Global Geometry of 3-Body Motions with Vanishing Angular Momentum I

Wu-Yi HSIANG*   Eldar STRAUME**

Abstract Following Jacobi's geometrization of Lagrange's least action principle, trajectories of classical mechanics can be characterized as geodesics on the configuration space M with respect to a suitable metric which is the conformal modification of the kinematic metric by the factor $(U + h)$, where U and h are the potential function and the total energy, respectively. In the special case of 3-body motions with zero angular momentum, the global geometry of such trajectories can be reduced to that of their moduli curves, which record the change of size and shape, in the moduli space of oriented m-triangles, whose kinematic metric is, in fact, a Riemannian cone over the shape space $M^* \simeq S^2(\frac{1}{2})$.

In this paper, it is shown that the moduli curve of such a motion is uniquely determined by its shape curve (which only records the change of shape) in the case of $h \neq 0$, while in the special case of $h = 0$ it is uniquely determined up to scaling. Thus, the study of the global geometry of such motions can be further reduced to that of the shape curves, which are time-parametrized curves on the 2-sphere characterized by a third order ODE. Moreover, these curves have two remarkable properties, namely the uniqueness of parametrization and the monotonicity, that constitute a solid foundation for a systematic study of their global geometry and naturally lead to the formulation of some pertinent problems.

Keywords 3-Body problem, Kinematic geometry, Reduction, Shape curves
2000 MR Subject Classification 70F07, 70F15, 51P05

1 Introduction

1.1 Local and global characterization of 3-body trajectories

The classical 3-body problem of celestial mechanics studies the local and global geometry of the trajectories of a 3-body system under the influence of the gravitational forces, or equivalently a conservative system with potential energy $-U$, where

$$U = \sum_{i<j} \frac{m_i m_j}{r_{ij}} \tag{1.1}$$

is the Newtonian potential function. Thus, when the particles have position vectors $\mathbf{a}_i = (x_i, y_i, z_i)$ with respect to an inertial frame, the trajectories are locally characterized by Newton's equation

$$m_i \ddot{\mathbf{a}}_i = \frac{\partial U}{\partial \mathbf{a}_i} = \frac{m_i m_j}{r_{ij}^3} (\mathbf{a}_j - \mathbf{a}_i) + \frac{m_i m_k}{r_{ik}^3} (\mathbf{a}_k - \mathbf{a}_i), \quad \{i, j, k\} = \{1, 2, 3\}, \tag{1.2}$$

where $r_{ij} = |\mathbf{a}_j - \mathbf{a}_i|$ are the mutual distances, and (m_1, m_2, m_3), $m_i > 0$, is the given mass distribution, assumed to be normalized so that $\sum m_i = 1$. Since the above equation is of order two, a trajectory is completely determined by the initial position and velocity of the particles — often referred to as the deterministic doctrine of classical mechanics.

We use the following notation

$$I = \sum m_i |\mathbf{a}_i|^2, \quad T = \frac{1}{2} \sum m_i |\dot{\mathbf{a}}_i|^2, \quad \boldsymbol{\Omega} = \sum m_i (\mathbf{a}_i \times \dot{\mathbf{a}}_i) \tag{1.3}$$

for the (polar) moment of inertia, kinetic energy and angular momentum, respectively. These are the basic kinematic quantities, and their interactions with the potential function U play a major role in the dynamics of the 3-body problem. For example, it is easy to deduce the classical conservation laws from the system (1.2), namely the conservation of total energy

$$h = T - U, \tag{1.4}$$

linear momentum $\sum m_i \dot{\mathbf{a}}_i$ and angular momentum $\boldsymbol{\Omega}$. As usual, the invariance of linear momentum allows us to choose the inertial reference frame with the origin at the center of mass. Moreover, by differentiation of I twice with respect to time t and using (1.2), we get

$$\ddot{I} = 4T + 2 \sum \mathbf{a}_i \cdot \frac{\partial U}{\partial \mathbf{a}_i} = 4T - 2U = 2(U + 2h), \tag{1.5}$$

where we have used the fact that U is homogeneous of degree -1 as a function of the vectors $\mathbf{a}_i$. This is the Lagrange-Jacobi equation.

On the other hand, trajectories can also be determined as solutions of a suitable boundary value problem, and the simplest and most basic one is, for example,

"For a given pair of points P, Q, what are those trajectories $\gamma(t)$, $t_0 \le t \le t_1$,

with $\gamma(t_0) = P$ and $\gamma(t_1) = Q$?"

Then solutions are found by applying an appropriate least action principle, which characterizes solutions as extremals of an action integral $J(\gamma)$ among those virtual motions $\gamma(t)$ with the given pair of end points, together with some additional constraints.

Here we shall focus attention on the two least action principles due to Lagrange and Hamilton, which are quite different but dual to each other:

$$\text{Lagrange} : J_1(\gamma) = \int_\gamma T \mathrm{d}t, \qquad \text{fixed energy } h, \tag{1.6}$$

$$\text{Hamilton} : J_2(\gamma) = \int_\gamma (T + U) \mathrm{d}t, \quad \text{fixed time interval } [t_0, t_1]. \tag{1.7}$$

The motions $t \to \gamma(t)$ are regarded as parametrized curves in the Euclidean configuration space

$$M = \left\{ (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3); \sum m_i \mathbf{a}_i = 0 \right\} \simeq \mathbb{R}^6. \tag{1.8}$$

Our aim, however, is to reduce the study of 3-body trajectories to a study of associated curves in a lower dimensional space, namely the interior configuration space, i.e., the moduli space

$$\overline{M} \simeq \mathbb{R}^6 / \mathrm{SO}(3) \approx \mathbb{R}^3_+ \subset \mathbb{R}^3. \tag{1.9}$$

With the appropriate assumptions, one expects that the least action principles, as well as Newton's differential equation, can be pushed down to the level of $\overline{M}$. In fact, one of our major results is that the study of 3-body motions with vanishing angular momentum further reduces to the analysis of specific curves on the shape space, which is the sphere $S^2 \subset \mathbb{R}^3$.

The final step of our program is, of course, the lifting procedure from the moduli curve $\overline{\gamma}(t)$ back to the actual trajectory $\gamma(t)$. But this is a purely geometric construction which is well understood (cf. e.g. [5] or the following subsection) and it will not be a topic here. Briefly, the curve in $\overline{M}$ determines the trajectory in M uniquely up to a global congruence.

1.2 Riemannian geometrization and reduction

Classical mechanics up to present time is largely based upon developments related to Hamilton's least action principle, involving Hamiltonian systems and canonical transformations. Geometrically speaking, the underlying structure is the symplectic geometry of the phase space. However, in this paper, we shall rather focus on the Riemannian geometric approach, based upon Jacobi's reformulation of Lagrange's least action principle. In his famous lectures (cf. [6]), Jacobi introduced the concept of a kinematic metric

$$\mathrm{d}s^2 = 2T\mathrm{d}t^2$$

on the configuration space M, in terms of the kinetic energy T of the mechanical system. For example, in the case of an n-body system with total mass $\sum m_i = 1$

$$\mathrm{d}s^2 = 2T\mathrm{d}t^2 = \sum_i m_i(\mathrm{d}x_i^2 + \mathrm{d}y_i^2 + \mathrm{d}z_i^2). \tag{1.10}$$

Now, for a system with kinematic metric $\mathrm{d}s^2$, potential function U and a given constant total energy h, set

$$M_{(U,h)} = \{p \in M; h + U(p) \geq 0\}, \tag{1.11}$$
$$\mathrm{d}s_h^2 = (h + U)\mathrm{d}s^2,$$

where $\mathrm{d}s_h^2$ is the associated dynamical metric. Then by writing

$$\mathrm{d}s_h = \sqrt{h + U}\,\mathrm{d}s = \sqrt{T}\,\mathrm{d}s = \sqrt{2}\,T\mathrm{d}t,$$

Jacobi transformed Lagrange's action integral (1.6) into an arc-length integral

$$J_1(\gamma) = \frac{1}{\sqrt{2}} \int_\gamma \mathrm{d}s_h, \tag{1.12}$$

and hence, in one stroke, the least action principle becomes the following simple geometric statement:

> "Trajectories with total energy h are exactly those geodesic curves in the space $M_{(U,h)}$ with the dynamical metric $\mathrm{d}s_h^2$."

Nowadays, such metric spaces are called Riemannian manifolds, and the global geometric study of geodesic curves is often referred to as the Morse theory of geodesics. In particular, we note that the dynamical metric $\mathrm{d}s_h^2$ is a conformal modification of the underlying kinematic metric $\mathrm{d}s^2$.

In this geometric setting, the notion of "congruence class" is defined by the action of the rotation group SO(3), fixing the center of gravity (= origin). It acts isometrically on the configuration space $(M, \mathrm{d}s^2)$ with the kinematic metric, and also on the modified metric space $(M_{(U,h)}, \mathrm{d}s_h^2)$ for any SO(3)-invariant potential function U. The corresponding SO(3)-orbit spaces inherit the structure of a (stratified) Riemannian manifold with the induced orbital distance metric, which we denote by

$$(\overline{M}, \mathrm{d}\overline{s}^2), \quad (\overline{M}_{(U,h)}, \mathrm{d}\overline{s}_h^2), \quad \mathrm{d}\overline{s}_h^2 = (h + U)\mathrm{d}\overline{s}^2, \tag{1.13}$$

and similar to (1.11), for h negative the geodesics must stay inside the Hill's region, namely the proper subset

$$\overline{M}_h = \overline{M}_{(U,h)} = \{\overline{p} \in \overline{M}; h + U(\overline{p}) \geq 0\}. \tag{1.14}$$

By definition, the projection map

$$\pi : M \to \overline{M} = M/\mathrm{SO}(3)$$

is a (stratified) Riemannian submersion, where the horizontal tangent vectors at $p \in M$ are those perpendicular to the SO(3)-orbit. They are mapped, by the tangent map $\mathrm{d}\pi$, isometrically to the tangent space of $\overline{M}$ at $\overline{p} = \pi(p)$. Via the map π there is a 1-1 correspondence between curves $\overline{\gamma}$ in $\overline{M}$ and their horizontal lifting γ in M (resp. $M_{(U,h)}$), up to congruence. In fact, for a (virtual) motion $\gamma(t)$, the property of being horizontal is equivalent to the vanishing of the angular momentum vector $\mathbf{\Omega}$.

On the other hand, the above metric $\mathrm{d}\overline{s}^2$ on $\overline{M}$ also has a kinematic interpretation in analogy with (1.10), namely

$$\mathrm{d}\overline{s}^2 = 2\overline{T}\mathrm{d}t^2, \quad \overline{T} = T - T^\omega, \tag{1.15}$$

where T^ω is the purely rotational energy and hence the difference $\overline{T}$, representing that of the change of size and shape, is naturally the kinetic energy at the level of $\overline{M}$. Therefore, we also refer to $\mathrm{d}\overline{s}^2$ (resp. $\mathrm{d}\overline{s}_h^2$) as the kinematic (resp. dynamical) metric on $\overline{M}$. Classical mechanics, indeed, tells us how the term T^ω can be calculated from $\mathbf{\Omega}$ via the so-called inertia operator of the system; in particular, it follows that $T^\omega = 0$ if and only if $\mathbf{\Omega} = 0$.

Now, assume $\mathbf{\Omega} = 0$ and let U be a nonnegative and SO(3)-invariant function on M. Then it is not difficult to see that both action principles (1.6), (1.7) can be pushed down to $\overline{M}$. In the first case, using Jacobi's reformulation (1.12), we arrive immediately at the following geometric statement similar to the one above:

> "Curves in $\overline{M}$ representing 3-body trajectories with total energy h
>
> (and $\mathbf{\Omega} = 0$) are exactly those geodesic curves in the moduli
>
> space $\overline{M}_{(U,h)}$ with the induced dynamical metric $\mathrm{d}\overline{s}_h^2$."

In the case of (1.7), the Lagrange function $L = T + U$ is also defined at the level of $\overline{M}$. Indeed, when $\mathbf{\Omega} = 0$, we can view $\overline{M}$ as the configuration space for a simple conservative classical mechanical system, namely with potential energy $-U$, kinetic energy T, and conserved total energy $h = T - U$. It is easy to calculate the associated Euler-Lagrange equations with respect to suitable coordinates in $\overline{M}$, as demonstrated in Subsection 3.2. Finally, the reduced Newton's differential equation on $\overline{M}$ can be calculated by the procedure described in [5], but we leave this topic here.

1.3 A brief survey of the main results

The results in this paper provide the foundations for the above Riemannian geometric approach to the 3-body problem with zero angular momentum. In Section 2, we present the kinematic geometric framework for the reduction method which we shall work out, consisting of the two successive reductions $M \to \overline{M} \to M^*$, where $\overline{M} \simeq \mathbb{R}^3$ is the (congruence) moduli space and $M^* \simeq S^2$ is the shape space. The second reduction uses the cone structure of $\overline{M}$ over M^* to eliminate the scaling variable $\rho = \sqrt{I}$ by radial projection to the sphere.

A trajectory $\gamma(t)$ in the configuration space M projects to its moduli curve $\overline{\gamma}(t)$ in $\overline{M}$, and away from the base point O the curve further projects to the shape curve $\gamma^*(t)$ on the sphere. Conversely, whereas $\gamma(t)$ is determined up to congruence by $\overline{\gamma}(t)$, the real power of the above reduction method rests upon the knowledge of the subtle relationship which, in fact, generally exists between $\overline{\gamma}(t)$ and the geometric (i.e. unparametrized) curve γ^*.

In reality, the complete study of the 3-body trajectories is hereby reduced to the study of the relative geometry between the shape curve γ^* and the gradient flow of U^*, namely the Newtonian potential function restricted to the 2-sphere M^*. The major results of the paper are divided into the following four main topics:

(1) The calculation of the reduced Newton's equations in $\overline{M}$ in several ways, such as the geodesic equations of the Riemannian space $(\overline{M}_h, \mathrm{d}\overline{s}_h)$, and the reformulation of the geodesic condition in terms of the curvature of the spherical shape curve. A suitable combination of these equations also yields a separation of the radial variable ρ and hence a third order ODE on the 2-sphere which describes all shape curves for any energy level h. The key step in this reduction is Lemma 3.1 which relates the intrinsic geometry of (γ^*, U^*) to a kinematic quantity of $\overline{\gamma}(t)$.

(2) The unique parametrization theorem (cf. Theorem 4.1) asserts that the time parametrized moduli curve $\overline{\gamma}(t)$ is (essentially) determined by the oriented geometric shape curve γ^*. Furthermore, the curve γ^* is in fact uniquely determined by the first two curvature coefficients at a generic point on the curve. The basic technique used here is the local analysis of solution curves via power series expansion.

(3) The monotonicity theorem (cf. Theorem 5.1), which describes a type of piecewise monotonic behavior of the shape curve γ^*. Namely, the mass-modified latitude is a strictly monotonic function along γ^* between two succeeding local maxima or minima, and they lie on opposite

hemispheres. In particular, the curve intersects the eclipse circle at a unique point between two such local extrema.

(4) Some initial applications to the study of triple collisions. In particular, simple geometric proofs of the fundamental theorems of Sundman and Siegel. Moreover, their asymptotic formulae for the derivatives of the moment of inertia I up to order two are extended to the derivatives of any order.

The present exposition is based upon previous works of the authors (cf. [3–5]) on the 3-body problem, exploring its "sphericality" as it manifests itself in various ways. The differential equations which describe the moduli curves of the 3-body trajectories, are elaborated in Sections 3 and 4, including a careful power series analysis and comparison of the initial value problems at the moduli space and the shape space level.

Section 5 is devoted to a geometric study of the gradient field of U^*, which also yields a simple and purely geometric proof of the monotonicity theorem. A similar type of monotonicity for the shape curves was first proved by R. Montgomery [9] with his "infinitely many syzygies" theorem, and later by T. Fujiwara et al [2].

In Section 6, we recall the classical results and clarify some issues on the work of Sundman and Siegel concerning triple collisions. Moreover, with the results obtained so far, many challenging problems, for example in the study of collisions and periodic motions, naturally present themselves for an in depth study of the global geometry of shape curves. Some of these open problems will be briefly discussed in Section 7.

2 Kinematic Geometry of m-Triangles

A 3-body motion with vanishing angular momentum is always confined to a fixed plane (for purely kinematic reasons), so the motions we shall study are always planary. Therefore, we choose a plane $\mathbb{R}^2 \subset \mathbb{R}^3$ with normal vectors $\pm\mathbf{n}$ and define an m-triangle to be a triple $\delta = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ of vectors $\mathbf{a}_i \in \mathbb{R}^2$ constrained by the center of mass condition in (1.8). Hence, for our purpose we shall modify the definition (1.8) of the configuration space by taking the subspace

$$M \simeq \mathbb{R}^4 \subset \mathbb{R}^6 : \sum_{i=1}^{3} m_i \mathbf{a}_i = 0 \tag{2.1}$$

which consists of the above m-triangles in the fixed plane $\mathbb{R}^2$. M has the natural action of the rotation group SO(2), and the moduli space, representing the SO(2)-congruence classes of m-triangles, is the orbit space

$$\overline{M} = M/\mathrm{SO}(2) \simeq \mathbb{R}^4/\mathrm{SO}(2) \supset \mathbb{R}^6/\mathrm{SO}(3). \tag{2.2}$$

The degenerate (or collinear) m-triangles constitute the eclipse subvariety E of M, and we say a nondegenerate m-triangle is positively (resp. negatively) oriented if $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{n})$ is a positive (resp. negative) frame of $\mathbb{R}^3$. Accordingly, we may write M and $\overline{M}$ as the disjoint union of

three subsets

$$M = M_+ \cup E \cup M_-, \quad \overline{M} = \overline{M}_+ \cup \overline{E} \cup \overline{M}_-. \tag{2.3}$$

Moreover, we observe that the moduli space of unoriented m-triangles would be

$$\overline{M}/\mathbb{Z}_2 \simeq \mathbb{R}^4/O(2) \simeq \overline{M}_\pm \cup \overline{E} \simeq \mathbb{R}^6/SO(3) = \mathbb{R}^6/O(3).$$

Remark 2.1 For a study of general (non-planary) 3-body motions, the natural configuration space $\widehat{M}$ consists of pairs $(\delta, \mathbf{n})$, where δ is an m-triangle in $\mathbb{R}^3$ and $\mathbf{n}$ is a unit vector perpendicular to all $\mathbf{a}_i$. $\widehat{M}$ is a 6-dimensional manifold (a 4-plane bundle over S^2) with the natural action of $SO(3)$, and now the moduli space coincides with the above one (cf. [5, Section 2]), namely

$$\widehat{M}/SO(3) \simeq M/SO(2) = \overline{M}.$$

We shall describe in more detail the topology and induced Riemannian structure of the above simple orbit spaces. Let δ (as above) and $\delta' = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ be m-triangles. The following $SO(2)$-invariant, but mass dependent inner product

$$\langle \delta, \delta' \rangle = \sum m_i \mathbf{a}_i \cdot \mathbf{b}_i \tag{2.4}$$

is just the kinematic metric $\mathrm{d}s^2$ of M defined by (1.10). In particular, the squared norm is the moment of inertia, $|\delta|^2 = I = \rho^2$. Let $S^3 \subset M$ be the unit sphere ($\rho = 1$) and denote its spherical metric by $\mathrm{d}u^2$. Then we can express M as the Riemannian cone over $(S^3, \mathrm{d}u^2)$

$$M = \mathbb{R}^4 = C(S^3) : \mathrm{d}s^2 = \mathrm{d}\rho^2 + \rho^2 \mathrm{d}u^2, \tag{2.5}$$

where ρ measures the distance from the base point (origin) of the cone.

A description similar to (2.5) applies to the moduli space

$$(\overline{M}, \mathrm{d}\overline{s}^2) = (M, \mathrm{d}s^2)/SO(2) \tag{2.6}$$

whose "unit sphere" $M^* = (\rho = 1)$ is called the shape space, namely it is the image of S^3 in $\overline{M}$

$$(M^*, \mathrm{d}\sigma^2) = (S^3, \mathrm{d}u^2)/SO(2), \tag{2.7}$$

with the induced metric denoted by $\mathrm{d}\sigma^2$. Thus, $\overline{M}$ also inherits the structure of a Riemannian cone over its "unit sphere"

$$\overline{M} = C(S^3)/SO(2) = C(S^3/SO(2)) = C(M^*) : \mathrm{d}\overline{s}^2 = \mathrm{d}\rho^2 + \rho^2 \mathrm{d}\sigma^2 \tag{2.8}$$

with ρ still measuring the distance from the base point O. The shape space is actually isometric to the 2-sphere of radius $\frac{1}{2}$, as follows from the well-known Hopf fibration construction

$$S^3 \rightarrow M^* = S^3/SO(2) = \mathbb{C}P^1 \simeq S^2\Big(\frac{1}{2}\Big). \tag{2.9}$$

As a cone over S^2, $\overline{M}$ is clearly homeomorphic to $\mathbb{R}^3$ with the origin at the base point O. Away from O they are even diffeomorphic, when $\overline{M}$ has the induced smooth functional structure as an orbit space of $\mathbb{R}^4$.

For the convenience of applying vector algebra, we also recall the Euclidean model of $\overline{M}$, where $\overline{M}$ is identified with $\mathbb{R}^3$, with Euclidean coordinates (x, y, z) and associated spherical coordinates (I, φ, θ), and the kinematic metric is expressed as the following conformal modification of the standard Euclidean metric:

$$d\overline{s}^2 = \frac{dx^2 + dy^2 + dz^2}{4\sqrt{x^2 + y^2 + z^2}} = d\rho^2 + \frac{\rho^2}{4}(d\varphi^2 + \sin^2 \varphi \, d\theta^2), \tag{2.10}$$

$$I^2 = \rho^4 = x^2 + y^2 + z^2, \quad 0 \le \varphi \le \pi, \ 0 \le \theta \le 2\pi.$$

Here (φ, θ) denotes any choice of spherical polar coordinates on the sphere

$$M^* = S^2 : x^2 + y^2 + z^2 = 1,$$

whose induced metric from the Euclidean 3-space is that of the round sphere of radius 1

$$S^2(1) : ds^2 = d\varphi^2 + \sin^2 \varphi \, d\theta^2, \tag{2.11}$$

whereas its induced (i.e. kinematic) metric as a submanifold of $(\overline{M}, d\overline{s}^2)$ is

$$d\sigma^2 = \frac{1}{4}ds^2. \tag{2.12}$$

By (1.15) and (2.10) the total kinetic energy can be written as

$$T = \overline{T} + T^\omega = \frac{1}{2}\dot{\rho}^2 + \frac{\rho^2}{8}(\dot{\varphi}^2 + (\sin^2 \varphi)\dot{\theta}^2) + T^\omega, \tag{2.13}$$

where the rotational term T^ω vanishes precisely when $\boldsymbol{\Omega} = 0$. Starting from Section 3 this is our standing assumption.

Remark 2.2 The radial distance function $\rho = \sqrt{I}$ is also referred to as the hyper-radius in the physics literature. For our purpose it is generally more convenient to use ρ rather than I as the scaling parameter, and we shall refer to (ρ, φ, θ) as a spherical coordinate system of $\overline{M}$. We refer to [5] for the relationships between spherical coordinates, individual moments of inertia (I_1, I_2, I_3), or mutual distances (r_{12}, r_{23}, r_{31}).

In the above Euclidean model of kinematic geometry the decomposition in (2.3) has a distinguished equator plane, namely the eclipse plane $\overline{E}$ which divides $\mathbb{R}^3$ into the two half-spaces $\overline{M}_\pm = \mathbb{R}^3_\pm$. We choose the Euclidean coordinates so that $\overline{E}$ is the xy-plane and the half-space $z > 0$ represents the congruence classes $\overline{\delta}$ of the positively oriented m-triangles. Finally,

$$E^* = \overline{E} \cap M^* : \ x^2 + y^2 = 1, \quad z = 0$$

is the distinguished equator or eclipse circle of the sphere $M^* = S^2$.

