t_2(k)$, corrected by aberration due to the finite velocity of the light c: $\tilde t_i(k) = \bar t_i - \rho_i(k)/c$ for $i = 1, 2$.

We have implemented this algorithm in FORTRAN 90 using *quadruple precision* for part of these computations, in particular the ones related to DFT and IDFT. This feature appeared necessary to obtain reliable results starting from our first numerical experiments.

5.3.2 Normal form of the problem

Another method to compute the solutions of (5.9) is based on a coordinate change to variables (ξ_1, ξ_2), that allows to perform easily the elimination of either ξ_1 or ξ_2. Let us set

$$p(\rho_1, \rho_2) = \sum_{i,j=0}^{20} p_{i,j}\rho_1^i\rho_2^j \,.$$

First we consider the affine transformation to intermediate variables (ζ_1, ζ_2)

$$\mathcal{T} : \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \to \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^{-1}\rho_1 - \tau_1 \\ \sigma_2^{-1}\rho_2 - \tau_2 \end{pmatrix} ,$$

where, to eliminate the linear terms in (5.4), we set

$$\sigma_1\tau_1 = -\frac{q_{1,0}}{2q_{2,0}} \stackrel{\text{def}}{=} \alpha \,, \qquad \sigma_2\tau_2 = -\frac{q_{0,1}}{2q_{0,2}} \stackrel{\text{def}}{=} \beta \,,$$

so that

$$q \circ \mathcal{T}^{-1}(\zeta_1, \zeta_2) = q_{2,0}\sigma_1^2\Big[\zeta_1^2 + \frac{q_{0,2}\sigma_2^2}{q_{2,0}\sigma_1^2}\zeta_2^2 + \frac{\kappa}{q_{2,0}\sigma_1^2}\Big] , \qquad \text{with } \kappa = q_{0,0} - \frac{q_{1,0}^2}{4q_{2,0}} - \frac{q_{0,1}^2}{4q_{0,2}} \,.$$

[1] The attributable elements are the same as spherical polar coordinates with their time derivatives: the coordinates are just reordered in such a way that the first four elements form the attributable, hence the name.

[2] Any other set of orbital elements in which the first four are defined by the two-body energy and angular momentum can be used, e.g. cometary elements $(p_d, e, I, \Omega, \omega, t_p)$ where p_d is the perihelion distance and t_p is the time of perihelion passage: this set would allow to handle also parabolic and hyperbolic orbits.

If we set, for an arbitrary $\sigma_2 \in \mathbb{R}$,

$$\sigma_1 = \gamma\sigma_2 \,, \qquad \gamma = \sqrt{-\frac{q_{0,2}}{q_{2,0}}}$$

we obtain

$$q \circ \mathcal{T}^{-1}(\zeta_1, \zeta_2) = q_{2,0}\sigma_1^2 \left[\zeta_1^2 - \zeta_2^2 - 2c_\star\right] , \qquad \text{with} \;\; c_\star = -\frac{\kappa}{2\, q_{2,0}\sigma_1^2} \;.$$

We already observed that, for $\bar{t}_2 - \bar{t}_1$ small enough, q_{02} and q_{20} have opposite signs, hence in this case the variable change $\mathcal{T}$ is real. However, in general, we have to consider $\mathcal{T}$ as a transformation of the complex domain $\mathbb{C}^2$. We also have

$$p \circ \mathcal{T}^{-1}(\zeta_1, \zeta_2) = \sum_{i,j=0}^{20} \tilde{p}_{i,j}\zeta_1^i\zeta_2^j \,,$$

where

$$\tilde{p}_{i,j} = \sigma_2^{i+j}\gamma^i \sum_{h=i}^{20}\sum_{k=j}^{20} p_{h,k} \begin{pmatrix} h \\ i \end{pmatrix} \begin{pmatrix} k \\ j \end{pmatrix} \alpha^{h-i}\beta^{k-j} \,, \tag{5.13}$$

and α, β, γ depend only on the coefficients of $q(\rho_1, \rho_2)$.
Now we apply a rotation of angle $\pi/4$ to pass to the (ξ_1, ξ_2) variables:

$$\mathcal{R} : \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \to \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{bmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix} \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} .$$

We have

$$\begin{aligned} q \circ \mathcal{T}^{-1} \circ \mathcal{R}^{-1}(\xi_1, \xi_2) &= 2\, q_{2,0}\sigma_1^2 \left[\xi_1\xi_2 - c_\star\right] , \\ p \circ \mathcal{T}^{-1} \circ \mathcal{R}^{-1}(\xi_1, \xi_2) &= \sum_{i,j=0}^{24} p^\star_{i,j}\xi_1^i\xi_2^j \,, \end{aligned}$$

where

$$p^\star_{i,j} = \begin{cases} \bar{p}^\star_{i+j,i} & \text{if } i+j \le 24 \\ 0 & \text{if } i+j > 24 \end{cases} \tag{5.14}$$

and

$$\begin{aligned} \bar{p}^\star_{m,n} &= \sum_{h+k=m} \tilde{p}_{h,k} \sum_{i+j=n} \begin{pmatrix} h \\ i \end{pmatrix} \begin{pmatrix} k \\ j \end{pmatrix} \frac{(-1)^j}{2^{(h+k)/2}} = \\ &= \sum_{h=0}^{m} \tilde{p}_{h,m-h} \sum_{i=0}^{n} \begin{pmatrix} h \\ i \end{pmatrix} \begin{pmatrix} m-h \\ n-i \end{pmatrix} \frac{(-1)^{n-i}}{2^{m/2}} = \\ &= \sum_{h=\max\{m-20,0\}}^{\min\{m,20\}} \tilde{p}_{h,m-h} \sum_{i=0}^{n} \begin{pmatrix} h \\ i \end{pmatrix} \begin{pmatrix} m-h \\ n-i \end{pmatrix} \frac{(-1)^{n-i}}{2^{m/2}} \,. \end{aligned}$$

The last equality is obtained taking into account that $\tilde{p}_{h,m-h} = 0$, for $h > 20$ or $m - h > 20$. Using the relation $\xi_1\xi_2 = c_\star$ we can consider in place of $p \circ \mathcal{T}^{-1} \circ$

$\mathcal{R}^{-1}(\xi_1, \xi_2)$ the polynomial

$$
\begin{aligned}
p^\star(\xi_1,\xi_2) &= \sum_{\substack{h,\,k\,=\,0\\ h\,>\,k}}^{24} p^\star_{h,k} c_\star^k \xi_1^{h-k} + \sum_{h=0}^{24} p^\star_{h,h} c_\star^h + \sum_{\substack{h,\,k\,=\,0\\ h\,<\,k}}^{24} p^\star_{h,k} c_\star^h \xi_2^{k-h} = \\
&= \sum_{j=1}^{24}\Bigg(\sum_{\substack{h,\,k\,=\,0\\ h\,-\,k\,=\,j}}^{24} p^\star_{h,k} c_\star^{h-j}\Bigg)\xi_1^j + \sum_{j=1}^{24}\Bigg(\sum_{\substack{h,\,k\,=\,0\\ k\,-\,h\,=\,j}}^{24} p^\star_{h,k} c_\star^{k-j}\Bigg)\xi_2^j + \sum_{h=0}^{24} p^\star_{h,h} c_\star^h = \\
&= A_{24}\,\xi_1^{24} + \ldots + A_1\xi_1 + A_0\,(\xi_2)\,,
\end{aligned}
$$

with

$$
A_j = \sum_{\substack{h,\,k\,=\,0\\ h\,-\,k\,=\,j}}^{24} p^\star_{h,k} c_\star^{h-j} = \sum_{h=j}^{24} p^\star_{h,h-j} c_\star^{h-j}\,, \qquad j = 1\ldots 24\,,
$$

$$
A_0(\xi_2) = B_{24}\,\xi_2^{24} + \ldots + B_1\,\xi_2 + B_0\,,
$$

$$
B_j = \sum_{\substack{h,\,k\,=\,0\\ k\,-\,h\,=\,j}}^{24} p^\star_{h,k} c_\star^{k-j} = \sum_{k=j}^{24} p^\star_{k-j,k} c_\star^{k-j}\,, \qquad j = 0\ldots 24\,.
$$

