Lobachevskii Journal of Mathematics
http://ljm.ksu.ru
ISSN 1818-9962
Vol. 25, 2007, 9–130

Wu-Yi Hsiang and Eldar Straume

KINEMATIC GEOMETRY OF TRIANGLES AND THE STUDY OF THE THREE-BODY PROBLEM

(submitted by V. V. Lychagin)

Contents

1. Introduction

The classical three-body problem studies the motion of a system with three point masses under the action of the Newtonian gravitational potential. Let $\mathbf{a}_i = \overrightarrow{OP_i}$, $i = 1, 2, 3$, be the position vectors of the points P_i with masses $m_i > 0$, with respect to a chosen inertial coordinate system for the Euclidean 3-space $\mathbb{R}^3$. Then a motion of the three point masses will be described as a curve in the (unrestricted) *configuration space*, namely the Euclidean space $\mathbb{R}^9$ consisting of all triples $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$.

Newton's equation of motion is the following second order system of three vector differential equations

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

where $\{i, j, k\} = \{1, 2, 3\}$, and

$$U = \frac{m_1 m_2}{r_{12}} + \frac{m_2 m_3}{r_{23}} + \frac{m_1 m_3}{r_{13}}, \quad r_{ij} = \|\mathbf{a}_i - \mathbf{a}_j\|, \tag{2}$$

is the Newtonian potential function. In brief, the central problem is to understand both the geometry and the analysis of the solutions of the above system (1), where each solution curve (or trajectory) is uniquely determined by the initial positions and velocities of the point masses.

1.1. **The classical conservation laws.** The equations (1) amount to solving a dynamical system in phase space of dimension 18. It is easily seen, however, that the classical 3-body problem (in fact, the n-body problem for all n) is invariant under Galilean transformations, namely the 10-dimensional Galilean group of 4-dimensional space-time $\mathbb{R}^3 \times \mathbb{R}$. Accordingly, there are 10 first integrals or *conservation laws,* and they actually reduce the integration problem to one of order $18 - 10 = 8$. These are the conservation of linear momentum, angular momentum and total energy, and they are easily deduced as follows. First, by adding the three vector equations (1) we have

$$m_1\ddot{\mathbf{a}}_1 + m_2\ddot{\mathbf{a}}_2 + m_3\ddot{\mathbf{a}}_3 = 0,$$

and hence the center of mass has the uniform motion

$$\mathbf{a}_{\mathrm{CM}} = \frac{1}{\bar{m}}(m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3) = \mathbf{p}_0 + t\mathbf{v}_0, \quad \bar{m} = \sum m_i,$$

where $\mathbf{p}_0$ and $\mathbf{v}_0$ are two constant vectors (and hence six scalar conservation laws) determined by the initial data. To make effective use of these we recall the Galilean principle of relativity of Newtonian mechanics, according to which one may choose an equivalent inertial frame of reference with origin at $\mathbf{a}_{\mathrm{CM}}$ and axes in the same directions as before (or rotated). This justifies using a center of mass inertial reference frame, which effectively reduces the actual configuration space to the following 6-dimensional Euclidean space

$$M_0 = \mathbb{R}^6 \subset \mathbb{R}^9 : \sum m_i\mathbf{a}_i = 0. \tag{3}$$

Next, the vector

$$\mathbf{\Omega} = m_1\mathbf{a}_1 \times \dot{\mathbf{a}}_1 + m_2\mathbf{a}_2 \times \dot{\mathbf{a}}_2 + m_3\mathbf{a}_3 \times \dot{\mathbf{a}}_3 \tag{4}$$

is the total *angular momentum* of the system. It varies covariantly with rotations of 3-space, and it is also constant along a trajectory since by (1)

$$\dot{\mathbf{\Omega}} = \sum_i m_i\mathbf{a}_i \times \ddot{\mathbf{a}}_i = \sum_i m_i\mathbf{a}_i \times (\sum_{j\neq i} \frac{m_j}{r_{ij}^3}(\mathbf{a}_j - \mathbf{a}_i)) = 0.$$

This is the law of conservation of angular momentum, whose geometric significance lies much deeper than that of conservation of linear momentum.

Finally, the total energy is defined to be

$$h = T - U, \tag{5}$$

where

$$T = \frac{1}{2}(m_1\left|\dot{\mathbf{a}}_1\right|^2 + m_2\left|\dot{\mathbf{a}}_2\right|^2 + m_3\left|\dot{\mathbf{a}}_3\right|^2) \tag{6}$$

is the *kinetic energy* and $-U$ is the *potential energy.* Straightforward differentiation using Newton's equation (1) gives

$$\dot{h} = \dot{T} - \dot{U} = \sum m_i \dot{\mathbf{a}}_i \cdot \ddot{\mathbf{a}}_i - \sum \frac{\partial U}{\partial \mathbf{a}_i} \cdot \dot{\mathbf{a}}_i = \sum \dot{\mathbf{a}}_i \cdot (m_i \ddot{\mathbf{a}}_i - \frac{\partial U}{\partial \mathbf{a}_i}) = 0.$$

Hence, h is constant along a solution curve of (1) and this is the law of conservation of energy.

1.2. **Least action principles.** Newton's equation provides a characterization of the motion from the "differential " viewpoint, but it is also useful to characterize the motion as a boundary value problem, namely we ask:
for a given pair $\{p,q\}$ in the configuration space and time interval (t_1,t_2), what are those trajectories $\Gamma(t), t_1 \leq t \leq t_2$, with $\Gamma(t_1) = p, \Gamma(t_2) = q$?

The idea of seeking solutions of the above problem as the extremals of a *variational principle* applied to virtual motions dates back to the 17th century, inspired by the success of Fermat's *principle of least time* in geometric optics. Thus, a type of least action principle for classical mechanics was proposed already by Leibniz, Euler and Maupertuis. It was, however, Lagrange who finally provided a precise mathematical formulation:
Lagrange's least action principle. *The solutions of the above boundary value problem are characterized by the variational principle of extremizing the action*

$$J_1[\Gamma] = \int_\Gamma T dt \tag{7}$$

among all virtual motions between a given pair of points and with the same constant total energy h.

It should be noted that time is allowed to vary in the above integral, that is, the limit of integration is not fixed. This awkwardness led Jacobi to reformulate the least action principle to the problem of determining the geodesics on a suitably defined Riemannian manifold (in modern terminology). In his famous lectures on mechanics [6], Jacobi essentially introduced the notion of *kinematic metric* on the configuration space M_0, namely the kinetic energy expression (6) defines a mass dependent Euclidean metric

$$ds^2 = 2T dt^2 = \sum m_i(dx_i^2 + dy_i^2 + dz_i^2), \tag{8}$$

where $\mathbf{a}_i = (x_i, y_i, z_i)$ is the position vector of the point P_i with mass m_i.

On the other hand, $T = U + h$ is also a function on M_0 for a fixed energy level h, and this explains the following step:

Jacobi's reformulation of Lagrange's least action principle. *The action integral*

$$\sqrt{2}J_1[\Gamma] = \sqrt{2}\int_\Gamma T dt = \int_\Gamma \sqrt{T} ds = \int_\Gamma \sqrt{U+h} ds = \int_\Gamma ds_h \qquad (9)$$

is the arc-length of the virtual motion Γ in the metric space (M_h, ds_h^2), namely in the subspace

$$M_h = \{p \in M_0; U(p) + h > 0\} \subset M_0 \qquad (10)$$

with the squared arc-length element

$$ds_h^2 = (U+h)ds^2. \qquad (11)$$

Thus, according to Jacobi's "geometrization trick", trajectories of Newton's equation are precisely the geodesics in the above Riemannian space (M_h, ds_h^2). This also demonstrates why Jacobi in his study of mechanics, in fact, anticipated the general notion of a Riemannian metric. For more information on these issues we refer to Lützen[8].

On the other hand, in 1840 Hamilton formulated another least action principle, also inspired by the results of geometric optics.

Hamilton's principle of least action. *The solutions of the above boundary value problem are characterized by the variational principle of extremizing the action integral*

$$J_2[\Gamma] = \int_{t_1}^{t_2} L dt\ , \quad L = T + U \qquad (12)$$

among all virtual motions $\Gamma(t)$ between a given pair of points, for a fixed time interval $[t_1, t_2]$.

In general, the validity of the "integral" viewpoint represented by any chosen variational principle is verified by calculating its infinitesimal limit, which must coincide with (or be equivalent to) Newton's equation. In the case of (9) and (12) respectively, this amounts to the calculation of the geodesic equations of the metric ds_h and the associated *Euler-Lagrange* equations of the above Lagrangian function L, respectively. In both cases it is easily checked that these are equivalent to Newton's equation.

1.3. **An alternative geometric approach.** Traditionally, the three-body problem is usually studied in the framework of Hamiltonian mechanics, canonical transformations and symplectic geometry, based on the least action principle of Hamilton and the Hamilton-Jacobi theory.

Moreover, the specific dynamics due to the Newtonian forces is usually assumed from the very beginning. Our present approach is, however, different from this, roughly for two major reasons:

- Firstly, we focus attention on the purely kinematic properties of virtual three-body motions in a Riemannian geometric setting and in the framework of equivariant differential geometry, and
- secondly, the Newtonian dynamics is introduced as the final step, involving geometric reduction and conformal modification of the kinematic Riemannian structure, based on the least action principle of Lagrange and Jacobi (cf. Jacobi[6], Lecture 6).

Guided by the above program, the first author initiated studies in 1993 and was joined by the second author in 1994. The basic material, covering their work up to the winter of 1995, was presented in the two preprints [3], [4], and further studies of the n-body problem continued in the following years. However, the two basic preprints were never published and, unfortunately, they have had a rather limited circulation in the mathematical community.

On the other hand, in the recent years new and beautiful results on the three-body problem, and the more general n-body problem as well, have appeared in the literature, some of which are deeply related to the above geometric approach. This clearly suggests that new and unsolved problems along these lines are now becoming more feasible. To further stimulate this trend we propose hereby a review of the works [3], [4] from 1994-95, and Chapter 1 -7 of the present paper is, indeed, merely a faithful presentation of their actual contents.

The exposition has been updated by some changes in notation and terminology, together with a restructuring of ideas and proofs in order to unify and enhance the readability of the presentation. For the convenience of the reader, central results are now formulated as main theorems, such as Theorem A, B, . . . ,F, G presented in Section 2.2.

In the final Chapter 8 we have included some additional and selected unpublished material from 1995, mostly concerning the moduli curves of triple collision motions in the special case of energy $h = 0$ (and, for technical reasons, the case of uniform mass distribution). The more general case is stated as an open problem in the last section. In this chapter we have done some explicit calculations which also serve as an illustration of how to apply the setting and the results up to Chapter 7.

The three-body problem has a vast literature with many excellent papers. In particular, the theoretical analysis of the problem and its influence during the last century is overwhelming. Our interest in the problem started with a study of Siegel's monumental analysis of triple collisions involving clever applications of canonical transformations in the Hamiltonian setting. However, we were also astounded by the lack of basic geometric reasoning more directly linked to the kinematic geometry of three-body configurations, and apparently, this seemed to be typical in the more recent literature (such as Marchal's book), of which we had only superficial knowledge. Perhaps it was, after all, worthwhile having a closer look at the underlying geometric structure of mass triangles and their motions in 3-space ?

With this ambition we started, hopefully with no prejudice due to the existing literature, and this lead us to the purely kinematic study described in the first five chapters of this paper, together with some preliminary investigations of the ensuing dynamics due to gravitational forces, most of which is presented in Chapter 7 and 8.

Thus, in our "blindfolded" study during 1994-95 we deliberately avoided and did not consult any paper on the three-body problem (except Siegel's work).This explains why the list of references were almost empty, and we must apologize for that. The present list purposely reflects this former state of affairs, but now there are at least some relevant titles which were available prior to 1995 and which may be useful for the reader. Some of these are also standard references of historical interest. In retrospect, various topics and results discussed in this paper are certainly more or less treated by the many authors who have contributed to the rich literature in classical or celestial mechanics. Therefore, we also apologize to those authors who may feel that we have failed to make the appropriate reference to their work appearing before 1995. On the other hand, in the present paper references newer than 1995 have not been considered.

2. The basic setting and a presentation of the Main Theorems

2.1. **Basic notions and terminology.** Let $\mathbb{R}^3$ denote Euclidean 3-space with the standard basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. A three-body system consists of three labeled point masses (P_i, m_i) in 3-space, and its geometric model is the triangle with vertices P_i and the mass $m_i > 0$ attached to P_i, which we shall refer to as an *m-triangle* (or simply a *triangle*) or *configuration.* An m-triangle will also be identified with its triple $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ of position vectors $\mathbf{a}_i = \overrightarrow{OP}_i$. These triples are usually denoted by boldface letters such as $\mathbf{X}, \mathbf{Y}$, but occasionally we also use the notation δ, δ_1 etc. It is tacitly assumed that a fixed mass distribution (m_1, m_2, m_3) is given, and it is normalized so that $m_1 + m_2 + m_3 = 1$.

Thus, the abundance of individual motions of three point masses jointly combine to a rich variety of *virtual motions* of m-triangles, through which the triangle changes its kinematic invariants such as size, shape and orientation (position) and velocity. This is our starting point for a systematic investigation of the kinematics of m-triangles, as a basis for a geometric approach to the three-body problem.

In addition to previously defined quantities, set

$$\begin{aligned}
&\alpha_j = \text{ the central angle (at origin) opposite to the vertex } P_j \\
&\omega_j = \text{ the (scalar) angular velocity of } \mathbf{a}_j \text{ for planary motion} \\
&I_j = m_j \left|\mathbf{a}_j\right|^2, \quad I = I_1 + I_2 + I_3 = \rho^2 \qquad (13) \\
&C_1 = -m_1 I_1 + m_2 I_2 + m_3 I_3 \ \text{ etc. (cyclic permutation of indices)} \\
&T_j = \frac{1}{2} m_j \left|\dot{\mathbf{a}}_j\right|^2, \qquad T = T_1 + T_2 + T_3, \ \text{ cf. (6)} \\
&\mathbf{\Omega}_j = m_j \mathbf{a}_j \times \dot{\mathbf{a}}_j, \qquad \mathbf{\Omega} = \mathbf{\Omega}_1 + \mathbf{\Omega}_2 + \mathbf{\Omega}_3, \ \text{ cf. (4)} \\
&\Delta = \text{the area of } \Delta(P_1, P_2, P_3), \ \ \Delta_1 = \text{the area of } \Delta(O, P_2, P_3) \text{ etc.}
\end{aligned}$$

where I (respectively I_j) is the total (respectively individual) polar moment of inertia, and similarly $T, \mathbf{\Omega}$ (respectively $T_j, \mathbf{\Omega}_j$) denote kinetic energy and angular momentum as in Section 1.1. See Figure 1.

Certain functions of the mass distribution appear frequently, so we introduce the notation

$$\begin{aligned}
&m_j^* = \frac{1}{2}(1 - m_j)\text{: the dual mass distribution, with } \sum m_j^* = 1 \qquad (14) \\
&\hat{m}_1 = m_2 m_3 \ , \ \hat{m}_2 = m_3 m_1 \ , \ \hat{m}_3 = m_1 m_2
\end{aligned}$$

and the basic elementary symmetric functions of the symbols m_i are

$$\sum m_i = 1, \quad \hat{m} = \sum \hat{m}_i, \ \ \bar{m} = m_1 m_2 m_3. \qquad (15)$$

2.1.1. *Vector algebra and kinematics in the Euclidean space M_0.* We will assume a center of mass reference frame as in Section 1.1, and hence an m-triangle $\mathbf{X}$ is a vector of the 6-dimensional Euclidean configuration space M_0, cf. (3). The zero vector $\mathbf{X} = 0$ represents the one-point triangle (or the "triple collision" configuration), and we say $\mathbf{X}$ is *collinear* or is an *eclipse* configuration (respectively is non-degenerate) if the subspace $\Pi(\mathbf{X}) \subset \mathbb{R}^3$ spanned by the position vectors $\mathbf{a}_i$ has dimension 1 (respectively 2). A *virtual motion* $\mathbf{X}(t)$ (or $\delta(t)$) is a time parametrized curve in M_0, assumed to be (piecewise) differentiable so that its kinetic energy (6) is defined. The *size* of an m-triangle $\mathbf{X}$ is naturally measured by the Euclidean length

$$\rho = \sqrt{I} = |\mathbf{X}| \tag{16}$$

in M_0 with the *kinematic metric* (8), equivalently given by the following inner product of Jacobi type

$$\mathbf{X} \cdot \mathbf{Y} = m_1 \mathbf{a}_1 \cdot \mathbf{b}_1 + m_2 \mathbf{a}_2 \cdot \mathbf{b}_2 + m_3 \mathbf{a}_3 \cdot \mathbf{b}_3, \tag{17}$$

where $\mathbf{X} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, $\mathbf{Y} = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$. For convenience, we define

$$\begin{aligned} \boldsymbol{\omega} \times \mathbf{X} &= (\boldsymbol{\omega} \times \mathbf{a}_1, \boldsymbol{\omega} \times \mathbf{a}_2, \boldsymbol{\omega} \times \mathbf{a}_3), \quad \boldsymbol{\omega} \in \mathbb{R}^3 \\ \mathbf{X} \times \mathbf{Y} &= \sum m_i (\mathbf{a}_i \times \mathbf{b}_i) \end{aligned} \tag{18}$$

and observe the general triple product identity

$$\boldsymbol{\omega} \times \mathbf{X} \cdot \mathbf{Y} = \boldsymbol{\omega} \cdot \mathbf{X} \times \mathbf{Y}, \quad \boldsymbol{\omega} \in \mathbb{R}^3. \tag{19}$$

The infinitesimal generators of the $SO(3)$-action on M_0 are the rotational (or Killing) vector fields

$$\mathbf{X} \rightarrow \boldsymbol{\omega} \times \mathbf{X}, \;\; \boldsymbol{\omega} \in \mathbb{R}^3 \simeq so(3)$$

of fixed angular velocity $\boldsymbol{\omega}$. These vectors are tangential to the $SO(3)$-orbits. Thus, at each $\mathbf{X}$ the tangent space $T_{\mathbf{X}} M_0 \simeq M_0$ has an orthogonal decomposition into *vertical* and *horizontal* vectors, where the vertical ones are the above Killing vectors $\boldsymbol{\omega} \times \mathbf{X}$ and the horizontal vectors $\mathbf{Y}$ are characterized by $\mathbf{X} \times \mathbf{Y} = 0$, due to (19).