On the other hand, the position of the various shapes δ^* of m-triangles on the sphere is uniquely determined by the position of the three binary collision points $\mathbf{b}_i$, $i = 1, 2, 3$, along the circle E^*, where $\mathbf{b}_1$ represents the shape of the degenerate m-triangle with $\mathbf{a}_2 = \mathbf{a}_3$ etc. We are still free to choose the cyclic ordering $\mathbf{b}_1 \to \mathbf{b}_2 \to \mathbf{b}_3$ either in the eastward or westward

direction. Moreover, the mass distribution and the relative positions of the three points $\mathbf{b}_i$ mutually determine each other. In fact, the angle β_1 between $\mathbf{b}_2$ and $\mathbf{b}_3$ is given by

$$\cos \beta_1 = \frac{m_2 m_3 - m_1}{m_2 m_3 + m_1}, \quad \text{etc.,} \tag{2.14}$$

and these formulae can be inverted.

Let $\{N, S\}$ be the north and south pole of S^2. It is often convenient to choose the spherical coordinate system (φ, θ) centered at N, namely $\varphi = 0$ at the pole N. Then the poles represent the m-triangles (congruent, but with opposite orientation) of maximal area for a fixed size ρ, and more generally, the area of an m-triangle is given by the formula

$$\Delta = \frac{\rho^2}{4\sqrt{m_1 m_2 m_3}} |\cos \varphi|. \tag{2.15}$$

For a normalized (i.e. $\rho = 1$) m-triangle of shape $\mathbf{p} \in S^2(1)$, we also recall the formula for the mutual distances (cf. (1.2))

$$r_{ij} = \frac{1}{2} \sqrt{\frac{1 - m_k}{m_i m_j}} |\mathbf{p} - \mathbf{b}_k| = \sqrt{\frac{1 - m_k}{m_i m_j}} \sin \sigma_k, \tag{2.16}$$

where $|\mathbf{p} - \mathbf{b}_k|$ (resp. $2\sigma_k$) is the Euclidean distance (resp. angle) between $\mathbf{p}$ and $\mathbf{b}_k$. For the proofs of (2.14)–(2.16), we refer to [4] or [5].

Remark 2.3 In this paper, we use both the kinematic and Euclidean model $S^2(r)$, $r = \frac{1}{2}$ or 1, of the shape space M^*. Their arc-length parameters are σ and $s = 2\sigma$, respectively (cf. (2.12)). Of course, the various geometric quantities, such as velocity, geodesic curvature, gradient etc., must also be scaled appropriately when passing from one model to the other.

3 Analysis on the Moduli Curves of 3-Body Motions with Zero Angular Momentum

In this chapter, we shall follow Jacobi's geometrization idea at the level of the moduli space $\overline{M} = \mathbb{R}^3$. This enables us to reduce the analysis of 3-body trajectories with zero angular momentum to that of the corresponding moduli curves. According to Jacobi, for a given energy level h the moduli curves can be interpreted as the geodesics of a specific Riemannian metric $\mathrm{d}\overline{s}_h^2$ on $\overline{M}$. Now, the standard procedure for the calculation of the geodesic equations amounts to the calculation of the Christoffel symbols of the metric, with respect to a suitably chosen coordinate system suggested by the geometry of the space, say. The resulting equations are ordinary differential equations whose solutions are curves parametrized by the arc-length.

The kinematic geometry describes $\overline{M}$ with the scaling and rotational symmetry of a Riemannian cone over a sphere, and therefore the spherical coordinates ρ, φ, θ present themselves as the most natural choice. In Subsection 3.1, we shall calculate the associated differential equations. However, the natural parameter for 3-body trajectories is the physical time t, and it is the effective usage of fixed energy that enables us to express the equations in terms of t as well (cf. (3.3)).

On the other hand, in the Hamiltonian least action principle the time interval is fixed and t is the natural parameter from the outset. The Euler-Lagrange equations for the Lagrange function L on $\overline{M}$ are calculated in Subsection 3.2, and this approach, in fact, yields the same system (3.3) in a much simpler way.

However, we also seek a differential equation purely at the shape space level, that is, with the scaling parameter ρ eliminated. This demands a deeper understanding of the relationship between ρ and the geodesic curvature of the shape curve. To this end we shall introduce an alternative geometric approach to the study of geodesics in $\overline{M}$, which takes the full advantage of the spherical symmetry and the cone structure of $\overline{M}$. This is the topic of Subsection 3.3.

Finally, in Subsection 3.4 we shall synthesize the results obtained in the previous subsections and, in particular, we explain how the moduli curve can be reconstructed from the shape curve.

3.1 Calculation of the standard geodesic equations in $\overline{M}_h$

We shall calculate the (standard) geodesic equations of $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ relative to the spherical coordinates (ρ, φ, θ), where (φ, θ) denotes a choice of spherical polar coordinates on the shape space $M^* = S^2$. The homogeneity of the Newtonian potential function (1.1) allows us to write

$$U = \frac{1}{\rho}U^*(\varphi, \theta), \tag{3.1}$$

where U^* is the shape potential function on S^2. By (1.13) and (2.10), the metric with the arc-length element $\overline{s}_h$ is

$$\mathrm{d}\overline{s}_h^2 = \Big(\frac{U^*}{\rho} + h\Big)\mathrm{d}\overline{s}^2 = \Big(\frac{U^*}{\rho} + h\Big)\Big(\mathrm{d}\rho^2 + \frac{\rho^2}{4}(\mathrm{d}\varphi^2 + \sin^2\varphi\,\mathrm{d}\theta^2)\Big), \tag{3.2}$$

and the standard procedure for the calculation of the geodesic equations via the Christoffel symbols, but expressed with respect to time t as the independent variable, yields the following system (cf. [13, (5.11)])

$$
\begin{aligned}
&(\,\mathrm{i}\,)\ \ 0 = \ddot{\rho} + \frac{\dot{\rho}^2}{\rho} - \frac{1}{\rho}\Big(\frac{1}{\rho}U^* + 2h\Big),\\[4pt]
&(\mathrm{ii})\ \ 0 = \ddot{\varphi} + 2\frac{\dot{\rho}}{\rho}\dot{\varphi} - \frac{1}{2}\sin(2\varphi)\dot{\theta}^2 - \frac{4}{\rho^3}U_\varphi^*,\\[4pt]
&(\mathrm{iii})\ \ 0 = \ddot{\theta} + 2\frac{\dot{\rho}}{\rho}\dot{\theta} + 2\cot(\varphi)\dot{\varphi}\dot{\theta} - \frac{4}{\rho^3}\frac{1}{\sin^2\varphi}U_\theta^*,
\end{aligned}
\tag{3.3}
$$

where the equation (3.3)(i) is just the Lagrange-Jacobi equation (1.5).

The calculation of the above system goes as follows. For simplicity, let us write

$$f^2 = U + h = T, \quad u = \overline{s}_h,$$

and $\frac{\mathrm{d}\xi}{\mathrm{d}u} = \xi'$, $\frac{\mathrm{d}\xi}{\mathrm{d}t} = \dot{\xi}$ for any function ξ. Then by (1.15) and (3.2)

$$\mathrm{d}u^2 = T\mathrm{d}\overline{s}^2 = 2T^2\mathrm{d}t^2 = 2f^4\mathrm{d}t^2,$$

and we deduce the useful identities relating arc-length and time derivatives

$$\xi' = \frac{\dot{\xi}}{\dot{u}}, \quad \xi'' = \frac{\dot{u}\ddot{\xi} - \ddot{u}\dot{\xi}}{\dot{u}^3}, \quad \frac{\ddot{u}}{\dot{u}} = 2\frac{\dot{f}}{f}, \tag{3.4}$$

$$2f^2 = \dot{\rho}^2 + \frac{\rho^2}{4}(\dot{\varphi}^2 + (\sin^2 \varphi)\dot{\theta}^2), \quad 2f^4 = \dot{u}^2. \tag{3.5}$$

Now, the first step is to calculate the geodesic equations with u as the independent variable, following the standard procedure, and the next step is to transform the equations by changing to t as the independent variable, using the identities in (3.4) and (3.5). With the notation $(x_1, x_2, x_3) = (\rho, \varphi, \theta)$ the equations in the first step can be stated as

$$\frac{\mathrm{d}^2 x_k}{\mathrm{d}u^2} + \sum_{i,j=1}^{3} \Gamma_{ij}^k \frac{\mathrm{d}x_i}{\mathrm{d}u} \frac{\mathrm{d}x_j}{\mathrm{d}u} = 0, \quad k = 1, 2, 3, \tag{3.6}$$

where the Christoffel symbols Γ_{ij}^k are defined by

$$\Gamma_{ij}^k = \frac{1}{2} \sum_m \left(\frac{\partial g_{jm}}{\partial x_i} + \frac{\partial g_{mi}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_m} \right) g^{km}.$$

Here, (g^{ij}) is the inverse of the matrix (g_{ij}) representing the metric $\mathrm{d}u^2$, in the sense that

$$\mathrm{d}u^2 = f^2 \left(\mathrm{d}\rho^2 + \frac{\rho^2}{4}(\mathrm{d}\varphi^2 + \sin^2 \varphi \, \mathrm{d}\theta^2) \right) = \sum g_{ij}\mathrm{d}x_i\mathrm{d}x_j.$$

Thus the matrices are diagonal and

$$g_{11} = f^2, \quad g_{22} = \frac{1}{4}f^2 x_1^2, \quad g_{33} = \frac{1}{4}f^2 x_1^2 \sin^2 x_2, \quad g^{ii} = g_{ii}^{-1}.$$

The first geodesic equation follows from (3.6) with $k = 1$, so by calculating the symbols Γ_{ij}^1 we obtain

$$\rho'' - \frac{U^*}{2f^2\rho^2}\rho'^2 + \left(\frac{U^*}{2f^2} - \rho \right)\frac{1}{4}(\varphi'^2 + (\sin^2 \varphi)\theta'^2) + \frac{\rho'}{f^2\rho}(U_\varphi^* \varphi' + U_\theta^* \theta') = 0.$$

Then, in the second step a straightforward calculation using (3.4) and (3.5) yields

$$\ddot{\rho} + \frac{\dot{\rho}^2}{\rho} + \left\{ -\frac{\dot{\rho}^2 U^*}{f^2\rho^2} - 2\frac{f^2}{\rho} + \frac{U^*}{\rho^2} + \left(\frac{\dot{U}^*}{f^2\rho} - 2\frac{\dot{f}}{f} \right)\dot{\rho} \right\} = 0,$$

where the bracket expression simplifies to $-\rho^{-1}(U + 2h)$. The final result is the first equation of (3.3).

Next, let us consider the case $k = 2$ where the first step yields the equation

$$0 = \varphi'' - \frac{2U_\varphi^*}{f^2\rho^3}\left(\rho'^2 + \frac{\rho^2}{4}(\varphi'^2 + (\sin^2 \varphi)\theta'^2) \right) + \frac{U_\varphi^*}{f^2\rho}\varphi'^2 - \frac{1}{2}(\sin 2\varphi)\theta'^2$$

$$+ \frac{1}{f^2}\left(-\frac{U^*}{\rho^2} + \frac{2f^2}{\rho} \right)\rho'\varphi' + \frac{U_\theta^*}{f^2\rho}\varphi'\theta'.$$

The second step leads to the equation

$$0 = \ddot{\varphi} + 2\frac{\dot{\rho}}{\rho}\dot{\varphi} - \frac{1}{2}\sin(2\varphi)\dot{\theta}^2 - \frac{4}{\rho^3}U_\varphi^* + \left\{ -2\frac{\dot{f}}{f}\dot{\varphi} + \frac{U_\varphi^*}{f^2\rho}\dot{\varphi}^2 - \frac{U^*}{f^2\rho^2}\dot{\rho}\dot{\varphi} + \frac{U_\theta^*}{f^2\rho}\dot{\varphi}\dot{\theta} \right\},$$

where the bracket expression simplifies to

$$\frac{\dot{\varphi}}{f^2}\left(\frac{\dot{U}^*}{\rho} - \frac{U^*}{\rho^2}\dot{\rho} - \frac{d}{dt}f^2\right) = 0.$$

This yields the second equation of (3.3). The last case $k = 3$ is similar to the previous one, so we have omitted the calculations.

3.2 An alternative derivation of the standard geodesic equations

Another way of deriving the system (3.3) is to calculate the Euler-Lagrange equations for Hamilton's least action principle (1.7), at the level of the moduli space $\overline{M}$. In fact, the Lagrange function $L = T + U$ descends to a function defined on the tangent bundle $T\overline{M}$ since $\boldsymbol{\Omega} = 0$, and

$$T = \frac{1}{2}\left|\frac{d\overline{\gamma}}{dt}\right|^2 = \frac{1}{2}\left(\frac{d\overline{s}}{dt}\right)^2$$

is actually the kinetic energy of a curve $\overline{\gamma}(t)$ in $\overline{M}$. Thus we have a simple classical conservative mechanical system with potential energy $-U$, kinetic energy T, and conserved total energy $h = T - U$. Therefore, in terms of the coordinates (ρ, φ, θ) the associated Lagrange system is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\rho}} = \frac{\partial L}{\partial \rho}, \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{\varphi}} = \frac{\partial L}{\partial \varphi}, \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} = \frac{\partial L}{\partial \theta}, \tag{3.7}$$

where by (2.10)

$$L = T + U = \left(\frac{1}{2}\dot{\rho}^2 + \frac{\rho^2}{8}(\dot{\varphi}^2 + \sin^2\varphi\,\dot{\theta}^2)\right) + \frac{1}{\rho}U^*(\varphi, \theta).$$

Now, the first equation of (3.7) reads

$$\frac{d}{dt}\dot{\rho} = \frac{\rho}{4}(\dot{\varphi}^2 + \sin^2\varphi\,\dot{\theta}^2) - \frac{1}{\rho^2}U^*, \tag{3.8}$$

and by substituting the expression

$$\frac{1}{4}(\dot{\varphi}^2 + \sin^2\varphi\,\dot{\theta}^2) = \frac{1}{\rho^2}(2T - \dot{\rho}^2) = \frac{1}{\rho^2}(2U + 2h - \dot{\rho}^2)$$

into (3.8) the equation transforms to the equation (3.3)(i).

Next, the second equation of (3.7) reads

$$\frac{d}{dt}\left(\frac{\rho^2}{4}\dot{\varphi}\right) = \frac{\rho^2}{4}\sin\varphi\cos\varphi\,\dot{\theta}^2 + \frac{1}{\rho}U^*_\varphi,$$

and after differentiation this becomes

$$\rho^2\ddot{\varphi} + 2\rho\dot{\rho}\dot{\varphi} - \frac{\rho^2}{2}\sin 2\varphi\,\dot{\theta}^2 - \frac{4}{\rho}U^*_\varphi = 0,$$

which is precisely the equation (3.3)(ii). Similarly, the third equation of (3.7) reads

$$\frac{d}{dt}(\rho^2\sin^2\varphi\,\dot{\theta}) = \frac{4}{\rho}U^*_\theta, \tag{3.9}$$

where the left side becomes

$$2\rho\dot{\rho}\sin^2\varphi\,\dot{\theta} + 2\rho^2\sin\varphi\cos\varphi\,\dot{\varphi}\dot{\theta} + \rho^2\sin^2\varphi\,\ddot{\theta}.$$

Then it is easily seen that the equation (3.9) becomes the equation (3.3)(iii).

3.3 Cone surfaces and geodesics in $\overline{M}_h$

To take full advantage of the cone structure of $\overline{M}$ and the scaling property (3.1) of the potential function U, one naturally seeks to reduce the analysis of the moduli curve to that of the shape curve. Therefore, in this subsection we shall study the geodesic equations with the shape curve in the forefront.

Definition 3.1 *Let $\overline{\gamma}$ be a curve in $\overline{M}$ not including the (triple collision) base point O. The cone surface spanned by $\overline{\gamma}$ consists of all rays emanating from O and intersecting $\overline{\gamma}$. The cone surface is denoted by $C(\overline{\gamma})$ or $C(\gamma^*)$.*

The intersection of the cone surface with the shape space M^* is the associated shape curve γ^*, and conversely, the cone surface is also uniquely determined by γ^*, which explains the notation $C(\gamma^*)$.

Let $\overline{\gamma}(t)$ be the moduli curve of a 3-body motion with zero angular momentum and $\gamma^*(t)$ the associated shape curve with arc-length parameter σ in $M^* \simeq S^2(\frac{1}{2})$. Endowed with the kinematic metric, the cone surface

$$C(\gamma^*) : \mathrm{d}\overline{s}^2 = \mathrm{d}\rho^2 + \rho^2 \mathrm{d}\sigma^2, \quad 0 \le \sigma \le \sigma_1 \tag{3.10}$$

is isometric to a (flat) Euclidean sector with polar coordinates (ρ, σ) and angular width σ_1.

Next, let the function $u(\sigma)$ denote the restriction of U^* along γ^*. There is a pair of conformally related metrics on the above sector, namely the flat metric (3.10) and the dynamical metric

$$\mathrm{d}\overline{s}_h^2 = \left(h + \frac{u(\sigma)}{\rho}\right)\mathrm{d}\overline{s}^2 \tag{3.11}$$

defined on the subregion

$$C(\gamma^*)_h = \left\{(\rho, \sigma); \rho \ge 0, 0 \le \sigma \le \sigma_1, h + \frac{u(\sigma)}{\rho} > 0\right\}. \tag{3.12}$$

On the one hand, $\overline{\gamma}$ is a geodesic curve of the surface $C(\gamma^*)_h$ with the metric (3.11), but on the other hand, it is also a geodesic of the ambient space $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ and hence the normal component of the curvature vector of $\overline{\gamma}$ in $\overline{M}_h$ also vanishes. Accordingly, it is natural to write the geodesic equations of $\overline{\gamma}$ in $\overline{M}_h$ as a pair of coupled ODEs expressing, respectively, the vanishing of the tangential and normal component of the curvature vector.

To analyze curvatures, let us fix some convention concerning orientation. We assume $\overline{\gamma}$ (and hence also γ^*) is oriented; they are curves in $\overline{M} \simeq \mathbb{R}^3$ and $M^* \simeq S^2$ respectively, and these spaces have their standard orientation. We choose a positive orthonormal frame (τ, η, ν) of $(\overline{M}, \mathrm{d}\overline{s}^2)$ along $\overline{\gamma}$, as follows. Let τ (resp. τ^*) be the positive unit tangent field of $\overline{\gamma}$ (resp. γ^*), and choose the unit normal field ν^* of γ^* so that (τ^*, ν^*) is a positive frame of M^*. Then $\nu = (\frac{1}{\rho})\nu^*$ is a unit normal field of the cone surface $C(\overline{\gamma})$ and hence orients the surface. Finally, η is the normal field of $\overline{\gamma}$ in $C(\overline{\gamma})$. For convenience, we write

$$\tau = \cos\alpha \frac{\partial}{\partial\rho} + \sin\alpha \frac{1}{\rho}\frac{\partial}{\partial\sigma}, \quad \eta = -\sin\alpha \frac{\partial}{\partial\rho} + \cos\alpha \frac{1}{\rho}\frac{\partial}{\partial\sigma}, \tag{3.13}$$

where α denotes the angle between the radial and tangential direction, that is, the angle between $\frac{\partial}{\partial \rho}$ and τ.

Let $\mathbf{n}$ be a vector normal to $\overline{\gamma}$, say η or ν as above, and let $\mathcal{K}(\mathbf{n})$ and $\mathcal{K}_h(\mathbf{n})$ denote the associated geodesic curvatures of $\overline{\gamma}$ with respect to the metrics $d\overline{s}^2$ and $d\overline{s}_h^2$, respectively. It follows from the first variation formula of arc-length that the curvatures of a given curve with respect to such a pair of conformally related metrics are linked by the following formula

$$\mathcal{K}_h(\mathbf{n}) = \mathcal{K}(\mathbf{n}) - \frac{1}{2}\frac{d}{d\mathbf{n}}\ln(h+U). \tag{3.14}$$

There are two cases to analyze, namely $\mathbf{n} = \eta$ or ν, and for simplicity we write $\mathcal{K} = \mathcal{K}(\eta)$, $\mathcal{K}^\perp = \mathcal{K}(\nu)$ etc., and we assume γ^* is not a single point.

First, the geodesic curvatures of $\overline{\gamma}$ in the surface $C(\gamma^*)$ with the two metrics are related by

$$\mathcal{K}_h = \mathcal{K} - \frac{1}{2}\frac{d}{d\eta}\ln(h+U) = \left(\frac{d\alpha}{d\overline{s}} + \frac{d\sigma}{d\overline{s}}\right) - \frac{1}{2}\frac{d}{d\eta}\ln\left(h + \frac{u(\sigma)}{\rho}\right), \tag{3.15}$$

where we recognize $(\alpha + \sigma)$ as the angle between the tangent line and a fixed reference ray in the Euclidean sector (3.10). Similarly, let $\mathcal{K}^\perp$ and $\mathcal{K}_h^\perp$ be the "surface normal" geodesic curvatures of $\overline{\gamma}$ with respect to the metrics $d\overline{s}^2$ and $d\overline{s}_h^2$, consequently

$$\mathcal{K}_h^\perp = \mathcal{K}^\perp - \frac{1}{2}\frac{d}{d\nu}\ln\left(h + \frac{U^*}{\rho}\right). \tag{3.16}$$

To find an expression for $\mathcal{K}^\perp$, it is a key observation that the principal curvatures of $C(\gamma^*)$ in $\overline{M}$ at a given point (ρ,σ) are the numbers 0 and $\frac{1}{\rho}\mathcal{K}^*$, where $\mathcal{K}^* = \mathcal{K}^*(\sigma)$ denotes the geodesic curvature of γ^* on the sphere $M^* \simeq S^2(\frac{1}{2})$ at the point $\gamma^*(\sigma)$. Therefore, by the classical Euler's formulas, the normal sectional curvature of $C(\gamma^*)$ in the tangential direction of $\overline{\gamma}$ is equal to

$$\mathcal{K}^\perp = \frac{1}{\rho}(\sin^2\alpha)\mathcal{K}^*(\sigma). \tag{3.17}$$

Finally, the geodesic condition for $\overline{\gamma}$ in $\overline{M}_h$ reads

$$\mathcal{K}_h = \mathcal{K}_h^\perp = 0,$$

which by (3.15)–(3.17) is neatly expressed by two scalar ODEs. In terms of the kinematic arc-length parameter $d\overline{s}$ of $\overline{\gamma}$, regarded either as a curve in $C(\gamma^*)$ or $\overline{M}$, we can state the final result as the following theorem.

Theorem 3.1 *Let $\overline{\gamma}(t)$ (resp. $\gamma^*(t)$) be the moduli (resp. shape) curve of a given 3-body motion with zero angular momentum and total energy h. Set σ to be the arc-length parameter of γ^* in the shape space $M^* \simeq S^2(\frac{1}{2})$, (ρ,σ) the polar coordinate system of the associated cone surface $C(\gamma^*)$, and $u(\sigma)$ the restriction of U^* along γ^*. If $\overline{\gamma}$ is not a ray solution, then it is characterized by the following pair of equations*

$$(\,i\,) \quad \frac{d\alpha}{d\overline{s}} + \frac{d\sigma}{d\overline{s}} - \frac{1}{2}\left(-\sin\alpha\frac{\partial}{\partial\rho} + \frac{\cos\alpha}{\rho}\frac{\partial}{\partial\sigma}\right)\ln\left(h + \frac{u(\sigma)}{\rho}\right) = 0, \tag{3.18}$$

$$(\,ii\,) \quad (\sin^2\alpha)\mathcal{K}^* - \frac{1}{2}\frac{\partial}{\partial\nu^*}\ln\left(h + \frac{U^*}{\rho}\right) = 0, \tag{3.19}$$

where α is the angle between the radial direction and the tangential direction and ν^* is the positive unit normal vector field of γ^* in M^*.