We consider the algebraic problem in normal form

$$
\left\{ \begin{array}{l} p^\star(\xi_1,\xi_2) = 0 \\ \xi_1\xi_2 - c_\star = 0 \end{array} \right. . \tag{5.15}
$$

In this case we have to consider all the solutions of (5.15), not only the ones with real and positive components.

If $c_\star = 0$, then the solutions (ξ_1, ξ_2) of (5.15) are of the form $(\xi_1(k), 0)$ or $(0, \xi_2(k))$, where $\xi_1(k), \xi_2(k)$, $k = 1\ldots 24$ are the roots of $A_{24}\,\xi_1^{24} + \ldots + A_1\xi_1 + B_0$ and $B_{24}\,\xi_2^{24} + \ldots + B_1\xi_2 + B_0$ respectively.

If $c_\star \neq 0$, using the relation $\xi_1\xi_2 = c_\star$ we can eliminate one variable, say ξ_1, from $p^\star$. Thus we obtain the univariate polynomial

$$
\mathfrak{p}(\xi_2) = \sum_{k=0}^{48} \mathfrak{p}_k \xi_2^k\,, \qquad \text{with} \ \ \mathfrak{p}_k = \left\{ \begin{array}{ll} A_{24-k} c_\star^{24-k}\,, & 0 \le k \le 23 \\ B_{k-24}\,, & 24 \le k \le 48 \end{array} \right. .
$$

We compute all the complex roots $\xi_2(k), k = 1\ldots 48$ of $\mathfrak{p}(\xi_2)$ by the algorithm in [3]; then for each k we define the other component of the solutions by

$$
\xi_1(k) = \frac{c_\star}{\xi_2(k)}\,.
$$

Given all the complex solutions of (5.15) we compute the corresponding points in the (ρ_1, ρ_2) plane by

$$
(\rho_1(k), \rho_2(k)) = \mathcal{T}^{-1} \circ \mathcal{R}^{-1}\,(\xi_1(k), \xi_2(k))\,, \qquad k = 1\ldots 48\,,
$$

discarding the ones with non-real or non-positive components. At this point the preliminary orbits can be computed following the same steps 5), 6), 7) of the algorithm explained in Subsection 5.3.1.

From a few experiments performed this method seems to require more than quadruple precision because of the complicated formulae defining the transformation used to obtain the normal form (5.15). Thus the advantage in the simple elimination of the variable ξ_1 must be balanced with the introduction of heavier computations.

5.3.3 Compatibility conditions

The knowledge of the angular momentum vector and of the energy at a given time allows us to compute the Keplerian elements

$$a, e, I, \Omega \ .$$

In fact the semimajor axis a and the eccentricity e can be computed from the energy and the size of the angular momentum through the relations

$$\mathcal{E} = -\frac{k^2}{2a}\,, \qquad \|\mathbf{c}\| = k\sqrt{a(1-e^2)}\,;$$

the longitude of the node Ω and the inclination I are obtained from the direction of the angular momentum

$$\hat{\mathbf{c}} = (\sin\Omega\sin I, -\cos\Omega\sin I, \cos I)\ .$$

The two attributables $\mathcal{A}_1, \mathcal{A}_2$ at epochs $\bar{t}_1, \bar{t}_2$ give 8 scalar data, thus the problem is over-determined. From a non-spurious pair $(\tilde{\rho}_1, \tilde{\rho}_2)$, solution of (5.9), we obtain the same values of a, e, I, Ω at both times $\tilde{t}_i, i = 1, 2$, but we must check that the orbit is indeed the same, that is check the compatibility conditions

$$\omega_1 = \omega_2\,, \qquad \ell_1 = \ell_2 + n(\tilde{t}_1 - \tilde{t}_2)\,, \tag{5.16}$$

where ω_1, ω_2 and ℓ_1, ℓ_2 are the arguments of perihelion and the mean anomalies of the body at times $\tilde{t}_1, \tilde{t}_2$ and $n = ka^{-3/2}$ is the mean motion, which is the same for the two orbits. The first of conditions (5.16) corresponds to the use of the fifth integral of the Kepler problem, related to Lenz-Laplace's integral vector

$$\mathbf{L} = \frac{1}{k^2}\dot{\mathbf{r}} \times \mathbf{c} - \frac{\mathbf{r}}{|\mathbf{r}|}\ .$$

Indeed the compatibility conditions (5.16) can not be exactly satisfied, due to both the errors in the observations and to the planetary perturbations. Actually the latter are important only when the observed body undergoes a close approach to some planet in the interval between $\tilde{t}_1$ and $\tilde{t}_2$. Thus we may be able to discard some solutions, for which the compatibility conditions are largely violated. Nevertheless, we need a criterion to assess whether smaller discrepancies from the exact conditions (5.16) are due to the measurement uncertainty or rather due to the fact that the two attributables do not belong to the same physical object. This will be introduced in the next section.

5.4 Covariance of the solutions

Given a pair of attributables $\mathbf{A} = (\mathcal{A}_1, \mathcal{A}_2)$ with covariance matrices $\Gamma_{\mathcal{A}_1}, \Gamma_{\mathcal{A}_2}$, we call $\mathbf{R} = (\rho_1, \dot{\rho}_1, \rho_2, \dot{\rho}_2)$ one of the solutions of the equation $\mathbf{\Phi}(\mathbf{R}; \mathbf{A}) = \mathbf{0}$, with

$$\mathbf{\Phi}(\mathbf{R}; \mathbf{A}) = \begin{pmatrix} \mathbf{D}_1\dot{\rho}_1 - \mathbf{D}_2\dot{\rho}_2 - \mathbf{J}(\rho_1, \rho_2) \\ \mathcal{E}_1(\rho_1, \dot{\rho}_1) - \mathcal{E}_2(\rho_2, \dot{\rho}_2) \end{pmatrix}\ . \tag{5.17}$$

We can repeat what follows for each solution of $\mathbf{\Phi}(\mathbf{R};\mathbf{A}) = \mathbf{0}$.