For any virtual motion $\mathbf{X}(t)$ in M_0, the velocity vector at each time t has the above type of splitting, namely

$$\dot{\mathbf{X}} = \frac{d}{dt} \mathbf{X} = \dot{\mathbf{X}}^{\omega} + \dot{\mathbf{X}}^{h} = (\boldsymbol{\omega} \times \mathbf{X}) + \dot{\mathbf{X}}^{h}, \tag{20}$$

where $\boldsymbol{\omega} = \boldsymbol{\omega}(t)$ is commonly referred to as the (instantaneous) *angular velocity* of the motion. Correspondingly, kinetic energy splits as the sum

$$T = \frac{1}{2} |\boldsymbol{\omega} \times \mathbf{X}|^2 + \frac{1}{2} \left| \dot{\mathbf{X}}^{h} \right|^2 = T^{\omega} + T^{h} \tag{21}$$

of purely rotational and horizontal kinetic energy, respectively. The motion is called *horizontal* if the velocity is always horizontal.

Using (19) we also deduce the following relationship between the angular momentum and angular velocity of a virtual motion, namely

$$\boldsymbol{\Omega} = \mathbf{X} \times \dot{\mathbf{X}} = \mathbf{X} \times \dot{\mathbf{X}}^{\omega} = \mathbf{X} \times (\boldsymbol{\omega} \times \mathbf{X}). \tag{22}$$

Indeed, to each m-triangle $\mathbf{X}$ is associated the *inertia operator*

$$\mathbb{I}_{\mathbf{X}} : \mathbb{R}^3 \to \mathbb{R}^3, \quad \boldsymbol{\omega} \to \mathbf{X} \times (\boldsymbol{\omega} \times \mathbf{X}) \tag{23}$$

relating the two vectors $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$. This operator is invertible when $\mathbf{X}$ is nondegenerate, whereas $\boldsymbol{\Omega}$ determines $\boldsymbol{\omega}$ modulo a summand along the line $\Pi(\mathbf{X})$ when $\mathbf{X}$ is an eclipse configuration. In any case, the rotational velocity component $\boldsymbol{\omega} \times \mathbf{X}$ in (20) is uniquely determined by $\boldsymbol{\Omega}$ and $\mathbf{X}$. Consequently, the motion is horizontal if and only if $\boldsymbol{\Omega}$ vanishes; in particular, such a motion must be planary (see Remark 23).

The above inertia operator corresponds uniquely to the associated *inertia tensor*

$$\begin{aligned} B_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) &= (\mathbf{u} \times \mathbf{X}) \cdot (\mathbf{v} \times \mathbf{X}) \\ &= \sum m_j (\mathbf{u} \times \mathbf{a}_j) \cdot (\mathbf{v} \times \mathbf{a}_j), \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^3 \end{aligned} \tag{24}$$

which is a bilinear symmetric form on Euclidean 3-space. They are related by the identity

$$B_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) = \mathbb{I}_{\mathbf{X}}(\mathbf{u}) \cdot \mathbf{v}.$$

For example, they provide an orthonormal eigenframe for each m-triangle, and hence a moving eigenframe for a motion of m-triangles, see Theorem D and Section 3.4.

2.1.2. *Oriented m-triangles and their configuration space M.* Since triangles in 3-space can be oriented we propose the following "refinement" of the notion of an m-triangle. Define an *oriented m-triangle* to be a pair $(\mathbf{X}, \mathbf{n})$, where $\mathbf{n} \in S^2 \subset \mathbb{R}^3$ is a unit vector perpendicular to $\Pi(\mathbf{X})$. In particular, a nondegenerate triangle can be oriented in two ways, namely we say the orientation is *positive* (respectively *negative*) if $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{n})$ is a right-handed (respectively left-handed) frame.

Clearly, for an m-triangle motion the orientation may be chosen so that the normals $\mathbf{n}(t)$ vary continuously along the motion, and then the orientation changes to the opposite one as the motion passes (transversely) through an *eclipse* configuration. In the study of *planary* motions we will (tacitly) assume the plane to be the xy-plane, with unit normals $\mathbf{n} = \pm \mathbf{k}$.

Set M_+ (respectively M_-) to be the set of positively (respectively negatively) oriented m-triangles, together with the set E of oriented eclipse configurations, where the latter includes the 2-sphere of orientations of the one-point triangle $\mathbf{X} = 0$. Then the *configuration space of oriented m-triangles* is the union

$$M = M_+ \cup M_- \subset M_0 \times S^2, \qquad E = M_+ \cap M_- \tag{25}$$

and a virtual motion of oriented m-triangles is a parametrized curve on M

$$\Gamma(t) = (\mathbf{X}(t), \mathbf{n}(t)) \tag{26}$$

As a submanifold of $M_0 \times \mathbb{R}^3 \simeq \mathbb{R}^9$, M inherits a Riemannian structure and a natural isometric action of $SO(3)$, also referred to as the *congruence group*. Let us have a closer look at this manifold and the two projection maps

$$M_0 \xleftarrow{\pi_1} M \xrightarrow{\pi_2} S^2,$$

where π_1 is a 2-fold covering over the non-degenerate m-triangles. On the other side, each $(\mathbf{X}, \mathbf{n})$ is $SO(3)$-equivalent to some $(\mathbf{Y}, \mathbf{k})$, where $\mathbf{Y}$ lies in the xy-plane $\mathbb{R}^2$ and hence belongs to

$$\mathbb{R}^4 \subset M_0 : \sum m_i \mathbf{a}_i = 0 \ , \ \mathbf{a}_i \in \mathbb{R}^2. \tag{27}$$

Then it is a useful observation (not mentioned in the 1994-95 preprints) that π_2 is the $SO(3)$-equivariant projection of a homogeneous 4-plane bundle

$$\mathbb{R}^4 \to M \simeq SO(3) \times_{SO(2)} \mathbb{R}^4 \to SO(3)/SO(2) = S^2, \tag{28}$$

where the 4-space in (27) is the fiber over the unit normal $\mathbf{k}$ of the xy-plane, and $SO(2)$ is the rotation group fixing the z-axis.

In (28) M is expressed as the $SO(2)$-orbit space of $SO(3) \times \mathbb{R}^4$, where $h \in SO(2)$ acts by $(g, \mathbf{Y}) \to (gh^{-1}, h\mathbf{Y})$, and the $SO(2)$-orbit of $(g, \mathbf{Y})$ is identified with the oriented m-triangle $(g\mathbf{Y}, g\mathbf{k})$. This is an example of a well known "twisted product" construction.

2.1.3. *Congruence moduli space and shape space.* The orbit space

$$\bar{M} = M/SO(3) = \bar{M}_+ \cup \bar{M}_- \ , \quad \bar{E} = \bar{M}_+ \cap \bar{M}_- \tag{29}$$

is the (congruence) *moduli space*, where the union corresponds to the $SO(3)$-invariant splitting (25) of M, and the *shape space* is the subspace

$$M^* = M^*_+ \cup M^*_- \ , \quad E^* = M^*_+ \cap M^*_- \tag{30}$$

corresponding to m-triangles of fixed size $\rho = 1$. Since an m-triangle (and its congruence classes) is scaled by the size function $\rho \geq 0$ (16), $\bar{M} = C(M^*)$ has a natural structure of a cone over M^*. So, we may regard $\bar{M}$ as the union of two identical cones or "half-spaces" $\bar{M}_\pm$ glued together along their common boundary $\bar{E}$. This will be made more precise below.

Let us first investigate the topology of the above spaces from a trigonometric viewpoint, using the quadratic form (68) representing the squared area of m-triangles. The triple (I_1, I_2, I_3) is, indeed, a complete congruence invariant for (unoriented) m-triangles. There is only one "half-space" in the unoriented case and we may express it as the following cone in $\{I_j\}$-coordinate 3-space

$$\bar{M}_{\pm} \simeq \{(I_1, I_2, I_3) \mid I_j \geq 0, Q(I_1, I_2, I_3) \geq 0\}, \tag{31}$$

where the corresponding shape space $M^*_{\pm}$ is cut out by the plane $I_1 + I_2 + I_3 = 1$. The eclipse variety $\bar{E}$, defined by the condition $Q = 0$, is the cone over E^*.

Now it is not difficult to see that $M^*_{\pm}$ is topologically a closed 2-disk with E^* as boundary circle, and consequently the full shape space (30) is a 2-sphere $M^* \approx S^2$ with a distinguished *equator* circle E^* separating the two hemispheres $M^*_{\pm}$. Note that the triple of central angles

$$(\alpha_{1,}\alpha_2, \alpha_3), \sum \alpha_j = 2\pi$$

is a complete system of invariants for the shape of unoriented m-triangles and hence these angles also yield coordinates for each of the hemispheres.

It follows that the full moduli space $\bar{M}$ is the cone over a 2-sphere and hence is homeomorphic to 3-space,

$$\bar{M} = C(M^*) \simeq C(S^2) = \mathbb{R}^3, \tag{32}$$

in such a way that $\bar{E} \simeq \mathbb{R}^2$ is the coordinate plane $z = 0$, $\bar{M}_+$ is the upper half space $z \geq 0$ and $\bar{M}_-$ is the lower half-space $z \leq 0$.

Finally, from the viewpoint of equivariant geometry, we observe that the pair $\bar{M} \supset M^*$ is the $SO(3)$-orbit space of the vector bundle (28) and its sphere bundle, namely

$$\begin{aligned} M^* &= (SO(3) \times_{SO(2)} S^3)/SO(3) = S^3/SO(2) \simeq S^2 \\ \bar{M} &= (SO(3) \times_{SO(2)} \mathbb{R}^4)/SO(3) = \mathbb{R}^4/SO(2) = C(S^3/SO(2)) \approx \mathbb{R}^3. \end{aligned} \tag{33}$$

For comparison reasons, if we only consider unoriented m-triangles, then the corresponding calculation of the moduli space as an orbit space will yield the closed half-space

$$\bar{M}_{\pm} = M_0/SO(3) = M_0/O(3) = \mathbb{R}^4/O(2) \approx \mathbb{R}^3_{\pm}. \tag{34}$$

2.2. **Statement of the Main Theorems.** In this summary we focus attention on six main topics, each of which is centered around one or two main theorems, labeled by $A, B, C1, C2 \ldots$

2.2.1. *Kinematic geometry of m-triangles and universal sphericality.* For a virtual 3-body motion with vanishing angular momentum, that is, a horizontal motion $\Gamma(t)$ in M, the kinetic energy T depends solely on the moduli curve $\bar{\Gamma}(t)$ in $\bar{M}$, namely in terms of the local coordinates (I_1, I_2, I_3) it is the following "differential" expression

$$\bar{T} = \frac{1}{8}\frac{\dot{I}^2}{I} + \frac{1}{2IQ}\left(\sum_{\text{i mod 3}} m_i I_{i+1} I_{i+2} \dot{I}_i^2 \; - C_i I_i \dot{I}_{i+1} \dot{I}_{i+2}\right), \tag{35}$$

where $Q = 16 m_1^2 m_2^2 m_3^2 \Delta^2$. This is, indeed, a positive definite quadratic form on the tangent bundle of the moduli space $\bar{M}$ and thus naturally defines a *kinematic Riemannian metric*

$$d\bar{s}^2 = 2\bar{T} dt^2. \tag{36}$$

For a general virtual motion the same expression (35) is, in fact, obtained from T by removing the rotational kinetic energy. Therefore, by (21), the horizontal kinetic energy

$$\bar{T} = T^h = \frac{1}{2}\left|\frac{d}{dt}\bar{\Gamma}(t)\right|^2 = T - T^\omega \tag{37}$$

may well be referred to as the kinetic energy in the moduli space.

Both the definition and the formula for the above metric (36) are dependent on the given mass distribution in a rather intricate manner. Therefore, it is a pleasant surprise that such a kinematically defined Riemannian structure $(\bar{M}, d\bar{s}^2)$ turns out to be not only independent of the mass distribution, but it is, in fact, isometric to the Riemannian cone of the Euclidean sphere of radius 1/2, namely

Theorem A *Let $I = \rho^2$ be the moment of inertia and M^* be the subspace of $\bar{M}$ with $I = 1$, and set $d\sigma^2 = d\bar{s}^2|_{M^*}$ to be the restriction of the kinematic metric. Then*

$$d\bar{s}^2 = d\rho^2 + \rho^2 d\sigma^2, \tag{38}$$

$$(M^*, d\sigma^2) \simeq S^3(1)/U(1) = \mathbb{C}P^1 \simeq S^2(1/2), \tag{39}$$

where $S^3(1)$ is the 3-sphere of radius 1 *and $S^3(1) \to \mathbb{C}P^1$ is the classical Hopf fibration.*

The surprising emergence of spherical symmetry in the kinematic Riemannian space $(\bar{M}, d\bar{s}^2)$ for arbitrary mass distribution naturally brings in the classical spherical geometry as a useful tool in the study of the three-body problem. We propose to call this fundamental fact the *universal sphericality* of the kinematic geometry of m-triangles.

The orientation reversing map $(\mathbf{X}, \mathbf{n}) \to (\mathbf{X}, -\mathbf{n})$ of oriented m-triangles induces an isometric involution of $(M^*, d\sigma^2)$ with E^* as its fixed point set, namely the distinguished equator which divides M^* into two hemispheres $M^*_\pm$. On this circle lie the three points $\mathfrak{p}_{ij}$ representing the shape of the three types of binary collisions, cf. (97). Indeed, their relative positions on the circle determine the mass distribution uniquely, see Section 4.4 and (136).

2.2.2. *Unique lifting property.* **Theorem B** *To a given curve $\bar{\Gamma}(t)$ in the moduli space $\bar{M} \setminus \{0\}$, together with a given constant vector Ω and initial configuration $\Gamma(t_0)$, there exists a unique curve $\Gamma(t)$ in M with $\bar{\Gamma}(t)$ as its moduli curve and with Ω as its conserved angular momentum. Moreover, the curve $\Gamma(t)$ can be computed in terms of the C^1-data of $\bar{\Gamma}(t)$.*

Consider the orbit map $\pi : M \to \bar{M}$, and observe that $SO(3)$ acts freely outside the sphere $\pi^{-1}(0) \simeq S^2$ and defines a principal bundle

$$\pi : M \setminus \pi^{-1}(0) \to \bar{M} \setminus \{0\}\,. \tag{40}$$

In particular, above the half-spaces $\bar{M}_\pm \simeq \mathbb{R}^3_\pm$ there are locally trivializing diffeomorphisms

$$SO(3) \times \bar{M}_\pm \to M_\pm. \tag{41}$$

Geometrically speaking, a motion of m-triangles can be represented by a time parametrized curve $\Gamma(t)$ in M (or $M_\pm$) which (locally) consists of two components, namely a *moduli curve* $\bar{\Gamma}(t)$ that records the change of size and shape of the oriented m-triangles, and a *position curve* $\gamma(t)$ in $SO(3)$ that records the change of position. The latter curve is, of course, constrained by the fixed angular momentum.

In the special case of $\mathbf{\Omega} = 0$, namely the horizontal lifting of moduli curves, the proof of Theorem B follows directly from the standard theory of principal G-bundles with a connection (cf. e.g. [5]), applied to the above principal $SO(3)$-bundle. Therefore, we shall rather focus on the general case with $\mathbf{\Omega} \neq 0$ and present two different proofs. The first proof involves the inertia operator (23), and the second proof is an application of Theorem D stated below. We refer to Section 5.2.2.

2.2.3. *Angular velocities and kinematic Gauss-Bonnet formula.*

Theorem C1 *For a planary motion of oriented m-triangles with normal vector* $\mathbf{k}$ *and angular momentum* $\mathbf{\Omega} = \Omega\mathbf{k}$, *let* $\omega_i = \dot{\phi}_i$ *be the individual (scalar) angular velocity of the position vector* $\mathbf{a}_i = \overrightarrow{OP}_i$. *Then*

$$\omega_i = \omega_i^0 + \frac{\Omega}{I}, \quad i = 1, 2, 3 \tag{42}$$

where ω_i^0 *is a "differential" expression purely at the moduli space level, namely*

$$\omega_1^0 = \frac{1}{I}(I_3\dot{\alpha}_2 - I_2\dot{\alpha}_3) \text{ etc. (cyclic permutation of indices)} \tag{43}$$

$$= \frac{1}{8m_1m_2m_3\Delta I}[(C_3I_3 - C_2I_2)\frac{\dot{I}_1}{I_1} - (C_1 + 2m_2I_3)\dot{I}_2 \tag{44}$$
$$+ (C_1 + 2m_3I_2)\dot{I}_3] \quad \text{etc.}$$

Remark 1. *The proof of Theorem C1 holds for non-planary motions as well, that is, the plane* $\Pi(t)$ *of the m-triangle is time dependent. Then* ω_i *stands for the (scalar) angular velocity of the velocity component of* $\mathbf{a}_i$ *in* $\Pi(t)$, *and in formula* (42) Ω *must be replaced by the normal component* $\mathbf{\Omega} \cdot \mathbf{n}(t)$ (*cf. Theorem* D). *We refer to Section* 3.2.1.

We introduce the following three *kinematic 1-forms* on the moduli space $\bar{M} - \{0\}$:

$$\Theta_1 = \frac{1}{I}(I_3 d\alpha_2 - I_2 d\alpha_3) \quad \text{etc. (cyclic permutation of indices)}$$
$$= \frac{1}{8m_1m_2m_3\Delta I}[(C_3I_3 - C_2I_2)\frac{dI_1}{I_1} - (C_1 + 2m_2I_3)dI_2 \tag{45}$$
$$+ (C_1 + 2m_3I_2)dI_3] \quad \text{etc.}$$

In fact, they are invariant under scaling and may therefore be regarded as 1-forms on the shape space M^* via the canonical retraction $\bar{M} \setminus \{0\} \to M^*$. They share the basic property

$$d\Theta_i = 2dA, \quad i = 1, 2, 3, \tag{46}$$

where dA is the area form of the 2-sphere $M^* \simeq S^2(1/2)$. Evidently, the 1-forms have singularities on the eclipse circle E^*.

By suitably combining the kinematic 1-forms on appropriate regions on M^* and applying Green's theorem, the *kinematic Gauss-Bonnet* version as described by the next theorem follows immediately from Theorem C1 and (46).