Remark 3.1 The exceptional case of ray solutions, that is, γ^* is a single point, can be settled directly from (3.14), where $\mathcal{K}(\mathbf{n}) = \mathcal{K}_h(\mathbf{n}) = 0$ and hence for each $\mathbf{n}$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{n}} \ln(h + U) = 0 \quad \text{or} \quad \nabla U \cdot \mathbf{n} = 0.$$

This is equivalent to $\nabla U^* = 0$, so γ^* is one of the five critical points of U^* on the 2-sphere, namely the two minima (called Lagrange points) and the three saddle points (called Euler points) lying on the equator (or eclipse) circle.

We also deduce the above result from the equations (3.3)(ii), (3.3)(iii), namely U^*_φ and U^*_θ must vanish if φ and θ are set to be constant. Ray solutions yield the simplest type of 3-body motions, namely the shape invariant or so-called homographic motions, which in the case of $\mathbf{\Omega} = 0$ are confined to a line. Then the only variable $\rho(t)$ is the solution of the 1-dimensional Kepler problem given by the Lagrange-Jacobi equation (i.e. (3.3)(i)).

Remark 3.2 (i) For a given value of total energy h, the influence of U^* on the geometry of the associated cone surface $C(\gamma^*)_h$ is via the function $u = U^*|_{\gamma^*}$, and the equation (3.18) is exactly the geodesic equation of $C(\gamma^*)_h$. Most of the geodesic curves of $C(\gamma^*)_h$ are, of course, not geodesics of the ambient space $\overline{M}_h$ since they are not moduli curves of actual 3-body motions.

(ii) Equation (3.19), on the other hand, is expressed in terms of the relative geometry of the inclusion $\gamma^* \subset M^*$, namely the geodesic curvature $\mathcal{K}^*$ and the normal derivative of U^* along γ^*.

Finally, recall from (3.13) that the scaling variable ρ and the angular variable α, which measures the radial inclination of the moduli curve in the cone $\overline{M}$, essentially determine each other via the relations

$$\cos\alpha = \frac{\mathrm{d}\rho}{\mathrm{d}\overline{s}}, \quad \sin\alpha = \rho\frac{\mathrm{d}\sigma}{\mathrm{d}\overline{s}}, \tag{3.20}$$

which tell us, for example, how to calculate ρ from the variation of α along the shape curve $\gamma^* \subset S^2(\frac{1}{2})$:

$$\rho(\sigma) = \rho(\sigma_0)e^{\int_{\sigma_0}^{\sigma} \cot\alpha(\sigma)\mathrm{d}\sigma} \quad \text{for } \rho(\sigma_0) \neq 0. \tag{3.21}$$

3.4 Synthesis of the analysis of the moduli curve and that of the shape curve

In the previous sections we have used geometric ideas to obtain differential equations in the moduli space $\overline{M} \simeq \mathbb{R}^3$ characterizing 3-body trajectories with zero angular momentum. Since $\overline{M}$ is a cone over the 2-sphere M^* defined by $\rho = 1$, it is natural to project the moduli curve $\overline{\gamma}$ down to its image curve γ^* on the sphere. However, unless one resolves the hidden interlocking between γ^* and the scaling variable ρ, implicitly described by the differential equations, one cannot reconstruct the moduli curve from its shape curve and thus fully utilize the reduction from $\overline{M}$ to M^*.

In order to separate the scaling variable ρ from the spherical variables (φ, θ), we shall proceed by combining the two systems of geodesic equations, (3.3) and (3.18)–(3.19), which

we derived in two different ways. Note that the natural parameter in mechanics is the time t, whereas the arc-length parameter is the natural parameter in metric geometry. This suggests a transformation of the latter system to equations with t as the independent variable, and as it turns out, this also provides a remarkable simple solution of the above separation problem.

3.4.1 Basic geometry of curves on the 2-sphere with a potential function

Since spherical curves play a crucial role in the present study, it is convenient to collect some basic formulae concerning the differential geometry of curves on $S^2(1)$, as well as the tangential and normal derivatives of a given (potential) function U^* on the sphere. We shall express them in terms of a chosen spherical polar coordinate system (φ, θ).

For a given oriented curve γ^*, let τ^* (resp. ν^*) be the unit tangent vector in the positive direction (resp. unit normal vector) such that (τ^*, ν^*) is a positively oriented frame of the sphere. We consider a (regular) time parametrized curve $\gamma^*(t) = (\varphi(t), \theta(t))$ and set $s = s(t) \geq 0$ to be the arc-length along the curve. As before, differentiation of a function f with respect to t or s are denoted by $\dot{f}$ and f' respectively, and clearly $\dot{f} = f'v$ where $v = \dot{s}$ is the speed of the curve. Then

$$\tau^* = \frac{\mathrm{d}\gamma^*}{\mathrm{d}s} = \frac{1}{v}\Big(\dot{\varphi}\frac{\partial}{\partial\varphi} + \dot{\theta}\frac{\partial}{\partial\theta}\Big), \quad \nu^* = \frac{1}{v}\Big(-\dot{\theta}\sin\varphi\frac{\partial}{\partial\varphi} + \dot{\varphi}\frac{1}{\sin\varphi}\frac{\partial}{\partial\theta}\Big), \tag{3.22}$$

and the velocity vector field of the curve is

$$\frac{\mathrm{d}\gamma^*}{\mathrm{d}t} = v\tau^*, \quad v = \sqrt{\dot{\varphi}^2 + (\sin^2\varphi)\dot{\theta}^2}. \tag{3.23}$$

The scalar acceleration

$$\dot{v} = \frac{\mathrm{d}}{\mathrm{d}t}v = \frac{1}{v}[\dot{\varphi}\ddot{\varphi} + (\sin\varphi\cos\varphi)\dot{\varphi}\dot{\theta}^2 + \sin^2(\varphi)\dot{\theta}\ddot{\theta}], \tag{3.24}$$

and its higher time derivatives are needed to express time derivatives of a function in terms of arc-length derivatives, using operators of increasing order

$$\frac{\mathrm{d}}{\mathrm{d}t} = v\frac{\mathrm{d}}{\mathrm{d}s}, \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2} = v^2\frac{\mathrm{d}^2}{\mathrm{d}s^2} + \dot{v}\frac{\mathrm{d}}{\mathrm{d}s}, \quad \text{etc.} \tag{3.25}$$

To calculate the geodesic curvature function K^*, let us first make use of Euclidean coordinates

$$x = \sin\varphi\cos\theta, \quad y = \sin\varphi\sin\theta, \quad z = \cos\varphi,$$

and write $\mathbf{x}(s) = (x(s), y(s), z(s))$ and use the formula

$$K^*(s) = \mathbf{x}(s) \times \mathbf{x}'(s) \cdot \mathbf{x}''(s),$$

where $\mathbf{x}'(s) = \tau^*$. This yields

$$K^* = (\cos\varphi)\theta'(1 + \varphi'^2) + \sin\varphi(\varphi'\theta'' - \theta'\varphi'')$$
$$= \frac{1}{v^3}\{(\cos\varphi)\dot{\theta}(v^2 + \dot{\varphi}^2) + \sin\varphi(\dot{\varphi}\ddot{\theta} - \dot{\theta}\ddot{\varphi})\}, \tag{3.26}$$

and its intrinsic first derivative is

$$K^{*\prime} = \frac{\mathrm{d}}{\mathrm{d}s}K^* = (-(\sin\varphi)\varphi'\theta' + (\cos\varphi)\theta'')(1 + \varphi'^2)$$
$$+ (\cos\varphi)\varphi'(\varphi'\theta'' - \theta'\varphi'') + \sin\varphi(\varphi'\theta''' - \theta'\varphi'''). \tag{3.27}$$

The gradient field of U^* is the following vector field on the sphere

$$\nabla U^* = U_\varphi^* \frac{\partial}{\partial\varphi} + \frac{U_\theta^*}{\sin^2\varphi}\frac{\partial}{\partial\theta}, \tag{3.28}$$

which allows us to calculate various derivatives of U^*. For example, the tangential and normal derivatives along the curve are, respectively,

$$U_\tau^* = \frac{\partial U^*}{\partial\tau^*} = \nabla U^* \cdot \tau^* = \frac{1}{v}(\dot\varphi U_\varphi^* + \dot\theta U_\theta^*), \tag{3.29}$$

$$U_\nu^* = \frac{\partial U^*}{\partial\nu^*} = \nabla U^* \cdot \nu^* = \frac{1}{v}\Big(-\dot\theta\sin\varphi U_\varphi^* + \dot\varphi\frac{1}{\sin\varphi}U_\theta^*\Big), \tag{3.30}$$

and the intrinsic first derivative of U_ν^* is

$$U_\nu^{*\prime} = \frac{\mathrm{d}}{\mathrm{d}s}U_\nu^* = \Big(\frac{U_\theta^*}{\sin\varphi}\Big)\varphi'' + (-\sin\varphi U_\varphi^*)\theta'' + \Big(-\frac{\cos\varphi U_\theta^*}{\sin^2\varphi}\Big)\varphi'^2$$
$$+ \Big(\frac{U_{\theta\theta}^*}{\sin\varphi} - U_{\varphi\theta}^*\sin\varphi\Big)\theta'^2 + \Big(\frac{U_{\theta\theta}^*}{\sin\varphi} - \sin\varphi U_{\varphi\varphi}^* - \cos\varphi U_\varphi^*\Big)\varphi'\theta'. \tag{3.31}$$

Finally, for convenience and later reference let us introduce the following definition.

Definition 3.2 *For a given curve γ^* and function U^* on S^2, the associated Siegel function along the curve is defined to be*

$$\mathfrak{S} = \frac{U_\nu^*}{K^*}. \tag{3.32}$$

We regard the function as undefined along geodesic arcs. Moreover, the function may have a singularity at isolated points where K^* vanishes. Note that $\mathfrak{S}$ is independent of the orientation of γ^*. Observe the following formula for the logarithmic derivative of $\mathfrak{S}$, as a function of s (and similarly for t as parameter)

$$\frac{\mathfrak{S}'}{\mathfrak{S}} = \frac{U_\nu^{*\prime}}{U_\nu^*} - \frac{K^{*\prime}}{K^*}. \tag{3.33}$$

Remark 3.3 We have named the above function after C. L. Siegel for the following reason. In his study (cf. [11]) of triple collisions in the 3-body problem, Siegel investigated the asymptotic behavior of the time derivatives $\dot I$, $\ddot I$ of the moment of inertia $I = \rho^2$. The major step in his proof was, indeed, to show that the expression $\sqrt{I}(2T - \frac{\dot I^2}{4I})$ tends to zero. It turns out that this expression equals $\mathfrak{S}$ whenever the latter is defined (see Lemma 3.1, where $\rho^3 v^2$ equals the above expression). In particular, it is an intrinsic quantity at the shape space level. We shall return to triple collisions and Siegel's approach in Section 6.

3.4.2 Reformulation of the geodesic equations in terms of the shape curve

For convenience, let us write $V_1 = \frac{\partial}{\partial\varphi}$ and $V_2 = \frac{\partial}{\partial\theta}$ for the basic coordinate vector fields of the spherical coordinate system on the unit sphere. By definition, the acceleration of the

spherical curve $\gamma^*(t) = (\varphi(t), \theta(t))$ is the expression

$$
\begin{aligned}
\ddot{\gamma}^* = \frac{D}{\mathrm{d}t}\gamma^* &= \ddot{\varphi}V_1 + \dot{\varphi}\nabla_{\dot{\gamma}^*}V_1 + \ddot{\theta}V_2 + \dot{\theta}\nabla_{\dot{\gamma}^*}V_2 \\
&= \ddot{\varphi}V_1 + \ddot{\theta}V_2 + \dot{\varphi}^2\nabla_{V_1}V_1 + \dot{\theta}^2\nabla_{V_2}V_2 + \dot{\varphi}\dot{\theta}(\nabla_{V_1}V_2 + \nabla_{V_2}V_1),
\end{aligned}
$$

which is the covariant derivative of the velocity along the curve, with respect to the metric (2.11). By definition,

$$
\nabla_{V_i}V_j = \Gamma_{ij}^1 V_1 + \Gamma_{ij}^2 V_2
$$

and the only nonzero Christoffel symbols of the metric are

$$
\Gamma_{12}^2 = \Gamma_{21}^2 = \cot\varphi, \quad \Gamma_{22}^1 = -\sin\varphi\cos\varphi.
$$

Consequently,

$$
\ddot{\gamma}^* = (\ddot{\varphi} - \dot{\theta}^2\sin\varphi\cos\varphi)\frac{\partial}{\partial\varphi} + (\ddot{\theta} + 2\dot{\varphi}\dot{\theta}\cot\varphi)\frac{\partial}{\partial\theta}. \tag{3.34}
$$

Now, take the above sphere and curve to be the shape space M^* and a shape curve $\gamma^*(t)$, respectively. Then it follows immediately from (3.28) and (3.34) that the equations (3.3)(ii) and (3.3)(iii) can be expressed neatly as the following coordinate-free vector equation on the 2-sphere,

$$
\ddot{\gamma}^* + \left(2\frac{\dot{\rho}}{\rho}\right)\dot{\gamma}^* - \frac{4}{\rho^3}\nabla U^* = 0. \tag{3.35}
$$

Here the scaling variable $\rho = \sqrt{I}$ of the cone $\overline{M}$ plays the role of an auxiliary function which couples the equation (3.35) to the equation (3.3)(i). The latter is the Lagrange-Jacobi equation (1.5), which for a given shape curve $\gamma^*(t)$ is a second order differential equation purely for $\rho(t)$. Another interpretation of the coefficient of the velocity in (3.35) follows from (3.20), namely we have

$$
2\frac{\dot{\rho}}{\rho} = v\cot\alpha. \tag{3.36}
$$

Remark 3.4 In fact, the differential equation (3.35) can be completely decoupled from the function $\rho(t)$ and hence it really becomes a differential equation on the 2-sphere, as summarized in Theorem 3.2 below.

3.4.3 Separation of the scaling function $\rho(t)$ from the shape space coordinates

The system (3.18)–(3.19) characterizes the moduli curves of 3-body trajectories with zero angular momentum, expressed in the language of kinematic geometry and, in particular, for that reason the natural parameter is the arc-length $\overline{s}$ of the moduli curve or the arc-length σ of its image shape curve on the sphere of radius $\frac{1}{2}$. Although the original mechanical system (1.2) naturally involves the physical parameter of time t, the latter is infinitesimally related to $\overline{s}$ by the identity

$$
\frac{\mathrm{d}\overline{s}}{\mathrm{d}t} = \sqrt{2T} = \sqrt{2}\sqrt{h+U}, \tag{3.37}
$$

and this enables us to express the system (3.18)–(3.19) in terms of t and hence combine it directly with the other system (3.3).

To this end, let us consider a shape curve $\gamma^*(t)$ on the sphere $S^2(\frac{1}{2})$. By (3.20),

$$\sin\alpha = \rho\frac{\frac{\mathrm{d}\sigma}{\mathrm{d}t}}{\frac{\mathrm{d}\bar{s}}{\mathrm{d}t}} = \frac{\rho}{\sqrt{2}}\left(h + \frac{U^*}{\rho}\right)^{-1/2}\left|\frac{\mathrm{d}\gamma^*}{\mathrm{d}t}\right|, \tag{3.38}$$

and then the equation (3.19) becomes

$$\frac{\rho^2}{2}\left|\frac{\mathrm{d}\gamma^*}{\mathrm{d}t}\right|^2\mathcal{K}^* = \frac{1}{2\rho}\frac{\mathrm{d}}{\mathrm{d}\nu^*}U^*. \tag{3.39}$$

The latter is not only considerably simpler than equation (3.19), but it also provides a simple formula to compute $\rho(t) = \rho(\overline{\gamma}(t))$ in terms of the geometry of the shape curve, namely

$$\rho^3 = \frac{\frac{\mathrm{d}}{\mathrm{d}\nu^*}U^*}{\mathcal{K}^*|\frac{\mathrm{d}\gamma^*}{\mathrm{d}t}|^2}. \tag{3.40}$$

In view of the integral formula (3.21) this is, indeed, a pleasant surprise which, in one stroke, shows how to reconstruct the moduli curve $\overline{\gamma}(t)$ from the shape curve $\gamma^*(t)$ by the simple formula (3.40).

The expression on the right-hand side of (3.40) refers to the kinematic geometry with $M^* = S^2(\frac{1}{2})$, and the whole product on this side would change by the factor $4r^2$ if we had worked in the sphere $S^2(r)$. Henceforth, we shall return to the unit sphere $S^2(1)$, and with the notation for speed, curvature and normal derivative from Subsection 3.4.1 we can restate (3.40) in the following way.

Lemma 3.1 *Let $\mathfrak{S}$ be the Siegel function (3.32) of (γ^*, U^*), which relates the intrinsic geometry of γ^* with the gradient field ∇U^* on the unit 2-sphere. If $\overline{\gamma}(t) = (\rho(t), \gamma^*(t))$ is the time-parametrized moduli curve of a 3-body motion with zero angular momentum, then the speed $v(t)$ of $\gamma^*(t)$ is related to $\rho(t)$ and $\mathfrak{S}(\gamma^*(t))$ by the identity*

$$\rho^3 = \frac{4}{v^2}\mathfrak{S} \quad or \quad \mathfrak{S} = \frac{1}{4}\rho^3 v^2. \tag{3.41}$$

In particular, $\mathfrak{S}$ is always nonnegative!

Now, returning to the equation (3.35), we set

$$P = 2\frac{\dot\rho}{\rho} = \frac{2}{3}\frac{\dot{\mathfrak{S}}}{\mathfrak{S}} - \frac{4}{3}\frac{\dot{v}}{v}, \tag{3.42}$$

$$Q = -\frac{4}{\rho^3} = -\frac{v^2}{\mathfrak{S}}, \tag{3.43}$$

where the rightmost identity in (3.42) follows from (logarithmic) differentiation of the identity (3.41). Then we can state the following result.

Theorem 3.2 *For 3-body motions with zero angular momentum and fixed total energy, the associated shape curves $\gamma^*(t)$ on the unit 2-sphere are characterized by the ODE*

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\gamma^* + P\frac{\mathrm{d}}{\mathrm{d}t}\gamma^* + Q\nabla U^* = 0, \tag{3.44}$$

where the first term is the (covariant) acceleration and the coefficients are the functions P, Q defined by (3.42), (3.43), which can be expressed purely in terms of φ, θ and their derivatives up to order 3.

Remark 3.5 The formula (3.41), or equivalently

$$K^* = \frac{4U_\nu^*}{v^2\rho^3} \tag{3.45}$$

expresses the geodesic curvature of the shape curve in terms of $\rho, \varphi, \theta, \dot{\varphi}, \dot{\theta}$. Therefore, since it only involves first order derivatives in the moduli space, it is not surprising to find that the same formula can, indeed, be derived more directly from the general spherical curvature formula (3.26) by elimination of the second order derivatives using the equations (3.3)(ii), (3.3)(iii).

3.4.4 Regular and irregular points and exceptional shape curves

The formula for ρ^3 in (3.41) involves the three quantities v, U_ν^* and K^* and the product may become indefinite when some of them vanish, namely we note the following implications

$$
\begin{aligned}
&[v = 0] \implies [U_\nu^* = 0] \impliedby [K^* = 0], \\
&[U_\nu^* = 0] \implies [v = 0] \quad \text{or} \quad [K^* = 0].
\end{aligned}
\tag{3.46}
$$

It is worthwhile having a closer look at the geometric interpretation and behavior of the shape curve due to the vanishing of any of these numbers, and accordingly we shall make some definitions to distinguish the various cases.

Formula (3.41) expresses the Siegel function $\mathfrak{S}$ of γ^* in two different ways, namely at a point $P_0 = \gamma^*(t_0)$ its value is

$$\lim_{t \to t_0} \frac{1}{4}\rho^3 v^2 = \mathfrak{S}_0 = \lim_{t \to t_0} \frac{U_\nu^*}{K^*}, \tag{3.47}$$

whenever any of the limits are defined, including $+\infty$ as a limiting value.

Definition 3.3 *P_0 is a regular point if $0 < \mathfrak{S}_0 < \infty$, and otherwise it is irregular.*

The irregular points consist of the cusps, the collision points (binary or triple), and for completeness we also include escape points:

$$
\begin{aligned}
\text{cusp at } P_0: \quad & \mathfrak{S}_0 = 0, \ v_0 = 0, \ \nabla U^*(P_0) \neq 0, \\
\text{triple collision at } P_0: \quad & \mathfrak{S}_0 = 0, \ \rho_0 = 0, \ \nabla U^*(P_0) = 0, \\
\text{binary collision at } P_0: \quad & \mathfrak{S}_0 = \infty, \ v_0 = \infty, \ P_0 = \mathbf{b}_i, \\
\text{escape at } P_0: \quad & \mathfrak{S}_0 = ?, \quad \rho_0 = \infty, \ v_0 = 0.
\end{aligned}
\tag{3.48}
$$

In Section 6, we shall return to triple collisions, but binary collisions and escape to infinity behavior will not be a topic in this paper.

The simplest curves on the 2-sphere are the geodesic circles, characterized by $K^* = 0$ at each point. If such a curve is the shape curve of a 3-body motion, then it follows from (3.47) that U_ν^* also vanishes. In fact, γ^* coincides with a gradient line segment if and only if it is a geodesic.

Definition 3.4 *The shape curve γ^* is called exceptional if it is confined to a gradient line (or a geodesic circle), or it consists of a single point.*

Clearly, a single point shape curve must be a fixpoint of the gradient flow, see Remark 3.1. Thus, apart from the exceptional shape curves, being a regular or irregular point is an intrinsic property, that is, it depends only on the geometric curve.

Remark 3.6 We omit the proof here, so we rather claim that the only exceptional shape curves (of length > 0) are those representing collinear motions or isosceles triangle motions. Their crucial property is that the shape curve lies on a circle fixed by an isometry (reflection) of the sphere which leaves U^* invariant. It also follows that a non-exceptional shape curve γ^* can only intersect an exceptional curve transversely, that is, neither tangentially nor with zero speed.

The equator circle E^* represents the collinear motions, of course, but isosceles motions exist only for special mass distributions, as follows. An isosceles m-triangle has (at least) two equal masses, say $m_1 = m_2$, and the mass m_3 lies on the symmetry axis of the triangle. Their shapes constitute the meridian through the north pole, the Euler point $\mathbf{e}_3$ and its antipodal point $-\mathbf{e}_3 = \mathbf{b}_3$. The latter point represents the collision of the two symmetric mass points somewhere on the symmetry axis. Thus, an isosceles triangle motion arises when the initial position and velocity have the above isosceles symmetry.

Henceforth, we shall assume the shape curve γ^* is not of exceptional type, unless otherwise stated. Consider the power series expansions

$$K^* = \sum_{i=0}^{\infty} K_i s^i, \quad U_\nu^* = \sum_{i=0}^{\infty} \omega_i s^i$$

at a regular point $P_0 = \gamma^*(t_0)$, where s is the arc-length measured from P_0. The value $\mathfrak{S}_0$ of $\mathfrak{S}$ at $s = 0$ can be calculated in two ways, possibly by the aid of l'Hospitals rule,

$$\frac{1}{4}\rho_0^3 v_0^2 = \mathfrak{S}_0 = \lim_{s \to 0} \frac{U_\nu^*}{K^*} = \frac{\lim \frac{\mathrm{d}}{\mathrm{d}s} U_\nu^*}{\lim \frac{\mathrm{d}}{\mathrm{d}s} K^*} = \cdots = \frac{\lim \frac{\mathrm{d}^k}{\mathrm{d}s^k} U_\nu^*}{\lim \frac{\mathrm{d}^k}{\mathrm{d}s^k} K^*} = \frac{\omega_k}{K_k} > 0, \tag{3.49}$$

where $k \geq 0$ is the smallest integer such that $K_k \neq 0$. Then $K_i = \omega_i = 0$ for $i < k$, $\omega_k \neq 0$, and we say P_0 is a regular point of order k. It is a finite number since otherwise U_ν^* and K^* vanishes identically and γ^* would be exceptional. In particular, we see that $\omega_0 \neq 0$ means $\gamma^*(t)$ is transversal to the gradient flow at $t = t_0$.