Let $\mathbf{R} = \mathbf{R}(\mathbf{A}) = (\mathcal{R}_1(\mathbf{A}), \mathcal{R}_2(\mathbf{A}))$, where $\mathcal{R}_i(\mathbf{A}) = (\rho_i(\mathbf{A}), \dot\rho_i(\mathbf{A}))$ for $i = 1, 2$. If both elements $(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))$, $(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A}))$ give negative two-body energy orbits, then we can compute the corresponding Keplerian elements at the times

$$\tilde t_i = \tilde t_i(\mathbf{A}) = \bar t_i - \frac{\rho_i(\mathbf{A})}{c}, \qquad i = 1, 2$$

through the transformation

$$(\alpha, \delta, \dot\alpha, \dot\delta, \rho, \dot\rho) = (\mathcal{A}, \mathcal{R}) \mapsto \mathcal{E}_{Kep}(\mathcal{A}, \mathcal{R}) = (a, e, I, \Omega, \omega, \ell)\,.$$

We have, for example, a smooth function

$$\omega_i = \omega_i(\mathbf{A}) = \omega(\mathcal{A}_i, \mathcal{R}_i(\mathbf{A}))\,, \qquad i = 1, 2$$

and similar functional relations for a, e, I, Ω, ℓ. Actually, by construction, we have $a_1 = a_2, e_1 = e_2, I_1 = I_2, \Omega_1 = \Omega_2$: we denote by a the common value of a_1 and a_2. We use the vector differences

$$\Delta_{1,2} = (\Delta\omega, \Delta\ell)\,,$$

where $\Delta\omega$ is the difference of the two angles ω_1 and ω_2, $\Delta\ell$ is the difference of the two angles ℓ_1 and $\ell_2 + \mathsf{n}(\tilde t_1 - \tilde t_2)$ and $\mathsf{n} = k\mathsf{a}^{-3/2}$ is the mean motion of both orbits. Here we compute the difference of two angles in such a way that it is a smooth function near a vanishing point; for example we define $\Delta\omega = [\omega_1 - \omega_2 + \pi (\mathrm{mod}\ 2\pi)] - \pi$. With this caution, the vector $\Delta_{1,2} = \Delta_{1,2}(\mathbf{A})$ represents the discrepancy in perihelion argument and mean anomaly of the two orbits, comparing the anomalies at the same time $\tilde t_1$. We introduce the map

$$\mathbf{\Psi} : \left([-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}^2 \right)^2 \longrightarrow [-\pi, \pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R} \times S^1 \times S^1$$

$$(\mathcal{A}_1, \mathcal{A}_2) = \mathbf{A} \mapsto \mathbf{\Psi}(\mathbf{A}) = (\mathcal{A}_1, \mathcal{R}_1, \Delta_{1,2})\ ,$$

giving the orbit $(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))$ in attributable elements at time $\tilde t_1$ (the epoch of the first attributable corrected by aberration), together with the difference $\Delta_{1,2}(\mathbf{A})$ in the angular elements, which are not constrained by the angular momentum and the energy integrals. By the covariance propagation rule we have

$$\Gamma_{\mathbf{\Psi}(\mathbf{A})} = \frac{\partial \mathbf{\Psi}}{\partial \mathbf{A}}\, \Gamma_{\mathbf{A}} \left[\frac{\partial \mathbf{\Psi}}{\partial \mathbf{A}}\right]^T , \tag{5.18}$$

where

$$\frac{\partial \mathbf{\Psi}}{\partial \mathbf{A}} = \begin{bmatrix} I & 0 \\ \dfrac{\partial \mathcal{R}_1}{\partial \mathcal{A}_1} & \dfrac{\partial \mathcal{R}_1}{\partial \mathcal{A}_2} \\ \dfrac{\partial \Delta_{1,2}}{\partial \mathcal{A}_1} & \dfrac{\partial \Delta_{1,2}}{\partial \mathcal{A}_2} \end{bmatrix} \qquad \text{and} \qquad \Gamma_{\mathbf{A}} = \begin{bmatrix} \Gamma_{\mathcal{A}_1} & 0 \\ 0 & \Gamma_{\mathcal{A}_2} \end{bmatrix},$$

so that the covariance of $\mathbf{\Psi}(\mathbf{A})$ is given by the 8×8 matrix

$$\Gamma_{\mathbf{\Psi}(\mathbf{A})} = \begin{bmatrix} \Gamma_{\mathcal{A}_1} & \Gamma_{\mathcal{A}_1, \mathcal{R}_1} & \Gamma_{\mathcal{A}_1, \Delta_{1,2}} \\ \Gamma_{\mathcal{R}_1, \mathcal{A}_1} & \Gamma_{\mathcal{R}_1} & \Gamma_{\mathcal{R}_1, \Delta_{1,2}} \\ \Gamma_{\Delta_{1,2}, \mathcal{A}_1} & \Gamma_{\Delta_{1,2}, \mathcal{R}_1} & \Gamma_{\Delta_{1,2}} \end{bmatrix},$$

where

$$\Gamma_{\mathcal{A}_1,\mathcal{R}_1} = \Gamma_{\mathcal{A}_1}\left[\frac{\partial \mathcal{R}_1}{\partial \mathcal{A}_1}\right]^T , \qquad \Gamma_{\mathcal{A}_1,\Delta_{1,2}} = \Gamma_{\mathcal{A}_1}\left[\frac{\partial \Delta_{1,2}}{\partial \mathcal{A}_1}\right]^T ,$$

$$\Gamma_{\mathcal{R}_1,\Delta_{1,2}} = \frac{\partial \mathcal{R}_1}{\partial \mathcal{A}_1}\Gamma_{\mathcal{A}_1}\left[\frac{\partial \Delta_{1,2}}{\partial \mathcal{A}_1}\right]^T + \frac{\partial \mathcal{R}_1}{\partial \mathcal{A}_2}\Gamma_{\mathcal{A}_2}\left[\frac{\partial \Delta_{1,2}}{\partial \mathcal{A}_2}\right]^T ,$$

$$\Gamma_{\mathcal{R}_1,\mathcal{A}_1} = \Gamma^T_{\mathcal{A}_1,\mathcal{R}_1} , \qquad \Gamma_{\Delta_{1,2},\mathcal{A}_1} = \Gamma^T_{\mathcal{A}_1,\Delta_{1,2}} , \qquad \Gamma^T_{\Delta_{1,2},\mathcal{R}_1} = \Gamma^T_{\mathcal{R}_1,\Delta_{1,2}} ,$$

and

$$\Gamma_{\mathcal{A}_1} = \frac{\partial \mathcal{A}_1}{\partial \mathbf{A}}\Gamma_{\mathbf{A}}\left[\frac{\partial \mathcal{A}_1}{\partial \mathbf{A}}\right]^T , \qquad \Gamma_{\mathcal{R}_1} = \frac{\partial \mathcal{R}_1}{\partial \mathbf{A}}\Gamma_{\mathbf{A}}\left[\frac{\partial \mathcal{R}_1}{\partial \mathbf{A}}\right]^T , \qquad \Gamma_{\Delta_{1,2}} = \frac{\partial \Delta_{1,2}}{\partial \mathbf{A}}\Gamma_{\mathbf{A}}\left[\frac{\partial \Delta_{1,2}}{\partial \mathbf{A}}\right]^T .$$

The matrices $\frac{\partial \mathcal{R}_i}{\partial \mathcal{A}_j}, i, j = 1, 2$, can be computed from the relation

$$\frac{\partial \mathbf{R}}{\partial \mathbf{A}}(\mathbf{A}) = -\left[\frac{\partial \mathbf{\Phi}}{\partial \mathbf{R}}(\mathbf{R}(\mathbf{A}), \mathbf{A})\right]^{-1} \frac{\partial \mathbf{\Phi}}{\partial \mathbf{A}}(\mathbf{R}(\mathbf{A}), \mathbf{A}) .$$