Theorem C2 *Let the shape curve of a piecewise differentiable motion of oriented m-triangles with* $\Omega = 0$ *constitute the oriented boundary of* *a region D in $M^* \simeq S^2(1/2)$. Then the total change of position of the triangle is a rotation of angle equal to twice the oriented area of D, namely*

$$\Delta\phi_i = \int_{t_0}^{t_1} \omega_i^0 dt = \int_{\partial D} \Theta_i = \iint_D 2dA. \tag{47}$$

The above type of integral (47) is an example of the *geometric phase* in the literature. Its value depends only on the shape curve and is independent of its parametrization. In the case of a planary motion with nonzero angular momentum, however, the total change of position in the above case (47) has an additional term called the *dynamical phase*, namely as a consequence of (42)

$$\Delta\phi_i = \iint_D 2dA + \int_{t_0}^{t_1} \frac{\Omega}{I} dt. \tag{48}$$

2.2.4. *Moving eigenframe and Euler equations for m-triangles.* Let $(\mathbf{X}, \mathbf{n})$ be a nondegenerate oriented m-triangle. Then we can choose eigenvectors of the inertia tensor (24) which constitute a positive orthonormal frame

$$(\mathbf{u}_1, \mathbf{u}_2, \mathbf{n}) \in SO(3),$$

where $\mathbf{u}_1, \mathbf{u}_2$ lie in the plane $\Pi(\mathbf{X})$ and $\mathbf{u}_1 \times \mathbf{u}_2 = \mathbf{n}$. By definition,

$$B_{\mathbf{X}}(\mathbf{u}_1, \mathbf{u}_1) = \lambda_1, \;\; B_{\mathbf{X}}(\mathbf{u}_2, \mathbf{u}_2) = \lambda_2, \quad B_{\mathbf{X}}(\mathbf{u}_1, \mathbf{u}_2) = 0 \tag{49}$$

where the two eigenvalues λ_i may be expressed (cf. Section 3.4) as

$$\{\lambda_1, \lambda_2\} = \frac{1}{2}(I \pm \sqrt{I^2 - 16m_1m_2m_3\Delta^2}) = \frac{I}{2}(1 \pm \sin\varphi), \tag{50}$$

using spherical polar coordinates (φ, θ) on the 2-sphere M^* centered at the north pole $\mathcal{N}$, where φ is the colatitude with $\varphi = 0$ at the pole. The eigenvalue in the normal direction $\pm\mathbf{n}$ is the largest eigenvalue

$$\lambda_3 = \lambda_1 + \lambda_2 = I.$$

To a continuous motion of oriented m-triangles we may choose such an eigenframe

$$\mathfrak{F}(t) = \{\mathbf{u}_1(t), \mathbf{u}_2(t), \mathbf{n}(t)\} \tag{51}$$

varying continuously with the motion. In particular, $t \to \mathfrak{F}(t)$ is also a parametrized curve in $SO(3)$.

Theorem D *Let* $\mathfrak{F}(t)$ in (51) *be a moving eigenframe attached to a differentiable motion $\Gamma(t)$ of m-triangles, with $\mathbf{\Omega}$ as the conserved angular momentum. Then the triple of inner products*

$$g_1 = \mathbf{\Omega} \cdot \mathbf{u}_1, \quad g_2 = \mathbf{\Omega} \cdot \mathbf{u}_2, \quad g_3 = \mathbf{\Omega} \cdot \mathbf{n} \tag{52}$$

satisfy the following system of ODE, namely

$$\begin{aligned}\dot{g}_1 &= g_2\left[(\frac{1}{\lambda_3}-\frac{1}{\lambda_2})g_3+\frac{1}{2}\dot{\theta}\cos\varphi\right]\\ \dot{g}_2 &= g_1\left[(\frac{1}{\lambda_1}-\frac{1}{\lambda_3})g_3-\frac{1}{2}\dot{\theta}\cos\varphi\right]\\ \dot{g}_3 &= g_1g_2(\frac{1}{\lambda_2}-\frac{1}{\lambda_1})\end{aligned} \tag{53}$$

where the numbers $\lambda_i(t)$ are the eigenvalues of the inertia tensor of $\Gamma(t)$ and depend solely on the moduli curve $\bar{\Gamma}(t) = (\rho(t), \varphi(t), \theta(t))$.

Corollary 2. *It follows from* (53) *that $g_1(t_0) = g_2(t_0) = 0$ at just one time t_0 implies that $g_1(t) = g_2(t) = 0$ for all time. Thus, such a motion is planary if and only if the angular momentum vector is perpendicular to the m-triangle at just one time t_0.*

Remark 3. *The system* (53) *is the exact generalization of the classical Euler equations for a rigid body, see e.g. Arnold* [1], *p.*143*, where M_i and the fixed numbers I_i correspond to our g_i and λ_i, respectively. In* (53) *the additional terms are due to the change of shape, and the system is singular where the motion passes through an eclipse configuration, say, with $\lambda_1 = 0$ and hence also $g_1 = 0$. In particular, the eclipse takes place along a line perpendicular to $\mathbf{\Omega}$.*

The triple (g_1, g_2, g_3) is the coordinate vector, with respect to the moving eigenframe, of the constant vector $\mathbf{\Omega} = \Omega\mathbf{k}$. It determines the position of the m-triangle, in particular its normal vector $\mathbf{n}$, up to a rotation around the $\mathbf{k}$-axis by a specific *precession angle* $\chi(t)$. This angle is calculated by quadrature from the formula

$$\dot{\chi} = \frac{\dot{\mathbf{n}}\cdot\mathbf{k}\times\mathbf{n}}{|\mathbf{k}\times\mathbf{n}|^2} = \frac{\Omega}{g_1^2+g_2^2}(\frac{g_1^2}{\lambda_1}+\frac{g_2^2}{\lambda_2}) , \qquad \text{cf. Section 5.2.1.} \tag{54}$$

It follows that the m-triangle motion $\Gamma(t)$ is largely described by two curves on the 2-sphere, namely the shape curve $\Gamma^*(t) = (\varphi(t), \theta(t))$ and the *precession curve*, that is, the curve traced out by the normal vector $\mathbf{n}(t)$. Thus, for the study of non-planar motions it is a basic problem to investigate the relationship between these two curves.

2.2.5. *The reduced Newton's equations.* Now, let us focus attention on the dynamics of three-body motions, namely the Newtonian equation of motion (1). Such motions are, of course, a very special subclass of all the virtual three-body motions considered before.

The Ω*-reduced Newton's equation* is a second order system of ODE for the moduli curves of three-body motions with a fixed angular momentum vector $\mathbf{\Omega}$, namely the three equations (with index i mod 3)

$$\ddot{I}_i = 4T_i - \left(\frac{m_i + 2m_{i+1}}{r_{i,i+1}^3} + \frac{m_i + 2m_{i+2}}{r_{i,i+2}^3}\right) I_i \tag{55}$$
$$- \left(\frac{1}{r_{i,i+1}^3} - \frac{1}{r_{i,i+2}^3}\right)(m_{i+1}I_{i+1} - m_{i+2}I_{i+2})$$

which are easily derived from the system (1). However, the individual kinetic energy terms T_i depend on $\mathbf{\Omega}$, of course, but are otherwise expressed solely at the level of the moduli space $\bar{M}$.

The two cases of planary and non-planary motions differ substantially in complexity, so we will consider them separately. In the following two theorems we assume the initial position $\Gamma(t_0)$ is not a collinear configuration (since otherwise the given initial data will be incomplete).

Theorem E1 *A planary three-body motion $\Gamma(t)$ is completely determined by its moduli curve $\bar{\Gamma}(t)$, initial position $\Gamma(t_0)$ and angular momentum vector. The curve $\bar{\Gamma}(t)$ is a solution of the Ω-reduced Newton's equation* (55), *with kinetic energy terms*

$$T_i = \frac{\dot{I}_i^2}{8I_i} + \frac{1}{2}\omega_i^2 I_i, \tag{56}$$

where ω_i is the i-th individual angular velocity (42).

Conversely, each solution curve of this Ω-reduced Newton's equation can be realized as the moduli curve of a three-body motion in the xy-plane, with a given initial position and normal vector $\Omega\mathbf{k}$ as the conserved angular momentum.

Remark 4. *The above Ω-reduced Newton's equation may, of course, be expressed purely in terms of the coordinates $\{I_j\}$ or the mutual distances $\{r_{ij}\}$, see* (64). *In fact, such an Ω-reduced Newton's equation in terms of coordinates $\{r_{ij}\}$ was derived by Lagrange* [7]. *We refer to Section* 4.3.1 *for another version in terms of spherical coordinates of $\bar{M} \simeq \mathbb{R}^3$ as a cone over the 2-sphere.*

For the statement of the general (e.g. non-planary) version of the above theorem, let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{n}\}$ denote a continuous eigenframe associated with the motion $\Gamma(t)$, and let

$$(g_1, g_2, g_3), \quad \text{where } g_1^2 + g_2^2 + g_3^2 = \Omega^2, \tag{57}$$

be the coordinate vector of $\mathbf{\Omega}$ relative to this frame, as in Theorem D. The individual kinetic energies depend on the components g_k, more precisely, they split into a tangential and normal component

$$T_i = T_i^\tau + T_i^\eta, \tag{58}$$

where the tangential term T_i^τ depends on the normal component g_3 and T_i^η depends on g_1 and g_2. We refer to Section 5.1 and 5.2.3 for a precise description of these quantities.

Theorem E2 *A general three-body motion* $\Gamma(t)$ *is completely determined by its moduli curve* $\bar{\Gamma}(t)$, *initial position* $\Gamma(t_0)$ *and angular momentum vector* $\mathbf{\Omega}$. *The curve* $\bar{\Gamma}(t)$ *is characterized by the* Ω*-reduced Newton's equations* (55) *with kinetic energy terms* T_i (58) *depending on the moving frame coordinates* (57) *of* $\mathbf{\Omega}$ *and are thus coupled with the Euler equations, namely the first order ODE* (53).

2.2.6. *Reduction of the least action principles.* Here we will only consider planary three-body motions and state the associated Ω-reduced least action principles, whose extremals are precisely the moduli curves of those planary three-body motions with a fixed angular momentum $\mathbf{\Omega} = \Omega\mathbf{k}$. In this case the total kinetic energy

$$T = \bar{T} + T^\omega = \bar{T} + \frac{\Omega^2}{2I} \tag{59}$$

and the Lagrange function $L = T + U$ are, in fact, defined at the level of the moduli space $\bar{M}$, and therefore the two action integrals

$$\bar{J}_{1,\Omega} = \int_{\bar{\Gamma}} T dt \ , \quad \bar{J}_{2,\Omega} = \int_{t_1}^{t_2} L dt \tag{60}$$

apply to moduli curves $\bar{\Gamma}(t)$.

Theorem F *The solution curves of the planary* Ω*-reduced Newton's equation can be characterized as the extremal curves of* $\bar{J}_{1,\Omega}$ (*respectively* $\bar{J}_{2,\Omega}$) *applied to curves* $\bar{\Gamma}$ *in* $\bar{M}$ *with fixed end points together with fixed energy* h (*respectively fixed time interval* $[t_1, t_2]$).

2.2.7. *Shape curves of triple collision trajectories.* In Chapter 8 we initiate a general study of the geometry of moduli curves in the vicinity of a triple collision. According to a classical result of Sundman and Siegel, towards the collision these curves $\bar{\Gamma}(t)$ approach a ray solution, which in the generic case has the shape $\pm\hat{\mathbf{p}}_0$ of an (oriented) equilateral triangle. In the simplest case of equal masses, $\pm\hat{\mathbf{p}}_0$ are the poles of the 2-sphere $M^* \simeq S^2$, and hence the corresponding shape curves approach one of the poles. Thus, it is natural to focus attention on the family of shape curves representing a triple collision with the limit shape of $\hat{\mathbf{p}}_0$. Here we state a theorem which is a simplified version of Theorem G_1 stated in Section 8.1.

Theorem G *Assume uniform mass distribution and zero total energy, and consider the family* $\mathfrak{S}$ *of arc-length parametrized shape curves* $\Gamma^*(s), s \geq 0$, *representing 3-body motions with a triple collision at* $s = 0$, *say* $\Gamma^*(0)$ *is the north pole of* S^2. *The family has the following properties:*

(i) *There is a unique curve* $\Gamma^*_{\theta_0}$ *for each initial longitude direction* θ_0 *at the pole, and* $\Gamma^*_{\theta_0}$ *and* $\Gamma^*_{\theta_0+\pi/3}$ *are congruent modulo a rotation of the sphere.*

(ii) *The six meridians representing isosceles triangles belong to the family* $\mathfrak{S}$. *They divide the sphere into six congruent sectors of angular width* $\pi/3$, *and each curve* $\Gamma^*_{\theta_0}$ *stays within a sector, at least until the first eclipse (at the equator).*

(iii) *The curves* $\Gamma^*_{\theta_0}(s)$ *are analytic in* s, *with no singularity before the first eclipse, and* $\Gamma^*_{\theta_0+\pi}(s) = \Gamma^*_{\theta_0}(-s)$.

(iv) *The sign of the curvature of the above shape curves is the same inside a sector, and the sign is the opposite in neighboring sectors.*

3. Basic geometric and kinematic invariants of m-triangles

3.1. **Ceva-type trigonometry.** In classical Greek geometry individual triangles - not their motions and kinematic relations - are the geometric objects of basic importance. A triangle $\Delta(P_1, P_2, P_3)$ is specified by its three vertices P_i, and its congruence properties involve the fundamental geometric concepts "side", "angle" and "area", whose relationships are expressed by trigonometric identities and congruence theorems. In our study, however, we are rather concerned with m-triangles, that is, a positive mass m_i is attached to P_i. Thus it is natural and useful to reformulate the usual trigonometry into a kind of *Ceva-trigonometry*, depending on the given mass distribution.

Let us first establish the following three identities (cf. (13)):

$$\textit{Ceva-area law: } \Delta_j = m_j \Delta, \tag{61}$$

$$\textit{Ceva-sine law: } \frac{\sin \alpha_i}{m_i \|\mathbf{a}_i\|} = \frac{2\Delta}{\|\mathbf{a}_1\| \, \|\mathbf{a}_2\| \, \|\mathbf{a}_3\|}, \; i = 1, 2, 3, \tag{62}$$

$$\textit{Ceva-cosine law } \quad 2\sqrt{m_i m_j}\sqrt{I_i I_j} \cos \alpha_k = -C_k. \tag{63}$$

By calculating cross products such as

$$\begin{aligned} 0 &= \mathbf{a}_1 \times \sum m_j \mathbf{a}_j = m_2 \mathbf{a}_1 \times \mathbf{a}_2 + m_3 \mathbf{a}_1 \times \mathbf{a}_3 \\ &= (2m_2\Delta_3 - 2m_3\Delta_2)\mathbf{n}, \end{aligned}$$

the first law (61) follows directly, and then the sine law (62) follows:

$$2m_3\Delta = 2\Delta_3 = |\mathbf{a}_1|\,|\mathbf{a}_2|\sin\alpha_3 \quad \Longrightarrow \frac{\sin\alpha_3}{m_3\,|\mathbf{a}_3|} = \frac{2\Delta}{\|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\,\|\mathbf{a}_3\|}.$$

Furthermore, consider the "small triangle" with side vectors $\{m_i\mathbf{a}_i\}$, say, with one vertex at the center of mass O and an adjacent side along OP_j for some j. The triangle has outer angles α_i, and by applying the usual cosine law to it we deduce the cosine law (63).

Next, by combining the usual cosine law and its Ceva version, the relationship between the mutual distances s_i and the moments of inertia I_i is

$$s_i^2 = r_{jk}^2 = \frac{(1-m_i)I - I_i}{m_j m_k}, \quad \{i,j,k\} = \{1,2,3\}\,, \tag{64}$$

from which we also deduce

$$I = \sum_{i<j} m_i m_j r_{ij}^2 \; = \frac{m_1^*}{m_2 m_3}C_1 + \frac{m_2^*}{m_3 m_1}C_2 + \frac{m_3^*}{m_1 m_2}C_3, \tag{65}$$

where the first identity is known as Lagrange's formula for the total moment of inertia with respect to the center of mass, and m_i^* are the dual masses (14).

Finally, consider the "Heron" quadratic form

$$H(a,b,c) = 2(ab+bc+ca) - (a^2+b^2+c^2) \tag{66}$$

and recall the classical Heron's formula for the area Δ

$$H(s_1^2, s_2^2, s_3^2) = 16\Delta^2. \tag{67}$$

Set

$$\begin{aligned} Q(I_1,I_2,I_3) = H(m_1I_1, m_2I_2, m_3I_3) &= 2\sum_{i<j} m_i m_j I_i I_j - \sum_j m_j^2 I_j^2 \qquad (68)\\ &= \sum_{i<j} C_i C_j = 4m_1m_2I_1I_2 - C_3^2 \;\text{ etc.} \end{aligned}$$

and consider again the "small triangle" with side vectors $m_i\mathbf{a}_i$. On the one hand, its area $\hat{\Delta}$ is related to Δ by

$$4\hat{\Delta}^2 = m_1^2 m_2^2\,|\mathbf{a}_1 \times \mathbf{a}_2|^2 = 4m_1^2 m_2^2 \Delta_3^2 = 4m_1^2m_2^2m_3^2\Delta^2$$

and on the other hand, $16\hat{\Delta}^2 = Q(I_1,I_2,I_3)$ by (67) and (68). Consequently, we obtain the

$$\textit{Ceva-Heron formula:}\; Q(I_1,I_2,I_3) = 16m_1^2m_2^2m_3^2\Delta^2. \tag{69}$$

3.1.1. *A simple torque formula.* As a simple application of the Ceva-area law (61) we prove the following result concerning the individual torques due to gravitational forces acting at the vertices P_i of a nondegenerate m-triangle $\Delta(P_1, P_2, P_3)$.

Lemma 5. *Let $\mathbf{t}_i$ be the torque of the Newtonian gravitational forces at $P_i, i = 1, 2, 3,$ with respect to the center of mass O. Then*

$$\mathbf{t}_i = \dot{\mathbf{\Omega}}_i = 2m_1m_2m_3\Delta(\frac{1}{r^3_{i,i+1}} - \frac{1}{r^3_{i,i+2}})\mathbf{n}, \qquad i \mod 3,$$

where $\mathbf{n}$ is the unit normal vector so that $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{n})$ is a right-handed frame.