The order k of a cusp at $P_0 = (\varphi_0, \theta_0)$ can be defined similarly by considering the limits in (3.49). The only difference is that $\omega_k = 0$ and $K_k \neq 0$ at the last step. Cusps arise when the moduli curve $\overline{\gamma}(t)$ in $\overline{M}$ is tangent to the ray at the point (ρ_0, P_0) and hence the projected curve $\gamma^*(t)$ on the 2-sphere "halts" at $t = t_0$. Geometrically, the curve γ^* near P_0 is a cusp consisting of two diverging branches which emanate from P_0, both with the initial direction of $\nabla U^*(P_0)$ and the initial curvature

$$K_0 = \frac{1}{3} \frac{\mathrm{d}}{\mathrm{d}\nu} \ln |\nabla U^*(P_0)|. \tag{3.50}$$

In Remark 3.8 we further describe the totality of cusps at P_0, at a fixed energy level h, as a specific family of curves γ^* parametrized by a number $c > 0$ (resp. $c \geq 0$) for $h \geq 0$ (resp. $h < 0$). In the case $c = 0$ the two branches coincide completely and $\overline{\gamma}(t_0) = (\rho_0, P_0)$ lies on the Hill's surface $\partial \overline{M}_h$ (cf. (1.14)).

3.4.5 On the initial value problem for the moduli curve and the shape curve

In the spherical coordinates (ρ, φ, θ) on $\overline{M} \simeq \mathbb{R}^3$, the time parametrized moduli curve and its associated shape curve are simply related by

$$\overline{\gamma}(t) = (\rho(t), \gamma^*(t)), \quad \gamma^*(t) = (\varphi(t), \theta(t)),$$

and conversely, formula (3.41) is the key to the lifting procedure, namely the reconstruction of $\overline{\gamma}(t)$ from its projection on the 2-sphere.

We choose an initial point $P_0 = (\varphi_0, \theta_0)$ on the shape curve, and assume (for simplicity) it is regular in the sense of Definition 3.3. In particular, the (local) lifting procedure is well defined, and the corresponding initial value problems for the ODEs (3.3) and (3.44) respectively are equivalent. Thus, on the one hand, the solution $\overline{\gamma}(t)$ is uniquely determined by its initial position and velocity

$$\overline{\gamma}(t_0) = (\rho_0, \varphi_0, \theta_0), \quad \frac{\mathrm{d}}{\mathrm{d}t}\overline{\gamma}(t_0) = (\rho_1, \varphi_1, \theta_1), \tag{3.51}$$

but on the other hand it is also determined by the corresponding initial data

$$(\varphi_0, \theta_0), (\varphi_1, \theta_1), (\varphi_2, \theta_2), \quad \text{where } \varphi_2 = \frac{1}{2}\ddot{\varphi}|_{t_0}, \ \theta_2 = \frac{1}{2}\ddot{\theta}|_{t_0}, \tag{3.52}$$

at the shape space level.

However, the above chosen initial data determine an energy level h, and conversely, if h is already given, the initial data (3.51) or (3.52) are acceptable only if the resulting energy level is h. To make this more transparent, let us rather state the initial value problem in the moduli space as a system with 4 equations

$$
\begin{aligned}
&\text{(i)} \quad 0 = \ddot{\rho} + \frac{\dot{\rho}^2}{\rho} - \frac{1}{\rho}\left(\frac{1}{\rho}U^* + 2h\right), \\[4pt]
&\text{(ii)} \quad 0 = \ddot{\varphi} + 2\frac{\dot{\rho}}{\rho}\dot{\varphi} - \frac{1}{2}\sin(2\varphi)\dot{\theta}^2 - \frac{4}{\rho^3}U_\varphi^*, \\[4pt]
&\text{(iii)} \quad 0 = \ddot{\theta} + 2\frac{\dot{\rho}}{\rho}\dot{\theta} + 2\cot(\varphi)\dot{\varphi}\dot{\theta} - \frac{4}{\rho^3}\frac{1}{\sin^2\varphi}U_\theta^*, \\[4pt]
&\text{(iv)} \quad 0 = \frac{1}{2}\dot{\rho}^2 + \frac{\rho^2}{8}(\dot{\varphi}^2 + \sin^2\varphi\,\dot{\theta}^2) - \frac{1}{\rho}U^*(\varphi, \theta) - h,
\end{aligned}
\tag{3.53}
$$

where the first order equation (iv) is the energy integral (1.4) for the fixed value h. In fact, any one of the four equations is redundant and can be deduced from the other three. For example, with h calculated from the initial data (3.51), the solution of (i)–(iii) also satisfies (iv).

More geometrically, if the initial data set (3.51) is a point on a specific level surface of type (iv) in the tangent bundle of $\overline{M}$, then the solution $\overline{\gamma}(t)$ of (3.53)(i)–(iii) must lie on the surface

(iv) for all t. But we can also determine the same solution $\overline{\gamma}(t)$ from (ii)–(iv). All this amounts to saying that for a given value of h the whole system (3.53) is of total order 5.

On the other hand, the ODE (3.44) on the 2-sphere is independent of h and its total order is 6. Its solutions are the projections γ^* of all solutions of the system (3.53) for any value of h. In general, they are divided into three disjoint classes, distinguished by the sign of h (positive, zero or negative). The exceptions to this subdivision are precisely the exceptional shape curves (cf. Definition 3.4), which can represent 3-body motions at any energy level (cf. Subsection 4.2).

Now, let us consider the problem of how to translate the initial data (3.51), at a given energy level h, into a set of suitable initial data consisting of five numbers depending only on the shape curve. A natural first choice would be

$$(\varphi_0, \theta_0), \quad (\psi_0, v_0), \quad \mathfrak{S}_0, \tag{3.54}$$

where the angle $\psi_0 \in [0, 2\pi)$ specifies the initial direction of γ^*, $v_0 > 0$ is the inital speed, and $\mathfrak{S}_0 = (U_\nu^* / K^*)_0 > 0$ is the initial value of the Siegel function.

Remark 3.7 The generic points of γ^* are the regular points of order $k = 0$, and hence γ^* is transversal to the gradient flow almost everywhere. For such points it is, perhaps, more natural to replace $\mathfrak{S}_0$ by K_0 in (3.54). In fact, the two choices — either K_0 or $\mathfrak{S}_0$ — are equivalent since the direction ψ_0 determines ω_0 when $k = 0$.

The pairs (ψ_0, v_0) and (φ_1, θ_1) evidently determine each other, namely they specify the initial velocity of γ^*. Clearly, the 5-tuple (3.54) is merely the data (3.51) with the pair (ρ_0, ρ_1) replaced by the single number $\mathfrak{S}_0$, which is a second order quantity at the shape space level. Indeed, we recover ρ_0 immediately from (3.54) by making use of (3.49).

Thus, it is clear that the initial data information given by (3.51) or (3.54) would be equivalent if we could also recover the radial speed ρ_1 from (3.54). However, the energy integral (3.53)(iv) determines only ρ_1^2; in fact, there is no solution at all if h is below the critical value

$$h_{\min} = \frac{1}{8}\rho_0^2 v_0^2 - \frac{1}{\rho_0}U^*(\varphi_0, \theta_0), \quad \text{where } \rho_0 = \left(\frac{4\mathfrak{S}_0}{v_0^2}\right)^{\frac{1}{3}}. \tag{3.55}$$

There is a unique solution if $h = h_{\min}$, with $\rho_1 = 0$, and for $h > h_{\min}$ there are two solutions which are distinguished by the sign of ρ_1.

A slightly different approach is to combine the identities (3.36) and (3.38), which yields a value of $\sin \alpha$, where $0 < \alpha < \pi$. In fact, $\alpha < \frac{\pi}{2}$ means $\rho_1 < 0$ and $\alpha > \frac{\pi}{2}$ means $\rho_1 > 0$. But, we are still left with the problem of how to determine the sign of ρ_1 from the initial data (3.54). Anyhow, up to now we have the following result as a summary of the above local analysis at a regular point.

Proposition 3.1 *Consider the 3-body motions with zero angular momentum and a given total energy h, whose oriented shape curves at a given regular point have the same initial direction, speed, and curvature (or Siegel number $\mathfrak{S}_0$, if the curvature K_0 vanishes). Then the number of solutions is, up to congruence, equal to $0, 1$ or 2 depending on whether $h < h_{\min}$*

(*resp. $h = h_{\min}$ or $h > h_{\min}$*), *where $h_{\min}$ is calculated from the given initial data by the formula* (3.55).

A 3-body motion is said to be expanding when $\dot{\rho}(t) > 0$, and it is contracting when $\dot{\rho}(t) < 0$. Accordingly, we say the expansion index at time t_0 is the sign $e(t_0) = 0, \pm 1$ of $\dot{\rho}(t_0) = \rho_1$.

Corollary 3.1 *A 3-body motion $\gamma(t)$ with zero angular momentum and a given total energy h is uniquely determined up to congruence by the oriented shape curve $\gamma^* \subset S^2$ as a subset (i.e. non-parametrized), together with the initial speed v_0 and expansion index $e(t_0)$ at a regular point $\gamma^*(t_0)$.*

Remark 3.8 The above corollary is, in fact, also true when the point $P_0 = \gamma^*(t_0)$ is a cusp. Indeed, for each fixed h, the family of possible cusps at P_0 is parametrized by the nonnegative numbers

$$c = \rho_1^2 = 2\left(\frac{U^*(P_0)}{\rho_0} + h\right),\tag{3.56}$$

as follows from the energy integral (3.53)(iv).

For $c = 0$ there is a unique moduli curve $\overline{\gamma}(t)$ starting out from the "rest point" (ρ_0, P_0) on the Hill's boundary $\partial \overline{M}_h$, and the corresponding shape curve γ^* is the simple cusp (i.e., with one branch) emanating from P_0. In general, the local geometry of γ^* at P_0 determines the number c. In fact, one can determine c from the first curvature coefficients K_i, $i > 0$, but note that K_0 is independent of c, by (3.50). Finally, one solves the initial value problem in $\overline{M}$ at (ρ_0, P_0), with the velocity component $\rho_1 = \pm\sqrt{c}$ selected according to our choice of $e(t_0)$.

For a given non-parametrized curve γ^* in M^*, the geodesic curves of the associated cone surface $C(\gamma^*)_h$ are characterized by the ODE of (3.18). The squared speed of such a curve is, of course, given by $2T = 2(U + h)$ and hence it is determined by the position. A geodesic $\overline{\gamma}$ is therefore uniquely determined by the initial position and direction at $t = t_0$, namely $\rho(t_0)$, $\alpha(t_0)$ and the point $\gamma^*(t_0)$. By (3.40) or (3.41), and elimination of the speed in the shape space using (3.38), it follows that the validity of the identity

$$\sin^2 \alpha = \frac{2\mathfrak{S}}{\rho h + U^*}\tag{3.57}$$

along the entire curve $\overline{\gamma}$, where $\mathfrak{S}$ is the Siegel function of γ^* in $S^2(1)$, is a necessary and sufficient condition for a geodesic of the cone surface to be a geodesic curve of $\overline{M}_h$ as well. This proves the following statement.

Corollary 3.2 *A non-parametrized (i.e. geometric) curve $\gamma^* \subset S^2(1)$ can be suitably parametrized as the shape curve of a 3-body motion with zero angular momentum and total energy h if and only if its cone surface $C(\gamma^*)_h$ has a geodesic curve satisfying (3.57) along the entire curve.*

Problem 3.1 *As indicated by the condition (3.57), only a very special kind of geometric curves on the 2-sphere S^2 can be suitably parametrized as the shape curve of a 3-body motion as above. How can they be characterized in a neat way? Can such a curve have different time parametrizations as the shape curve of 3-body motions?*

Problem 3.2 *Is the initial speed and expansion index also determined by the geometric shape curve in Corollary* 3.1? (*The case* $h = 0$ *turns out to be special.*)

4 On the Analysis of Moduli and Shape Curves via Power Series

We continue to use the notation and terminology from Section 3. Consider a time parametrized moduli curve $t \to \overline{\gamma}(t)$ in $\overline{M} \simeq \mathbb{R}^3$ which represents a 3-body motion with vanishing angular momentum, and let $t \to \gamma^*(t)$ be the associated shape curve, namely its projection in the 2-sphere S^2. In this section we shall investigate the possibility of reconstructing the parametrized curve $\overline{\gamma}(t)$ solely from the oriented geometric shape curve. Moreover, there is the question of how much geometric information about the curve γ^* is really needed for such a lifting procedure. At the end we shall also answer the question concerning the uniqueness of the time parametrization.

4.1 Generation of recursive relations and intrinsic geometric invariants

In the local analysis of the moduli and the shape curve, and their interaction with the potential function U^*, we shall distinguish between two types of variables or quantities. Namely, on the one hand there are the intrinsic quantities which depend only on γ^* as an oriented geometric (i.e. unparametrized) curve and U^* as a function on S^2, and on the other hand there are the variable quantities, defined along the curve $\overline{\gamma}$ or γ^*, which depend on the scaling function ρ in the moduli space $\overline{M}$ or the time parametrization of the curves. The basic intrinsic quantities are the gradient field ∇U^* (or the tangential and normal derivatives U_τ^*, U_ν^*), the unit tangent field of γ^*, and the geodesic curvature function K^* of γ^*. Moreover, we shall assume that γ^* is not exceptional and hence the linkage between γ^* and U^* is also neatly encoded into the intrinsic Siegel function $\mathfrak{S} = U_\nu^*/K^*$ (cf. (3.32) and Subsection 3.4.4).

Let s be the arc-length parameter of γ^* measured in the positive direction from a given regular point P_0 of order $k \geq 0$. Then the coefficients of the power series expansions of the above functions, such as

$$
\begin{aligned}
K^* &= K_0 + K_1 s + K_2 s^2 + \cdots, \\
U^* &= u_0 + \overline{u}_1 s + \overline{u}_2 s^2 + \cdots, \\
U_\nu^* &= \omega_0 + \omega_1 s + \omega_2 s^2 + \cdots, \\
\mathfrak{S} &= \mathfrak{S}_0 + \mathfrak{S}_1 s + \mathfrak{S}_2 s^2 + \cdots,
\end{aligned}
\tag{4.1}
$$

yield intrinsic quantities localized at the point P_0. Note that the expansion of the tangential derivative of U^* is

$$
U_\tau^* = \nabla U^* \cdot \tau = \frac{\mathrm{d}}{\mathrm{d}s} U^* = \overline{u}_1 + 2\overline{u}_2 s + 3\overline{u}_3 s^2 + \cdots,
\tag{4.2}
$$

and the coefficients $\mathfrak{S}_n$ are expressible as rational functions of K_i and ω_i, $k \leq i \leq n + k$. Let us say that the order of a coefficient is the highest order of derivatives of local coordinates in its expression. Thus, the pair (φ_0, θ_0) and u_0 are the intrinsic (geometric) data of order 0 at P_0. Next, the triple $\overline{u}_1$, ω_0, and the unit tangent vector at P_0 represent the intrinsic data of order 1 at P_0, and ω_n and $\overline{u}_{n+1}$ (resp. K_n and $\mathfrak{S}_n$) has order $n + 1$ (resp. $n + 2$).

We choose a spherical polar coordinate system (φ, θ), with P_0 different from any of the "poles" $\varphi = 0$ or π, and for a given moduli curve $\overline{\gamma}(t) = (\rho(t), \varphi(t), \theta(t))$ we shall expand the coordinate functions, as well as U^* and its partial derivatives, as power series with respect to t :

$$
\begin{aligned}
\rho &= \rho_0 + \rho_1 t + \rho_2 t^2 + \rho_3 t^3 + \cdots, \\
\varphi &= \varphi_0 + \varphi_1 t + \varphi_2 t^2 + \varphi_3 t^3 + \cdots, \\
\theta &= \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \cdots, \\
v &= v_0 + v_1 + v_2 t + v_2 t^2 + \cdots, \\
U^* &= u_0 + u_1 t + u_2 t^2 + u_3 t^3 + \cdots, \\
U_\varphi^* &= \mu_0 + \mu_1 t + \mu_2 t^2 + \mu_3 t^3 + \cdots, \\
U_\theta^* &= \eta_0 + \eta_1 t + \eta_2 t^2 + \eta_3 t^3 + \cdots.
\end{aligned}
\tag{4.3}
$$

For convenience, we also write

$$
\begin{aligned}
\sin(2\varphi) &= f_0 + f_1 t + f_2 t^2 + \cdots, \\
\sin^2(\varphi) &= g_0 + g_1 t + g_2 t^2 + \cdots,
\end{aligned}
$$

and list some of the initial coefficiens:

$$
\begin{aligned}
u_0 &= U^*(\varphi_0, \theta_0), \quad u_1 = \mu_0 \varphi_1 + \eta_0 \theta_1, \quad \text{etc.,} \\
f_0 &= \sin(2\varphi_0), \qquad f_1 = 2\cos(2\varphi_0)\varphi_1, \quad \text{etc.,} \\
g_0 &= \sin^2(\varphi_0), \qquad g_1 = f_0 \varphi_1, \quad \text{etc.,} \\
v_1 &= \frac{1}{v_0}[2\varphi_1 \varphi_2 + \sin(\varphi_0)\cos(\varphi_0)\varphi_1 \theta_1^2 + 2\sin^2(\varphi_0)\theta_1 \theta_2],
\end{aligned}
\tag{4.4}
$$

where the expression for v_1 follows from (3.24). We shall regard μ_0, η_0 as intrinsic data, but they depend on the coordinate system, of course.

Below we shall investigate dependence relations among the coefficients $\rho_i, \varphi_j, \theta_k$ of the coordinate functions in (4.3) and various other coefficients. Some of them are directly expressible in terms of the intrinsic data and hence regarded as constants, whereas the others are the variables.

Definition 4.1 *The following list of coefficients from* (4.3)

$$
\rho_0, v_0; \quad \rho_1, \varphi_1, \theta_1; \quad \rho_2, \varphi_2, \theta_2
\tag{4.5}
$$

will be referred to as the variables of order ≤ 2. The variables of order n are $\rho_n, \varphi_n, \theta_n$ when $n > 0$, and ρ_0, v_0 are the only variables of order zero.

Henceforth, assume that the above moduli curve $\overline{\gamma}(t)$ is a solution of the ODE system (3.53). By inserting the power series into the equations (3.53)(i)–(iv) and applying the method of undetermined coefficients, we arrive at the following scheme of recursive relations for the

variables of increasing order $0, 1, 2, \cdots$:

$$E_{10} : 0 = 2\rho_0^2\rho_2 + \rho_0\rho_1^2 - 2h\rho_0 - u_0,$$

$$E_{20} : 0 = 2\rho_0^3\varphi_2 + 2\rho_0^2\rho_1\varphi_1 - \frac{1}{2}\rho_0^3 f_0\theta_1^2 - 4\mu_0,$$

$$E_{30} : 0 = 2g_0\rho_0^3\theta_2 + 2g_0\rho_0^2\rho_1\theta_1 + \rho_0^3 f_0\varphi_1\theta_1 - 4\eta_0,$$

$$E_{40} : 0 = \rho_0\rho_1^2 + \frac{1}{4}\rho_0^3(\varphi_1^2 + g_0\theta_1^2) - 2u_0 - 2h\rho_0,$$

$$E_{11} : 0 = 6\rho_0^2\rho_3 + 8\rho_0\rho_1\rho_2 + \rho_1^3 - 2h\rho_1 - u_1,$$

$$E_{21} : 0 = 6\rho_0^3\varphi_3 + 10\rho_0^2\rho_1\varphi_2 + 4(\rho_0^2\rho_2 + \rho_0\rho_1^2)\varphi_1 - 2f_0\rho_0^3\theta_1\theta_2$$
$$- \frac{1}{2}(f_1\rho_0^3 + 3f_0\rho_0^2\rho_1)\theta_1^2 - 4\mu_1,$$

$$E_{31} : 0 = 6g_0\rho_0^3\theta_3 + (10g_0\rho_0^2\rho_1 + 2g_1\rho_0^3 + 2f_0\rho_0^3\varphi_1)\theta_2 + 2f_0\rho_0^3\theta_1\varphi_2$$
$$+ (f_1\rho_0^3 + 3f_0\rho_0^2\rho_1)\varphi_1\theta_1 + (4g_0\rho_0^2\rho_2 + 4g_0\rho_0\rho_1^2 + 2g_1\rho_0^2\rho_1)\theta_1 - 4\eta_1,$$

(4.6)

(4.7)

and in general

$$E_{1n} : 0 = (n+2)(n+1)\rho_0^2\rho_{n+2} + \cdots,$$

$$E_{2n} : 0 = (n+2)(n+1)\rho_0^3\varphi_{n+2} + \cdots,$$

$$E_{3n} : 0 = (n+2)(n+1)\rho_0^3\theta_{n+2} + \cdots,$$

(4.8)

where the remaining terms are of less order since they involve $\rho_i, \varphi_i, \theta_i$ for $i < n + 2$. For example, the coefficients u_n, μ_n, η_n occur in (4.8) and their order is n. The equations E_{4n} for $n > 0$ are omitted since they do not lead to additional (algebraic independent) relations.

Now, let us select some independent and recursive relations from the above ones, but first we take the basic identity (3.41) and the expression (3.23) for the speed in the spherical metric, whose zero order terms yield the two identities:

$$E_0 : \rho_0^3 v_0^2 = 4\mathfrak{S}_0, \quad \mathfrak{S}_0 = \frac{\omega_0}{K_0},$$

(4.9)

$$E_0' : v_0 = \sqrt{\varphi_1^2 + g_0\theta_1^2}.$$

(4.10)

We shall use the symbols J_1, J_2 etc. to denote various expressions which are of intrinsic type. By using (4.9) the identities E_{10} and E_{40} can be restated as

$$E_1 : \rho_0(\rho_1^2 - 2h) = J_1, \quad J_1 = 2u_0 - \mathfrak{S}_0,$$

$$E_4 : \rho_0^2\rho_2 = J_4, \quad J_4 = \frac{1}{2}(-u_0 + \mathfrak{S}_0).$$

(4.11)

Next, the direction ψ_0 of γ^* at the point (φ_0, θ_0) is intrinsic; it is also conveniently represented by the unit tangent vector

$$\tau^* = \frac{1}{v_0}\left(\varphi_1\frac{\partial}{\partial\varphi} + \theta_1\frac{\partial}{\partial\theta}\right) = J_\varphi\frac{\partial}{\partial\varphi} + J_\theta\frac{\partial}{\partial\theta}.$$

The coefficients J_φ, J_θ are intrinsic functions, depending on the coordinate system, and they are related by the identity

$$J_\varphi^2 + g_0 J_\theta^2 = 1.$$

(4.12)

Therefore, we adjoin to our list (4.11) the two identities

$$E_2 : \varphi_1 = J_\varphi v_0, \quad E_3 : \theta_1 = J_\theta v_0. \tag{4.13}$$

Still, we have not used all zero order relations, namely E_{20} and E_{30}, and now we state them as

$$\begin{aligned}
E_5 &: \rho_0^3 \varphi_2 + \rho_0^2 \rho_1 \varphi_1 = J_5, \quad J_5 = 2\mu_0 + f_0 J_\theta^2 \mathfrak{S}_0, \\
E_6 &: \rho_0^3 \theta_2 + \rho_0^2 \rho_1 \theta_1 = J_6, \quad J_6 = \frac{2\eta_0}{g_0} - 2\frac{f_0}{g_0} J_\varphi J_\theta \mathfrak{S}_0.
\end{aligned} \tag{4.14}$$

By continuing this way, we obtain for each $n > 0$ three new relations

$$\begin{aligned}
E_{3n+1} &: 0 = \rho_0^2 \rho_{n+2} + \cdots, \\
E_{3n+2} &: 0 = \rho_0^3 \varphi_{n+2} + \cdots, \\
E_{3n+3} &: 0 = \rho_0^3 \theta_{n+2} + \cdots,
\end{aligned} \tag{4.15}$$

involving at each step the new triple $\rho_{n+2}, \varphi_{n+2}, \theta_{n+2}$ of variables of order $n + 2$.