We also have

$$\frac{\partial \Delta\omega}{\partial \mathcal{A}_1} = \frac{\partial \omega}{\partial \mathcal{A}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A})) + \frac{\partial \omega}{\partial \mathcal{R}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))\frac{\partial \mathcal{R}_1(\mathbf{A})}{\partial \mathcal{A}_1} - \frac{\partial \omega}{\partial \mathcal{R}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A}))\frac{\partial \mathcal{R}_2(\mathbf{A})}{\partial \mathcal{A}_1} ,$$

$$\frac{\partial \Delta\omega}{\partial \mathcal{A}_2} = \frac{\partial \omega}{\partial \mathcal{R}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))\frac{\partial \mathcal{R}_1(\mathbf{A})}{\partial \mathcal{A}_2} - \frac{\partial \omega}{\partial \mathcal{A}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A})) - \frac{\partial \omega}{\partial \mathcal{R}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A}))\frac{\partial \mathcal{R}_2(\mathbf{A})}{\partial \mathcal{A}_2}$$

and

$$\begin{aligned}
\frac{\partial \Delta\ell}{\partial \mathcal{A}_1} &= \frac{\partial \ell}{\partial \mathcal{A}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A})) + \frac{\partial \ell}{\partial \mathcal{R}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))\frac{\partial \mathcal{R}_1(\mathbf{A})}{\partial \mathcal{A}_1} - \frac{\partial \ell}{\partial \mathcal{R}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A}))\frac{\partial \mathcal{R}_2(\mathbf{A})}{\partial \mathcal{A}_1} + \\
&+ \frac{3}{2}\frac{\mathsf{n}}{\mathsf{a}}\left[\frac{\partial a}{\partial \mathcal{A}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A})) + \frac{\partial a}{\partial \mathcal{R}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))\frac{\partial \mathcal{R}_1(\mathbf{A})}{\partial \mathcal{A}_1}\right][\tilde{t}_1(\mathbf{A}) - \tilde{t}_2(\mathbf{A})] + \\
&+ \frac{\mathsf{n}}{c}\left[\frac{\partial \rho_1}{\partial \mathcal{A}_1}(\mathbf{A}) - \frac{\partial \rho_2}{\partial \mathcal{A}_1}(\mathbf{A})\right] ,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Delta\ell}{\partial \mathcal{A}_2} &= \frac{\partial \ell}{\partial \mathcal{R}}(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))\frac{\partial \mathcal{R}_1(\mathbf{A})}{\partial \mathcal{A}_2} - \frac{\partial \ell}{\partial \mathcal{A}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A})) - \frac{\partial \ell}{\partial \mathcal{R}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A}))\frac{\partial \mathcal{R}_2(\mathbf{A})}{\partial \mathcal{A}_2} + \\
&+ \frac{3}{2}\frac{\mathsf{n}}{\mathsf{a}}\left[\frac{\partial a}{\partial \mathcal{A}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A})) + \frac{\partial a}{\partial \mathcal{R}}(\mathcal{A}_2, \mathcal{R}_2(\mathbf{A}))\frac{\partial \mathcal{R}_2(\mathbf{A})}{\partial \mathcal{A}_2}\right][\tilde{t}_1(\mathbf{A}) - \tilde{t}_2(\mathbf{A})] + \\
&+ \frac{\mathsf{n}}{c}\left[\frac{\partial \rho_1}{\partial \mathcal{A}_2}(\mathbf{A}) - \frac{\partial \rho_2}{\partial \mathcal{A}_2}(\mathbf{A})\right] .
\end{aligned}$$

5.4.1 Identification of attributables

One important step is to decide if trying to link $\mathcal{A}_1$, $\mathcal{A}_2$ has produced at least one reliable orbit, so that we can state the two sets of observations defining the $\mathcal{A}_i, i = 1, 2$ may belong to the one and the same solar system body. Neglecting the unavoidable errors in the observations and the approximations made both with the interpolation to compute $\mathcal{A}_1$, $\mathcal{A}_2$ and with the use of a two-body model, if the observations belong to the same solar system body, then $\Delta_{1,2}(\mathbf{A}) = \mathbf{0}$. We need to check whether the failure of this condition is within the acceptable range of

values which is statistically expected to be generated by the errors in the available observations.
The marginal covariance matrix of the compatibility conditions is

$$\Gamma_{\Delta_{1,2}} = \frac{\partial \Delta_{1,2}}{\partial \mathbf{A}} \Gamma_{\mathbf{A}} \left[\frac{\partial \Delta_{1,2}}{\partial \mathbf{A}} \right]^T \ .$$

The inverse matrix $C^{\Delta_{1,2}} = \Gamma_{\Delta_{1,2}}^{-1}$ defines a norm $\| \cdot \|_\star$ in the $(\Delta\omega, \Delta\ell)$ plane, allowing to test an identification between the attributables $\mathcal{A}_1, \mathcal{A}_2$: the test is

$$\|\Delta_{1,2}\|_\star^2 = \Delta_{1,2} C^{\Delta_{1,2}} \Delta_{1,2}^T \leq \chi_{max}^2 \ , \tag{5.19}$$

where χ_{max} is a control parameter. The value of the control could be selected on the basis of χ^2 tables, if we could assume that the observations errors are Gaussian and their standard deviations, mean values and correlations were known. Since this hypothesis is not satisfied in practice, the control value χ_{max} needs to be selected on the basis of large scale tests. Note that, for each pair of attributables, more than one preliminary orbit computed with the method of Section 5.3 could pass the control (5.19); thus we can have alternative preliminary orbits.

5.4.2 Uncertainty of the orbits

The methods explained in Section 5.3 also allow to assign an uncertainty to the preliminary orbits that we compute. A solution $(\mathcal{A}_1, \mathcal{R}_1(\mathbf{A}))$, in attributable elements, has the marginal covariance matrix

$$\begin{bmatrix} \Gamma_{\mathcal{A}_1} & \Gamma_{\mathcal{A}_1,\mathcal{R}_1} \\ \Gamma_{\mathcal{R}_1,\mathcal{A}_1} & \Gamma_{\mathcal{R}_1} \end{bmatrix} \ .$$

The preliminary orbits obtained by the other available algorithms do not produce a nondegenerate covariance matrix: this is usually computed in the differential correction step of the orbit determination procedure. With the algorithm of [41] a covariance matrix may be defined, but it is not positive definite (see (4.19)). The advantage of having a covariance matrix already from the preliminary orbit step could be important in two ways. First, the covariance matrix describes a confidence ellipsoid where a two-body orbit, compatible with the observations and their errors, can be found. The size of this ellipsoid can provide useful hints on the difficulty of the differential corrections procedure. Second, even if the differential corrections are divergent, the covariance matrix of the preliminary orbit can be used to compute a prediction with confidence region, allowing for a planned recovery, for assessment of impact risk, and so on.

Chapter 6

Numerical experiments

We present some numerical experiments to test the algorithms introduced in the previous chapter: in Section 6.1 we show the results of a test case, illustrating the computation of the preliminary orbits for a numbered asteroid whose orbit is well known, while in Section 6.2 we investigate the performance of the method on a large database of simulated observations. Finally, in Section 6.3 we expose the results of a validation test for the case of the space debris.