Proof. Let $\mathbf{F}_{12}$ and $\mathbf{F}_{13}$ be the gravitational forces due to the mass points P_2 and P_3 acting on P_1, namely

$$\mathbf{F}_{12} = \frac{m_1m_2}{r^3_{12}}(\mathbf{a}_2 - \mathbf{a}_1), \ \ \mathbf{F}_{13} = \frac{m_1m_3}{r^3_{13}}(\mathbf{a}_3 - \mathbf{a}_1).$$

Then, by definition of torque and the area law (61)

$$\begin{aligned}\mathbf{t}_1 = \mathbf{a}_1 \times (\mathbf{F}_{12} + \mathbf{F}_{13}) &= \frac{m_1m_2}{r^3_{12}}(\mathbf{a}_1 \times \mathbf{a}_2) + \frac{m_1m_3}{r^3_{13}}(\mathbf{a}_1 \times \mathbf{a}_3) \\ = (\frac{m_1m_2}{r^3_{12}}2m_3\Delta)\mathbf{n} - (\frac{m_1m_3}{r^3_{13}}2m_2\Delta)\mathbf{n} &= 2m_1m_2m_3\Delta(\frac{1}{r^3_{12}} - \frac{1}{r^3_{13}})\mathbf{n}\end{aligned}$$

and similarly at the other two vertices. ∎

Corollary 6. *Corollary $\mathbf{t}_1 = 0$ if and only if $r_{12} = r_{13}$, and all $\mathbf{t}_i = 0$ if and only if the triangle is regular (i.e. equilateral).*

3.2. **Analysis of angular velocities and kinetic energies.** In the orthogonal splitting (20) of the velocity of a virtual motion $\mathbf{X}(t) = (\mathbf{a}_1(t), \mathbf{a}_2(t), \mathbf{a}_3(t))$, the horizontal component further splits into two summands

$$\dot{\mathbf{X}}^h = \dot{\mathbf{X}}^\rho + \dot{\mathbf{X}}^\sigma = \frac{\dot{\rho}}{\rho}\mathbf{X} + \dot{\mathbf{X}}^\sigma \tag{70}$$

representing the change of size and shape, respectively, and correspondingly the total kinetic energy splits as

$$T = T^\omega + T^h = T^\omega + (T^\rho + T^\sigma) = \frac{1}{2}\left|\boldsymbol{\omega} \times \mathbf{X}\right|^2 + (\frac{1}{2}\dot{\rho}^2 + T^\sigma). \tag{71}$$

In this chapter we will show that T^h actually equals the expression in (35), and in particular it depends only on the velocity of the image curve in $\bar{M}$. This will justify our definition of T^h as the kinetic energy $\bar{T}$ of the moduli curve, hence also our definition of the kinematic Riemannian metric on $\bar{M}$

$$d\bar{s}^2 = 2\bar{T}dt^2 = d\rho^2 + 2T^\sigma dt^2. \tag{72}$$

Our first proof of Theorem A is by showing that the metric (72) actually transforms to the metric (138).

The differential expression (35), as a function on the tangent bundle of $\bar{M}$, is calculated by eliminating from T its dependence on the angular momentum, namely the rotational energy. In fact, it suffices to consider a class of virtual motions whose term T^ω is easy to calculate and hence eliminate. For this single purpose we could as well assume $T^\omega = 0$ from the outset and simply express T at the moduli space level. However, it is also illuminating to analyze the class of planary motions with a broader perspective.

3.2.1. *Kinematics of planary motions and proof of Theorem C1 and C2.* We assume the motion takes place in the xy-plane and write

$$\boldsymbol{\Omega} = \Omega\mathbf{k}, \quad \boldsymbol{\omega} = \omega\mathbf{k}.$$

In this case (71) reads

$$T = T^\omega + T^\rho + T^\sigma = \frac{1}{2}\frac{\Omega^2}{I} + \frac{1}{2}\dot{\rho}^2 + T^\sigma. \tag{73}$$

On the other hand, from the orthogonal decomposition of each $\dot{\mathbf{a}}_j$ into its rotational and radial component

$$\dot{\mathbf{a}}_j = \omega_j(\mathbf{k} \times \mathbf{a}_j) + \frac{\dot{\rho}_j}{\rho_j}\mathbf{a}_j\ , \quad \text{where } \rho_j^2 = I_j, \tag{74}$$

the total kinetic energy also adds up to

$$T = \frac{1}{2}\sum I_i\omega_i^2 + \frac{1}{8}\sum \frac{\dot{I}_j^2}{I_j}. \tag{75}$$

Therefore, by combining (73) and (75) the "intricate" energy term T^σ, responsible for the change of shape, is given by

$$T^\sigma = \frac{1}{2}\sum I_i\omega_i^2 + (\frac{1}{8}\sum \frac{\dot{I}_j^2}{I_j} - \frac{1}{2}\dot{\rho}^2) - \frac{1}{2}\frac{\Omega^2}{I}. \tag{76}$$

Now, start from the above expression to express T^σ purely in terms of the individual moments of inertia I_j.

Lemma 7. *Let ω_j be the (scalar) angular velocity of $\mathbf{a}_j$. Then*

$$\omega_1 = \frac{1}{I}(I_3\dot{\alpha}_2 - I_2\dot{\alpha}_3) + \frac{\Omega}{I} \quad \textit{etc.} \quad (\textit{cyclic permutation of indices}) \tag{77}$$

$$\dot{\alpha}_1 = \frac{1}{8m_1m_2m_3\Delta}(-2m_1\dot{I}_1 + C_3\frac{\dot{I}_2}{I_2} + C_2\frac{\dot{I}_3}{I_3}) \quad etc. \tag{78}$$

Proof. Set θ_j to be the angle of $\mathbf{a}_j$ with respect to a chosen reference direction in the plane. Then $\omega_j = \dot{\theta}_j$ and

$$\alpha_i = \theta_{i+2} - \theta_{i+1}, \quad \dot{\alpha}_i = \omega_{i+2} - \omega_{i+1} \qquad i \mod 3.$$

The total (scalar) angular momentum sums up to

$$\begin{aligned}\Omega &= \sum m_i(\mathbf{a}_j \times \dot{\mathbf{a}}_j)\cdot \mathbf{k} = \sum \Omega_j = \sum I_j\omega_j \\ &= \mathbf{X} \times (\boldsymbol{\omega} \times \mathbf{X}) \cdot \mathbf{k} = I\omega.\end{aligned}$$

Consequently,

$$I_1\omega_1 = \Omega - I_2\omega_2 - I_3\omega_3 = \Omega - I_2\dot{\alpha}_3 + I_3\dot{\alpha}_2 - (I_2\omega_1 + I_3\omega_1)$$

and this proves formula (77).

Next, by differentiating the Ceva-cosine formula (63) for α_1 with respect to t and use the expression for $\sin\alpha_1$ from the Ceva-sine formula (62) for α_1, we obtain the formula (78). ∎

Substitution of the expressions (78) into (77) also leads to the following formula involving only I_j's, namely

$$\begin{aligned}\omega_1 = \frac{1}{8m_1m_2m_3\Delta I}&[(C_3I_3 - C_2I_2)\frac{\dot{I}_1}{I_1} - (C_1 + 2m_2I_3)\dot{I}_2 \\ &+ (C_1 + 2m_3I_2)\dot{I}_3] + \frac{\Omega}{I} \quad \text{etc.}\end{aligned} \tag{79}$$

and this completes the proof of Theorem C1.

Furthermore, using either (77), (78) and the relation $\dot{\alpha}_1 + \dot{\alpha}_2 + \dot{\alpha}_3 = 0$, or using (79) directly, we calculate

$$\begin{aligned}\frac{1}{2}\sum I_i\omega_i^2 - \frac{\Omega^2}{2I} &= \frac{1}{2I}(I_1I_2\dot{\alpha}_3^2 + I_2I_3\dot{\alpha}_1^2 + I_3I_1\dot{\alpha}_2^2) \\ = \frac{1}{8I}&\left(\sum_{\text{i mod 3}}(\frac{4I_1I_2I_3m_i}{Q} + I_i - I\ )\frac{\dot{I}_i^2}{I_i} + (2 - \frac{4C_iI_i}{Q})\dot{I}_{i+1}\dot{I}_{i+2}\right)\end{aligned} \tag{80}$$

and finally by insertion into (76) we deduce the formula

$$T^\sigma = \frac{1}{2IQ}\left(\sum_{\text{i mod 3}} m_iI_{i+1}I_{i+2}\dot{I}_i^2 \ - C_iI_i\dot{I}_{i+1}\dot{I}_{i+2}\right). \tag{81}$$

Consequently, the metric (72) on $\bar{M}$ may be written

$$d\bar{s}^2 = d\rho^2 + \rho^2 d\sigma^2, \tag{82}$$

where

$$d\sigma^2 = \frac{1}{I^2 Q}\left(\sum_{\text{i mod 3}} m_i I_{i+1} I_{i+2} dI_i^2 - C_i I_i dI_{i+1} dI_{i+2}\right)$$

is the induced metric on the shape space $M^* = (I = 1)$. Indeed, the metric expression $d\sigma^2$ is a tensor on M^* since it is invariant under scaling in $\bar{M}$.

On the other hand, on M^* the relation $I_1 + I_2 + I_3 = 1$ implies $dI_1 + dI_2 + dI_3 = 0$, and therefore the above metric on M^* can be restated as

$$d\sigma^2 = \frac{1}{Q^*}\left\{ \begin{array}{c} [-I_2^2 + (1-m_2)I_2]\, dI_1^2 + [-I_1^2 + (1-m_1)I_1]\, dI_2^2 \\ -\,[2I_1I_2 - (1-m_2)I_1 - (1-m_1)I_2 + m_3]\, dI_1 dI_2 \end{array} \right\}, \quad (83)$$

where Q^* denotes the restriction of Q to M^* and we have used the mass normalization $m_1 + m_2 + m_3 = 1$.

Lemma 8. *The area form of* $(M^*, d\sigma^2)$ *is*

$$dA = \frac{1}{2\sqrt{Q^*}} dI_1 \wedge dI_2.$$

Proof. As usual, the area form expresses as

$$dA = \sqrt{D} dI_1 \wedge dI_2,$$

where

$$D = \frac{1}{Q^{*2}}\left\{ \begin{array}{c} I_1 I_2 (1 - m_1 - I_1)(1 - m_2 - I_2) \\ -\frac{1}{4}\left[2I_1I_2 - (1-m_2)I_1 - (1-m_1)I_2 + m_3\right]^2 \end{array} \right\} = \frac{1}{4Q^*}$$

is the determinant of the metric (83). ∎

Finally, we turn to the kinematic 1-forms (45) on $\bar{M}$, whose definition is suggested by the expressions (79) for the individual angular velocities. Regarded as 1-forms on M^* they are related to the area form by

$$\begin{aligned} d\Theta_1 &= dI_3 \wedge d\alpha_2 - dI_2 \wedge d\alpha_3 \\ &= \frac{1}{2\sqrt{Q^*}}\left\{ \begin{array}{c} dI_3 \wedge (-2m_2 dI_2 + \frac{C_1}{I_3} dI_3 + \frac{C_3}{I_1} dI_1) \\ -dI_2 \wedge (-2m_3 dI_3 + \frac{C_2}{I_1} dI_1 + \frac{C_1}{I_2} dI_2) \end{array} \right\} \\ &= \frac{1}{2\sqrt{Q^*}}\left(\frac{C_2 + C_3 + 2(m_2+m_3)I_1}{I_1}\right) dI_1 \wedge dI_2 \\ &= \frac{1}{\sqrt{Q^*}} dI_1 \wedge dI_2 = 2dA. \end{aligned}$$

This proves formula (46) and, as observed in Section 2.2.3, this also completes the proof of Theorem C2.

3.2.2. *A purely kinematic proof of Theorem A.* From the metric expression (82) it follows that $\bar{M}$ is a Riemannian cone over the shape space $(M^*, d\sigma^2)$, expressed in (30) as the union of two isometric disks along their common boundary circle E^*. Both disks are parametrized by the region $Q^* \geq 0$ in the (I_1, I_2)-plane, where

$$Q^*(I_1, I_2) = Q(I_1, I_2, 1 - I_1 - I_2), \quad 0 \leq I_i \leq 1 \tag{84}$$

is the quadratic form (68) with $I_3 = 1 - I_1 - I_2$. In the following we will describe our original calculations in [4] leading to the discovery of the *universal sphericality* .

At first glance, the mass distribution $\{m_i\}$ is intricately involved in the formula (83) of $d\sigma^2$, so we will focus attention on the mass dependent quadratic form Q^*. The major step of the proof is, in fact, the algebraic approach of seeking better coordinates by transforming the metric tensor $d\sigma^2$ into a simpler one. The geometric proof using the Hopf bundle (see Section 3.2.3 below) is, in fact, our *second* proof.

Intuitively, one expects that optimal simplicity and maximal symmetry is achieved by a suitable affine transformation of the (I_1, I_2)-plane which transforms the region $Q^* \geq 0$ into the unit disk and makes the metric more "transparent". This simple idea was, indeed, the key leading to such a remarkable coordinate transformation.

As indicated in Figure 2, $Q^* = 0$ defines an ellipse which is tangent to the triple of lines given by $I_1 = 0, I_2 = 0$ and $I_3 = 1 - I_1 - I_2 = 0$. It is easy to see that its center of symmetry is the point (m_1^*, m_2^*), so we first set

$$\tilde{I}_1 = I_1 - m_1^*, \quad \tilde{I}_2 = I_2 - m_2^* \tag{85}$$

and obtain

$$Q^* = m_1 m_2 m_3 - (1 - m_2)^2 \tilde{I}_1^2 - (1 - m_1)^2 \tilde{I}_2^2 - 2(m_3 - m_1 m_2)\tilde{I}_1 \tilde{I}_2.$$

This suggests a rotation through the angle

$$\psi_0 = \frac{1}{2} \tan^{-1} \frac{2(m_1 m_2 - m_3)}{(m_1 - m_2)(1 + m_3)}$$

and new coordinates $\tilde{x}, \tilde{y}$ defined by

$$\tilde{I}_1 = \tilde{x} \cos \psi_0 - \tilde{y} \ \sin \psi_0, \tilde{I}_2 = \tilde{x} \sin \psi_0 + \tilde{y} \ \cos \psi_0.$$

Then

$$Q^* = m_1 m_2 m_3 - \mu_1 \tilde{x}^2 - \mu_2 \tilde{y}^2, \tag{86}$$

where

$$\mu_1 = \frac{1}{2}((1-m_1)^2 + (1-m_2)^2) + \frac{1}{2}(m_1 - m_2)(1+m_3)\cos 2\psi_0$$
$$+ (m_3 - m_1 m_2)\sin 2\psi_0$$
$$\mu_2 = \frac{1}{2}((1-m_1)^2 + (1-m_2)^2) - \frac{1}{2}(m_1 - m_2)(1+m_3)\cos 2\psi_0$$
$$- (m_3 - m_1 m_2)\sin 2\psi_0$$

and we notice the identity

$$(-(1-m_2)^2 + (1-m_1)^2)\sin 2\psi_0 + 2(m_3 - m_1 m_2)\cos 2\psi_0 = 0.$$

Thus, by setting

$$\tilde{x} = \sqrt{\frac{m_1 m_2 m_3}{\mu_1}}x \ , \quad \tilde{y} = \sqrt{\frac{m_1 m_2 m_3}{\mu_2}}y$$

the expression (86) transforms to

$$Q^* = m_1 m_2 m_3 (1 - x^2 - y^2). \tag{87}$$

Therefore, the following combined transformation

$$I_1 = \cos\psi_0 \sqrt{\frac{m_1 m_2 m_3}{\mu_1}}x - \sin\psi_0 \sqrt{\frac{m_1 m_2 m_3}{\mu_2}}y + m_1^* \tag{88}$$
$$I_2 = \sin\psi_0 \sqrt{\frac{m_1 m_2 m_3}{\mu_1}}x + \cos\psi_0 \sqrt{\frac{m_1 m_2 m_3}{\mu_2}}y + m_2^*$$

will transform the formula (83) of $d\sigma^2$ into

$$d\sigma^2 = \frac{1}{4}\frac{(1-y^2)dx^2 + (1-x^2)dy^2 + 2xydxdy}{1 - x^2 - y^2}. \tag{89}$$

From here, we simply set

$$x = \sin\varphi\cos\theta \ , \quad y = \sin\varphi\sin\theta \tag{90}$$

which will transform (89) into the metric (107). This proves that

$$(M^*, d\sigma^2) \simeq S^2(1/2).$$

3.2.3. *The Hopf fibration and a geometric proof of Theorem A.* The moduli space $\bar{M}$ is, by definition, an $SO(3)$-orbit space with the induced differential structure, and according to (33) it is also the orbit space

$$\bar{M} = M/SO(3) \simeq \mathbb{R}^4/SO(2) \approx \mathbb{R}^3 = C(S^2) \tag{91}$$

of the orthogonal transformation group $(SO(2), \mathbb{R}^4)$. As a quotient of a Riemannian space by a compact group of isometries $\bar{M}$ has the induced *orbital distance metric* which measures the distance between orbits in $\mathbb{R}^4$ (or M).

Let $S^3 = S^3(1) \subset \mathbb{R}^4$ be the unit sphere and recall the well known classical Hopf fibration, which in the above metric setting reads

$$SO(2) \to S^3 \to S^3/SO(2) = \mathbb{C}P^1 \simeq S^2(1/2), \tag{92}$$

where the projection is a Riemannian submersion and the quotient space is the round 2-sphere of radius 1/2. Combined with (91) we have an isometry

$$\bar{M} \simeq \mathbb{R}^4/SO(2) = C(S^3/SO(2)) = C(S^2(1/2))$$

of Riemannian cones over the 2-sphere $M^* \simeq S^2(1/2)$. The cone $\bar{M}$ is homeomorphic to $\mathbb{R}^3$, but they are only diffeomorphic away from the cone vertex (or base point O) which corresponds to the origin $0 \in \mathbb{R}^3$.

Finally, to complete the proof of Theorem A it remains to observe that the above orbital distance metric actually coincides with the kinematically defined one. The two metrics are, for example, determined by the kinetic energy they associate to "motions" in $\bar{M}$. These are the image curves of virtual m-triangle motions $\mathbf{X}(t)$, which can always be chosen with vanishing angular momentum, namely they are *horizontal* (cf. Section 2.1.1). These motions are planar, say $\mathbf{X}(t)$ is a curve in $\mathbb{R}^4 \subset M_0$. Horizontal curves are those perpendicular to the $SO(2)$-orbits, and at a point $\mathbf{X} \neq 0$ the horizontal tangent vectors constitute the subspace $\mathcal{H}(\mathbf{X}) \simeq \mathbb{R}^3$ consisting of all $\mathbf{Y}$ such that $\mathbf{X} \times \mathbf{Y} = 0$.

Now, the orbital distance metric on $\bar{M}$ is defined by demanding the projection $\pi : \mathbb{R}^4 \to \mathbb{R}^4/SO(2) = \bar{M}$ to be a Riemannian submersion, that is, that the tangent map $d\pi$ takes $\mathcal{H}(\mathbf{X})$ isometrically to the tangent space of $\bar{M}$ at $\pi(\mathbf{X})$. Equivalently, the kinetic energy associated to a moduli curve is the same as the kinetic energy of a horizontal lifting. On the other hand, the kinematic metric (36) also associates to a moduli curve the kinetic energy $\bar{T}$ of a lifting with vanishing angular momentum. Consequently, the two metrics on $\bar{M}$ are identical.