Claim 4.1 It is possible to solve the above recursive relations for the variables (4.5) completely in terms of the intrinsic local geometric data in the shape space.

This will be finally settled at the end of the subsection. At this point, we have altogether $3n + 8$ variables

$$\rho_0, v_0; \ \rho_1, \varphi_1, \theta_1; \ \rho_2, \varphi_2, \theta_2; \ \cdots; \ \rho_{n+2}, \varphi_{n+2}, \theta_{n+2};$$

involved in $3n + 8$ recursive relations, and the first eight involve only the variables up to order 2. However, E_0', E_2 and E_3 are obviously algebraic dependent due to the identity (4.12), so let us search for one more independent relation among the variables of order ≤ 2. Since we expect such a relation to involve local intrinsic quantities of order (at least) 3, a natural approach is to differentiate the basic identity (3.41) involving the Siegel function. Thus evaluation of the resulting identity (3.42) at $t = t_0$ yields

$$3\frac{\rho_1}{\rho_0 v_0} + 2\frac{v_1}{v_0^2} = J_7, \quad J_7 = \frac{\mathfrak{S}_1}{\mathfrak{S}_0} = \left(\frac{\omega_{k+1}}{\omega_k} - \frac{K_{k+1}}{K_k}\right). \tag{4.16}$$

Using the expression in (4.4) for v_1 we can restate the above identity as

$$3\frac{\rho_1}{\rho_0 v_0} + \frac{4}{v_0^3}\left[\varphi_1 \varphi_2 + \frac{1}{4} f_0 \varphi_1 \theta_1^2 + g_0 \theta_1 \theta_2\right] = J_7. \tag{4.17}$$

By simple calculation and substitution using some of the previous relations E_i,

$$\begin{aligned}
&\rho_0^3\left[\varphi_1 \varphi_2 + \frac{1}{4} f_0 \varphi_1 \theta_1^2 + g_0 \theta_1 \theta_2\right] \\
&= \varphi_1(J_5 - \rho_0^2 \rho_1 \varphi_1) + \varphi_1\left(\frac{1}{4} f_0 J_\theta^2 \rho_0^3 v_0^2\right) + \theta_1(g_0 J_6 - g_0 \rho_0^2 \rho_1 \theta_1) \\
&= -\frac{\rho_1}{\rho_0}\rho_0^3(\varphi_1^2 + g_0 \theta_1^2) + \varphi_1(J_5 + f_0 J_\theta^2 \mathfrak{S}_0) + \theta_1 g_0 J_6 \\
&= -4\frac{\rho_1}{\rho_0}\mathfrak{S}_0 + v_0[J_\varphi J_5 + f_0 J_\varphi J_\theta^2 \mathfrak{S}_0 + g_0 J_\theta J_6]
\end{aligned}$$

and by substitution into (4.17), using the identity $\rho_0^3 v_0^2 = 4\mathfrak{S}_0$ and the expressions for J_5, J_6 in (4.14), this leads to our new identity

$$E_1' : \frac{\rho_1}{\rho_0 v_0} = J_8, \quad J_8 = 2\mathfrak{S}_0^{-1}(J_\varphi \mu_0 + J_\theta \eta_0) - J_7 = \frac{1}{\mathfrak{S}_0}(2\overline{u}_1 - \mathfrak{S}_1), \tag{4.18}$$

where $\overline{u}_1$ is the tangential derivative of U^* at P_0 (cf. (4.2)).

From the system of algebraic equations

$$E_0, E_0', E_1, E_1', E_2, E_3, E_4, \cdots,$$

we can now solve recursively and thus determine the variables

$$\rho_0, v_0, \rho_1, \varphi_1, \theta_1, \rho_2, \varphi_2, \theta_2, \cdots$$

successively in terms of the intrinsic data. In fact, this is obvious from the structure of the equations, once we have determined ρ_0, v_0, ρ_1, namely using the three equations E_0, E_1, E_1':

$$\rho_0^3 v_0^2 = 4\mathfrak{S}_0, \quad \rho_0(\rho_1^2 - 2h) = J_1, \quad \frac{\rho_1}{\rho_0 v_0} = J_8. \tag{4.19}$$

It follows that

$$2h\rho_0 = -J_1 + J_8^2 \rho_0^3 v_0^2 = -J_1 + 4J_8^2 \mathfrak{S}_0 \tag{4.20}$$

and consequently, for $h \neq 0$,

$$\rho_0 = \frac{1}{h}\left[\frac{\mathfrak{S}_0}{2}(4J_8^2 + 1) - u_0\right], \quad v_0 = \frac{2}{\rho_0^{3/2}}\sqrt{\mathfrak{S}_0}, \quad \rho_1 = 2J_8\sqrt{\frac{\mathfrak{S}_0}{\rho_0}}. \tag{4.21}$$

In the case $h = 0$, the identity (4.20) merely tells us that

$$u_0 = \frac{1}{2}(4J_8^2 + 1)\mathfrak{S}_0, \tag{4.22}$$

and we can freely choose any initial size ρ_0 of ρ and then calculate v_0 and ρ_1 from (4.21). In particular, we calculate the pair (φ_1, θ_1) using (4.13), where the pair (J_φ, J_θ) represents the initial direction and hence is intrinsic. Now, we are able to calculate successively each new triple $(\rho_n, \varphi_n, \theta_n)$, $n = 2, 3, 4, \cdots$, expressed in terms of the intrinsic data, as claimed above.

Remark 4.1 The initial direction, (J_φ, J_θ), is the only basic intrinsic data with no invariant description, that is, independent of the coordinate frame. However, from the recursive procedure it follows that $u_n, v_n, \rho_n, n \geq 0$, come out with coordinate free expressions involving only the coefficients in (4.1). In fact, we can calculate v_{n-1} in terms of ρ_i ($i < n$) and v_j ($j < n-1$), by repeated differentiation of (3.41), next we calculate u_n by applying differential operators such as (3.25) to U^*, and finally ρ_{n+2} is calculated using equation E_{1n}. The beginning terms are

$$u_1 = v_0 \overline{u}_1, \quad u_2 = v_0^2 \overline{u}_2 + \frac{1}{2}v_1 \overline{u}_1, \quad v_1 = \frac{4\mathfrak{S}_1 - 3\rho_0^2 v_0 \rho_1}{2\rho_0^3}.$$

4.2 Some basic results on the shape curves of 3-body motions with vanishing angular momentum

Let us first review some of the above facts from the local analysis and draw a few immediate but important consequences. The above power series developments amount to the explicit calculation of the solution $t \to \overline{\gamma}(t)$ of the system (3.53) in the moduli space with the initial data (3.51). In doing so we started from the following 5-tuple

$$(\varphi_0, \theta_0), \quad (\psi_0, \mathfrak{S}_0, \mathfrak{S}_1), \tag{4.23}$$

where $\mathfrak{S}_0, \mathfrak{S}_1$ may be replaced by K_0, K_1 (or K_k, K_{k+1}, for the smallest k with $K_k \neq 0$), which consists of three specific intrinsic local geometric invariants at the point (φ_0, θ_0) on the shape curve. In particular, we also recover the time parametrized shape curve $\gamma^*(t)$ by projecting $\overline{\gamma}(t)$ to the 2-sphere.

Actually, since $\overline{\gamma}(t)$ is uniquely determined by the initial value problem, (3.51) and (3.53), it suffices to recover (3.51) from (4.23), namely the missing information in (3.51) is ρ_0 and ρ_1. This turns out to be possible when $h \neq 0$, but for a "good" reason (see below) it is impossible when $h = 0$ since in this case the shape curve only controls the product $\rho_0^3 v_0^2 = 4\mathfrak{S}_0$. In any case, with the quantity $\mathfrak{S}_1$ we can actually determine ρ_1 and, in particular, the question in Problem 3.2 concerning the expansion index is settled.

The general 3-body problem has the 10 classical conservation laws (linear and angular momentum, and energy) due to its invariance under the Galilean symmetry group. All of them have been used and, in particular, the set of solutions is invariant under time translation, $t \to t + t_0$, as well as reversal of time ($t \to -t$) which reverses the direction of the trajectory. However, there is also an additional 1-parameter size/time scaling symmetry group, whose induced action on parametrized moduli curves sends $\overline{\gamma}(t) = (\rho(t), \gamma^*(t))$ to

$$\overline{\gamma}_{(r)}(t) = (\rho_{(r)}(t), \gamma^*_{(r)}(t)) = (r\rho(r^{-\frac{3}{2}}t), \gamma^*(r^{-\frac{3}{2}}t)), \quad \forall\, r > 0, \tag{4.24}$$

and changes the energy from h to $h_{(r)} = r^{-1}h$. In particular, although scaling and time translation leaves the oriented shape curve geometrically unchanged, its time parametrization is subject to an affine transformation

$$t \to at + t_0, \quad a = r^{-\frac{3}{2}} > 0. \tag{4.25}$$

Since the energy level $h = 0$ is scaling invariant, this also explains why the reconstruction of a unique initial size ρ_0 fails when $h = 0$.

Remark 4.2 For any $n \geq 1$, the Newtonian n-body problem has the above 1-parameter symmetry group $\{\Phi_r\}$, acting on size and time but leaves the shape invariant. For example, for a periodic motion with period $P_{(1)}$ and average (or initial) size $\rho_{(1)}$, the group sweeps out a periodic motion with the same shape, and the ratio $P_{(r)}^2/\rho_{(r)}^3$ is independent of r. The case $n = 1$ means the restricted case $n = 2$ with one of the masses (e.g. a planet) infinitesimal small, in which case there is only one shape (a point) and the above ratio depends only on the large

mass (the sun). This gives Kepler's third law, so the above symmetry group is essentially the generalization of this law.

On the other hand, for 3-body motions with vanishing angular momentum, the identity $\rho^3 v^2 = 4\mathfrak{S}$ of (3.41) gives another quantity, $\rho^3 v^2$, which is invariant under the above symmetry group.

Definition 4.2 *A time reparametrization of $\overline{\gamma}(t)$ or $\gamma^*(t)$ by an affine transformation (4.25), for any $a \neq 0$, is called canonical, otherwise it is called exceptional.*

Of course, in order to stay at a given nonzero energy level a canonical reparametrization must have $a = \pm 1$, and moreover, the orientation of the curve is reversed if $a < 0$. Now we can state the following basic unique parametrization theorem.

Theorem 4.1 *A 3-body motion with zero angular momentum is, up to congruence and canonical reparametrization, uniquely determined by its oriented shape curve on the 2-sphere. In fact, it suffices to know the direction and the first two Siegel numbers $\mathfrak{S}_0, \mathfrak{S}_1$ at any regular point on the shape curve. In particular, there are no exceptional reparametrizations.*

As shown before, the theorem still holds with $(\mathfrak{S}_0, \mathfrak{S}_1)$ replaced by the curvature numbers (K_0, K_1) if the point is regular of order 0, and this is, indeed, the generic type of points.

A curve on the 2-sphere which is the shape curve of a motion with total energy h can also be the shape curve for some motion with any other energy h' of the same sign as h. Indeed, we find the other motions by suitable canonical reparametrizations of the given motion, and by (4.21) it also follows that the sign of h (viewed as a number $0, \pm 1$) is an intrinsic invariant at the shape space level. More precisely, we have the following quantitative measurement of the energy type.

Theorem 4.2 *Let γ^* be a geometric curve on the 2-sphere, with the Siegel function $\mathfrak{S}$ (with respect to U^*, as usual), and consider the function*

$$\Delta = \frac{\mathfrak{S}}{2}\left(4\left[\frac{1}{\mathfrak{S}}\frac{\mathrm{d}}{\mathrm{d}\tau}(2U^* - \mathfrak{S})\right]^2 + 1\right) - U^*$$

along the curve, where $\frac{\mathrm{d}}{\mathrm{d}\tau}$ denotes the tangential derivative. If γ^* can be realized as the shape curve of a 3-body motion with vanishing angular momentum, then the sign of Δ is constant along the curve (whenever Δ is defined), namely equal to the sign of the total energy h of the motion.

Corollary 4.1 *A given oriented (geometric) curve on the 2-sphere can be time parametrized in at most one way, up to canonical reparametrization, as the shape curve $t \to \gamma^*(t)$ of a 3-body motion with zero angular momentum. Moreover, the sign of the total energy of such a motion is determined by the local relative geometry of (γ^*, U^*) at a (regular) point.*

The above uniqueness property of time parametrization of geometric shape curves, the minimal amount of geometric information needed to determine the shape curve, and the monotonicity theorem which we shall discuss in Section 5, are our basic tools for the understanding of both the local and global picture of shape curves representing 3-body motions with vanishing angular

momentum. The monotonicity property tells us the m-latitude function is monotonic increasing or decreasing until the curve turns back somewhere in the opposite hemisphere. Thus the curve resembles an "oscillating motion" between the upper and lower hemisphere which never stops, unless it ends at a triple collision or escapes to infinity. The curve crosses the equator circle transversely, or it goes to a binary collision and bounces back (via regularization) to the same hemisphere. In Subsection 7.5 we describe the problem of how to construct such a curve by linking together its maximal monotonic segments.

5 The Monotonicity Theorem for Shape Curves

5.1 A closer look at the gradient vector field of U^*

The analysis of trajectories, moduli curves or shape curves describing 3-body motions depends, of course, ultimately on the function U^*, whose behavior is largely reflected by the geometry of its gradient field. In this subsection, some useful facts are established which are beyond those simpler statements concerning the critical or singular points of U^*.

We shall apply vector algebra in the Euclidean model for the moduli space, namely with $\overline{M} = \mathbb{R}^3$ as the Euclidean space (cf. Section 2) and vectors denoted by boldface letters. Thus the shape space $M^* = S^2(1)$ consists of unit vectors $\mathbf{p} = (x, y, z)$, $|\mathbf{p}|^2 = x^2 + y^2 + z^2$, and $\mathbf{p} \cdot \mathbf{q}$ denotes the usual inner product. Set

$$\widehat{m} = \sum \widehat{m}_i, \quad \overline{m} = m_1 m_2 m_3, \quad \widehat{m}_1 = m_2 m_3 \quad \text{etc., and} \quad \sum m_i = 1,$$

$$k_i = 2\frac{\widehat{m}_i^{\frac{3}{2}}}{\sqrt{1 - m_i}}, \quad \beta_i = \cos^{-1}\left(\frac{\widehat{m}_i - m_i}{\widehat{m}_i + m_i}\right) = \sin^{-1}\left(\frac{2\sqrt{\overline{m}}}{\widehat{m}_i + m_i}\right), \tag{5.1}$$

$$\mathbf{b_1} = (1, 0, 0), \quad \mathbf{b_2} = (\cos\beta_3, \sin\beta_3, 0), \quad \mathbf{b_3} = (\cos\beta_2, -\sin\beta_2, 0),$$

where $0 < \beta_i < \pi$ is the angle between the binary collision points $\mathbf{b}_j$ and $\mathbf{b}_k$, $\{i, j, k\} = \{1, 2, 3\}$. The Newtonian shape potential function is the restriction of U to the above 2-sphere,

$$U^* = \sum_{i=1}^{3} \frac{\widehat{m}_i}{s_i} = \sum_{i=1}^{3} \frac{k_i}{|\mathbf{p} - \mathbf{b}_i|} \quad \text{(cf. (2.16))}, \tag{5.2}$$

where the mutual distances $s_i = r_{jk}$ are normalized to $I = 1$, and hence by a formula of Lagrange

$$I = \sum \widehat{m}_i s_i^2 = 1. \tag{5.3}$$

The basic behavior of U^* is, of course, given by the 8 special points on the 2-sphere, namely

$$\mathbf{b_1}, \mathbf{e_3}, \mathbf{b_2}, \mathbf{e_1}, \mathbf{b_3}, \mathbf{e_2}; \ \mathbf{p_0}, \mathbf{p_0'}, \tag{5.4}$$

where the first six are cyclically ordered (eastward) along the equator circle E^* representing degenerate m-triangles. The $\mathbf{b}_i$ are poles where U^* tends to ∞, and the Euler points $\mathbf{e}_i$ are the saddle points. Finally, the remaining two are the minima, say $\mathbf{p_0}$ lies on the northern hemisphere and $\mathbf{p_0'}$ is the symmetric (mirror) image with respect to the equator plane.

We can also use the points $\mathbf{b}_i$ to describe the gradient field, as follows. Let $d\mathbf{x}$ be an arbitrary infinitesimal vector perpendicular to $\mathbf{p}$ (i.e. $d\mathbf{x} \in T_{\mathbf{p}}S^2$). Then, on the one hand

$$\nabla U^*(\mathbf{p}) \cdot d\mathbf{x} \equiv U^*(\mathbf{p} + d\mathbf{x}) - U^*(\mathbf{p}) \pmod{|d\mathbf{x}|^2} \tag{5.5}$$

and on the other hand,

$$U^*(\mathbf{p} + d\mathbf{x}) - U^*(\mathbf{p}) = \sum_{i=1}^{3} k_i((2 - 2(\mathbf{p} + d\mathbf{x}) \cdot \mathbf{b}_i)^{-\frac{1}{2}} - (2 - 2\mathbf{p} \cdot \mathbf{b}_i)^{-\frac{1}{2}})$$

$$\equiv \Big(\sum_{i=1}^{3} \frac{k_i \mathbf{b}_i}{|\mathbf{p} - \mathbf{b}_i|^3} \Big) \cdot d\mathbf{x} \pmod{|d\mathbf{x}|^2}. \tag{5.6}$$

Set

$$\mathbf{B} = \mathbf{B}(\mathbf{p}) = \sum_{i=1}^{3} \frac{k_i \mathbf{b}_i}{|\mathbf{p} - \mathbf{b}_i|^3}. \tag{5.7}$$

Then it follows from (5.5) and (5.6) that the gradient of the function U^* is the orthogonal projection of the vector $\mathbf{B}$ to the tangent plane of the sphere at $\mathbf{p}$, namely

$$\nabla U^* = \mathbf{B} - (\mathbf{B} \cdot \mathbf{p})\mathbf{p}. \tag{5.8}$$

The characterization of the critical points of U^* is, of course, well known. However, with the following lemma we also like to establish the identity (5.9).

Lemma 5.1 *Let $\mathbf{p}_0$ and $\mathbf{p}_0'$ represent the pair of equilateral m-triangles with $I = 1$ and with opposite orientations. Then $\mathbf{p}_0$ and $\mathbf{p}_0'$ are the minima of U^*, and moreover*

$$\mathbf{b}_i \cdot (\mathbf{b}_i - \mathbf{p}_0) = \mathbf{b}_i \cdot (\mathbf{b}_i - \mathbf{p}_0') = \frac{2m_j m_k}{(1 - m_i)\widehat{m}}. \tag{5.9}$$

Proof The determination of the critical points away from the equator follows readily by Lagrange's multiplier method in $\overline{M}$ with the constraint $I = 1$. As coordinates we can, for example, use the individual moments of inertia I_j, but the calculations are simplest in terms of the mutual distances $r_{ij} = s_k$ using (5.2) and (5.3). This shows that the minimum (s_1^0, s_2^0, s_3^0) of U^* (on any of the hemispheres) satisfies the following set of equations with a multiplier λ

$$\frac{1}{2(s_i^0)^3} = \lambda, \quad i = 1, 2, 3, \tag{5.10}$$

and hence, by (5.3), all the sides s_i^0 are equal to $\frac{1}{\sqrt{\widehat{m}}}$. Moreover, by (5.2)

$$\frac{k_i}{|\mathbf{p}_0 - \mathbf{b}_i|} = \frac{\widehat{m}_i}{s_i^0} = m_j m_k \sqrt{\widehat{m}} \Longrightarrow 2 - 2\mathbf{p}_0 \cdot \mathbf{b}_i = |\mathbf{p}_0 - \mathbf{b}_i|^2 = \frac{4m_j m_k}{(1 - m_i)\widehat{m}}$$

$$\Longrightarrow \mathbf{p}_0 \cdot \mathbf{b}_i = 1 - \frac{2m_j m_k}{(1 - m_i)\widehat{m}} \tag{5.11}$$

and this gives (5.9).

Remark 5.1 It is also straightforward to check the identities

$$\frac{k_i}{|\mathbf{p}_0 - \mathbf{b}_i|^3} = \frac{m_j m_k \sqrt{\widehat{m}}}{|\mathbf{p}_0 - \mathbf{b}_i|^2} = \frac{1}{4}\widehat{m}^{\frac{3}{2}}(1 - m_i), \tag{5.12}$$

$$\sum_{i=1}^{3}(1 - m_i)\mathbf{b}_i = 0. \tag{5.13}$$

To simplify the notation below, let us write

$$\psi_i(t) = \left(1 - \frac{t}{c_i}\right)^{-\frac{3}{2}}, \quad c_i = \frac{m_j m_k}{(1 - m_i)\widehat{m}} \quad (\text{cf. }(5.9)), \tag{5.14}$$

$$\Psi_i(t) = \frac{1}{4}\widehat{m}^{\frac{3}{2}}(1 - m_i)\psi_i(t), \tag{5.15}$$

where ψ_i is defined for $t < c_i, i = 1, 2, 3$, and observe that the derivative of ψ_i is strictly positive. The following two lemmas will be useful.

Lemma 5.2 *The expression* (5.7) *of* B *can be restated as*

$$B(\mathbf{p}) = \sum_{i=1}^{3}\Psi_i(\xi_i)\mathbf{b}_i, \tag{5.16}$$

where

$$\xi_i = \mathbf{b}_i \cdot (\mathbf{p} - \mathbf{p}_0). \tag{5.17}$$

Proof   By (5.12),

$$\mathbf{B} = \sum_{i=1}^{3}\frac{k_i\mathbf{b}_i}{|\mathbf{p} - \mathbf{b}_i|^3} = \sum_{i=1}^{3}\frac{k_i}{|\mathbf{p}_0 - \mathbf{b}_i|^3}\left(\frac{|\mathbf{p}_0 - \mathbf{b}_i|}{|\mathbf{p} - \mathbf{b}_i|}\right)^3\mathbf{b}_i$$

$$= \sum_{i=1}^{3}\frac{\widehat{m}^{\frac{3}{2}}}{4}(1 - m_i)\left(\frac{\mathbf{b}_i \cdot (\mathbf{b}_i - \mathbf{p}_0)}{\mathbf{b}_i \cdot (\mathbf{b}_i - \mathbf{p})}\right)^{\frac{3}{2}}\mathbf{b}_i$$

$$= \frac{\widehat{m}^{\frac{3}{2}}}{4}\sum_{i=1}^{3}(1 - m_i)\left(1 - \frac{\mathbf{b}_i \cdot (\mathbf{p} - \mathbf{p}_0)}{\mathbf{b}_i \cdot (\mathbf{b}_i - \mathbf{p}_0)}\right)^{-\frac{3}{2}}\mathbf{b}_i$$

$$= \frac{\widehat{m}^{\frac{3}{2}}}{4}\sum_{i=1}^{3}(1 - m_i)\psi_i(\xi_i)\mathbf{b}_i = \sum_{i=1}^{3}\Psi_i(\xi_i)\mathbf{b}_i. \tag{5.18}$$

Lemma 5.3 *Let* $\mathbf{p}$ *be a unit vector different from* $\mathbf{p}_0$ *and* $\mathbf{p}_0'$. *Then* $\mathbf{B} \cdot (\mathbf{p} - \mathbf{p}_0)$ *is strictly positive.*

Proof   By the mean value theorem, there exists $0 < \varepsilon_i < 1$,

$$\Psi_i(\xi_i) = \Psi_i(\xi_i) - \Psi_i(0) = \Psi_i'(\varepsilon_i\xi_i)\xi_i, \quad 1 \leq i \leq 3,$$

and we recall that the derivative of Ψ_i is strictly positive. By (5.16) and (5.17)

$$\mathbf{B}(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{p}_0) = \left(\sum_{i=1}^{3}\Psi_i(\xi_i)\mathbf{b}_i\right) \cdot (\mathbf{p} - \mathbf{p}_0) = \sum_{i=1}^{3}\Psi_i(\xi_i)\xi_i = \sum_{i=1}^{3}\Psi_i'(\varepsilon_i\xi_i)\xi_i^2 > 0.$$

Note that $\mathbf{B}$ is, by definition, a linear combination of $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, hence $\mathbf{B}$ lies in the xy-plane and consequently

$$\mathbf{B} \cdot (\mathbf{p} - \mathbf{p}_0) = \mathbf{B} \cdot (\overline{\mathbf{p}} - \overline{\mathbf{p}}_0),$$

where $\overline{\mathbf{p}}$ (resp. $\overline{\mathbf{p}}_0$) is the orthogonal projection of $\mathbf{p}$ (resp. $\mathbf{p}_0$) in the xy-plane. Clearly, the geometric meaning of the positivity of the inner product in the lemma is that the angle between $\mathbf{B}$ and the vector from $\overline{\mathbf{p}}_0$ to $\overline{\mathbf{p}}$ is strictly less than $\frac{\pi}{2}$.