6.1 A test case

We show a test of the linkage procedure using the attributables

$$\begin{aligned} \mathcal{A}_1 &= (0.2872656, 0.1106342, -0.00375115, -0.00167695)\,, \\ \mathcal{A}_2 &= (0.2820817, 0.1086542, 0.00514465, 0.00215975) \end{aligned}$$

of the asteroid (101878) 1999 NR_{23} at epochs $\bar{t}_1 = 54000$, $\bar{t}_2 = 54109$ respectively (time in MJD). The values of the components of $\mathcal{A}_1, \mathcal{A}_2$ are in radians/radians per day. They have been computed from two groups of observations, separated by more than 100 days, made from two different observatories: Mauna Kea (568) and Mt. Lemmon Survey (G96). From the known nominal orbit of this asteroid we obtain the values

$$\rho_1 = 1.0419\,, \qquad \rho_2 = 2.0485$$

of the topocentric distance (in AU) at the two mean epochs of the observations.

In Figure 6.1 we show the intersections between the curves defined by $p(\rho_1, \rho_2)$ and $q(\rho_1, \rho_2)$. By solving the corresponding problem (5.9) with the method described in Subsection 5.3.1 we find the 6 positive pairs of solutions (ρ_1, ρ_2) displayed in Table 6.1.

After removing solution 1 (with both components very small), the spurious solution 6 (not satisfying (5.7)) and the spurious solutions 3 and 5 (not satisfying (5.6)), we are left with the values labeled 2 and 4 in Table 6.1. Note that, even if solutions 2 and 3 look close, they are far apart enough to select only one of them as a good solution. For the left solutions 2 and 4 we succeed in computing Keplerian orbits, that we list in Table 6.2. The values of a, e, I, Ω are the same for each pair: this is due to the fact that each pair of orbits shares the same angular momentum and the

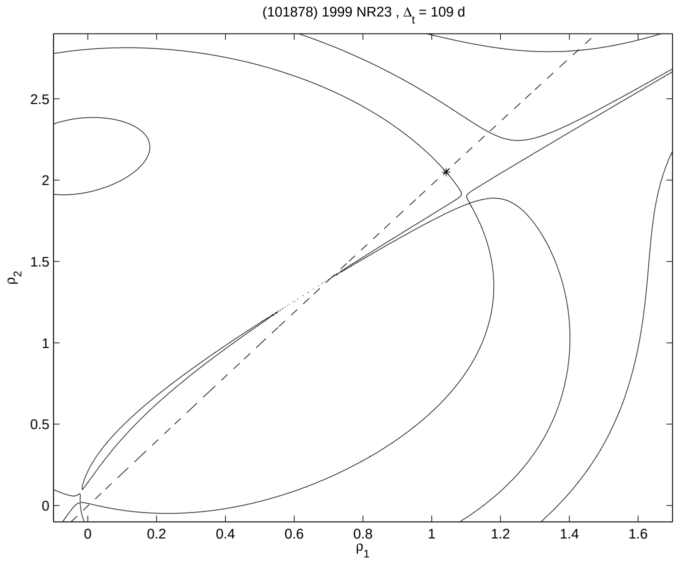

Figure 6.1. Intersections of the curves $p = 0, q = 0$ (solid and dashed, respectively) in the plane ρ_1, ρ_2 for the asteroid (101878) 1999 NR_{23}: the asterisk corresponds to the true orbit.

	ρ_1	ρ_2
1	0.0059	0.0097
2	0.7130	1.4100
3	0.7045	1.3933
4	1.0409	2.0517
5	1.1659	2.2952
6	1.4246	2.7968

Table 6.1. Solutions of the system (5.9) for (101878) 1999 NR_{23}.

	1		2	
$\\|\Delta_{1,2}\\|_\star$	487.65806		0.19505	
a	6.87384	6.87384	2.25828	2.25828
e	0.81798	0.81798	0.19787	0.19787
I	0.51733	0.51733	0.59995	0.59995
Ω	156.55521	156.55521	156.42531	156.42531
ω	144.68146	321.78289	144.39580	145.26330
ℓ	4.66178	355.27766	47.75173	78.65378
t (MJD)	53999.8205	54109.1368	53999.8186	54109.1331

Table 6.2. Keplerian elements (angles in degrees) corresponding to the pairs (ρ_1, ρ_2) labelled with 2 and 4 in Table 6.1. The value of $\|\Delta_{1,2}\|_\star$ is shown for the two solutions.

same energy. The value of the identification norm $\|\Delta_{1,2}\|_\star$, also shown in Table 6.2, strongly suggests to select the second solution. The results appear pretty good, in fact the differences with the true solution are of the order of 3×10^{-5} AU and the errors in the Keplerian elements are comparable with the planetary perturbations; this is intrinsically bound to the use of a two-body approximation.

6.2 Numerical experiments with simulated observations

We have tested our identification method with the DFT algorithm, explained in Subsection 5.3.1, using simulated observations of objects in a solar system model. The data have been given to us by R. Jedicke and L. Dennau from the Institute of Astronomy, University of Hawaii, and the data quality resemble the one which should be achieved by the Pan-STARRS telescope when it will be fully operative. The RMS of the observations vary from 0.01 to 0.02 *arcsec*, that is rather optimistic for the current surveys. The current astrometric data quality of the Pan-STARRS 1 telescope is such that the RMS of the residuals for well determined asteroid orbits is between 0.11 and 0.13 *arcsec*. Better results should be achieved when the astrometric reduction of asteroid detections will be performed with respect to a catalogue generated by the Pan-STARRS survey itself.
The simulated observations cover 31 observing nights, in three consecutive lunations and are grouped into *tracklets*. Each tracklet is composed by observations presumably belonging to the same object and covering a short arc: some of them are false (e.g. join observations of different objects). From each tracklet we can compute an attributable. We have first applied to the database of tracklets the identification procedures defined in [41], [31]. Then we have tested our method on the leftover database, for which the previous procedures have failed. These remaining observations corresponds to 19441 objects, and 24590 tracklets, but only 4132 objects have at least two tracklets, that is a necessary requirement for the application of our method. The hyperbolic orbits have been removed from the solar system model: in fact our current method does not search for them, but it could be easily modified to include their orbit determination. To each accepted preliminary orbit obtained from a pair of attributables we apply the differential corrections, using all the observations at our disposal, to compute a least squares orbit with its covariance matrix.[1]

To reduce the computational complexity, we need to define a filter to select the pairs of attributables which we try to link. In Subsection 6.2.1 we describe the two filters we have used in processing the simulated data.

6.2.1 Filtering pairs of attributables

First filter: guessing the second angular position.