3.3. **Linear motions of m-triangles.** According to Newton's inertia law, in a center of mass reference frame and in the absence of forces, the trajectory of the three-body system in the Euclidean configuration space M_0 will be a geodesic, namely a *linear motion*

$$\delta(t) = (1-t)\delta_1 + t\delta_2, \tag{93}$$

where $\delta_1 = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3), \delta_2 = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ are appropriate m-triangles. Such motions are also characterized by having constant velocity $(\delta_2 - \delta_1)$, and the motion (93) has constant angular momentum

$$\mathbf{\Omega} = \delta_1 \times \delta_2. \tag{94}$$

Moreover, twice the action integral (7) of the motion from δ_1 to δ_2 is the squared distance

$$2\int T dt = |\delta_1 - \delta_2|^2 = \sum m_i \left|\mathbf{a}_i - \mathbf{b}_i\right|^2 . \tag{95}$$

On the other hand, it is clear that the two m-triangles δ_i lie in a common plane if the vector (94) is zero, namely the linear motion (93) has vanishing angular momentum. Then the motions (93) provide a useful tool in analyzing the kinematic geometry of m-triangles since their moduli curves are exactly the geodesics in the moduli space $(\bar{M}, d\bar{s}^2)$. Their shape curves will be arcs along great circles (geodesics) on the round sphere $M^* = S^2$.

Let us collect some simple facts, assuming δ_1 and δ_2 are m-triangles in the xy-plane (with normal vector $\mathbf{k}$) and $\mathbf{\Omega} = 0$.

Example 9. *If δ_1 and δ_2 have the same orientation, then the distance between their congruence classes $\bar{\delta}_i$ in $\bar{M}$ is*

$$dist(\bar{\delta}_1, \bar{\delta}_2) \ = |\delta_1 - \delta_2| . \tag{96}$$

Example 10. *Consider two congruence classes $\bar{\delta}_1$, $\bar{\delta}_2$ in $\bar{M}$ whose shapes δ_i^* are different and not antipodal points on S^2. It is not difficult to see that representative m-triangles δ_i can be chosen in exactly two ways, modulo a rotation of the xy-plane. The two choices are (δ_1, δ_2) and $(\delta_1, -\delta_2)$ for suitable δ_1 and δ_2.*

The shape curves of the corresponding linear motions (93), for $0 \leq t \leq 1$, are the two geodesic arcs $\Gamma_\pm^*$ between δ_1^* and δ_2^* whose union is a great circle. Each of the shape curves extends (as $t \to \pm\infty$) to the whole circle, minus the limit point $(\delta_1 \mp \delta_2)^*$ as $|t| \to \infty$, which lies on the opposite arc $\Gamma_\mp^*$.

We also remark that the relative position of δ_1 and δ_2 can be calculated from the line integrals of the kinematic 1-forms Θ_i along the geodesic arc $\Gamma_\pm^*$, see Theorem C1 *and* (45)*.*

For easy reference, the shape of the three types of binary collisions are the following three points on E^*

$$\mathfrak{b}_{ij} : I_i = \frac{m_i m_k}{1 - m_k}, \quad I_j = \frac{m_j m_k}{1 - m_k}, \quad I_k \ = (1 - m_k), \tag{97}$$

where $\mathfrak{b}_{ij} = \mathfrak{b}_{ji}$ represents an m-triangle $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ with $\mathbf{a}_i = \mathbf{a}_j$ and $I = 1$, and $\{i, j, k\} = \{1, 2, 3\}$. There are two more points on the sphere M^* which are of kinematic importance, namely the north pole and south pole $\{\mathcal{N}, \mathrm{S}\}$. Each pole is, of course, the *geometric center* of the corresponding pole $M_\pm^*$, and we refer to Corollary 13 for an intrinsic geometric characterization of their shape.

Lemma 11. *The north pole* (*respectively south pole*) *is the shape of the positively* (*respectively negatively*) *oriented m-triangle whose normalized individual moments of inertia equal the dual masses, namely*

$$\mathcal{N} \ (or\ \mathrm{S}) : I_j = m_j^* = \frac{1}{2}(1 - m_j), \ \ j = 1, 2, 3. \tag{98}$$

Proof. Let $\delta_0 = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ be an m-triangle with I_j as in (98). It suffices to show that the three points (97) on the equator E^* have the same distance in M^* to the point $\delta_0^* \in M_\pm^*$. Let $\delta_1 = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ be the (unit size) m-triangle with $\mathbf{b}_1 = \sqrt{2}\mathbf{a}_1$ and $\mathbf{b}_2 = \mathbf{b}_3 = -\frac{m_1}{1-m_1}\mathbf{b}_1$. It is easily checked that $\mathbf{\Omega} = 0$ in (94), that is, the linear motion between δ_0 and δ_1 has vanishing angular momentum.

The image of δ_1 in $\bar{M}$ is the point $\delta_1^* = \mathfrak{b}_{23}$ on E^*, and according to (96)

$$|\delta_0 - \delta_1| = \sqrt{\sum m_j \left|\mathbf{a}_j - \mathbf{b}_j\right|^2} = \sqrt{2 - \sqrt{2}} = 2 \sin \frac{\pi}{8} \tag{99}$$

equals the distance between δ_0^* and δ_1^* in $\bar{M}$. Therefore, their (spherical) distance in M^* is equal to $\pi/4$. On the other hand, it is clear from the above calculation (or by symmetry) that similar choices of δ_1 with $\delta_1^* = \mathfrak{b}_{12}$ or $\mathfrak{b}_{31}$ lead to the same distance $\pi/4$. ∎

3.4. **Eigenvalues and eigenframe of the inertia tensor.** The bilinear form $B_{\mathbf{X}}$ defined by (24) is identical to the well known *inertia tensor* in classical mechanics, for the special case of an m-triangle $\mathbf{X}$ viewed as a rigid body. This is useful in the kinematic study of non-planary motions of m-triangles. The geometric interpretation of the quadratic form is that it calculates the moment of inertia $I_{\boldsymbol{\omega}}$ of the body with respect to the central axis through the vector $\boldsymbol{\omega}$, hence also the rotational kinetic energy due to the angular velocity $\boldsymbol{\omega}$, namely

$$2T^{\omega} = B_{\mathbf{X}}(\boldsymbol{\omega}, \boldsymbol{\omega}) = \left|\boldsymbol{\omega} \times \mathbf{X}\right|^2 = \left|\boldsymbol{\omega}\right|^2 I_{\boldsymbol{\omega}}.$$

By an *eigenframe* of $\mathbf{X}$ we mean an orthonormal basis in 3-space consisting of eigenvectors (i.e., along the principal axes) of $B_{\mathbf{X}}$, whose *eigenvalues* are the associated moments of inertia. We will use the following notation for the eigenvalues and associated eigenframe,

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \ \longleftrightarrow (\mathbf{u}_1, \mathbf{u}_2, \mathbf{n}), \tag{100}$$

where $\mathbf{u}_1$ and $\mathbf{u}_2$ span the plane $\Pi(\mathbf{X})$ if the triangle is nondegenerate, whereas in the collinear case $\lambda_1 = 0$ and $\{\mathbf{u}_2, \mathbf{n}\}$ can be any orthonormal basis of the normal plane $\Pi(\mathbf{X})^\perp$.

Lemma 12. *The eigenvalues of $B_{\mathbf{X}}$ are related by*

$$\lambda_1 + \lambda_2 = \lambda_3 = I, \qquad \lambda_1\lambda_2 = 4m_1m_2m_3\Delta^2,$$

and consequently

$$\lambda_i = \frac{1}{2}(I \pm \sqrt{I^2 - 16m_1m_2m_3\Delta^2}), \;\; i = 1, 2. \tag{101}$$

Proof. We consider the case that $\mathbf{X}$ is nondegenerate, and then I is the moment of inertia with respect to the normal direction. Let $\{\mathbf{u}_1, \mathbf{u}_2\}$ be an orthonormal frame of $\Pi(\mathbf{X})$ consisting of eigenvectors of $B_{\mathbf{X}}$, namely

$$B_{\mathbf{X}}(\mathbf{u}_1, \mathbf{u}_1) = \lambda_1, \quad B_{\mathbf{X}}(\mathbf{u}_2, \mathbf{u}_2) = \lambda_2, \quad B_{\mathbf{X}}(\mathbf{u}_1, \mathbf{u}_2) = 0.$$

Then

$$\begin{aligned}\lambda_1 + \lambda_2 &= |\mathbf{u}_1{\times}\mathbf{X}|^2 + |\mathbf{u}_2{\times}\mathbf{X}|^2 = \sum m_j(2\,|\mathbf{a}_j|^2 - (\mathbf{u}_1{\cdot}\mathbf{a}_j)^2 - (\mathbf{u}_2{\cdot}\mathbf{a}_j)^2)\\ &= \sum m_j\,|\mathbf{a}_j|^2 = I.\end{aligned}$$

Set

$$\mathbf{a}_1 = a_{11}\mathbf{u}_1 + a_{12}\mathbf{u}_2, \quad \mathbf{a}_2 = a_{21}\mathbf{u}_1 + a_{22}\mathbf{u}_2. \tag{102}$$

Then on the one hand

$$B_{\mathbf{X}}(\mathbf{a}_1, \mathbf{a}_1)B_{\mathbf{X}}(\mathbf{a}_2, \mathbf{a}_2) - B_{\mathbf{X}}(\mathbf{a}_1, \mathbf{a}_2)^2 = \tag{103}$$

$$|\mathbf{a}_1 \times \mathbf{X}|^2 \cdot |\mathbf{a}_2 \times \mathbf{X}|^2 - [(\mathbf{a}_1 \times \mathbf{X}){\cdot}(\mathbf{a}_2 \times \mathbf{X})]^2 \;\; = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}^2 \lambda_1\lambda_2,$$

where by the Ceva-area law (61)

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}^2 = |\mathbf{a}_1 \times \mathbf{a}_2|^2 = 4m_3^2\Delta^2,$$

and on the other hand,

$$\begin{aligned} B_{\mathbf{X}}(\mathbf{a}_1, \mathbf{a}_1) &= m_2\,|\mathbf{a}_1 \times \mathbf{a}_2|^2 + m_3\,|\mathbf{a}_1 \times \mathbf{a}_3|^2 = 4m_2m_3(m_2 + m_3)\Delta^2\\ B_{\mathbf{X}}(\mathbf{a}_2, \mathbf{a}_2) &= 4m_1m_3(m_1 + m_3)\Delta^2\\ B_{\mathbf{X}}(\mathbf{a}_1, \mathbf{a}_2) &= -4m_1m_2m_3\Delta^2.\end{aligned} \tag{104}$$

When the expressions (104) are substituted into (103) we obtain

$$\lambda_1\lambda_2 = 4m_1m_2m_3\Delta^2$$

and then formula (101) follows. ∎

Corollary 13. *The poles* (98) *are the shapes uniquely characterized by any of the two equivalent conditions:*
(i) $\lambda_1 = \lambda_2$ *(i.e. "umbilical" shape),*
(ii) the m-triangle attains the maximal area

$$\Delta_{\max} = \frac{I}{4\sqrt{m_1 m_2 m_3}} \tag{105}$$

among all m-triangles with the same moment of inertia I*.*

Proof. Clearly, $\lambda_1 = \lambda_2$ if and only if the area Δ is given by the formula of (105). On the other hand, let us maximize the area function $Q = C_1C_2 + C_2C_3 + C_3C_1$ (cf. (65), (68)), using Lagrange's multiplier method subject to the constraint $I = 1$. It follows that

$$C_i = -m_i m_i^* \ + m_j m_j^* + m_k m_k^*$$

or equivalently $I_j = m_j^*$, by (13). ∎

The following result will also be useful. Briefly, it says that a linear motion whose shape curve is a meridian arc from a pole to the equator, has a constant eigenframe.

Lemma 14. *Let* $\delta_0 = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ *be an m-triangle with the shape of a pole (98), and let* $\{\mathbf{u}_1, \mathbf{u}_2\}$ *be an orthonormal frame of the plane* $\Pi(\delta_0)$*. Moreover, let* $\delta_1 = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ *be a degenerate m-triangle satisfying* $\delta_0 \times \delta_1 = 0$ *and* $\mathbf{u}_2 \cdot \mathbf{b}_j = 0$ *for all* j*. Then*

$$B_t(\mathbf{u}_1, \mathbf{u}_2) = 0$$

holds along the linear motion $\delta_t = (1-t)\delta_0 + t\delta_1$*, where* B_t *is the inertia tensor of* δ_t*, cf.* (24)*. Hence,* $\{\mathbf{u}_1, \mathbf{u}_2\}$ *is an eigenframe for* δ_t *for each* t*.*

Proof. From the above Corollary it follows that

$$B_0(\mathbf{u}_1, \mathbf{u}_2) = B_1(\mathbf{u}_1, \mathbf{u}_2) = 0.$$

Moreover, by the assumptions

$$\begin{aligned}(\mathbf{u}_2 \times \delta_0) \cdot (\mathbf{u}_1 \times \delta_1) &= \sum_j m_j (\mathbf{u}_2 \times \mathbf{a}_j) \cdot (\mathbf{u}_1 \times \mathbf{b}_j) \\ &= \sum_j m_j \left[(\mathbf{u}_1 \cdot \mathbf{u}_2)(\mathbf{a}_j \cdot \mathbf{b}_j) - (\mathbf{u}_2 \cdot \mathbf{b}_j)(\mathbf{u}_1 \cdot \mathbf{a}_j) \right] = 0.\end{aligned}$$

Therefore

$$\begin{aligned}
B_t(\mathbf{u}_1, \mathbf{u}_2) &= (\mathbf{u}_1 \times [(1-t)\delta_0 + t\delta_1]) \cdot (\mathbf{u}_2 \times [(1-t)\delta_0 + t\delta_1]) \\
&= (1-t)^2 B_0(\mathbf{u}_1, \mathbf{u}_2) + t^2 B_1(\mathbf{u}_1, \mathbf{u}_2) \\
&+ (1-t)t[(\mathbf{u}_1 \times \delta_0) \cdot (\mathbf{u}_2 \times \delta_1) \pm (\mathbf{u}_2 \times \delta_0) \cdot (\mathbf{u}_1 \times \delta_1)] \\
&= (1-t)t(\mathbf{u}_1 \times \mathbf{u}_2) \cdot (\delta_0 \times \delta_1) = 0.
\end{aligned}$$

∎

4. The spherical representation of shape space M^*

By the *spherical representation* we refer to an identification of M^* with a round 2-sphere, with a distinguished (northern) hemisphere M^*_+ whose natural orientation induces the positive orientation of the equator E^* and hence the (eastward) direction of increasing longitude. We also assume the (cyclic) ordering $\mathfrak{b}_{23}, \mathfrak{b}_{31}, \mathfrak{b}_{12}$ of the three binary collision points (lying on E^*) is in the positive direction. Finally, the correspondence should represent the kinematic geometry and hence is an isometry

$$M^* \to S^2(1/2) \tag{106}$$

which identifies each shape δ^* with a specific point on the sphere. We will develop methods enabling us to express the spherical coordinates in terms of intrinsic invariants of δ^*, and conversely.

Let (r, θ) denote *polar coordinates* on $S^2(1/2)$, where r is the *polar distance* which measures the spherical distance from δ^* to the north pole $\mathcal{N} \in M^*_+$ and θ is the longitude angle. For convenience, we also introduce spherical coordinates (φ, θ), where the angle $\varphi = 2r$ is the *colatitude* with $\varphi = 0$ at the north pole. In these coordinates the Riemannian metric of the sphere M^* expresses as

$$d\sigma^2 = dr^2 + \frac{1}{4}\sin^2(2r)d\theta^2 = \frac{1}{4}(d\varphi^2 + \sin^2\varphi\, d\theta^2), \tag{107}$$
$$0 \le r \le \frac{\pi}{2}, \;\; 0 \le \varphi \le \pi, \;\; 0 \le \theta \le 2\pi.$$

Remark 15. *The choice of the zero meridian $\theta = 0$ is a matter of convenience, and until further notice our convention is that it passes through $\mathfrak{b}_{23}$. Indeed, only longitude differences $(\theta - \theta')$ is an intrinsic property of shapes of m-triangles, see Section* 4.2 *and* 4.3.

4.1. **Geometric interpretation of the polar distance r.** Let δ be a nonzero positively oriented m-triangle, that is, $r \le \pi/4$. We will investigate the relationship between the polar distance r and the geometric invariants of δ. Let δ_1^* be the intersection point between the equator E^* and the meridian passing through δ^*. Choose unit size representatives

$$\delta_0 = (\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3), \quad \delta_1 = (\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3),$$

of the pole $\mathcal{N}$ and δ_1^* with $\delta_0 \times \delta_1 = 0$, and observe that the above meridian is the shape curve of the linear motion $\delta_t = (1-t)\delta_0 + t\delta_1$, $0 \leq t \leq 1$. Henceforth, let t be the unique value such that $\delta_t^* = \delta^*$.

Let C be the cone surface (cf. also Definition 42) in $\bar{M}$ spanned by the rays through points on the above meridian between $\mathcal{N}$ and δ_1^*. C is isometric to a Euclidean sector of angular width $\pi/4$, see Figure 3. The cord distance between $\mathcal{N}$ and δ_1^* is the number in (99). On the other hand, the cord distance between $\mathcal{N}$ and $\bar{\delta}_t$ can be computed in two different ways, namely

$$dist(\mathcal{N},\bar{\delta}_t) = 2t\sin\frac{\pi}{8} = \sin\frac{\pi}{8} - \cos\frac{\pi}{8}\tan(\frac{\pi}{8} - r).$$

Hence,

$$t = \frac{1}{2}(1 - \cot\frac{\pi}{8}\tan(\frac{\pi}{8} - r)) \tag{108}$$

and, moreover,

$$dist(O,\bar{\delta}_t) = \frac{\cos\frac{\pi}{8}}{\cos(\frac{\pi}{8} - r)}. \tag{109}$$

Next, let us compute the eigenvalues of $B_t = B_{\delta_t}$, namely

$$\lambda_1' = B_t(\mathbf{u}_1,\mathbf{u}_1), \quad \lambda_2' = B_t(\mathbf{u}_2,\mathbf{u}_2),$$

where by Lemma 14 we have chosen an orthonormal frame $\{\mathbf{u}_1,\mathbf{u}_2\}$ of the plane $\Pi(\delta_0)$ with $\mathbf{u}_2 \cdot \mathbf{b}_j = 0$ for all j. It follows that

$$\begin{aligned}\lambda_1' &= B_t(\mathbf{u}_1,\mathbf{u}_1) = |\mathbf{u}_1 \times \delta_t|^2 = |\mathbf{u}_1 \times (1-t)\delta_0|^2 \\ &= (1-t)^2 B_0(\mathbf{u}_1,\mathbf{u}_1) = \frac{1}{2}(1-t)^2.\end{aligned}$$

Therefore, since δ^* and $\bar{\delta}_t$ differ by the scaling factor (109), the eigenvalues of B_{δ^*} are

$$\lambda_1 = \frac{\cos\frac{\pi}{8}}{\cos(\frac{\pi}{8} - r)}\lambda_1' = \frac{1}{2}(1-t)^2\frac{\cos\frac{\pi}{8}}{\cos(\frac{\pi}{8} - r)} \tag{110}$$

and $\lambda_2 = 1 - \lambda_1$.