5.2 The relative geometry between ∇U^* and the conjugate pair of co-axial families of circles associated to $\{\mathbf{p}_0, \mathbf{p}_0'\}$

In spherical geometry, circles are the simplest kind of curves and they are characterized by the constancy of their geodesic curvature. Associated to a given pair of points, such as the minima $\{\mathbf{p}_0, \mathbf{p}_0'\}$ of U^*, there are two co-axial families of circles, namely the family of circles passing through the two given points and its dual family consisting of those circles which are orthogonal to all circles of the former family. We shall denote the conjugate pair of coaxial families of circles associated to $\{\mathbf{p}_0, \mathbf{p}_0'\}$ by $\mathcal{F}$ and $\mathcal{F}'$.

To a given $\mathbf{p} \in S^2(1)$ other than $\mathbf{p}_0, \mathbf{p}_0'$, let us denote the unique circle of $\mathcal{F}$ (resp. $\mathcal{F}'$) passing through $\mathbf{p}$ by Γ_p (resp. Γ_p'). In fact, Γ_p is simply the intersection of $S^2(1)$ and the plane spanned by the triple $\{\mathbf{p}_0, \mathbf{p}_0', \mathbf{p}\}$. Therefore, the tangent line of Γ_p at $\mathbf{p}$ is the intersection of the tangent plane $T_\mathbf{p} S^2(1)$ and the above plane. In fact, the following pair of vectors

$$\mathbf{T} = \mathbf{p} \times [(\mathbf{p} - \mathbf{p}_0) \times (\mathbf{p} - \mathbf{p}_0')], \quad \mathbf{T}' = \mathbf{p} \times \mathbf{T} \tag{5.19}$$

constitutes a positively oriented orthogonal basis $\{\mathbf{T}, \mathbf{T}'\}$ for the tangent plane at $\mathbf{p}$ such that $\mathbf{T}$ (resp. $\mathbf{T}'$) is tangent to Γ_p (resp. Γ_p'). The vector $\mathbf{T}$ defines the southward direction, that is, away from $\mathbf{p}_0$, whereas $\mathbf{T}'$ defines the eastward direction along Γ_p'.

For example, in the case of uniform mass distribution, $\{\mathbf{p}_0, \mathbf{p}_0'\} = \{N, S\}$ are the north and south pole of the sphere, whose conjugate pair of coaxial families of circles are the usual longitude circles (or meridians) and latitude circles. In this case

$$\mathbf{T} = 2[(\mathbf{p} \cdot \mathbf{k})\mathbf{p} - \mathbf{k}], \quad \mathbf{T}' = -2\mathbf{p} \times \mathbf{k}$$

and these are positive multiples of the coordinate vectors $\frac{\partial}{\partial \varphi}$ and $\frac{\partial}{\partial \theta}$ associated with spherical polar coordinates (φ, θ) centered at the pole N.

Proposition 5.1 *The inner product between the "southward" vector $\mathbf{T}$ and the gradient vector ∇U^* at a point $\mathbf{p}$ on the sphere is*

$$\mathbf{T} \cdot \nabla U^* = \mathbf{T} \cdot \mathbf{B} = [\mathbf{p} \cdot (\mathbf{p}_0 - \mathbf{p}_0')][\mathbf{B} \cdot (\mathbf{p} - \mathbf{p}_0)]. \tag{5.20}$$

In particular, on the northern (resp. southern) hemisphere the angle β between $\mathbf{T}$ and ∇U^* is in the range $0 \leq \beta < \frac{\pi}{2}$ (resp. $\frac{\pi}{2} < \beta \leq \pi$).

Proof   By (5.8), $\mathbf{T} \cdot \nabla U^* = \mathbf{T} \cdot \mathbf{B}$, and clearly $\mathbf{B} \cdot \mathbf{p}_0 = \mathbf{B} \cdot \mathbf{p}_0'$, consequently

$$\mathbf{T} \cdot \mathbf{B} = \mathbf{B} \cdot \{\mathbf{p} \times [\mathbf{p} \times (\mathbf{p}_0 - \mathbf{p}_0')] + \mathbf{p} \times (\mathbf{p}_0 \times \mathbf{p}_0')\}$$

$$= \begin{vmatrix} \mathbf{B} \cdot \mathbf{p} & 1 \\ \mathbf{B} \cdot (\mathbf{p}_0 - \mathbf{p}_0') & \mathbf{p} \cdot (\mathbf{p}_0 - \mathbf{p}_0') \end{vmatrix} + \begin{vmatrix} \mathbf{B} \cdot \mathbf{p}_0 & \mathbf{p} \cdot \mathbf{p}_0 \\ \mathbf{B} \cdot \mathbf{p}_0 & \mathbf{p} \cdot \mathbf{p}_0' \end{vmatrix}$$

$$= (\mathbf{B} \cdot \mathbf{p})(\mathbf{p} \cdot (\mathbf{p}_0 - \mathbf{p}_0')) + (\mathbf{B} \cdot \mathbf{p}_0)(\mathbf{p} \cdot (\mathbf{p}_0' - \mathbf{p}_0))$$

$$= [\mathbf{p} \cdot (\mathbf{p}_0 - \mathbf{p}_0')][\mathbf{B} \cdot (\mathbf{p} - \mathbf{p}_0)]. \tag{5.21}$$

Finally, by Lemma 5.3 the second factor in (5.20) is always positive, whereas the first factor is positive on the northern hemisphere and vanishes precisely on the equator circle.

5.3 The monotone m-latitude theorem

Spherical polar coordinates (φ, θ) centered at the north pole N parametrize, of course, the latitude and longitude (meridian) circles, which constitute the pair of coaxial families of circles associated to the pair $\{N, S\}$ of geometric centers of the two hemispheres. However, instead of using the colatitude φ let us rather parametrize the latitude circles by the latitude in radians, $-\frac{\pi}{2} \leq \lambda \leq \frac{\pi}{2}$, namely $\lambda = \frac{\pi}{2} - \varphi$ and hence λ is positive on the northern hemisphere.

For equal masses the pair of minima $\{\mathbf{p}_0, \mathbf{p}_0'\}$ of U^* happens to coincide with the pair $\{N, S\}$, but this does not hold for non-equal masses. However, there exists a unique Möbius transformation which maps $\mathbf{p}_0$ to N, $\mathbf{p}_0'$ to S and the equator circle E^* to itself. Such a Möbius transformation maps $\mathcal{F}$ (resp. $\mathcal{F}'$) to the family of meridians (resp. latitude circles).

Definition 5.1 *For a given mass distribution $\{m_1, m_2, m_3\}$, the m-modified latitude of $\mathbf{p} \in \mathbf{S}^2(1)$ is defined to be the latitude in radians of the image of $\mathbf{p}$ under the above Möbius transformation, and it is denoted by $\lambda(\mathbf{p})$.*

For example, $\lambda(\mathbf{p}_0) = \frac{\pi}{2}$, $\lambda(\mathbf{p}_0') = -\frac{\pi}{2}$, and $\lambda(\mathbf{p}) = 0$ if and only if $\mathbf{p} \in E^*$. Moreover, $\lambda(\mathbf{p}) = -\lambda(\mathbf{p}')$ for any pair $\{\mathbf{p}, \mathbf{p}'\}$ representing similar m-triangles of opposite orientations.

For a given (smooth) curve $\gamma^*(t)$ on the sphere we shall consider the associated function

$$\lambda_{\gamma^*}(t) = \lambda(\gamma^*(t)), \tag{5.22}$$

which records the m-modified latitude along the curve. It turns out that for shape curves representing 3-body motions with zero angular momentum this function has a remarkable monotonicity property. Namely, it oscillates between local maxima where it is positive and local minima where it is negative, and between two such extremals it is monotonic. The only exceptions arise when the function is a constant, as described by the following lemma.

Lemma 5.4 *If the m-latitude function λ_{γ^*} is constant along γ^*, then γ^* is an exceptional shape curve which is either a single point or is confined to the equator circle (cf. Definition 3.4).*

Proof   Assume γ^* is not a single point and is confined to an m-modified latitude circle Γ' different from the equator. We choose a spherical polar coordinate system (φ, θ) centered at the geometric center of Γ'; hence φ is also constant along γ^*. By the equation (3.3)(ii), U_φ^* must

be negative along Γ', that is, the gradient ∇U^* is pointing inward along the circle. However, Γ' encloses $\mathbf{p}_0$ (or $\mathbf{p}_0'$) and Proposition 5.1 tells us that ∇U^* is pointing outward, so this is a contradiction.

Now, let us assume γ^* is not exceptional as in the above lemma. We shall state and prove the Monotone m-latitude theorem.

Theorem 5.1 *Let $\gamma^*(t)$, $a \leq t \leq b$, be a segment of the associated shape curve of a 3-body trajectory with vanishing angular momentum, and let λ_{γ^*} as in (5.22) be the function recording the m-modified latitude along $\gamma^*[a, b]$. Suppose that $a \leq t_0 \leq b$ is a critical point of λ_{γ^*} (i.e. $\lambda'_{\gamma^*}(t_0) = 0$) or is possibly a singularity. Then $\lambda_{\gamma^*}(t_0)$ must be a local maximum (resp. minimum) when $\gamma^*(t_0)$ lies on the northern (resp. southern) hemisphere.*

Proof We may assume that the point $\mathbf{q}_0 = \gamma^*(t_0)$ on the m-modified latitude circle $\Gamma' = \Gamma'_{q_0}$ is strictly inside either the northern or southern hemisphere, since γ^* crosses the equator transversely or it hits a binary collision point and bounces back into the same hemisphere (by regularization). Moreover, by using the reflectional symmetry which reverses orientation we may reduce the proof to the case that $\mathbf{q}_0$ lies on the northern hemisphere.

There are two cases to consider; either $\mathbf{q}_0$ is a cusp, that is, the speed v_0 of γ^* vanishes, or $\mathbf{q}_0$ is a regular point ($v_0 \neq 0$) and hence the curvature function of $\gamma^*(t)$ is smooth at t_0.

If a cusp is encountered at $\mathbf{q}_0$, it cannot be the critical point $\mathbf{p}_0$ of U^*. In general, the nonzero vector $\nabla U^*(\mathbf{q}_0)$ actually gives the outgoing direction of the cusp, which by Proposition 5.1 is directed "southward" and hence λ_{γ^*} is strictly increasing (resp. decreasing) when γ^* approaches (resp. leaves) $\mathbf{q}_0$.

In the other case, γ^* and Γ' are tangent to each other at $\mathbf{q}_0$. By reversal of time if necessary, we may assume that the velocity vector of γ^* at $\mathbf{q}_0$ points in the "eastward" direction of Γ', that is, the positive direction of Γ' as the oriented boundary of the circular cap containing $\mathbf{p}_0$. Geometrically speaking, a local maximum of λ_{γ^*} at t_0 means exactly that the geodesic curvature K_0 of γ^* at $\mathbf{q}_0$ is strictly less than that of Γ', which is a positive constant k_0. Thus, it suffices to show $K_0 < k_0$, and we claim, in fact, that $K_0 \leq 0$.

Suppose to the contrary that K_0 is at least positive, and recall Theorem 3.1 and its identity (3.19), but scaled with M^* as the unit sphere. Then, on the one hand

$$(\sin^2 \alpha)K_0 \geq 0,$$

and on the other hand, the positive normal vector to γ^* is

$$\nu^* = \frac{-\mathbf{T}(\mathbf{q_0})}{|\mathbf{T}(\mathbf{q_0})|}$$

and, by Proposition 5.1, evaluation at $\mathbf{q_0}$ yields

$$-\frac{d}{d\nu^*} \ln(U + h) = -\frac{1}{U + h}\nabla U^* \cdot \nu^* = \frac{1}{(U + h)}\nabla U^* \cdot \frac{\mathbf{T}}{|\mathbf{T}|} > 0.$$

This implies that

$$2(\sin^2 \alpha)K_0 - \frac{d}{d\nu^*} \ln(U + h) > 0$$

and hence contradicts the identity (3.19). Consequently, K_0 cannot be positive and, in particular, we conclude that $\lambda_{\gamma^*}(t_0)$ is a local maximum.

Corollary 5.1 *Suppose that $\gamma^*(t)$ is the associated shape curve of a 3-body motion with vanishing angular momentum, without ever encountering a triple collision or escape to infinity. Then it contains infinitely many eclipse points (resp. local maxima and local minima for λ_{γ^*}), and they occur at alternating sequences of times.*

6 The Asymptotic Behavior at a Triple Collision

3-Body motions leading to a triple collision have vanishing angular momenum, and their moduli curves $\overline{\gamma}(t)$ are exactly those geodesic curves in $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ leading to the base point $O = (\rho = 0)$ as the limit. Therefore, it is also natural to review and study their basic asymtotic properties at the triple collision.

The classical works of Sundman and Siegel tell us that the triple collision is the only essential singularity of 3-body motions, and their asymptotic theorem, briefly stated as Theorem 6.1 below, gives a qualitative description of the behavior at the singularity. We mention here some major works in the classical literature which have contributed to the understanding of the collision motions, namely Sundman [14, 15], Levi-Civita [8], Siegel [10, 11], Siegel-Moser [12] and Wintner [16]. Unfortunately, the proofs one finds in the above literature are rather long and difficult, and thus it is worthwhile to provide simpler proofs, as well as improvements of their results, in the setting of kinematic geometry.

The asymptotic theorem is, in fact, a direct consequence of the asymptotic estimates of I and its lowest derivatives $\dot{I}$ and $\ddot{I}$, and since $\rho = \sqrt{I}$ is the kinematic distance to the triple collision (base) point, these lower order asymptotic estimates are needed somehow for any proof of the above theorem. On the other hand, the theorem can actually be regarded as the geometric interpretation of such estimates.

Along the way, we shall also give remarks on the works of Sundman and Siegel, and in the final subsection we shall apply Wintner's idea of using a logarithmic time scale to deduce the asymptotic formulae for the time derivatives of I of any order (cf. Theorem 6.2).

6.1 Ray solutions as a model for the asymptotic behavior at a triple collision

We begin with some vector algebra in the Euclidean space $\mathbb{R}^3_{(1)} \oplus \mathbb{R}^3_{(2)} \oplus \mathbb{R}^3_{(3)} = \mathbb{R}^9$ of all triples $\delta = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, or rather in the subspace of m-triangles defined by $\sum m_i \mathbf{a}_i = 0$, equipped with the Jacobi metric (2.4), cross product and exterior product

$$\delta \times \delta' = \sum m_i \mathbf{a}_i \times \mathbf{b}_i \in \mathbb{R}^3, \quad \delta \wedge \delta' = \sum \mathbf{a}_i \wedge \mathbf{b}_j \in \wedge^2 \mathbb{R}^9,$$

where the standard basis vectors $\mathbf{e}_r \wedge \mathbf{e}_s \in \mathbb{R}^3_{(i)} \wedge \mathbb{R}^3_{(j)}$ has length $\sqrt{m_i m_j}$. Some useful relationships between these operations are expressed by

$$|\delta \wedge \delta'|^2 = \det \begin{vmatrix} \delta \cdot \delta & \delta \cdot \delta' \\ \delta \cdot \delta' & \delta' \cdot \delta' \end{vmatrix} = |\delta|^2 |\delta'|^2 - (\delta \cdot \delta')^2 \geq |\delta \times \delta'|^2,$$

$$|\delta \wedge \delta'|^2 = \sum m_i^2 |\mathbf{a}_i \times \mathbf{b}_i|^2 + \sum_{i<j} m_i m_j \mu_{ij}, \quad \mu_{ij} \geq 0.$$

In particular, for a motion $\delta(t)$ with velocity $\dot{\delta}(t)$ and individual angular momenta $\mathbf{\Omega}_i$, we deduce the relations

$$|\delta \wedge \dot{\delta}|^2 = \sum |\mathbf{\Omega}_i|^2 + |\mathbf{\Omega}_{\mathrm{mix}}|^2 = 2IT - \frac{1}{4}\dot{I}^2 \geq |\mathbf{\Omega}|^2. \tag{6.1}$$

These can also be interpreted in tems of the splitting of kinetic energy

$$T = T^\rho + (T^\sigma + T^\omega) = \frac{1}{8}\frac{\dot{I}^2}{I} + \frac{|\delta \wedge \dot{\delta}|^2}{2I},$$

where T^ρ, T^σ or T^ω is due to radial motion, change of shape, or (rigid) rotational motion, respectively. In the case of planary motions, equality holds in (6.1) since $T^\omega = \frac{|\mathbf{\Omega}|^2}{2I}$. In particular, equality holds if $\mathbf{\Omega} = 0$ since in that case the motion is planary, e.g., by a simple geometric argument.

Ray solutions provide, of course, the simplest examples of triple collision motions. Here each particle moves along a fixed line through the center of gravity (origin) and hence $\mathbf{\Omega}_i = 0$ for each i, and also $T^\sigma = T^\omega = 0$. By Remark 3.1, the shape of the ray is a critical point of U^* on the 2-sphere M^*, namely a Lagrange point or an Euler point. In other words, the motion is either a homothetic deformation of an equilateral triangle or a degenerate triangle which is an Euler configuration. Let μ be the value of U^* at the above critical point. Then the Lagrange-Jacobi equation (1.5) reads

$$\ddot{I} = 2\mu\frac{1}{\sqrt{I}} + 4h \tag{6.2}$$

and hence the only variable of the problem, $I(t)$, is the solution of a 1-dimensional Kepler problem.

The equation (6.2) can be solved explicitly, but we seek the solutions with the (singular) initial condition $I(0) = 0$. In the special case of $h = 0$,

$$I(t) = Kt^{\frac{4}{3}}, \quad K = \left(\frac{9\mu}{2}\right)^{\frac{2}{3}}, \tag{6.3}$$

and for general h there is a formula $t = F_h(I)$ which can be inverted and, for example, this yields a series development of type

$$I(t) = t^{\frac{4}{3}}\left(K + \sum_{i=1}^{\infty} k_i t^{\frac{2i}{3}}\right).$$

To facilitate our study of asymptotic estimates in general, let us introduce the commonly used notation

$$f \sim g \Longleftrightarrow \lim \, f/g = 1, \quad \text{as } t \to 0. \tag{6.4}$$

Then all ray solutions have the same asymptotic behavior at $t = 0$, in the sense that their time derivatives of $I(t)$ at $t = 0$ yield the same asymptotic formulae, beginning with

$$I(t) \sim Kt^{\frac{4}{3}}, \quad \dot{I}(t) \sim \frac{4}{3}Kt^{\frac{1}{3}}, \quad \ddot{I} \sim \frac{4}{9}Kt^{-\frac{2}{3}}, \tag{6.5}$$

and clearly the higher order asymtotic formulae follow the same pattern, namely

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}I(t) \sim \frac{\mathrm{d}^k}{\mathrm{d}t^k}Kt^{\frac{4}{3}}, \quad k = 0, 1, 2, 3, \cdots. \tag{6.6}$$

For a general triple collision motion, it was first realized that $\mathbf{\Omega}$ must vanish, and hence also T^ω vanishes. On the other hand, although T^ρ was found to be the dominating kinetic energy, T^σ cannot vanish for a non-radial motion and may perhaps tend to infinity at some lower order of magnitude. However, although the general asymptotic behavior is certainly more involved due to the change of shape, it turns out that the estimates (6.6) still hold, by Theorem 6.2. Moreover, a general triple collision motion has one of the above simple ray solutions as its asymptotic limit, according to Theorem 6.1.

Exact information on the limiting behavior of the shape is not really needed to derive the asymptotic formulae (6.6) for the motion in the radial direction. The theory of Sundman and Siegel establishes the formulae only up to $k = 2$, namely the asymptotic estimates (6.5). To proceed from $k = 0, 1$ to $k = 2$ Siegel introduced the following function $g(t)$ and proved its crucial property

$$g(t) = (8IT - \dot{I}^2)t^{-\frac{2}{3}} \to 0, \quad \text{as } t \to 0 \quad (\text{cf. [11, Chapter III, §1]}),$$

which in our setting can be reformulated as

$$\frac{|\delta \wedge \dot{\delta}|^2}{\rho} = 2\rho T^\sigma = \frac{1}{4}\rho^3 v^2 \to 0, \quad \text{as } t \to 0, \tag{6.7}$$

where v is the speed of the shape curve. From our viewpoint, we recognize the expression in (6.7) as the Siegel function $\mathfrak{S}$ of the associated shape curve $\delta^*(t)$ on the 2-sphere (cf. (3.41)).

6.2 The results of Sundman and Siegel

The following basic fact on the vanishing of the angular momentum of 3-body motions leading to triple collision had already been stated by Weierstrass when Sundman first proved the following classical statement at the beginning of the 20th century.

Lemma 6.1 (Sundman) *The angular momentum* $\mathbf{\Omega}$ *is necessarily zero for a triple collision motion.*

Proof   First, by translation and (possibly) reversal of time, we shall rather assume (in all Section 6) there is a triple explosion at $t = 0$. Using the Lagrange-Jacobi equation (1.5), it follows from $I \to 0$ that $\ddot{I} \to \infty$ and hence $\dot{I} > 0$ for $t \in (0, t_0)$ and t_0 suitably small. Sundman discovered and made use of the rightmost inequality in (6.1). Namely, for a given value $|\mathbf{\Omega}| > 0$ it is not so difficult to see that $I(t)$ has a positive lower bound.

However, we shall proceed with a slightly different proof since (6.1) also involves the individual momenta $\mathbf{\Omega}_i$, and this will enable us to prove a stronger version of the lemma (see Corollary 6.1). For that purpose, we set

$$C(t) = |\delta \wedge \dot{\delta}|^2, \quad C_0 = \inf C \quad \text{for } t \in (0, t_0).$$

Now, multiplying the Lagrange-Jacobi equation by $\dot{I}$ gives the inequality

$$\dot{I}\ddot{I} = \dot{I}(2T + 2h) = \frac{\dot{I}}{I}C + 2h\dot{I} + \frac{1}{4I}\dot{I}^3 \geq \frac{\dot{I}}{I}C + 2h\dot{I},$$

which by integration yields

$$\int_t^{t_0} \dot{I}\ddot{I}\,\mathrm{d}t \geq \int_t^{t_0} \frac{\dot{I}}{I}C\mathrm{d}t + 2h(I_0 - I),$$

$$\frac{1}{2}\dot{I}_0^2 \geq \frac{1}{2}(\dot{I}_0^2 - \dot{I}^2) \geq C_0 \ln\left(\frac{I_0}{I}\right) + 2h(I_0 - I),$$

$$C_0 \leq \frac{2hI - 2hI_0 + \frac{1}{2}\dot{I}_0^2}{\ln(\frac{I_0}{I})} \to 0, \quad \text{as } t \to 0.$$

Hence, $C_0 = 0$ and the constant sum $\boldsymbol{\Omega} = \sum \boldsymbol{\Omega}_i$ is zero.