A first simple way to discard pairs of attributables at epochs $\bar{t}_1, \bar{t}_2$ is to constrain the time span $\delta t = \bar{t}_2 - \bar{t}_1$: we require

$$\delta t_{min} \leq \delta t \leq \delta t_{max} \tag{6.1}$$

for suitable positive constants $\delta t_{min}, \delta t_{max}$. In our experiment we have used $\delta t_{max} =$ 99 days and $\delta t_{min} = 0.5$ days, that practically means we have tried to link attributa-

[1]We use the preliminary orbit at time $\tilde{t}_1$ as starting guess for the differential corrections. We could also use the orbit at time $\tilde{t}_2$, or an 'average orbit' at time $(\tilde{t}_1 + \tilde{t}_2)/2$.

bles obtained in different nights. For each given pair of attributables at epochs $\bar{t}_1, \bar{t}_2$ fulfilling (6.1) we consider for $i = 1, 2$ the corresponding proper motions η_i and the mobile bases $\{\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{v}}_i, \hat{\mathbf{n}}_i\}$, defined in Section 5.1. We want to use one of the proper motions, say η_1, to bound the region in the sky where we could recover the object at the other time $\bar{t}_2$.

Let us form the orthogonal matrices $V_1 = [\hat{\boldsymbol{\rho}}_1|\hat{\mathbf{v}}_1|\hat{\mathbf{n}}_1]$ and $V_2 = [\hat{\boldsymbol{\rho}}_2|\hat{\mathbf{v}}_2|\hat{\mathbf{n}}_2]$: these are rotation matrices to the mobile bases $\{\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{v}}_i, \hat{\mathbf{n}}_i\}$, $i = 1, 2$. Let $R_{\phi\,\hat{\mathbf{e}}}$ denote the rotation of an angle ϕ around the unit vector $\hat{\mathbf{e}}$. Then $R_{\eta_1\delta t\,\hat{\mathbf{n}}_1} = V_1\, R_{\eta_1\delta t\,\hat{\mathbf{z}}}\, V_1^T$ ($\hat{\mathbf{z}}$ is the third unit vector of the reference frame defining our rectangular coordinates) is the parallel transport matrix along the geodesic on the unit sphere defined by $\mathcal{A}_1$ to time $\bar{t}_2$; hence $\hat{\boldsymbol{\rho}}_{12} = R_{\eta_1\delta t\,\hat{\mathbf{n}}_1}\hat{\boldsymbol{\rho}}_1$ is the predicted observation direction at time $\bar{t}_2$, assuming the trajectory is a great circle and the proper motion is constant. By exchanging the order of the two attributables we can compute $R_{-\eta_2\delta t\,\hat{\mathbf{n}}_2} = V_2\, R_{-\eta_2\delta t\,\hat{\mathbf{z}}}\, V_2^T$ and $\hat{\boldsymbol{\rho}}_{21} = R_{-\eta_2\delta t\,\hat{\mathbf{n}}_2}\, \hat{\boldsymbol{\rho}}_2$, that is the prediction at time $\bar{t}_1$. We use the metric

$$d(\hat{\boldsymbol{\rho}}_1, \hat{\boldsymbol{\rho}}_2) = \min\{\widehat{\hat{\boldsymbol{\rho}}_{12}, \hat{\boldsymbol{\rho}}_2}, \widehat{\hat{\boldsymbol{\rho}}_{21}, \hat{\boldsymbol{\rho}}_1}\}\,,$$

that is the minimum between the two angular differences, discarding pairs of attributables that give rise to a large value of this metric.

Note that the proper motion does not vary too much in the time interval between the two attributables provided δt_{max} is small enough; thus, if we want to use large values of δt, we have also to allow large values of the metric d.

Second filter: symmetric LLS fit.

Given the two attributables $\mathcal{A}_1, \mathcal{A}_2$ at times $\bar{t}_1, \bar{t}_2$ we perform a quadratic approximation of the apparent motion on the celestial sphere S^2 by using a Linear Least Squares (LLS) fit. The apparent motion is given by the functions $\alpha(t), \delta(t)$. We approximate $\alpha(t), \delta(t)$ with second degree polynomials whose coefficients are derived from a least squares fit. We denote the approximating quadratic functions as

$$\alpha(t) = \alpha_q + \dot{\alpha}_q(t - \bar{t}) + \frac{1}{2}\ddot{\alpha}_q(t - \bar{t})^2\,, \qquad \delta(t) = \delta_q + \dot{\delta}_q(t - \bar{t}) + \frac{1}{2}\ddot{\delta}_q(t - \bar{t})^2\,, \quad (6.2)$$

where $\bar{t} = \frac{1}{2}(\bar{t}_1 + \bar{t}_2)$ is the mean of the times of the attributables. The corresponding time derivatives are

$$\dot{\alpha}(t) = \dot{\alpha}_q + \ddot{\alpha}_q(t - \bar{t})\,, \qquad \dot{\delta}(t) = \dot{\delta}_q + \ddot{\delta}_q(t - \bar{t})\,.$$

We want to determine the 6 quantities $\alpha_q, \dot{\alpha}_q, \ddot{\alpha}_q, \delta_q, \dot{\delta}_q, \ddot{\delta}_q$ using the data coming from the attributables. The vector of residuals is

$$\boldsymbol{\xi} = \left(\mathcal{A}_1 - \mathcal{A}(\bar{t}_1)\,, \mathcal{A}_2 - \mathcal{A}(\bar{t}_2)\right)^T\,,$$

with $\mathcal{A}(t) = (\alpha(t), \delta(t), \dot{\alpha}(t), \dot{\delta}(t))$.
Given the covariance matrices $\Gamma_{\mathcal{A}_1}, \Gamma_{\mathcal{A}_2}$ associated to the attributables we use them to *weight* the residuals in the definition of the target function:

$$Q(\boldsymbol{\xi}) = \frac{1}{8}\boldsymbol{\xi}\cdot W\boldsymbol{\xi}\,, \qquad \text{where} \quad W^{-1} = \begin{pmatrix} \Gamma_{\mathcal{A}_1} & 0 \\ 0 & \Gamma_{\mathcal{A}_2} \end{pmatrix}.$$

We introduce the notation

$$\mathbf{x} = (\alpha_q, \dot{\alpha}_q, \ddot{\alpha}_q, \delta_q, \dot{\delta}_q, \ddot{\delta}_q)^T \,, \quad \vec{\lambda} = (\mathcal{A}_1, \mathcal{A}_2)^T \,.$$

The value of $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{x})$ that minimizes the target function is obtained by solving the normal equation

$$C\mathbf{x} = -B^T W \vec{\lambda}, \qquad \text{where } B = \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}, \quad C = B^T W B \,,$$

and the matrix B has the form

$$B = -\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \quad \text{with} \quad B_i = \begin{pmatrix} 1 & (\bar{t}_i - \bar{t}) & \frac{1}{2}(\bar{t}_i - \bar{t})^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & (\bar{t}_i - \bar{t}) & \frac{1}{2}(\bar{t}_i - \bar{t})^2 \\ 0 & 1 & (\bar{t}_i - \bar{t}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & (\bar{t}_i - \bar{t}) \end{pmatrix},$$

for $i = 1, 2$. Once the value of $\mathbf{x}$ is given, we compute the residuals $\boldsymbol{\xi}(\mathbf{x})$ and use the norm $\sqrt{Q(\boldsymbol{\xi})}$ to decide which are the pairs of attributables $(\mathcal{A}_1, \mathcal{A}_2)$ to discard. We also discard the pairs giving rise to a large value of the quantity

$$\kappa_q \eta_q^2 = \frac{1}{\eta_q} \left[(\ddot{\delta}_q \dot{\alpha}_q - \ddot{\alpha}_q \dot{\delta}_q) \cos \delta_q + \dot{\alpha}_q (\eta_q^2 + \dot{\delta}_q^2) \sin \delta_q \right] ,$$

where $\eta_q = \sqrt{\dot{\delta}_q^2 + \dot{\alpha}_q^2 \cos^2 \delta_q}$ and κ_q is the geodesic curvature (see [37], Chapter 9).