Finally, we can use (108), (110) and the identity

$$\lambda_1\lambda_2 = 4m_1m_2m_3\Delta^2$$

to solve for r as a function of the area Δ of δ^*. We state the final result as follows:

Lemma 16. *The polar distance r for an arbitrary given positively oriented m-triangle δ is given by the formula*

$$\cos(2r) = 4\sqrt{m_1 m_2 m_3}\frac{\Delta}{I}, \tag{111}$$

where $\sum m_i = 1$ and Δ (respectively I) is the area (respectively moment of inertia) of δ.

Equivalently, by (101) there is the formula

$$\sin 2r = \frac{1}{I}\left|\lambda_1 - \lambda_2\right|. \tag{112}$$

4.2. **Geometric interpretation of the longitude angle θ.** Let δ_0 be an oriented m-triangle whose shape δ_0^* is the pole $\mathcal{N}$ or S (i.e. $r = 0$ or $\pi/2$). Recall from Lemma 14, it is possible to deform δ_0, through a linear motion with zero angular momentum, to the shape of any given degenerate m-triangle. Then the shape curve will be the meridian from the pole to a point δ_1^* on the equator circle E^*. Moreover, the line spanned by the final configuration δ_1 is uniquely determined by δ_0, and there is a constant eigenframe throughout the deformation.

Now, let us consider two points δ_1^*, δ_2^* on E^* and seek an interpretation of their spherical distance in $M^* = S^2(1/2)$.

Theorem 17. *Let $\delta_0 = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ be an m-triangle, of maximal area for a fixed moment of inertia, and let δ_1, δ_2 be degenerate (but nonzero) m-triangles satisfying the vanishing angular momentum condition*

$$\delta_0 \times \delta_1 = \delta_0 \times \delta_2 = 0$$

for the linear motions from δ_0 to $\delta_i, i = 1, 2$. Then the angle ψ between the lines spanned by δ_1 and δ_2 is equal to the distance between the associated points δ_1^* and δ_2^* in the shape space M^*, namely

$$\psi = \frac{1}{2}\left|\theta_2 - \theta_1\right|.$$

Proof. We may assume all m-triangles are confined to the xy-plane, the shape δ_0^* is the north pole and δ_i^* has longitude angle $\theta_i, i = 1, 2$. Consider the piecewise linear motion whose associated shape curve is the spherical triangle D in M^* with vertices $\delta_1^*, \delta_2^*, \delta_0^*$. Starting from δ_1, the motion passes successively through the m-triangles $\delta_1, \delta_2', \delta_0, \delta_1'$. Here δ_2' is congruent to δ_2 and is situated in the same line as δ_1, whereas δ_1' is congruent to δ_1 but is actually situated in a line making the angle ψ with the original line.

Next, let us apply the Gauss-Bonnet formula (47) to the region D, thus obtaining an equality between twice the area of D and the angle ψ.

Finally, we simply combine this with the fact that, as a geodesic triangle on the sphere of radius 1/2, the area of D equals one quarter of its angle at δ_0^*, namely the angle $|\theta_2 - \theta_1|$. ∎

In general, it turns out that the longitude angle θ of an m-triangle is determined by the relative position and size of the eigenframe and normalized area Δ/I respectively. To make this relationship precise, let $\delta = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ be a non-degenerate, positively oriented m-triangle in the xy-plane with normal vector

$$\frac{\mathbf{a}_1 \times \mathbf{a}_2}{|\mathbf{a}_1 \times \mathbf{a}_2|} = \mathbf{k}$$

and assume the shape δ^* is not the pole $\mathcal{N}$.

Theorem 18. *Let $\{\mathbf{u}_1, \mathbf{u}_2\}$ be an eigenframe of δ, where $\mathbf{u}_1$ is the eigenvector of the inertia tensor B_δ associated with the smallest eigenvalue λ_1, and let ψ_i be the (oriented) angle from $\mathbf{a}_1$ to $\mathbf{u}_i$. Then the following identity*

$$\tan\frac{\theta}{2} = -\frac{1+\sin\varphi}{\cos\varphi}\tan\psi_1 = \frac{1+\sin\varphi}{\cos\varphi}\cot\psi_2 \tag{113}$$

relates the angle ψ_i to the spherical coordinates (φ, θ) of the shape δ^* on the 2-sphere M^*, where φ is the colatitude and θ is the longitude (eastward, with $\theta = 0$ at $\mathfrak{b}_{23}$).

Remark 19. *The above formula holds for any choice of eigenframe since ψ_1 changes by $\pm\pi$ if $\mathbf{u}_1$ is replaced by $-\mathbf{u}_1$.*

Proof. The first step of the proof is to derive a formula which expresses ψ_1 solely in terms of intrinsic invariants of the m-triangle, together with a simple recipe for calculating this angle. Then, by applying the Gauss-Bonnet formula (cf. Theorem C2) we shall deduce formula (113).

Since $\psi_2 = \psi_1 \pm \pi/2$ we need only prove the first identity in (113). First of all, in order to have the angle ψ_1 uniquely defined we must specify the choice of eigenframe. Namely, let $\{\mathbf{u}_1, \mathbf{u}_2\}$ be a positive frame and hence $\mathbf{u}_1 \times \mathbf{u}_2 = \mathbf{k}$, and moreover, we assume $\mathbf{u}_1$ chosen so that

$$-\frac{\pi}{2} \le \psi_1 < \frac{\pi}{2}. \tag{114}$$

Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be the orthonormal frame derived from $\{\mathbf{a}_1, \mathbf{a}_2\}$ by the Gram-Schmidt algorithm, with $\mathbf{e}_1 = \mathbf{a}_1 / |\mathbf{a}_1|$. Using the expressions (104) it is not difficult to deduce

$$\begin{aligned}
A &= B_\delta(\mathbf{e}_1, \mathbf{e}_1) = 4m_1m_2m_3(m_2+m_3)\frac{\Delta^2}{I_1} \\
B &= B_\delta(\mathbf{e}_1, \mathbf{e}_2) = ((m_2+m_3)C_3 - 2m_1m_2I_1)\frac{\Delta}{I_1} \\
C &= B_\delta(\mathbf{e}_2, \mathbf{e}_2) = I - A.
\end{aligned} \tag{115}$$

By writing

$$\begin{aligned}
\mathbf{u}_1 &= \cos\psi_1\mathbf{e}_1 + \sin\psi_1\mathbf{e}_2 \\
\mathbf{u}_2 &= -\sin\psi_1\mathbf{e}_1 + \cos\psi_1\mathbf{e}_2
\end{aligned} \tag{116}$$

and inserting these expressions into $B_\delta(\mathbf{u}_1, \mathbf{u}_2) = 0$, we deduce the formula

$$\tan 2\psi_1 = \frac{2B}{A-C} = \frac{2B}{2A-I} \tag{117}$$

and similarly

$$\lambda_1 - \lambda_2 = B_\delta(\mathbf{u}_1, \mathbf{u}_1) - B_\delta(\mathbf{u}_2, \mathbf{u}_2) = (A-C)\cos 2\psi_1 + 2B\sin 2\psi_1. \tag{118}$$

How is ψ_1 determined from (117) and (118)? Assume first $B = 0$. With the normalization $\sum m_1 = 1, \sum I_i = 1$, this happens when

$$I_1(m_3 - m_1m_2) + I_2(1-m_1)^2 = (1-m_1)m_3.$$

For example, with uniform mass distribution this holds for the isosceles triangle with $|\mathbf{a}_2| = |\mathbf{a}_3|$. Note that $\{\mathbf{e}_1, \mathbf{e}_2\}$ is also a positive eigenframe, and the eigenvalues λ_i equals A and C. Since $\lambda_1 < \lambda_2$, by assumption, we deduce from (118) that $A > C$ implies $\psi_1 = \frac{\pi}{2}$, and $A < C$ implies $\psi_1 = 0$.

Next, assume $B \neq 0$. By combining (117) and (118) we eliminate $\cos 2\psi_1$ and obtain the expression

$$\lambda_1 - \lambda_2 = \frac{(A-C)^2 + 4B^2}{2B}\sin 2\psi_1. \tag{119}$$

Consequently, $\sin 2\psi_1$ has the opposite sign of B, namely

$$0 < \psi_1 < \frac{\pi}{2} \text{ if } B < 0, \quad -\frac{\pi}{2} < \psi_1 < 0 \text{ if } B > 0. \tag{120}$$

Finally, we turn to the proof of formula (113). On the sphere M^*, let δ_1^* be the intersection point of E^* and the meridian passing through δ^*, and consider the spherical triangle on M^* with vertices $\mathfrak{b}_{23}, \delta_1^*, \delta^*$, whose area is denoted by $\bar{\Delta}$. The right angle at the vertex δ_1^* has adjacent edges of length

$$s = \frac{|\theta|}{2} \leq \frac{\pi}{2} \text{ and } \frac{\pi}{4} - r,$$

and by applying the spherical sine law and area formula to the magnified triangle on $S^2(1)$ with area $\tilde{\Delta} = 4\bar{\Delta}$, we have

$$\tan\frac{\tilde{\Delta}}{2} = \frac{\sin 2s \sin(\frac{\pi}{2} - 2r)}{1 + \cos 2s + \cos(\frac{\pi}{2} - 2r) + \cos 2s \cos(\frac{\pi}{2} - 2r)}$$
$$= \frac{\sin 2s}{1 + \cos 2s}\frac{\sin(\frac{\pi}{2} - 2r)}{1 + \cos(\frac{\pi}{2} - 2r)} = \tan s \tan(\frac{\pi}{4} - r).$$

On the other hand, applying the Gauss-Bonnet formula (Theorem C2) to the triangle in Figure 4, it is not difficult to see that the total rotation of the vector $\mathbf{a}_1$ is through the angle ψ_1, hence

$$\pm\psi_1 = 2\bar{\Delta} = \frac{\tilde{\Delta}}{2}$$

and consequently

$$\tan\psi_1 = \pm\tan\frac{\theta}{2}\tan(\frac{\pi}{4} - r). \tag{121}$$

We claim that the sign to be used in (121) is -1. This can be seen by considering the situation where δ_1^* lies between $\mathfrak{b}_{23}$ and $\mathfrak{b}_{31}$, by observing that ψ_1 decreases as θ increases (i.e. when δ_1^* approaches $\mathfrak{b}_{31}$). This completes the proof of formula (113). ∎

4.3. **Intrinsic form of the spherical representation.** We will focus attention on the "inverse" of the correspondence (106), namely

$$(\varphi, \theta) \to (I_1, I_2, I_3), \quad \sum I_j = 1.$$

The first step in this direction was, in fact, our kinematic proof of Theorem A (cf. Section 3.2.2). Namely, by substituting (90) into (88) and considering the special case of $m_3 = m_1 m_2$, it is easy to verify that the expressions (88) simplify to

$$I_1 = m_1^*(1 + \sin\varphi\cos\theta), \;\; I_2 = m_2^*(1 + \sin\varphi\sin\theta). \tag{122}$$

These formulas are, indeed, a special case of a general intrinsic description of the spherical representation, purely in terms of geometric concept.

The *longitude distance* between $\delta_1^* = (\varphi_1, \theta_1)$ and $\delta_2^* = (\varphi_2, \theta_2)$ is, by definition, the angle $|\theta_1 - \theta_2| \mod 2\pi$. Then it is easy to check that (122) can be stated as

$$I_i = m_i^*(1 + \sin\varphi\cos\tilde{\theta}_i), \quad i = 1, 2, 3, \tag{123}$$

where $\tilde{\theta}_1$ (respectively $\tilde{\theta}_2$ or $\tilde{\theta}_3$) is the longitude distance between $\delta^* = (\varphi, \theta)$ and the binary collision point $\mathfrak{b}_{23}$ (respectively $\mathfrak{b}_{31}$ or $\mathfrak{b}_{12}$).

On the other hand, consider the three distance functions on $M^* = S^2(1/2)$

$$\sigma_i = \sigma_i(\delta^*) = dist(\delta^*, \mathfrak{b}_{i+1,i+2}) \qquad (i \mod 3) \tag{124}$$

which measure the (spherical) distances to the points $\mathfrak{b}_{ij}$. By the spherical cosine law applied to the triangle with vertices $\mathcal{N}, \mathfrak{b}_{ij}, \delta^*$, it follows that

$$\cos 2\sigma_i = \sin\varphi \cos\tilde{\theta}_i, \;\; i = 1, 2, 3$$

and consequently (123) has the invariant form

$$I_i = m_i^*(1 + \cos 2\sigma_i), \;\; i = 1, 2, 3. \tag{125}$$

This may be stated as

$$\cos 2\sigma_i = \frac{I_i}{m_i^*} - 1 = \frac{1}{m_i^*}\tilde{I}_i, \text{ cf. } (85), \tag{126}$$

or equivalently

$$\sigma_i = \arccos\sqrt{\frac{I_i}{1 - m_i}} \quad (\sum I_i = 1). \tag{127}$$

Thus, in order to establish (125) or (123) as a general formula it suffices to verify formula (127) in general. Again, the basic idea we use is to construct a suitable linear motion in M_0, as we did in Chapter 3, namely we consider the linear motion whose shape curve is the (shortest) geodesic from δ^* to $\mathfrak{b}_{ij}$. We shall calculate the length of this curve in the following way.

Let $\delta = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ be a given m-triangle, normalized with $I = 1$, and consider the linear motion of vanishing angular momentum

$$\delta(t) = (1-t)\delta + t(\sqrt{\frac{1-m_1}{I_1}}\mathbf{a}_1, -\frac{m_1}{\sqrt{(1-m_1)I_1}}\mathbf{a}_1, -\frac{m_1}{\sqrt{(1-m_1)I_1}}\mathbf{a}_1)$$

between $\delta = \delta(0)$ and the normalized m-triangle $\delta_1 = \delta(1)$ with the shape $\delta_1^* = \mathfrak{b}_{23}$. Its shape curve is the desired geodesic.

The length $|\delta - \delta_1|$ of the segment in M_0 from δ to δ_1 is also the length $\bar{\sigma}_1$ of the chord in $\bar{M}$ between δ^* and δ_1^*, see (96) and Figure 4. By applying the Ceva-cosine law (63) to the calculation of inner products $\mathbf{a}_i \cdot \mathbf{a}_j$ we arrive at the expression

$$\bar{\sigma}_1^2 = |\delta - \delta_1|^2 = 2(1 - \sqrt{\frac{I_1}{1-m_1}}),$$

and consequently

$$\sigma_1 = 2\arcsin\frac{\bar{\sigma}_1}{2} = \arccos\sqrt{\frac{I_1}{1-m_1}}.$$

4.4. **The reduced Newton's equation in spherical coordinates.** Let us utilize the structure of $\bar{M}$ as a cone over the 2-sphere to express the reduced Newton's equation of Theorem E1 in terms of the spherical coordinate system (ρ, φ, θ), where $\rho = \sqrt{I}$ measures the distance from the base point O (cone vertex). The relationship between coordinate functions $\{I_j\}, \{r_{ij}\}$ and $\{\rho, \varphi, \theta\}$ is expressed by (64) and (123), thus enabling us to transform the equation to a system purely in terms of ρ, φ, θ. This change of variable is, however, rather messy, but an equivalent system can be worked out in several ways. For example, we obtain the following system of three second-order equations

$$\begin{aligned} (i) \quad & 0 = \ddot{\rho} + \frac{\dot{\rho}^2}{\rho} - \frac{1}{\rho}(U + 2h) \\ (ii) \quad & 0 = \ddot{\varphi} + 2\frac{\dot{\rho}}{\rho}\dot{\varphi} - \frac{1}{2}\sin(2\varphi)\dot{\theta}^2 - \frac{4}{\rho^2}\frac{\partial U}{\partial \varphi} \\ (iii) \quad & 0 = \ddot{\theta} + 2\frac{\dot{\rho}}{\rho}\dot{\theta} + 2\cot(\varphi)\dot{\varphi}\dot{\theta} - \frac{4}{\rho^2}\frac{1}{\sin^2\varphi}\frac{\partial U}{\partial \theta} \end{aligned} \tag{128}$$

valid for planary three-body motions with a fixed energy level h. The angular momentum Ω constant is not implicit in these equations since it is an integration constant defined by the initial value problem. In fact, we have also the equations

$$T - U = h \qquad \ddot{I} = 2T + 2h,$$

namely the energy equation and the Lagrange-Jacobi equation (cf. (217)). The latter is precisely equation (i) in (128), and the energy integral

$$U + h - \frac{\Omega^2}{2\rho^2} = \frac{1}{2}\dot{\rho}^2 + \frac{\rho^2}{8}(\dot{\varphi}^2 + (\sin\varphi)^2\,\dot{\theta}^2)$$

makes any of the two equations (ii) or (iii) superfluous.

4.5. **Ceva-type relations in the spherical representation.** The classical Ceva theorem tells us that the lines from the vertices to the center of mass of an m-triangle δ divide the triangle into subtriangles whose areas are in the proportion

$$\Delta_1 : \Delta_2 : \Delta_3 = m_1 : m_2 : m_3.$$

On the other hand, a point δ^* on a hemisphere $M^*_\pm$ divides it into three spherical triangles with areas A_i, with the common vertex δ^* and the binary collision points $\mathfrak{b}_{12}$, $\mathfrak{b}_{23}$, $\mathfrak{b}_{31}$ as the other vertices, cf. Figure 5. In this way, various (normalized) geometric invariants of δ such as sides, areas, angles (cf. Figure 1) have their spherical "dual" counterparts, although the dual quantity may be of a different type. There are, for example, the three central angles α_i (respectively π_i) of δ (respectively at the shape point δ^*) with $\sum \alpha_i = \sum \pi_i = 2\pi$. According to the following lemma, the areas A_i are dual to the angles α_i, and later we also show the areas Δ_i are dual to certain sides of the spherical triangles.

Lemma 20. *Let $\delta = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ be an m-triangle with central angle α_i opposite to the vector $\mathbf{a}_i$. Then the area of the spherical triangle in M^* with vertices $\mathfrak{b}_{12}, \delta^*, \mathfrak{b}_{31}$ equals*

$$A_1 = \frac{1}{2}(\pi - \alpha_1).$$

Proof. Let L_2, L_3 be the lines spanned by the vectors $\mathbf{a}_2$ and $\mathbf{a}_3$ respectively. There is an obvious piecewise linear motion with $\mathbf{\Omega} = 0$ which starts at δ and collapses the triangle to a degenerate configuration of shape $\mathfrak{b}_{31}$ along L_2, and continues along L_2 until the shape of $\mathfrak{b}_{12}$ is reached. This motion keeps the direction of $\mathbf{a}_2$ unaltered.