Remark 6.1 Clearly, the above proof also gives inf $|\boldsymbol{\Omega}_i| = 0$ for each i, but one cannot yet conclude that $\boldsymbol{\Omega}_i \to 0$. See Corollary 6.1 for this last step.

The major results of Sundman and Siegel concerning a general triple collision motion $\delta(t)$ can be summarized as follows.

(1) (Sundman) $t^{-\frac{2}{3}}\overline{\delta}(t)$ tends to a limit whose shape is that of an equilateral triangle or an Euler configuration.

(2) (Siegel) The magnified or "big triangle" $t^{-\frac{2}{3}}\delta(t)$ approaches a fixed m-triangle $\widetilde{\delta}_0$ in the Euclidean configuration space M.

Standard references for proofs of these statements are [11, 12]. One finds that results proved by Sundman are, typically, seen to express properties at the moduli space level, that is, statements about the moduli curve $\overline{\delta}(t)$. Siegel improved his results by lifting them up to the configuration space level, where he studied the motion of the "big triangle" in the Hamiltonian setting and performed a series of successive canonical transformations to simplify the analysis.

A major step was to establish the validity of the above asymptotic estimates (6.5), and with the following proposition we shall provide a proof of this — in the spirit of Sundman and Siegel. Moreover, for the sake of completeness, in the last subsection we shall extend the proof to the higher order asymptotic estimates, $k > 2$ in (6.6), using ideas due to Wintner.

Proposition 6.1 *For a 3-body motion $\delta(t)$ with $I(t) \to 0$ as $t \to 0$, the asymptotic estimates in (6.5) hold, where K is the expression in (6.3) with $\mu = \lim U^*(\delta^*(t))$.*

Proof For a ray solution, $I = \rho^2 \sim Kt^{\frac{4}{3}}$ and $\rho\dot{\rho}^2 \sim \frac{4}{9}K^{\frac{3}{2}}$. This suggests a study of the asymptotic behavior of $\rho\dot{\rho}^2$ for triple collisions in general, using equation (1.5) and kinematic geometry. Now, $\boldsymbol{\Omega} = 0$ and the kinetic energy has the splitting

$$T = \frac{1}{2}\left(\frac{\mathrm{d}\overline{s}}{\mathrm{d}t}\right)^2 = T^\rho + T^\sigma = \frac{1}{2}\dot{\rho}^2 + \frac{1}{8}\rho^2 v^2 \quad \text{(cf. (2.10))}, \tag{6.8}$$

and we set

$$\mathfrak{R}(t) = \rho T^\rho, \quad \widetilde{\mathfrak{S}}(t) = \frac{1}{2}\mathfrak{S}(\mathfrak{t}) = \rho T^\sigma, \tag{6.9}$$

where $\mathfrak{S}$ is the Siegel function (see (3.32), (3.41)). By the energy integral $U^* = \rho T - \rho h$, we can also write

$$U^* \sim \mathfrak{R}(t) + \widetilde{\mathfrak{S}}(t), \quad \text{as } t \to 0. \tag{6.10}$$

Our first claim is that

$$\mathfrak{R}(t) \to \mu > 0 \quad \text{and} \quad \widetilde{\mathfrak{S}}(t) \to 0, \tag{6.11}$$

as $t \to 0$. Let us differentiate $\mathfrak{R}$ and substitute for $\ddot{I}$ using (1.5), or equivalently (3.3)(i), to obtain

$$\dot{\mathfrak{R}} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\rho}{2}\dot{\rho}^2\right) = -\frac{1}{2}\dot{\rho}^3 + \dot{\rho}(U + 2h) = \dot{\rho}(T^\sigma + h),$$

$$\int_t^{t_0} \dot{\mathfrak{R}} \, \mathrm{d}t = \mathfrak{R}(t_0) - \mathfrak{R}(t) = \int_t^{t_0} \dot{\rho} T^\sigma \mathrm{d}t + h(\rho(t_0) - \rho(t)). \tag{6.12}$$

Since $\mathfrak{R} \geq 0$ and the integral on the right side is ≥ 0, $\lim \mathfrak{R}(t) = \mu \geq 0$ must exist, that is, $\mathfrak{R} = \mu + o(1)$ for small t.

Suppose $\mu = 0$. Since $\min U^* > 0$, the equation (3.3)(i) implies

$$\ddot{\rho}\rho^2 = U^* + 2h\rho - 2\mathfrak{R} = U^* + o(1) \geq C > 0$$

and consequently

$$\dot{\rho}\ddot{\rho} = \frac{\dot{\rho}}{\rho^2}(\ddot{\rho}\rho^2) = \frac{\dot{\rho}}{\rho^2}(U^* + o(1)) \geq C\frac{\dot{\rho}}{\rho^2},$$

$$\dot{\rho}(t_0)^2 - \dot{\rho}(t)^2 = 2\int_t^{t_0} \dot{\rho}\ddot{\rho}\mathrm{d}t \geq 2C\int_t^{t_0} \frac{\dot{\rho}}{\rho^2}\mathrm{d}t = 2C\left(\frac{1}{\rho(t)} - \frac{1}{\rho(t_0)}\right) \to \infty.$$

This is clearly impossible, so we conclude $\mu > 0$.

Next, we deduce successively

$$\frac{1}{2}\rho\dot{\rho}^2 = \mu + o(1) \implies \dot{\rho}^2 = \frac{1}{\rho}(2\mu + o(1)) \implies \dot{\rho} = \sqrt{2\mu}\,\rho^{-\frac{1}{2}} + o(\rho^{-\frac{1}{2}}), \tag{6.13}$$

$$\frac{2}{3}\rho^{\frac{3}{2}} = \int_0^\rho \sqrt{\rho}\,\mathrm{d}\rho = \int_0^t \dot{\rho}\sqrt{\rho}\,\mathrm{d}t = \int_0^t (\sqrt{2\mu} + o(1))\mathrm{d}t = \sqrt{2\mu}\,t + o(t) \tag{6.14}$$

and hence by (6.13) and (6.14)

$$I = \rho^2 \sim Kt^{\frac{4}{3}}, \quad K = \left(\frac{9}{2}\mu\right)^{\frac{2}{3}}, \tag{6.15}$$

$$\dot{I} = 2\rho\dot{\rho} \sim \frac{4}{3}Kt^{\frac{1}{3}}, \quad T^\rho = \frac{1}{2}\dot{\rho}^2 \sim \frac{\mu}{\rho}. \tag{6.16}$$

Next we show $\widetilde{\mathfrak{S}}(t) \to 0$. The integral on the right side of (6.12) exists, but the integrand is

$$\dot{\rho}T^\sigma = \frac{\dot{\rho}}{\rho}\widetilde{\mathfrak{S}} \sim \frac{2}{3}\frac{\widetilde{\mathfrak{S}}}{t}$$

and the integral of $\frac{1}{t}$ is divergent, hence $\liminf \widetilde{\mathfrak{S}}(t) = 0$.

It is, however, more difficult to show $\limsup \widetilde{\mathfrak{S}}(t) = 0$, but let us apply an idea of Siegel [11]. Namely, suppose to the contrary, that $\limsup \widetilde{\mathfrak{S}}(t) > 0$. Then, for a given $\epsilon > 0$ there is

an infinite decreasing sequence of numbers in $(0, t_0)$, $\epsilon > t_1 > t_2 > \cdots > t_k > 0$, $\lim t_i = 0$, so that

$$\epsilon \leq \widetilde{\mathfrak{S}}(t) \leq 3\epsilon \quad \text{for } t \in J_k = [t_{2k}, t_{2k-1}],$$
$$\widetilde{\mathfrak{S}}(t_{2k}) = \epsilon, \quad \widetilde{\mathfrak{S}}(t_{2k-1}) = 3\epsilon, \tag{6.17}$$
$$|\mathfrak{R}(t_{2k}) - \mathfrak{R}(t_{2k-1})| \leq \epsilon.$$

By (6.10), in each interval J_k, U^* and hence also the norm of ∇U^* are bounded by the same constant C_1, and then it is not difficult to show

$$\dot{T} = \dot{U} = O(t^{-\frac{5}{3}}) \quad \text{for } t \in J_k.$$

But for small t we also have

$$T = U + h = \frac{1}{\rho}(U^* + h\rho) \leq \frac{C_2}{\rho} \tag{6.18}$$

for a suitable constant C_2, consequently

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}(\rho T)\right| = |\rho\dot{T} + \dot{\rho}T| \leq \frac{C_2}{t} \quad \text{for } t \in J_k. \tag{6.19}$$

By (6.17) and (6.19),

$$2\epsilon = \widetilde{\mathfrak{S}}(t_{2k-1}) - \widetilde{\mathfrak{S}}(t_{2k}) = (\rho T - \mathfrak{R})_{t=t_{2k-1}} - (\rho T - \mathfrak{R})_{t=t_{2k}} \leq C_2 \int_{t_{2k}}^{t_{2k-1}} \frac{\mathrm{d}t}{t} + \epsilon,$$

which implies

$$\int_{t_{2k}}^{t_{2k-1}} \frac{\widetilde{\mathfrak{S}}(t)}{t}\,\mathrm{d}t \geq \epsilon \int_{t_{2k}}^{t_{2k-1}} \frac{\mathrm{d}t}{t} \geq \frac{\epsilon^2}{C_2} \quad \text{for each } k, \tag{6.20}$$

and hence the sum of the integrals is infinite. On the other hand,

$$\frac{\widetilde{\mathfrak{S}}(t)}{t} = \frac{\rho T^\sigma}{t} \sim \frac{3}{2}\dot{\rho}T^\sigma,$$

and the integral of $\dot{\rho}T^\sigma$ on $[0, t_0]$ exists by (6.12), so this is a contradiction.

Having proved that $\widetilde{\mathfrak{S}}(t) \to 0$, it follows from (6.10) that $U^* \to \mu$, and now the Lagrange-Jacobi equation yields

$$\ddot{I} \sim 2T \sim 2T^\rho = \dot{\rho}^2 \sim \frac{4}{9}Kt^{-\frac{2}{3}}, \tag{6.21}$$

where K is the expression from (6.15). This completes the proof.

Corollary 6.1 *The quantity $|\delta \wedge \dot{\delta}|^2$ in (6.1) tends to zero at the triple collision. In particular, the individual angular momenta $\mathbf{\Omega}_i$ as well as the "mixed" momentum term $\mathbf{\Omega}_{\mathrm{mix}}$ tend to zero.*

Since $\mathfrak{S}(t) \to 0$, the above statement follows immediately from

$$|\delta \wedge \dot{\delta}|^2 = 2IT - \frac{1}{4}\dot{I}^2 = 2IT^\sigma = 2\rho\widetilde{\mathfrak{S}} \to 0 \quad (\text{cf. (6.1)}). \tag{6.22}$$

By "infinite magnification" at the triple collision the solution $\delta(t)$ coincides with one of the ray solutions in Subsection 6.1. This is the idea behind the classical asymptotic theorem, and now we give a simple proof of this in the setting of kinematic geometry.

Theorem 6.1 (Sundman-Siegel) *Any triple collision orbit is asymptotic to one of the ray solutions.*

Proof   By (3.20) and (6.21),

$$2T = \left(\frac{\mathrm{d}\overline{s}}{\mathrm{d}t}\right)^2 \sim \frac{4}{9}Kt^{-\frac{2}{3}} \sim \left(\frac{\mathrm{d}\rho}{\mathrm{d}t}\right)^2 = 2T^\rho,$$

$$\cos^2\alpha = \left(\frac{\mathrm{d}\rho}{\mathrm{d}\overline{s}}\right)^2 \to 1, \quad \text{as } t \to 0, \tag{6.23}$$

which simply means that the moduli curve $\overline{\delta}(t)$ of the given triple collision motion $\delta(t)$ is tangent to a ray or, equivalently, the limit of its infinite magnification exists. It also follows that the limit ray must itself be a geodesic in $\overline{M}_h$, namely one of those rays representing the shape of a Lagrange or Euler configuration.

However, the claim is also that $\delta(t)$ itself approaches a ray in M. To see this, consider as above the angle $\widetilde{\alpha}$ between the radial and tangential direction in M, that is, the angle between the vectors $\delta(t)$ and $\dot{\delta}(t)$. It follows that

$$\cos\widetilde{\alpha} = \frac{\delta \cdot \dot{\delta}}{|\delta||\dot{\delta}|} = \frac{\dot{\rho}}{|\dot{\delta}|} = \frac{\dot{\rho}}{\sqrt{2T}} \to 1, \quad \text{as } t \to 0. \tag{6.24}$$

The limit ray in M projects to a geodesic ray in $\overline{M}_h$, namely a ray consisting of the homothetic images of either a fixed equilateral triangle or a fixed degenerate triangle of Euler's type. As shown in Subsection 6.1, these are the rays which admit triple collision motions, and thus the given motion $\delta(t)$ will be asymptotic to the corresponding limit ray solution with the same energy h.

Corollary 6.2 *The shape curve $\delta^*(t)$ converges to a Lagrange or Euler point δ_0^* on the 2-sphere, and the "big triangle" $t^{-\frac{2}{3}}\delta(t)$ converges to an m-triangle $\widetilde{\delta}_0$ with the shape δ_0^* and moment of inertia*

$$\widetilde{I}_0 = K = \left(\frac{9\mu}{2}\right)^{\frac{2}{3}}, \quad \mu = U^*(\delta_0^*). \tag{6.25}$$

Remark 6.2 Actually, a limiting shape of Euler's type cannot be reached unless the whole 3-body motion itself is collinear (cf. e.g. [12, §13]), which refines and improves the classical Sundman-Siegel approach. The latter is described in detail in Siegel's lectures (cf. [11]) of about 240 pages. In [11, p. 138], he writes: "The difficulty of the problem consists in the fact that we cannot yet prove (this will be proved only at the end) that the big triangle referred to a fixed coordinate system has a limiting position as $t \to 0$; all that we have proved so far is the existence of a limiting configuration relative to a rotating coordinate system. The triangle itself may go on rotating about its centre of gravity, $\cdots$"

However, although finiteness of the rotation of the "big triangle" was proved via the convergence of the "big triangle", neither an estimate of the actual angle of rotation nor its precise definition was addressed in the above studies of the triple collision. In reality, the "big triangle" is approaching its final shape and position quite fast and in a monotonic way. To make this precise, we propose to measure how much the equilateral limiting triangle $\widetilde{\delta}_0$ deviates in

position from some natural reference equilateral m-triangle ζ, depending on the given collision motion $\delta(t)$, but also $\zeta = \zeta(t)$ will be a function depending on the chosen time interval $[0, t]$ under consideration. This goes as follows.

We may assume the shape curve $\delta^*(t)$ is on the northern hemisphere and hence starts at the Lagrange point $\mathbf{p}_0 = \delta_0^*$. Let $\delta_1 = \delta(t_1)$ be the m-triangle at a given time $t_1 > 0$ and write $\delta_1^* = \delta^*(t_1)$. Then there is a unique linear m-triangle motion

$$Z(t) = \frac{(t_1 - t)}{t_1}\zeta_1 + \frac{t}{t_1}\delta_1, \quad t \in [0, t_1],$$

with vanishing angular momentum, connecting δ_1 to some equilateral m-triangle $\zeta_1 = \zeta(t_1)$ (cf. [5, Subsection 3.3]). The shape curve of this (virtual) motion is the geodesic arc on the sphere $S^2(1)$ from the Lagrange point $\mathbf{p}_0$ to the point δ_1^*, and together with the curve segment of $\delta^*(t)$ from δ_1^* to $\mathbf{p}_0$ they constitute a closed curve C_1 on the sphere. We define the rotation angle $\psi(t_1)$ of $\delta(t)$ at the triple collision, measured from time $t = t_1$, to be half of the signed area

$$\psi(t_1) = \frac{1}{2}\mathrm{Area}(D_1) = \int_{C_1} \omega \tag{6.26}$$

of the region D_1 enclosed by C_1. This is motivated by the kinematic Gauss-Bonnet theorem (cf. [5]) for 3-body motions with zero angular momentum, where traversal of a loop on the 2-sphere amounts to a net rotation (i.e. a geometric phase) of the m-triangle in the configuration space, which can be calculated as the line integral of a kinematic 1-form ω (depending on the mass distribution and region of S^2). Moreover, $\mathrm{d}\omega = \frac{1}{2}\mathrm{d}A$, where $\mathrm{d}A$ is the area form of the unit sphere.

Now, for t_1 not too large, the shape curve $\delta^*(t)$ will stay on one side of the geodesic arc since its curvature will have a fixed sign. So the rotation angle (6.26) decreases monotonically to zero as $t_1 \to 0$, and hence the kinematic geometric approach explains Siegel's angle of rotation and yields as well a recipe for how to measure it quantitatively.

6.3 Higher order asymptotic estimates at a triple collision

The asymptotic formulae for the energy functions $\Xi = T, T^\rho, U$ and their time derivatives up to order k can be developed inductively together with those formulae for I up to order $k+2$. However, from the three identities

$$T^\rho = \frac{\dot{I}^2}{8I}, \quad T = U + h = \frac{1}{2}\ddot{I} - h, \tag{6.27}$$

it is easy to show that the three cases of Ξ, for a given order k, yield the same asymptotic formula. Therefore, the final description of the asymptotic behavior of the above quantities can be stated as follows.

Theorem 6.2 *For a 3-body motion $\delta(t)$, $t \geq 0$, with a triple collision at $t = 0$, the following asymptotic estimates hold as $t \to 0$:*

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}I \sim \frac{\mathrm{d}^k}{\mathrm{d}t^k}(Kt^{\frac{4}{3}}), \quad \frac{\mathrm{d}^k}{\mathrm{d}t^k}T \sim \frac{\mathrm{d}^k}{\mathrm{d}t^k}\left(\frac{\mu}{\rho}\right) \sim \frac{\mathrm{d}^k}{\mathrm{d}t^k}\left(\frac{2}{9}Kt^{-\frac{2}{3}}\right) \quad \textit{for all } k \geq 0,$$

where $K = (\frac{9\mu}{2})^{\frac{2}{3}}$, $\mu = U^*(\delta_0^*)$, and δ_0^* is the limiting shape.

From the initial cases $k = 0, 1, 2$, proved in the previous subsection, we shall complete the proof of the above theorem for $k > 2$ by deducing the following equivalent formulae for the behavior of $\rho = \sqrt{I}$,

$$\frac{d^k}{dt^k}\rho \sim \frac{d^k}{dt^k}(\sqrt{K}t^{\frac{2}{3}}), \quad k \geq 3. \tag{6.28}$$

In fact, they will follow inductively as a rather direct consequence of Newton's equation (1.2) and its energy integral, namely,

$$\nabla U(\delta) = \frac{d^2}{dt^2}\delta, \quad T = U + h, \tag{6.29}$$

but only after an appropriate transformation of space and time. This is the composition of a time dependent space transformation and a pure time transformation, as follows:

(1) Magnification of the motion $\delta(t) = (\mathbf{a}_1(t), \mathbf{a}_2(t), \mathbf{a}_3(t))$ by the time factor $t^{-\frac{2}{3}}$, as in the works of Sundman and Siegel, to assure convergence at $t = 0$ of the magnified motion. We use the notation

$$\widetilde{\delta}(t) = t^{-\frac{2}{3}}\delta(t) = (\widetilde{\mathbf{a}}_1, \widetilde{\mathbf{a}}_2, \widetilde{\mathbf{a}}_3), \quad \widetilde{\mathbf{a}}_i(t) = t^{-2/3}\mathbf{a}_i(t), \tag{6.30}$$

$$f(t) = f(\delta(t)), \quad \widetilde{f}(t) = f(\widetilde{\delta}(t)), \quad \widetilde{\delta}_0 = \lim_{t \to 0}\widetilde{\delta}(t), \tag{6.31}$$

where f is any (homogeneous) function on M or its tangent bundle which we shall evaluate along the trajectory.

(2) A logarithmic transformation of time; set

$$u = -\log t \quad (\text{or } t = e^{-u})$$

and hence $t \to 0$ means $u \to \infty$. This transforms a function $g(t)$ to the function $\breve{g}(u) = g(e^{-u})$.

The composition of the two transformations yields the motion $u \to \widehat{\delta}(u)$ in M, and we write

$$\widehat{\delta}(u) = \widetilde{\delta}(e^{-u}) = (\widehat{\mathbf{a}}_1, \widehat{\mathbf{a}}_2, \widehat{\mathbf{a}}_3), \quad \widehat{\mathbf{a}}_i(u) = e^{\frac{2}{3}u}\mathbf{a}_i(e^{-u}),$$
$$\widehat{f}(u) = f(\widehat{\delta}(u)), \quad \widehat{\rho}(u) = |\widehat{\delta}(u)| = e^{\frac{2}{3}u}\rho(e^{-u}), \quad \widehat{\delta}_0 = \widetilde{\delta}_0. \tag{6.32}$$

This motion is, of course, a solution of the transformed equations in (6.29), which can be stated as

$$\nabla \widehat{U} = \frac{d^2}{du^2}\widehat{\delta} - \frac{1}{3}\frac{d}{du}\widehat{\delta} - \frac{2}{9}\widehat{\delta}, \tag{6.33}$$

$$\widehat{T} = \widehat{U} + he^{-\frac{2}{3}u} - \frac{2}{9}\widehat{\rho}^2 + \frac{1}{3}\widehat{\rho}\frac{d}{du}\widehat{\rho}, \tag{6.34}$$

with the appropriate interpretation of $\widehat{T}$, $\widehat{U}$ and $\nabla\widehat{U}$ (cf. (6.37), (6.38)). For example, from the above definitions

$$\widehat{U}(u) = U(\widehat{\delta}(u)) = e^{-\frac{2}{3}u}U(\delta(e^{-u})) = t^{\frac{2}{3}}U(t).$$

To derive the above equations and prepare for its usage, we shall make a few more definitions and establish some useful identities for differential operators generated by $\frac{d}{dt}$ and $\frac{d}{du}$. For functions of t (or u) it is convenient to write

$$f_1 \approx f_2 \iff (f_1 - f_2) = o(1), \quad \text{as } t \to 0 \ (\text{or } u \to \infty),$$

and, for example, since the magnified motion converges,

$$\widehat{\delta} \approx \widetilde{\delta}_0, \quad \widehat{\rho} \approx |\widetilde{\delta}_0| = \sqrt{K}. \tag{6.35}$$

We say $f(t)$ has order q at $t = 0$ if

$$\frac{f(t)}{t^q} \to f_0 \neq 0, \quad \text{as } t \to 0,$$

and then the notation

$$\widetilde{f}(t) = t^{-q} f(t) \approx f_0 \tag{6.36}$$

is consistent with (6.31) since a homogeneous function g of degree d on M, with $g(\widetilde{\delta}_0) \neq 0$, has order $q = \frac{2d}{3}$ at $t = 0$.

The transformed potential function, kinetic energy, and gradient are given by

$$\widehat{U} = U(\widehat{\delta}) = \sum_{i<j} \frac{m_i m_j}{|\widehat{\mathbf{a}}_i - \widehat{\mathbf{a}}_j|}, \quad \widehat{T} = \frac{1}{2} \sum m_i \left| \frac{d}{du} \widehat{\mathbf{a}}_i \right|^2, \tag{6.37}$$

$$\nabla \widehat{U} = \left(\frac{1}{m_1} \frac{\partial \widehat{U}}{\partial \widehat{\mathbf{a}}_1}, \cdots \right) = t^{\frac{4}{3}} \nabla U, \tag{6.38}$$

and by substituting these expressions together with

$$\frac{d^2}{dt^2} \delta = \frac{d^2}{dt^2} (t^{\frac{2}{3}} \widetilde{\delta}) = t^{-\frac{4}{3}} \left(-\frac{2}{9} \widehat{\delta} - \frac{1}{3} \frac{d}{du} \widehat{\delta} + \frac{d^2}{du^2} \widehat{\delta} \right)$$

into the equations (6.29) one obtains the system (6.33)–(6.34).