6.2.2 Results

The accuracy of the linkage method can be measured by the number of true identifications over the total number of identifications found. The total number is 3625 and the true ones (that may be related to the same object if it has more than 2 tracklets) are 2908, i.e. 80.2% of the total. We could eliminate almost half of the 717 false identifications by lowering from 0.15 to 0.0625 *arcsec* the control on the RMS for acceptable orbits after differential corrections: but this would make us lose 95 true identifications.

In Table 6.3 we show the efficiency of the linkage procedure, that is we write the number of objects for which at least a pair of tracklets has been correctly linked, giving the details for the MB (Main Belt) and the NEO (Near Earth Object) class. As expected, the efficiency appears greater if there are three tracklets that can be pairwise linked. We stress that we have tested our method with data for which the other available methods in [41], [31] could not perform the linkage.

An interesting feature that comes out from our numerical experiments is that this method appears to work also when the time span between the two attributables is large, hence it can be used in cases where the other linkage methods fail. Another feature of our linkage procedure is that it allows to compute a nondegenerate covariance matrix for each preliminary orbit. Therefore this method can be important for two kinds of applications: 1) to recover objects whose orbit could not be computed with either the classical or the modern known algorithms; 2) to design the scheduler of new surveys planning a smaller number of observations for each object.

The number of alternative solutions of the problem deserves a deeper investigation, however we expect that the acceptable ones should often be much less than 48,

with 2 tracklets in 2 nights	Total	Found		Lost	
all	1074	951	88.5%	123	11.5%
MB	1038	936	90.2%	102	9.8%
NEO	19	9	47.4%	10	52.6%
with 3 tracklets in 3 nights	Total	Found		Lost	
All	214	205	95.8%	9	4.2%
MB	197	194	98.5%	3	1.5%
NEO	3	2	66.7%	1	33.3%

Table 6.3. Efficiency of the identification procedure.

the total degree of the polynomial system (5.9). Moreover the performance of the second algorithm to solve (5.9), described in Subsection 5.3.2, has not been tested yet: we would like to perform further experiments to decide if it allows to decrease the computation time.

6.3 Space debris

In the context of orbit determination of satellites and space debris the word used to indicate the analogous of the problem of identification for the asteroids is *correlation*.

The method and the algorithms described so far can be used without major changes for the case of the space debris: it suffices to indicate with the vector $\mathbf{r}$ the geocentric position, instead of the heliocentric one.

We have asked for the opportunity to use an existing data set of observations from ESA Optical Ground Station (OGS) to test our algorithms. The provided data have been obtained at Teide Observatory (Canary Island) in the year 2007.

In the special case of a survey of the geosynchronous region the observations can be taken by stopping the telescope motor, that means in a reference frame fixed with respect to the Earth. The stars appear as long trails, the nearly geostationary objects as very short ones or even points, the other debris as medium to long trails. The ends of all trails are measured. The ones of the moving objects are converted into two positions taken at the beginning and at the end of the exposure and form a *tracklet*. From any tracklet we can compute an attributable by a linear fit, as explained in Subsection 4.2.2.

We have runned the algorithm described in Section 5.3.1 on the 3172 tracklets from the year 2007 OGS observations. We have limited the time interval to $|\bar{t}_2-\bar{t}_1| \leq$ 10 days, to avoid excessive accumulation of perturbations making the two-body preliminary orbit a poor approximation; we do not yet know what is the maximum usable time span.

We have found 363 correlations of 2 tracklets, with 378 accepted orbits. These need to be confirmed. Moreover, there are 15 cases with two significantly different orbits, where we do not know how to choose among the two.

The *correlation confirmation* can be obtained by looking for a third tracklet which can be correlated to both the tracklets of a couple: this process is called *attribution*. To do it, we take the 2-tracklets orbit and predict the attributable at the time $\bar{t}_3$ of the third tracklet. Then we compare the obtained attributable with the one computed from the third tracklet. Both the attributables have a covariance matrix, then we can compute the weighted norm of the difference, to be compared with a suitable bound. If this test is passed, then we proceed with the differential corrections. After a set of three tracklets has been correlated we can try to search for a fourth tracklet to add and so on.

At the end of this recursive procedure we have a lot of many-tracklets orbits, but there can be duplicates, corresponding to the same attributions made in a different order. We call the procedure to remove the duplicates *correlation management*. We can have duplicates in the form of correlations with exactly the same tracklets in different order, but we can also have inferior correlations: this is the case when a correlation has a subset of tracklets with respect to another one. In the latter case we have to remove the correlations with less tracklets. After removing duplicates the output catalog is said to be normalized. At this point we may try to put together two correlations with some tracklets in common.

After the correlation management, the output of the test included 206 correlations, with 220 orbits. Of these, 112 were not confirmed, that is they were limited to two tracklets (see Table 6.4).

T	2	3	4	5	6	7	8	9	10	12
C	112	40	29	10	3	5	3	1	1	1

Table 6.4. C is the number of correlations found with T tracklets.

Out of 3172 input tracklets, 464 have been correlated, 2708 left uncorrelated. However, we have no way to know how many should have been correlated, that is how many physically distinct objects are there: in particular, objects re-observed at intervals longer than 10 days have escaped correlation.

Bibliography

[1] S. Aubry, R. S. Mackay, and C. Baesens. Equivalence of uniform hyperbolicity for symplectic twist maps and phonon gap for Frenkel-Kontorova models. *Physica D*, 56:123–134, 1992.

[2] R. R. Bate, D. D. Mueller, and J. E. White. *Fundamentals of Astrodynamics*. Dover publications, 1971.

[3] D. A. Bini. Numerical computation of polynomial zeros by means of Aberth method. *Numer. Algorithms*, 13(3-4):179–200, 1997.

[4] S. V. Bolotin. Nonintegrability of the n-centre problem for $n > 2$. *Mosc. Univ. Mech. Bull.*, 39(3):24–28, 1984. Translation from Vestn. Mosk. Univ., Ser.I(3):65–68, 1984.

[5] S. V. Bolotin. The effect of singularities of the potential energy on the integrability of mechanical systems. *Journal of Applied Math. and Mech.*, 48(3):255–260, 1985. Translation from Prikl. Matem. Mekhan., 48(3):356–362, 1984.

[6] S. V. Bolotin. Second species periodic orbits of the elliptic 3-body problem. *Celestial Mechanics & Dynamical Astronomy*, 93:343–371, 2005.

[7] S. V. Bolotin. Shadowing chains of collision orbits. *Discrete and Continuous Dynamical Systems*, 14(2):235–260, 2006.

[8] S. V. Bolotin and R. S. Mackay. Periodic and chaotic trajectories of the second species for the n-centre problem. *Celestial Mechanics & Dynamical Astronomy*, 77:49–75, 2000.

[9] S. V. Bolotin and R. S. Mackay. Nonplanar second species periodic and chaotic trajectories for the circular restricted three-body problem. *Celestial Mechanics & Dynamical Astronomy*, 94:433–449, 2006.