On the other hand, the linear motion with $\mathbf{\Omega} = 0$ which collapses δ to a configuration of shape $\mathfrak{b}_{12}$ along L_3, will rotate $\mathbf{a}_2$ to a vector along L_3 which lies opposite to $\mathbf{a}_3$. Hence, the total change of position when δ is deformed according to the above piecewise linear motion whose shape curve encloses the spherical triangle, is equal to the angle $\pi - \alpha_1$. Finally, by the Gauss-Bonnet formula, this is twice the area A_1 of the triangle. ∎

Next, we turn to the mutual distances, that is, the sides of the m-triangle δ

$$s_1 = r_{23} = |\mathbf{a}_2 - \mathbf{a}_3| \quad \text{etc.}$$

and ask for their spherical counterpart, namely the spherical distances σ_i from δ^* to the binary collision points. The quantities s_i have, indeed, a nice geometric interpretation in the vector algebra representation described below, see (141). But first, by combining (64) and (126), the identity

$$s_i^2 = \frac{1 - m_i - I_i}{\hat{m}_i} = \frac{1 - m_i}{\hat{m}_i}(1 - \frac{I_i}{1 - m_i}) = \frac{1 - m_i}{\hat{m}_i} \sin^2 \sigma_i \tag{129}$$

holds, where we have assumed normalization $I = 1$ and (as usual) $\sum m_i = 1$, cf. also (14) for notation. By summation over i the condition $I = 1$ reads

$$1 = \sum \hat{m}_i s_i^2 = \sum (1 - m_i) \sin^2 \sigma_i \quad \text{or} \quad \sum m_i^* \cos 2\sigma_i = 0, \tag{130}$$

where the first identity is just the normalized version of Lagrange's formula (65).

Now, let δ_0 be an m-triangle whose shape is a pole $\delta_0^* = \mathcal{N}$ or S, and let

$$\{\alpha_1, \alpha_2, \alpha_3\}\,, \quad \{\beta_1, \beta_2, \beta_3\} \tag{131}$$

denote the central angles α_i of δ_0 and the central angles $\beta_i = \pi_i$ at δ_0^*, respectively, see Figure 5. In particular, β_1 is the longitude distance between $\mathfrak{b}_{31}$ and $\mathfrak{b}_{12}$, or equivalently, $\beta_1/2$ is their distance in M^*. The relationship between the two triples in (131) follows from the above lemma, namely

$$\beta_i = 4A_i = 2\pi - 2\alpha_i. \tag{132}$$

On the other hand, the triple of angles and the (normalized) mass distribution uniquely determine each other. In one direction, we apply the Ceva-cosine law to δ_0 and obtain

$$\cos\alpha_1 = \frac{-\sqrt{m_2m_3}}{\sqrt{1-m_2}\sqrt{1-m_3}}, \quad \cos\beta_1 = \frac{m_2m_3 - m_1}{(1-m_2)(1-m_3)}, \quad \text{etc.} \tag{133}$$

In particular, the angles are in the range

$$\frac{\pi}{2} < \alpha_i < \pi \qquad 0 < \beta_i < \pi. \tag{134}$$

We also note that $\cos\beta_i$ can be calculated by applying a distance function $\sigma_j, j \neq i$, see (124). For example, $\cos\beta_1 = \cos(2\sigma_3(\mathfrak{b}_{31}))$, and by (126) and the fact that $\mathfrak{b}_{31}$ has $I_3 = \frac{m_2m_3}{1-m_2}$, we deduce again the above expression for $\cos\beta_1$.

In the other direction, we would like to know which angles in the range (134) are actually realizable for some mass distribution. The condition on the angles is

$$2\sin\beta_i < \sum\sin\beta_j, \;\text{ for each } i, \tag{135}$$

and the corresponding (normalized) mass distribution is defined by

$$m_1 = 1 - \frac{2\sin\beta_1}{\sum\sin\beta_j} = \frac{1+\sum\cos\beta_j}{-1+\sum\cos\beta_j - 2\cos\beta_1}\,, \quad \text{etc.} \tag{136}$$

We omit the simple proof of this, remarking that the realizability condition (135) also has a nice geometric interpretation. Namely, consider the three binary collision points $\mathfrak{b}_{12}, \mathfrak{b}_{23}, \mathfrak{b}_{31}$ as the vertices of a triangle in the Euclidean disk (of radius 1/2) with the equator circle E^* as boundary. In general, let us call a triangle *central* if its circumcenter O lies in its interior. Then the realizability condition simply says that the above triangle must be central. For another property of this triangle we also refer to Lemma 21.

4.6. **The vector algebra representation of the kinematic geometry.** Since the kinematic study of m-triangles and their motions essentially involves spherical geometry, we are naturally led to the vector algebra in the Euclidean 3-space $\mathbb{R}^3$, where inner products and determinants are the basic invariants. Therefore, it is sometimes convenient to represent M^* by the sphere $S^2(1)$ of unit vectors in $\mathbb{R}^3$ and hence its cone $\bar{M}$ becomes the whole Euclidean space 3-space. Thus we introduce the *vector algebra representation* of the moduli space $\bar{M}$ by constructing the following transformation between Riemannian cones

$$\begin{aligned} \bar{M} = C(S^2(1/2)) &\xrightarrow{\Psi} C(S^2(1)) = \mathbb{R}^3, \\ \Psi : (\rho, r, \theta) &\to (\rho^2, 2r, \theta) = (I, \varphi, \theta), \end{aligned} \tag{137}$$

which magnifies M^* to a sphere of radius 1 and squares the distance to the origin. It is a diffeomorphism away from the base point (or origin) O, namely the class of the triple collision $\rho = 0$.

In (137), (I, φ, θ) are the usual spherical coordinates in 3-space associated with the Euclidean coordinates (x, y, z), that is,

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

where $0 \leq \varphi \leq \pi$, $0 \leq \theta \leq 2\pi$. The kinematic metric (38) on $\bar{M}$, expressed as a metric on $\mathbb{R}^3$, now becomes the following conformal modification of the Euclidean metric, namely

$$d\bar{s}^2 = d\rho^2 + \rho^2 d\sigma^2 = \frac{dx^2 + dy^2 + dz^2}{4\sqrt{x^2 + y^2 + z^2}}. \tag{138}$$

A variable point on M^* will be represented by a unit vector

$$\delta^* \to \mathbf{p} = (x, y, z) \in S^2(1) \subset \mathbb{R}^3$$

and we fix the following notation and location of binary collision points (cf. (133) and Figure 6)

$$\begin{aligned} \mathfrak{b}_{23} &\to \hat{\mathbf{b}}_1 = (1, 0, 0), \\ \mathfrak{b}_{31} &\to \hat{\mathbf{b}}_2 = (\cos\beta_3, \sin\beta_3, 0), \\ \mathfrak{b}_{12} &\to \hat{\mathbf{b}}_3 = (\cos\beta_2, -\sin\beta_2, 0). \end{aligned} \tag{139}$$

Moreover, (126) now reads

$$\mathbf{p} \cdot \hat{\mathbf{b}}_i = \cos 2\sigma_i = \frac{1}{m_i^*}\tilde{I}_i = \begin{cases} x & i = 1 \\ x\cos\beta_3 + y\sin\beta_3 & i = 2 \\ x\cos\beta_2 - y\sin\beta_2 & i = 3 \end{cases} \tag{140}$$

Recall that $2\sigma_i$ is the spherical distance between $\mathbf{p}$ and $\hat{\mathbf{b}}_i$; hence, by (129) there is the simple formula

$$s_i = \frac{1}{2}\sqrt{\frac{1-m_i}{\hat{m}_i}}\left|\mathbf{p}-\hat{\mathbf{b}}_i\right| \tag{141}$$

which expresses the three mutual distances s_i for the normalized m-triangle δ as the Euclidean distances (modulo a fixed factor) from $\mathbf{p} \in S^2$ to the three fixed points $\hat{\mathbf{b}}_i$ lying on the circle $x^2+y^2=1, z=0$.

Lemma 21. *There is a unique mass distribution* (m_1', m_2', m_3') *such that* $(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \hat{\mathbf{b}}_3)$ *becomes an m-triangle with center of mass at origin, that is,*

$$m_1'\hat{\mathbf{b}}_1 + m_2'\hat{\mathbf{b}}_2 + m_3'\hat{\mathbf{b}}_3 = 0,$$

namely the dual masses $m_i' = m_i^*$*, cf.* (14).

Proof. Since the triangle is central the origin lies in its interior, so there are barycentric coordinates $q_i > 0$, unique up to a common multiple, such that $\sum q_i\hat{\mathbf{b}}_i = 0$. On the other hand, by combining (130) and (140), $\sum m_i^*\hat{\mathbf{b}}_i \cdot \mathbf{p} = 0$ holds for all $\mathbf{p}$ and consequently $\sum m_i^*\hat{\mathbf{b}}_i = 0$. ∎

4.7. **An integral formula for the distance function on M^*.** The kinematic Riemannian metric $d\sigma^2$, expressed in terms of coordinates I_1, I_2 as in (83), may be viewed as the infinitesimal version of an integral formula for the distance function $\sigma(p,p')$ on each hemisphere $M_\pm^*$ of $M^* = S^2(1/2)$. Our calculation of such an integral formula will be based upon a special type of coordinates; namely, we choose the points $\{\mathfrak{b}_{23}, \mathfrak{b}_{31}\}$ as a *bipolar system* whose associated distance functions $\{\sigma_1, \sigma_2\}$ constitute a coordinate system on each hemisphere $M_\pm^*$. In our final formula, however, we shall express the distance in terms of I_1, I_2, and more simply in terms of their translates $\tilde{I}_i = I_i - m_i^*$.

To this end, it is convenient to apply the above vector algebra representation, where we use the unit sphere $S^2(1)$ representation of M^* rather than the sphere $S^2(1/2)$, and the collision points $\mathfrak{b}_{23}, \mathfrak{b}_{31}$ and variable points p, p' are replaced by the unit vectors $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \mathbf{p}, \mathbf{p}'$, respectively. Thus, $\sigma(p,p')$ is half of the spherical distance between $\mathbf{p}$ and $\mathbf{p}'$ on the unit sphere.

We shall reduce the calculation of the distance function σ to a simple vector algebra involving determinants and Lagrange's formula:

$$\det(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \mathbf{p})\det(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \mathbf{p}') = \begin{vmatrix} 1 & \hat{\mathbf{b}}_1\cdot\hat{\mathbf{b}}_2 & \hat{\mathbf{b}}_1\cdot\mathbf{p}' \\ \hat{\mathbf{b}}_2\cdot\hat{\mathbf{b}}_1 & 1 & \hat{\mathbf{b}}_2\cdot\mathbf{p}' \\ \mathbf{p}\cdot\hat{\mathbf{b}}_1 & \mathbf{p}\cdot\hat{\mathbf{b}}_2 & \mathbf{p}\cdot\mathbf{p}' \end{vmatrix}. \tag{142}$$

By writing the left side as

$$\left[\det(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \mathbf{p})^2 \det(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \mathbf{p}')^2\right]^{1/2}$$

and applying Lagrange's formula to each square, the right side of (142) equals the product

$$\begin{vmatrix} 1 & \hat{\mathbf{b}}_1 \cdot \hat{\mathbf{b}}_2 & \hat{\mathbf{b}}_1 \cdot \mathbf{p} \\ \hat{\mathbf{b}}_2 \cdot \hat{\mathbf{b}}_1 & 1 & \hat{\mathbf{b}}_2 \cdot \mathbf{p} \\ \mathbf{p} \cdot \hat{\mathbf{b}}_1 & \mathbf{p} \cdot \hat{\mathbf{b}}_2 & 1 \end{vmatrix}^{1/2} \begin{vmatrix} 1 & \hat{\mathbf{b}}_1 \cdot \hat{\mathbf{b}}_2 & \hat{\mathbf{b}}_1 \cdot \mathbf{p}' \\ \hat{\mathbf{b}}_2 \cdot \hat{\mathbf{b}}_1 & 1 & \hat{\mathbf{b}}_2 \cdot \mathbf{p}' \\ \mathbf{p}' \cdot \hat{\mathbf{b}}_1 & \mathbf{p}' \cdot \hat{\mathbf{b}}_2 & 1 \end{vmatrix}^{1/2} \tag{143}$$

and this gives us an identity where $\hat{\mathbf{b}}_1 \cdot \hat{\mathbf{b}}_2 = \cos\beta_3$ is a constant, the inner product

$$\mathbf{p} \cdot \mathbf{p}' = \cos 2\sigma(p, p')$$

appears linearly, and the other non-constant entries $\hat{\mathbf{b}}_i \cdot \mathbf{p} = \cos 2\sigma_i$, $\hat{\mathbf{b}}_i \cdot \mathbf{p}' = \cos 2\sigma_i'$ in the determinants are of the type (126) (or (140)).

Let D, D' be the determinants in (143). Solving the above determinant identity with respect to $\mathbf{p} \cdot \mathbf{p}'$ gives

$$\cos 2\sigma(p, p') = \frac{1}{\sin^2\beta_3}(\sqrt{DD'} + F), \tag{144}$$

where F is a bilinear form of both vectors $(\tilde{I}_1, \tilde{I}_2)$ and $(\tilde{I}_1', \tilde{I}_2')$, and moreover, $F = 0$ and $D' = \sin^2\beta_3$ when p' is the pole P of the hemisphere. The latter observation implies

$$\cos 2\sigma(p, \mathrm{P}) = \cos 2r = 4\Delta\sqrt{m_1 m_2 m_3} = \frac{\sqrt{D}}{\sin\beta_3}, \quad \text{cf. (111).}$$

Consequently, by the Ceva-Heron formula (69)

$$D = \frac{\sin^2\beta_3}{m_1 m_2 m_3} Q^* = \frac{\sin^2\beta_3}{m_1 m_2 m_3}(m_1 m_2 m_3 - Q_0^*(\tilde{I}_1, \tilde{I}_2)),$$

where $Q^* = Q|_{M^*}$ is the restriction of the quadratic form (68), and

$$Q_0^* = (1 - m_2)^2 \tilde{I}_1^2 + (1 - m_1)^2 \tilde{I}_2^2 + 2(m_3 - m_1 m_2)\tilde{I}_1 \tilde{I}_2. \tag{145}$$

On the other hand, taking $p = p'$ in (144) gives

$$\frac{F}{\sin^2\beta_3} = \frac{m_1 m_2 m_3 - Q^*}{m_1 m_2 m_3} = \frac{Q_0^*}{m_1 m_2 m_3},$$

and therefore, F as a function of $(\tilde{I}_1, \tilde{I}_2, \tilde{I}_1', \tilde{I}_2')$, is a constant times the polarization of Q_0^* in (145). This establishes the general spherical distance formula

$$\sigma(p, p') = \frac{1}{2}\arccos\left(\frac{\sqrt{Q^* Q'^*} + \frac{1}{2}Pol(Q_0^*)}{m_1 m_2 m_3}\right), \tag{146}$$

where

$$\frac{1}{2}Pol(Q_0^*) = (1-m_2)^2\tilde{I}_1\tilde{I}_1' + (1-m_1)^2\tilde{I}_2\tilde{I}_2' + (m_3 - m_1m_2)(\tilde{I}_1\tilde{I}_2' + \tilde{I}_1'\tilde{I}_2).$$

Remark 22. *By spherical trigonometry it is easy to see that the polarization term in (146) can be expressed in polar coordinates as*

$$\sin\varphi\sin\varphi'\cos(\theta-\theta').$$

5. Motions of m-triangles with conserved angular momentum

In the previous chapters we have investigated the kinematic quantities, their general relationships, and the resulting kinematic identities are valid for any motion of m-triangles, with no explicit assumption on the invariance of the angular momentum vector $\mathbf{\Omega}$. In dynamics, however, the motion is governed by a potential function and one is primarily interested in the trajectories of the equations of motion, which is a second order ODE. In these cases the invariance of $\mathbf{\Omega}$ is generally seen as a consequence of (rotational) symmetry properties of the potential function, as in the Newtonian case discussed in the introductory chapter. From this viewpoint, the linear motions (cf. Section 3.3) are the trajectories in the trivial case of a constant potential function, but still we have found them to be useful in our survey of the kinematic geometry of the moduli space $\bar{M}$.

In this chapter we turn to the study of virtual m-triangle motions, in full generality except with the explicit assumption that the angular momentum is conserved. Our aim is also to complete the proofs of the Main Theorems B, D, E1, E2 and F stated in Section 2.2.

Remark 23. *It is a classical result, dating* (*at least*) *back to Weierstrass, that a 3-body motion with* $\mathbf{\Omega} = 0$ *must be planar. There is a simple and purely kinematic proof of this fact. Namely, if* $(\mathbf{X}(t), \mathbf{n}(t))$ *is a motion of oriented m-triangles and* $\mathbf{X} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ *is nondegenerate, then* $\dot{\mathbf{n}} = 0$ *if and only if* $q_1 = q_2 = 0$*, where* $q_i = \mathbf{n}\cdot\dot{\mathbf{a}}_i$*. Therefore, the proof follows from the identity*

$$\pm(\mathbf{a}_1\times\mathbf{a}_2)\times\mathbf{\Omega} = \tilde{q}_1\mathbf{a}_1 + \tilde{q}_2\mathbf{a}_2,$$

where for $\{i,j\} = \{1,2\}$*,* $\tilde{q}_i$ *is a linear combination of* q_i *and* q_j *with coefficients* $2\Delta m_i(1-m_j)$ *and* $2\Delta m_i m_j$*, respectively.*

5.1. **Moving eigenframe and intrinsic decomposition of velocities.** Consider an m-triangle motion $\mathbf{X}(t) = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ with a (continuous) eigenframe $\mathfrak{F}(t) = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{n})$ of its inertia tensor $B_{\mathbf{X}}$ (24). For convenience, let us assume $\mathbf{X}(t)$ is nondegenerate (say, for some time interval) and hence $\mathbf{X}$ spans a plane $\Pi(\mathbf{X}) = lin\,\{\mathbf{u}_1, \mathbf{u}_2\}$. Any vector $\mathbf{v}$ in 3-space has an orthogonal splitting, $\mathbf{v} = \mathbf{v}^\tau + \mathbf{v}^\eta$, where $\mathbf{v}^\tau$ is the *tangential* component lying in the plane $\Pi(\mathbf{X})$ and $\mathbf{v}^\eta$ is the *normal* component.