Lemma 6.2 *If $f(t)$ has order q at $t = 0$, with $\widetilde{f}(t) \approx f_0$ and $q \notin \{0, 1, 2, \cdots\}$, then there is the equivalence*

$$\left[1 \leq k \leq m, \ t^k \frac{d^k}{dt^k} \widetilde{f}(t) \approx 0 \right] \iff \left[1 \leq k \leq m, \ \frac{d^k}{dt^k} f(t) \sim \frac{d^k}{dt^k} (f_0 t^q) \right].$$

Proof By applying the Leibniz formula

$$\frac{d^m}{dt^m} \widetilde{f}(t) = \frac{d^m}{dt^m} (t^{-q} f(t)) = \sum_{i=0}^{m} \binom{m}{i} \frac{d^{m-i}}{dt^{m-i}} (t^{-q}) \frac{d^i}{dt^i} f(t),$$

one proves the above equivalence by induction on m. We refer to [13, Lemma 6.1] for a detailed proof.

Moreover, using the operator identity

$$t^k \frac{d^k}{dt^k} = (-1)^k \left(n_{k,1} \frac{d}{du} + \cdots + n_{k,k} \frac{d^k}{du^k} \right), \quad n_{k,k} = 1, \ n_{k,i} \in \mathbb{Z}, \tag{6.39}$$

associated with the logarithmic time change, $t \to u = -\log t$, one can verify the following equivalence

$$\left[1 \le k \le m, \ t^k \frac{\mathrm{d}^k}{\mathrm{d}t^k} g \approx 0\right] \Longleftrightarrow \left[1 \le k \le m, \ \frac{\mathrm{d}^k}{\mathrm{d}u^k} \breve{g} \approx 0\right]. \tag{6.40}$$

The reason for introducing the change of variable $t \to u$ is the following useful lemma of Tauberian type.

Lemma 6.3 (cf. [16, #363]) *Let $f(u)$ be defined for $u > 0$ and assume $f(u)$ has a limit and $\frac{\mathrm{d}^2}{\mathrm{d}u^2} f(u)$ is bounded as $u \to \infty$. Then $\frac{\mathrm{d}}{\mathrm{d}u} f(u) \to 0$ as $u \to \infty$.*

Finally, we turn to the proof of the asymptotic formulae (6.28). The "initial" data needed to start up are provided by Proposition 6.1 and (6.35), which by Lemma 6.2 and (6.40) can be restated as

$$\widehat{\rho} \approx \sqrt{K}, \quad \widehat{U} \approx \frac{\mu}{\sqrt{K}}, \quad \frac{\mathrm{d}}{\mathrm{d}u} \widehat{\rho} \approx \frac{\mathrm{d}^2}{\mathrm{d}u^2} \widehat{\rho} \approx 0. \tag{6.41}$$

Then, by the equation (6.34), we first deduce $\widehat{T} \approx 0$, or equivalently $\frac{\mathrm{d}}{\mathrm{d}u} \widehat{\mathbf{a}}_i \approx 0$ for each i. Moreover, by (6.41) each $\widehat{\mathbf{a}}_i$ is bounded, $\widehat{U}$ is bounded and hence $|\widehat{\mathbf{a}}_i - \widehat{\mathbf{a}}_j|$ has a lower bound when $i \ne j$. It follows that all partial derivatives of $\widehat{U}$, with respect to components of $\widehat{\mathbf{a}}_i$ and of any order, are bounded (as functions of u). In particular, in the equation (6.33) $\nabla \widehat{U}$ is bounded and hence also $\frac{\mathrm{d}^2}{\mathrm{d}u^2} \widehat{\delta}$ is bounded.

Now, apply the operator $\frac{\mathrm{d}}{\mathrm{d}u}$ repeatedly to the equation (6.33) and deduce

$$\frac{\mathrm{d}^k}{\mathrm{d}u^k} \nabla \widehat{U} \text{ is bounded } \Longrightarrow \frac{\mathrm{d}^{k+1}}{\mathrm{d}u^{k+1}} \widehat{\delta} \text{ is bounded}, \quad \text{for all } k \ge 1.$$

By the above Tauberian lemma, it follows that

$$\frac{\mathrm{d}^k}{\mathrm{d}u^k} \widehat{\delta} \approx 0 \quad \text{for all } k \ge 1.$$

Similarly, apply $\frac{\mathrm{d}}{\mathrm{d}u}$ successively to the equation (6.34) and deduce that the highest derivative of $\widehat{\rho}$ is always bounded. Hence by the Tauberian lemma

$$\frac{\mathrm{d}^k}{\mathrm{d}u^k} \widehat{\rho} \approx 0 \quad \text{for all } k \ge 3. \tag{6.42}$$

By the equivalence (6.40) and Lemma 6.2, the statement (6.42) is equivalent to the statement of (6.28), and this completes the proof of Theorem 6.2.

Remark 6.3 The asymptotic estimates in Theorem 6.2 are also valid for a general collision (i.e. total collapse) of an n-body motion, for any $n \ge 2$. The proof is esssentially the same as above and the previous subsection, since we have used only the Riemannian cone structure of the moduli space $\overline{M}$ and, for example, the angle α is similarly defined for any $n > 2$. In fact, the actual structure of the shape space is irrelevant as far as the asymptotic behavior of the radial motion is concerned. Moreover, the exponent ν of t in the formula $I \sim Kt^\nu$ is independent of n, but depends on the degree $-e$ of homogenity of the potential function, $U \sim \frac{1}{r^e}$, namely $\nu = \frac{4}{2+e}$ where we assume $0 < e < 2$, and $e = 1$ is the Newtonian case. We refer to [13].

7 A Brief Discussion of Some Open Problems

In this concluding section, we shall formulate and explain some natural open problems in the present geometric setting. Recall that the trajectories of 3-body motions with zero angular momentum are already uniquely determined up to congruence by their associated moduli curves, which can be characterized (geometrically) as geodesic curves in $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$. Furthermore, these geodesics together with their time evolution are essentially determined by their shape curve on the 2-sphere. Finally, we recall two major results concerning shape curves, namely the unique parametrization theorem and the monotone m-latitude theorem (cf. Subsections 4.2 and 5.3). Therefore, from now on shape curve means geometric shape curve, unless otherwise specified, and we also assume they are oriented.

The geometric behavior of these spherical curves raises many interesting questions for an in depth understanding of 3-body motions. Here we propose a few natural problems of basic importance.

7.1 Shape curves of periodic motions with vanishing angular momentum

The study of periodic orbits is naturally a central topic of the 3-body problem as a whole. Clearly, the moduli curve and the shape curve of such a motion are periodic, but the converse may not be true. Therefore, we say the 3-body motion is congruence periodic or shape periodic if the moduli curve or shape curve, respectively, is periodic with respect to time. However, it is an important consequence of the unique parametrization property that the time parametrization of these curves is dictated by the geometry of the shape curve, whenever the latter is non-exceptional. Then the notion of congruence periodic is the same as shape periodic, and this means the shape curve is periodic in a geometric sense which we explain as follows.

Since it is natural to allow binary collision points, periodic shape curves can be characterized as the topologically closed shape curves. Namely, the curve is either the immersion of a circle, and we call it circular periodic, or the immersion of a closed interval (of length > 0) and is contractible, and we call it string periodic. In the latter case the curve is a "string" with two end points which are either a reversing cusp (i.e. at the Hill's boundary) or a collision point, and in order to qualify as a periodic curve it is tacitly assumed that we take two copies of the "string" with the opposite orientation.

Remark 7.1 A shape curve consisting of a single point p is an exceptional case, and there are additional string periodic moduli curves of the fixed shape p. Namely, p must be a Lagrange or an Euler point, and the ray solution with negative energy, starting at rest from (ρ_0, p) on the Hill's boundary (cf. Subsection 7.4), leads directly to the triple collision point O. By traversing this ray segment in both directions we obtain a string periodic moduli curve.

Next, we shall describe the distinction between periodic 3-body motions and shape periodic motions (i.e. of circular or string type). Let γ^* be a closed (piecewise smooth) curve on the 2-sphere which is the shape curve of a motion $\gamma(t)$, $t_0 \leq t \leq t_1$, of m-triangles with vanishing angular momentum, and assume γ^* is periodic as above. Then $\gamma(t_0)$ and $\gamma(t_1)$ are congruent

m-triangles and hence differ only by a rotation angle $\Delta\psi$, which we can calculate as a line integral along the shape curve, according to the kinematic Gauss-Bonnet theorem (see (6.26)).

In particular, $\Delta\psi$ is zero if the shape curve is string periodic, and hence the given motion $\gamma(t)$ must also be periodic. Thus the notions of "periodic" and "shape periodic" are identical in this case. On the other hand, a circular periodic curve γ^* encloses a signed area ΔA (depending on orientation and self-intersections), and the above line integral can also be expressed as a surface integral which yields

$$\Delta\psi = \frac{1}{2}\Delta A. \tag{7.1}$$

Therefore, the motion $\gamma(t)$ is periodic if and only if the angle (7.1) is a rational multiple of 2π, say $\Delta\psi = (\frac{p}{q})2\pi$ with $(p, q) = 1$, and hence the number $q(t_1 - t_0)$ is the period of the motion.

Thus the study of periodic 3-body trajectories is completely reduced to the study of closed shape curves on the 2-sphere, and it is a challenge to describe or characterize the various types of these curves in terms of simple geometric invariants. For example, due to the monotonicity theorem it is natural to regard the number of eclipse points (counted with multiplicity) as a measure of the complexity of the curve, and hence the simplest curves are characterized by a small number of eclipse points.

Problem 7.1 *What is the minimal number of eclipse points on a (string or circular) periodic shape curve? What even numbers can be realized? What are those periodic curves with a small number of eclipse points, say up to* 10 *?*

The homotopy classes of closed curves inside $P = S^2 - \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ are elements of the fundamental group $\pi_1(P)$, namely the free group of two generators.

Problem 7.2 *What are those homotopy classes of closed curves in P which can be represented by circular periodic shape curves?*

Definition 7.1 *We propose to define the chaoticity of γ^* to be the following value*

$$\mathrm{ch}(\gamma^*) = \frac{\mathrm{Area}(D(\gamma^*))}{\mathrm{Area}(S^2)},$$

where $D(\gamma^*) \subset S^2$ is the closure of the set γ^*, and we say γ^* is chaotic or non-chaotic if $\mathrm{ch}(\gamma^*) > 0$ or $\mathrm{ch}(\gamma^*) = 0$, respectively.

Problem 7.3 *What are the possible values of chaoticity for shape curves representing motions with $\mathbf{\Omega} = 0$?*

Problem 7.4 *What are the non-chaotic shape curves other than the periodic ones?*

7.2 Triple collisions

The works of Sundman and Siegel show that triple collisions is the only type of essential singularity of 3-body motions, while the binary collisions can be regularized analytically (cf. [15, 8]). Recall that the singularities of the Newtonian potential function U in the moduli space $\overline{M} \simeq \mathbb{R}^3$ consists of the triple of rays $\{\overrightarrow{O\mathbf{b}_i}, i = 1, 2, 3\}$, where the base point (or origin)

O represents the triple collision and the rays represent the three types of binary collisions. Moreover, the moduli curves of triple collision motions with total energy h are exactly those geodesic curves in $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ with the point O as a limit. The following are some pertinent problems on the geometry of such geodesics.

Problem 7.5 *The existence (resp. uniqueness) problem on the shortest path in $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ linking a given point p in $\overline{M}_h$ to the base point O.*

By the scaling symmetry we may assume $h = 0, \pm 1$, where the case of $h = -1$ is most difficult and also most interesting. The existence of such a shortest geodesic curve between O and a given point p in $\overline{M}_h$ can be proved by Hilbert's direct method when $h = 0$ or 1, whereas for the case of $h = -1$ the existence will depend on the position of p in the Hill's region $\overline{M}_h$. Note that Hilbert's direct method also applies for those p with

$$d(p, O) < d(p, \partial \overline{M}_h) + d(\partial \overline{M}_h, O).$$

The uniqueness problem is, however, much more interesting than the existence problem, but it is also much more difficult and subtle. For a geodesic $\overline{\gamma}$ starting from O, the question is how far out $\overline{\gamma}$ is the unique shortest geodesic from O. We remark that for points lying in the eclipse plane, the shortest geodesic is not in the eclipse plane, and hence the limiting shape at O of the shortest geodesic must be a Lagrange point, say $\mathbf{p}_0$. By the monotonicity of its shape curve $\overline{\gamma}$ will eventually reach the eclipse plane, but after the first eclipse $\overline{\gamma}$ ceases to be of shortest length. Hence, the best we can hope for is uniqueness up to the first eclipse point.

We propose to investigate first the case of $h = 0$, due to the scaling invariance of this energy level. Then the general uniqueness of a shortest geodesic between any point p and O reduces to the uniqueness for eclipse points p lying at the distance $\rho = 1$ from O, namely for points on the eclipse circle E^*. Thus the problem is reduced from the moduli space $\overline{M}$ to the shape space M^*, namely we ask about the uniqueness of such shape curves between points on E^* and $\mathbf{p}_0$.

Problem 7.6 *For the case of energy level $h = 0$ and for a given mass distribution, let S be the set of triple collision moduli curves emanating from O, whose shape curve starts out from the Lagrange point $\mathbf{p}_0$ on the upper hemisphere of $M^* = S^2$. Let S^* be the initial arcs of the shape curves from $\mathbf{p}_0$ (but $\mathbf{p}_0$ not included) to their first point on the equator circle E^*. Is the family of curves S^* a foliation of the punctured upper hemisphere $S_+^2 - \{\mathbf{p}_0\}$?*

In the case that there exists a unique shortest geodesic in $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ linking a given m-triangle δ to the collapsed m-triangle O, it is certainly interesting to actually estimate its length, initial direction and total rotation angle of the triangle in terms of its geometric invariants. See also the last part of Subsection 6.2.

7.3 Binary collisions and nearby trajectories

The base point O is, of course, the only singularity for $(\overline{M}, \mathrm{d}\overline{s}^2)$. But the Riemannian manifold $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ has another kind of singularity along the triple of rays $\{\overrightarrow{O\mathbf{b}_i}, i = 1, 2, 3\}$ minus the initial point O, say, of binary collision type. One expects that understanding of the

geometry of geodesic curves in the vicinity of this type of singularity will be an important topic in the study of the global geometry of geodesics on $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$.

Problem 7.7 *What kind of local analysis will enable us to provide an effective control on the local geometry of geodesic curves in the vicinity of a singular ray of a given binary collision type?*

Let us make some further remarks. In the vicinity of the ray $\overrightarrow{O\mathbf{b}}_i$ the gradient vector field ∇U is closely approximated by the field ∇U_i, where

$$U = \sum U_i, \quad U_i = \frac{m_j m_k}{r_{jk}} = \frac{(m_j m_k)^{\frac{3}{2}}}{\sqrt{1 - m_i}} \frac{1}{d_i}$$

and $d_i = d_i(x) = \rho \sin \sigma_i$ is the distance in $(\overline{M}, \mathrm{d}\overline{s}^2)$ between x and the ray $\overrightarrow{O\mathbf{b}}_i$ (cf. (2.16)). Thus, there is a suitable rotationally symmetric metric which provides a good approximation of $\mathrm{d}\overline{s}_h^2$ when we are close to such a ray singularity. Application of Noether's theorem to this simpler metric yields a first integral of its geodesic equation which is almost constant along a geodesic segment of $(\overline{M}_h, \mathrm{d}\overline{s}_h^2)$ near $\overrightarrow{O\mathbf{b}}_i$. This will serve as a useful auxiliary function whose analysis will provide an effective control on the above local geometry.

7.4 Trajectories starting at the boundary of the Hill's region

In the case of negative energy, say $h = -1$, the variety $\overline{M}_h$ is the Hill's region, namely the proper subset of the moduli space from which the moduli curves of the 3-body motions cannot leave. The region is enclosed by its boundary, namely Hill's surface $\partial \overline{M}_h$ which is the smooth surface defined by $\rho = U^*(\varphi, \theta)$, with the (kinematic) gradient field ∇U as a normal field.

Here, we shall focus on those geodesics of the metric $\mathrm{d}\overline{s}_h^2$ starting at the surface $\partial \overline{M}_h$, where the metric becomes identically zero. Hence, a curve lying on $\partial \overline{M}_h$ has zero length, and a minimizing curve containing a segment on $\partial \overline{M}_h$ is only virtual and cannot, of course, represent an actual trajectory of a 3-body motion. Therefore, in the study of variational problems of this kind one often needs a certain estimate or geometrical control of those geodesic curves starting at $\partial \overline{M}_h$, that is, the moduli curves of those 3-body motions with no kinetic energy at $t = t_0$. They constitute a family $\{\overline{\gamma}_{x_0}\}$ of geodesics parametrized by their initial points $x_0 \in \partial \overline{M}_h$.

Following Jacobi, it is natural to study the variational vector fields along $\overline{\gamma}_{x_0}$ with respect to variations within the above family of geodesics. These vector fields are solutions of the Jacobi equation along $\overline{\gamma}_{x_0}$ with their initial vectors belonging to $T_{x_0}(\partial \overline{M}_h)$.

Problem 7.8 *Let x_0 be a generic point of $\partial \overline{M}_h, h = -1$, and let $\overline{\gamma}_{x_0}$ be the geodesic curve with initial point x_0. How do we obtain an effective (that is, simple and useful) lower bound estimate of the distance between x_0 and the first zero point of Jacobi vector fields of the above type, in terms of the geometric invariants at x_0?*

7.5 On the problem of fundamental segments

For a fixed energy level $h = 0, \pm 1$, consider the family $\Sigma(h)$ of all oriented geometric shape curves, with the exceptional ones removed, of 3-body motions with zero angular momentum.

According to the monotone m-latitude theorem the curve γ^* can be viewed as a union of its segments $C_i = (p_i, q_i)$, between two consecutive points p_i and q_i of extremal m-latitude. We shall refer to them as the fundamental segments. Thus the end points p_i, q_i lie on opposite hemispheres, unless one of them is a binary collision point (and hence lies on the equator circle), and moreover, the m-latitude is strictly monotonic along the segment. Clearly, a global shape curve can be regarded as being pieced together by such fundamental segments, and a periodic shape curve has only a finite number of them.

Conversely, we may try to construct curves by connecting C_i to C_{i+1} in a "smooth" way. Here C_i and C_{i+1} belong to $\Sigma(h)$, so "smooth" means their union also belongs to $\Sigma(h)$. For simplicity, assume we are using only regular fundamental segments $C = (p, q)$, that is, p and q are regular points. Observe that C is tangential to the m-latitude circle at p, so its direction will be completely specified by an index $\varepsilon = 0, 1$ representing "eastward" or "westward" respectively. Thus we can associate to the starting point $p = (\varphi, \theta)$ the following 5-tuple of numbers

$$[p] = (\varphi, \theta, \mathfrak{S}_0, \mathfrak{S}_1, \varepsilon), \tag{7.2}$$

which determines C completely and therefore also the 5-tuple $[q]$ associated to its end point. In (7.2) $\mathfrak{S}_0, \mathfrak{S}_1$ are the Siegel numbers of C at p, as explained in Subsections 4.1 and 4.2.

Roughly speaking, the relationship between the initial data and terminal data for a fundamental segment with regular end points provides a type of correspondence

$$[p] = (\varphi, \theta, \mathfrak{S}_0, \mathfrak{S}_1, \varepsilon) \to [q] = (\pi - \varphi', \theta', \mathfrak{S}_0', \mathfrak{S}_1', \varepsilon')$$

on a dense open set of $S^2 \times \mathbb{R}^+ \times \mathbb{R} \times \{0, 1\}$. Moreover, since p and q are points on opposite hemispheres, let us compose the above correspondence with the reflectional symmetry with respect to the equator circle, namely we replace $[q]$ by $[p'] = (\varphi', \theta', \mathfrak{S}_0', \mathfrak{S}_1', \varepsilon')$. Finally, we assume p (and hence also p') lies on the upper hemisphere, thus arriving at the fundamental correspondence

$$\Theta : \mathfrak{T}_0 \cup \mathfrak{T}_1 \to \mathfrak{T}_0 \cup \mathfrak{T}_1, \quad [p] \to [p'], \tag{7.3}$$

where the $\mathfrak{T}_i$ are identical copies of the 4-dimensional space $\mathfrak{T} = S_+^2 \times \mathbb{R}_+ \times \mathbb{R}$. The correspondence is defined on a dense, open set, where it is also invertible. In fact, with some more labour it would be possible to extend the fundamental correspondence to include irregular points (i.e. cusps and collisions) as well.

Remark 7.2 A periodic shape curve is the assemblage of a finite number of fundamental segments whose initial data constitute a periodic orbit of the above correspondence (7.3). Namely, if $[p]$ has even order $2k$, then the orbit of $[p]$ defines $2k$ fundamental segments which join together to a periodic curve. On the other hand, if the order is $2k + 1$, then the end of the curve lies in the southern hemisphere, so by running through the orbit twice the order will be $4k + 2$, and the associated curve will be periodic.

Thus, the correspondence (7.3) provides a natural way to a systematic study of the geometry of global shape curves.

References

[1] Euler, L., De motu rectilineo trium corporum se mutuo attahentium, *Novi Commun. Acad. Sci. Imp. Petrop.*, **11**, 1767, 144–151.

[2] Fujiwara, T., Fukuda, H., Kameyama, A., Ozaki, H. and Yamada, M., Synchronized similar triangles for three-body orbits with zero angular momentum, *J. Phys. A: Math. Gen.*, **37**, 2004, 10571–10584.

[3] Hsiang, W. Y., Geometric Study of the Three-Body Problem, I, Center for Pure and Applied Math., Univ. of Calif., Berkeley, PAM-620, 1994.

[4] Hsiang, W. Y. and Straume, E., Kinematic Geometry of Triangles with Given Mass Distribution, Center for Pure and Applied Math., Univ. of Calif., Berkeley, PAM-636, 1995.

[5] Hsiang, W. Y. and Straume, E., Kinematic geometry of triangles and the study of the three-body problem, *Lobachevskii J. of Math.*, **25**, 2007, 9–130.

[6] Jacobi, C. G. J., Vorlesungen über Dynamik, Reimer, Berlin, 1866.

[7] Lagrange, J. L., Essai sur le problème de trois corps, *Ouvres*, **6**, 1772, 229–324.

[8] Levi-Civita, T., Sur la régularisation du problème de trois corps, *Acta Math.*, **42**, 1920, 99–144.

[9] Montgomery, R., Infinitely many syzygies, *Arch. Rat. Mech. Anal.*, **164**, 2002, 311–340.

[10] Siegel, C. L., Der Dreierstoss, *Ann. of Math.*, **42**, 1941, 127–168.

[11] Siegel, C. L., Lectures on the Singularities of the Three-body Problem, Tata Institute of Fundamental Research Lectures on Mathematics, No. 42, Tata Institute of Fundamental Research, Bombay, 1967.

[12] Siegel, C. L. and Moser, J., Lectures on Celestial Mathematics, Die Grundlehren der Mathematischen Wissenschaften, Vol. 187, Springer-Verlag, New York, 1971.

[13] Straume, E., On the geometry and behavior of n-body motions, *Internat. J. of Math. and Math. Sci.*, **28**(12), 2001, 689–732.

[14] Sundman, K. F., Recherches sur le problème de trois corps, *Acta Soc. Sci. Fennicae*, **34**, 1907, 144–151.

[15] Sundman, K. F., Mémoire sur le problème de trois corps, *Acta Math.*, **36**, 1912, 105–179.

[16] Wintner, A., The Analytic Foundations of Celestial Mechanics, Princeton Univ. Press, Princeton, 1947.