[10] S. V. Bolotin and P. Negrini. Regularization and topological entropy for the spatial n-centre problem. *Ergodic Theory & Dynamical Systems*, 21:383–399, 2001.

[11] S. V. Bolotin and P. Negrini. Chaotic behavior in the 3-centre problem. *Journal of Differential Equations*, 190:539–558, 2003.

[12] H. Bruns. Über die Integrale des Vielkörper-Problems. *Acta Math.*, 11, 1887.

[13] A. Celletti and G. Pinzari. Four classical methods for determining planetary elliptic elements: a comparison. *Celestial Mechanics & Dynamical Astronomy*, 93:1–52, 2005.

[14] A. Celletti and G. Pinzari. Dependence on the observational time intervals and domain of convergence of orbital determination methods. *Celestial Mechanics & Dynamical Astronomy*, 95:327–344, 2006.

[15] C. L. Charlier. *Die Mechanik des Himmels*, volume I, II. Verlag Von Veit & Comp., Leipzig, 1902.

[16] D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, Varieties and Algorithms*. Springer, 1996.

[17] B. Deng. The Sil'nikov problem, exponential expansion, strong λ-lemma, C^1-linearization, and homoclinic bifurcation. *Journal of Differential Equations*, 79(2):189–231, 1989.

[18] Y. Duan and J. Yuan. Periodic orbits of the hydrogen molecular ion. *The European Physical Journal D*, 6:319–326, 1999.

[19] Y. Duan, J. Yuan, and C. Bao. Periodic orbits of the hydrogen molecular ion and their quantization. *Physical Review A*, 52(5):3497–3502, 1995.

[20] J. Font, A. Nunes, and C. Simó. Consecutive quasi-collisions in the planar circular RTBP. *Nonlinearity*, 15:115–142, 2002.

[21] J. Font, A. Nunes, and C. Simó. A numerical study of the orbits of second species of the planar circular RTBP. *Celestial Mechanics & Dynamical Astronomy*, 103:143–162, 2009.

[22] C. F. Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Reprinted by Dover Publications, 1963, or. ed. 1809.

[23] G. F. Gronchi. On the stationary points of the squared distance between two ellipses with a common focus. *SIAM Journ. Sci. Comp.*, 24(1):61–80, 2002.

[24] G. F. Gronchi. Multiple solutions in preliminary orbit determination from three observations. To appear in Celestial Mechanics & Dynamical Astronomy, 2009.

[25] S. Herrick. *Astrodynamics*, volume I. Van Nostrand Reinhold, London, 1971.

[26] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Cambridge University Press, 1995.

[27] A. Knauf and I. A. Taimanov. On the integrability of the n-centre problem. *Mathematische Annalen*, 331:631–649, 2005.

[28] V. V. Kozlov. Topological obstacles to the integrability of natural mechanical systems. *Doklady Akademii Nauk SSSR*, 249(6):1299–1302, 1980.

[29] V. V. Kozlov and O. A. Harin. Kepler's problem in constant curvature spaces. *Celestial Mechanics & Dynamical Astronomy*, 54(4):393–399, 1992.

[30] L. K. Kristensen. Orbit Determination by Four Observations. *Astron. Nachr.*, 316(4):261–266, 1995.

[31] J. Kubica, L. Denneau, T. Grav, J. Heasley, R. Jedicke, J. Masiero, A. Milani, A. Moore, D. Tholen, and R. J. Wainscoat. Efficient intra- and inter-night linking of asteroid detections using kd-trees. *Icarus*, 189:151–168, 2007.

[32] P. Kustaanheimo and E. Stiefel. Perturbation theory of Kepler motion based on spinor regularization. *Journal für die Reine und Angewandte Mathematik*, 218:204–219, 1965.

[33] P. S. Laplace. Mém. Acad. R. Sci. Paris (1780). In Laplace's collected works, *10*, 93–146.

[34] A. Leuschner. On the Laplacean orbit methods. In *Proceedings of the ICM*, pages 209–217, 1912.

[35] T. Levi Civita. Sur la résolution qualitative du probléme restreint des trois corps. *Acta Math.*, 30:305–327, 1906.

[36] G. Merton. A modification of Gauss's Method for the Determination of Orbits. *Monthly Notices of the Royal Astronomical Society*, 85:693–731, 1925.

[37] A. Milani and G. F. Gronchi. *Theory of Orbit Determination*. Cambridge University Press, 2009.

[38] A. Milani, G. F. Gronchi, M. de'Michieli Vitturi, and Z. Knežević. Orbit Determination with Very Short Arcs I. Admissible Regions. *Celestial Mechanics & Dynamical Astronomy*, 90:59–87, 2004.

[39] A. Milani, G. F. Gronchi, D. Farnocchia, Z. Knežević, R. Jedicke, L. Denneau, and F. Pierfederici. Topocentric Orbit Determination: algorithms for the next generation surveys. *Icarus*, 195:474–492, 2008.

[40] A. Milani, G. F. Gronchi, and Z. Knežević. New Definition of Discovery for Solar System Objects. *Earth Moon Planets*, 100:83–116, 2007.

[41] A. Milani, G. F. Gronchi, Z. Knežević, M. E. Sansaturio, and O. Arratia. Orbit Determination with Very Short Arcs II. Identifications. *Icarus*, 79:350–374, 2005.

[42] A. Milani, M. E. Sansaturio, and S. R. Chesley. The Asteroid Identification Problem IV: Attributions. *Icarus*, 151:150–159, 2001.

[43] J. Milnor. *Morse theory*. Princeton University Press, 1963.

[44] O. F. Mossotti. Nuova analisi del problema di determinare le orbite dei corpi celesti. Scritti, Pisa, Domus Galileiana, 1816-1818.

[45] P. Painlevé. Mémoire sur les intégrales premières du problème des n corps. *Bull. Astron.*, 15, 1898.

[46] H. C. Plummer. *An introductory treatise on Dynamical Astronomy*. Cambridge University press, 1918. Reprinted by Dover publications, New York (1960).

[47] H. Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Math.*, 13, 1890.

[48] H. Poincaré. *Les Méthodes Nouvelles de la Mécanique Céleste*, volume III. Gauthiers-Villars, Paris, 1899.

[49] H. Poincaré. Sur la détermination des orbites par la méthode de Laplace. *Bull. Astron.*, 23:161–187, 1906.

[50] C. L. Siegel. Über die algebraischen Integrale des restringierten Dreikörperproblems. *Transactions of the American Mathematical Society*, 39:225–233, 1936.

[51] V. Szebehely. *Theory of orbits. The Restricted Problem of Three Bodies.* Academic Press, 1967.

[52] L. G. Taff and D. L. Hall. The use of angles and angular rates. I - Initial orbit determination. *Celestial Mechanics & Dynamical Astronomy*, 16(4):481–488, 1977.

[53] L. G. Taff and D. L. Hall. The use of angles and angular rates. II - Multiple Observation Initial orbit determination. *Celestial Mechanics & Dynamical Astronomy*, 21(3):281–290, 1980.

[54] T. G. Vozmishcheva. Integrable problems of Celestial Mechanics in spaces of constant curvature. *Journal of Mathematical Sciences*, 125(4):419–532, 2005.

[55] E. T. Whittaker. *A treatise on the analytical dynamics of particles and rigid bodies. With an introduction to the problem of three bodies* . Cambridge University Press, fourth edition, 1960.