We will combine the splitting (20) of the velocity of $\mathbf{X}(t)$ with its tangential and normal decomposition, namely we decompose the individual velocities $\dot{\mathbf{a}}_i$ into their tangential and normal parts and write

$$\dot{\mathbf{X}} = \dot{\mathbf{X}}^\tau + \dot{\mathbf{X}}^\eta = (\dot{\mathbf{a}}_1^\tau, \dot{\mathbf{a}}_2^\tau, \dot{\mathbf{a}}_3^\tau) + (\dot{\mathbf{a}}_1^\eta, \dot{\mathbf{a}}_2^\eta, \dot{\mathbf{a}}_3^\eta). \tag{147}$$

Recall the roles of the (instantaneous) angular velocity vector $\boldsymbol{\omega} = \boldsymbol{\omega}(t)$ and the angular momentum $\boldsymbol{\Omega}$, which by the inertia operator (23) essentially determine each other. They are responsible for the purely rotational (or rigid) motion of the m-triangle, and in accordance with (147) the rotational velocity has the splitting

$$\dot{\mathbf{X}}^\omega = \boldsymbol{\omega} \times \mathbf{X} = \boldsymbol{\omega}^\eta \times \mathbf{X} + \boldsymbol{\omega}^\tau \times \mathbf{X} \ = (\dot{\mathbf{X}}^{\boldsymbol{\omega}})^{\boldsymbol{\tau}} + (\dot{\mathbf{X}}^\omega)^\eta. \tag{148}$$

Lemma 24. *The normal velocity component of the m-triangle motion is purely rotational, that is,*

$$\dot{\mathbf{X}}^\eta = (\dot{\mathbf{X}}^\omega)^\eta = \boldsymbol{\omega}^\tau \times \mathbf{X},$$

where $\boldsymbol{\omega}^\tau$ is the tangential component of $\boldsymbol{\omega}$, and moreover,

$$\dot{\mathbf{n}} = \boldsymbol{\omega} \times \mathbf{n} = \boldsymbol{\omega}^\tau \times \mathbf{n}.$$

Proof. Consider the 2-parameter family of degenerate m-triangles

$$(\lambda \mathbf{u}_1 + \mu \mathbf{u}_2) \times \mathbf{X} = (c_1 \mathbf{n}, \mathbf{c}_2 \mathbf{n}, c_3 \mathbf{n}),$$

where the constants $c_i = c_i(\lambda, \mu)$ are linear combinations of λ and μ with the relation $\sum m_i c_i = 0$. The last identity also expresses the range of the linear transformation $(\lambda, \mu) \to (c_1, c_2, c_3)$.

On the other hand, for certain constants k_i

$$\dot{\mathbf{X}}^\eta = (k_1 \mathbf{n}, k_2 \mathbf{n}, k_3 \mathbf{n}), \quad \sum m_i k_i = 0,$$

and consequently the system of equations $c_i = k_i, i = 1, 2, 3$, has a unique solution (λ, μ), that is, there is a unique vector $\boldsymbol{\omega}'$ such that $\dot{\mathbf{X}}^\eta = \boldsymbol{\omega}' \times \mathbf{X}$. Clearly, $\boldsymbol{\omega}'$ is just the tangential component $\boldsymbol{\omega}^\tau$ of $\boldsymbol{\omega}$. Finally, $\mathbf{n}$ is a multiple of $\mathbf{a}_1 \times \mathbf{a}_2$ and $\dot{\mathbf{n}}$ is a tangential vector, so it is a simple vector algebra calculation to verify that $\dot{\mathbf{n}} = \boldsymbol{\omega}^\tau \times \mathbf{n}$. ∎

It follows that the tangential velocity version of (20) reads

$$\dot{\mathbf{X}}^{\tau} = \boldsymbol{\omega}^{\eta} \times \mathbf{X} + \dot{\mathbf{X}}^{h}$$

and, in particular, the horizontal velocity of an m-triangle motion is always tangential, whereas the rotational velocity (148) in general has a tangential and normal component.

Now, we turn to the angular momentum

$$\boldsymbol{\Omega} = \mathbf{X} \times \dot{\mathbf{X}}^{\boldsymbol{\omega}} = \mathbf{X} \times (\boldsymbol{\omega} \times \mathbf{X})$$

which we assume is a fixed vector along the z-axis, say, and the expansion

$$\boldsymbol{\Omega} \;= \Omega \mathbf{k} = \boldsymbol{\Omega}^{\tau} + \boldsymbol{\Omega}^{\eta} = (g_1 \mathbf{u}_1 + g_2 \mathbf{u}_2) + g_3 \mathbf{n} \tag{149}$$

defines its (time dependent) coordinate vector (g_1, g_2, g_3) relative to the moving frame $\mathfrak{F}(t)$. The inner product of $\boldsymbol{\Omega}$ with a vector $\mathbf{v}$ may be written as

$$\boldsymbol{\Omega} \cdot \mathbf{v} = (\mathbf{X} \times \boldsymbol{\omega}) \times \mathbf{X} \cdot \mathbf{v} = (\mathbf{X} \times \boldsymbol{\omega}) \cdot (\mathbf{X} \times \mathbf{v}) = B_{\mathbf{X}}(\boldsymbol{\omega}, \mathbf{v}).$$

Hence, by letting $\mathbf{v} = \mathbf{v}_i$ be any of the vectors from $\mathfrak{F}$,

$$g_i = \boldsymbol{\Omega} \cdot \mathbf{v}_i = B_{\mathbf{X}}(\boldsymbol{\omega}, \mathbf{v}_i) = \lambda_i \boldsymbol{\omega} \cdot \mathbf{v}_i, \;\; i = 1, 2, 3,$$

and we obtain the expansion

$$\boldsymbol{\omega} = \boldsymbol{\omega}^{\tau} + \boldsymbol{\omega}^{\eta} = (\frac{g_1}{\lambda_1} \mathbf{u}_1 + \frac{g_2}{\lambda_2} \mathbf{u}_2) + \frac{g_3}{I} \mathbf{n}, \tag{150}$$

In particular, the rotational kinetic energy can be expressed as

$$T^{\omega} = \frac{1}{2} \left| \dot{\mathbf{X}}^{\omega} \right|^2 = \frac{1}{2} B_{\mathbf{X}}(\boldsymbol{\omega}, \boldsymbol{\omega}) = \frac{1}{2} (\frac{g_1^2}{\lambda_1} + \frac{g_2^2}{\lambda_2} + \frac{g_3^2}{I}) \tag{151}$$

with tangential and normal parts

$$(T^{\omega})^{\tau} = \frac{1}{2} \frac{g_3^2}{I}, \qquad (T^{\omega})^{\eta} = \frac{1}{2} (\frac{g_1^2}{\lambda_1} + \frac{g_2^2}{\lambda_2}).$$

Now, let us also have a closer look at the individual velocities in (147) and their splitting, namely

$$\dot{\mathbf{a}}_i = \dot{\mathbf{a}}_i^{\tau} \; + \dot{\mathbf{a}}_i^{\eta} = \left(\omega_i (\mathbf{n} \times \mathbf{a}_i) + \frac{\dot{\rho}_i}{\rho_i} \mathbf{a}_i \right) + (\boldsymbol{\omega}^{\tau} \times \mathbf{a}_i), \;\text{ cf. (74)}, \tag{152}$$

where ω_i is the scalar tangential angular velocity of $\mathbf{a}_i$. We claim that

$$\omega_i = \omega_i^0 + \frac{g_3}{I} \;, \;\text{ cf. Remark 1}, \tag{153}$$

where ω_i^0 is the scalar angular velocity in the case of vanishing angular momentum, and a formula is given in Theorem C1, see (44). The proof of (153) is really the same as in Section 3.2.1, if we only consider velocity components in the plane $\Pi(\mathbf{X})$ and replace $\boldsymbol{\Omega}$ by its normal component

$\mathbf{\Omega}^\eta$. The identities in Section 3.2.1, in fact, expresses kinematic relationships valid at each moment t, and there is no need to assume $\mathbf{\Omega}$ is a constant.

Finally, according to (152) the individual kinetic energy terms expresses as

$$T_i = T_i^\tau + T_i^\eta = (\frac{1}{2}\omega_i^2 I_i + \frac{\dot{I}_i^2}{8I_i}) + \frac{1}{2}m_i \left|\boldsymbol{\omega}^\tau \times \mathbf{a}_i\right|^2 \tag{154}$$

and hence they depend, in fact, only on the moduli curve $\bar{\Gamma}(t)$ of the motion and the moving frame coordinates g_i of the angular momentum.

5.2. Final proof of the Main Theorems D, B, E1,E2.

5.2.1. *The Euler equations and proof of Theorem D.* Consider the horizontal (i.e. with vanishing angular momentum) m-triangle motion $\mathbf{X}^h(t) = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$, with the same moduli curve $\bar{\Gamma}(t)$ as $\mathbf{X}(t)$ and with moving eigenframe

$$\mathfrak{F}^h(t) = (\mathbf{u}_1^h(t), \mathbf{u}_2^h(t), \mathbf{n}(t_0)),$$

subject to the initial conditions

$$\mathbf{X}^h(t_0) = \mathbf{X}(t_0), \quad \mathfrak{F}^h(t_0) = \mathfrak{F}(t_0). \tag{155}$$

The existence of this motion is the statement of Theorem B in the simple case of vanishing angular momentum (cf. Section 2.2.2).

Let $\Gamma^*(t) = (\varphi(t), \theta(t))$ be the associated shape curve on $M^* = S^2$, expressed in the usual spherical coordinates, and consider the two nearby points

$$\mathbf{p} = (\varphi(t_0), \theta(t_0)), \qquad \mathbf{p}' = (\varphi(t_0 + \Delta t), \theta(t_0 + \Delta t))$$

with the longitude difference $\Delta\theta$. The meridians through $\mathbf{p}$ and $\mathbf{p}'$ intersect the equator circle E^* in the points $\mathbf{e}$ and $\mathbf{e}'$ respectively, and there is the piecewise geodesic closed path

$$\mathbf{p} \rightarrow \mathbf{e} \rightarrow \mathbf{e}' \rightarrow \mathbf{p}' \rightarrow \mathbf{p}$$

enclosing the shaded region as indicated in Figure 7, whose area (on the unit sphere) is

$$\Delta A \equiv \cos\varphi\ \Delta\theta$$

modulo higher orders of $\Delta\theta$.

It follows from Theorem C2 and Theorem 17 applied to the above path, with a piecewise linear motion in the plane perpendicular to $\mathbf{n}(t_0)$, that

$$\mathbf{u}_1^h(t_0 + \Delta t) \equiv \cos(\Delta\psi)\mathbf{u}_1(t_0) + \sin(\Delta\psi)\mathbf{u}_2(t_0) \tag{156}$$

$$\mathbf{u}_2^h(t_0 + \Delta t) \equiv -\sin(\Delta\psi)\mathbf{u}_1(t_0) + \cos(\Delta\psi)\mathbf{u}_2(t_0)$$

modulo higher orders of $\Delta\theta$, where

$$\Delta\psi = \frac{1}{2}\cos\varphi\ \Delta\theta. \tag{157}$$

Consequently, we infer from (156) and (157)

$$\dot{\mathbf{u}}_1^h(t_0) = (\frac{1}{2}\dot{\theta}\cos\varphi)\mathbf{u}_2|_{t_0},\ \ \dot{\mathbf{u}}_2^h(t_0) = (-\frac{1}{2}\dot{\theta}\cos\varphi)\mathbf{u}_1|_{t_0}. \tag{158}$$

Next, consider the following intrinsic frame version of (20)

$$\dot{\mathbf{u}}_i = \boldsymbol{\omega}\times\mathbf{u}_i + \dot{\mathbf{u}}_i^h, i=1,2; \quad \dot{\mathbf{n}} = \boldsymbol{\omega}\times\mathbf{n}, \tag{159}$$

By taking the inner product with $\mathbf{\Omega}$ on both sides of these identities, using (149) and (150), we perform the following calculations:

$$\begin{aligned}
\mathbf{\Omega}\cdot\dot{\mathbf{u}}_1(t_0) &= \mathbf{\Omega}\cdot(\boldsymbol{\omega}\times\mathbf{u}_1 + \dot{\mathbf{u}}_1^h)|_{t_0} \\
&= \mathbf{\Omega}\cdot\left((\frac{g_1}{\lambda_1}\mathbf{u}_1 + \frac{g_2}{\lambda_2}\mathbf{u}_2 + \frac{g_3}{I}\mathbf{n})\times\mathbf{u}_1 + (\frac{1}{2}\dot{\theta}\cos\varphi)\mathbf{u}_2\right)|_{t_0} \\
&= \mathbf{\Omega}\cdot\left(-\frac{g_2}{\lambda_2}\mathbf{n} + (\frac{g_3}{I} + \frac{1}{2}\dot{\theta}\cos\varphi)\mathbf{u}_2\right)|_{t_0} \\
&= g_2\left((\frac{1}{I} - \frac{1}{\lambda_2})g_3 + \frac{1}{2}\cos\varphi\ \dot{\theta}\right)|_{t_0}
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Omega}\cdot\dot{\mathbf{u}}_2(t_0) &= \mathbf{\Omega}\cdot(\boldsymbol{\omega}\times\mathbf{u}_2 + \dot{\mathbf{u}}_2^h)|_{t_0} \\
&= \mathbf{\Omega}\cdot\left((\frac{g_1}{\lambda_1}\mathbf{u}_1 + \frac{g_2}{\lambda_2}\mathbf{u}_2 + \frac{g_3}{I}\mathbf{n})\times\mathbf{u}_2 - (\frac{1}{2}\dot{\theta}\cos\varphi)\mathbf{u}_1\right)|_{t_0} \\
&= \mathbf{\Omega}\cdot\left(\frac{g_1}{\lambda_1}\mathbf{n} + (-\frac{g_3}{I} - \frac{1}{2}\dot{\theta}\cos\varphi)\mathbf{u}_1\right)|_{t_0} \\
&= g_1\left((\frac{1}{\lambda_1} - \frac{1}{I})g_3 - \frac{1}{2}\cos\varphi\ \dot{\theta}\right)|_{t_0}
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Omega}\cdot\dot{\mathbf{n}} = \mathbf{\Omega}\cdot(\boldsymbol{\omega}\times\mathbf{n}) &= \mathbf{\Omega}\cdot\left((\frac{g_1}{\lambda_1}\mathbf{u}_1 + \frac{g_2}{\lambda_2}\mathbf{u}_2 + \frac{g_3}{I}\mathbf{n})\times\mathbf{n}\right)|_{t_0} \\
&= \mathbf{\Omega}\cdot\left(-\frac{g_1}{\lambda_1}\mathbf{u}_2 + \frac{g_2}{\lambda_2}\mathbf{u}_1\right)|_{t_0} = (\frac{1}{\lambda_2} - \frac{1}{\lambda_1})g_1g_2|_{t_0}.
\end{aligned}$$

Since $\mathbf{\Omega}$ is a constant vector and time t_0 is arbitrary, the above three identities amount precisely to the ODE (53), and this completes the proof of Theorem D.

Finally, we turn to the precession angle $\chi(t)$ which records the motion of the normal vector $\mathbf{n}$ around the z-axis, that is, the fixed $\mathbf{\Omega}$-axis. For example, using spherical coordinates $(\tilde{\varphi},\tilde{\theta})$ on the unit sphere in Euclidean 3-space, with $\tilde{\varphi}=0$ at the north pole $\mathbf{k}$ and $\chi(t)=\tilde{\theta}(t)$, it is a simple exercise to deduce the first equality in (54), and by substituting $\dot{\mathbf{n}} = \boldsymbol{\omega} \times \mathbf{n}$ the second expression in (54) follows by calculating cross products in the frame $\mathfrak{F}$.

5.2.2. *The lifting problem and proof of Theorem B.* To complete the proof in the general case $\boldsymbol{\Omega} \neq 0$, let us also choose a horizontal lifting $\mathbf{X}^h(t)$ of $\bar{\Gamma}(t)$. Then it follows from the identity (20) that the lifting $\Gamma(t)$ is the motion $\mathbf{X}(t) = (\mathbf{a}_1(t), \mathbf{a}_2(t), \mathbf{a}_3(t))$ determined by the following initial value problem

$$\frac{d}{dt}\mathbf{X} = (\boldsymbol{\omega} \times \mathbf{X}) + \frac{d}{dt}\mathbf{X}^h, \quad \mathbf{X}(t_0) = \Gamma(t_0). \tag{160}$$

Here the vector $\boldsymbol{\omega}$ is a function of $\mathbf{X}$ and the constant vector $\boldsymbol{\Omega}$, namely for a fixed and nondegenerate $\mathbf{X}$, $\boldsymbol{\omega}$ is found by inverting the inertia operator on 3-space:

$$\mathbb{I}_{\mathbf{X}} : \boldsymbol{\omega} \to \mathbf{X} \times (\boldsymbol{\omega} \times \mathbf{X}) = \boldsymbol{\Omega}.$$

The matrix of this operator is

$$B = |\mathbf{X}|^2 \, Id - ADA^t \sim diag(\lambda_1, \lambda_2, \lambda_3),$$

where Id is the identity, the vectors $\mathbf{a}_i$ are the columns of A, and

$$D = diag(m_1, m_2, m_3).$$

The eigenvalues λ_i of B are listed in Section 2.2.4, and one of them vanishes when $\mathbf{X} \neq 0$ is degenerate. However, in that case it is easy to check that the indeterminacy of $\boldsymbol{\omega}$ is a summand along the line $\Pi(\mathbf{X})$ and consequently the summand $\boldsymbol{\omega} \times \mathbf{X}$ in (160) is still well defined as a function of $\mathbf{X}$ and $\boldsymbol{\Omega}$.

For another proof of Theorem B, more directly related to the construction of a position curve $\gamma(t)$ in $SO(3)$, we use either Theorem C1 or D. The first theorem applies to planary motions and calculates a position curve $\gamma(t)$ in $SO(2)$, recording the rotation of the vectors $\mathbf{a}_i$. In the nonplanary case the position of the m-triangle is represented by the position of its moving eigenframe $\mathfrak{F}$, and the latter is determined by the coordinate vector (g_1, g_2, g_3) of $\boldsymbol{\Omega}$ relative to $\mathfrak{F}$ together with the precession angle χ (of the normal vector $\mathbf{n}$) in the "invariant" plane perpendicular to $\boldsymbol{\Omega}$. The four functions g_i and χ are the solution of an initial value problem depending only on the moduli curve $\bar{\Gamma}(t)$, as explained by Theorem D and formula (54).

5.2.3. *Geometric reduction of Newton's equation and proof of Theorem E1 and E2.* To derive the reduced Newton's equations from the Newton's equations (1), we differentiate the kinematic quantities I_i up to second order, for example, $\dot{I}_1 = 2m_1\mathbf{a}_1 \cdot \dot{\mathbf{a}}_1$ and

$$\begin{aligned} \ddot{I}_1 &= 2m_1 \left|\dot{\mathbf{a}}_1\right|^2 + 2m_1\mathbf{a}_1 \cdot \ddot{\mathbf{a}}_1 \\ &= 4T_1 + 2m_1\mathbf{a}_1 \cdot \left(\frac{m_2}{r_{12}^3}(\mathbf{a}_2 - \mathbf{a}_1) + \frac{m_3}{r_{13}^3}(\mathbf{a}_3 - \mathbf{a}_1)\right). \end{aligned} \tag{161}$$

Then we use the Ceva-cosine law (63), stated in the form

$$\mathbf{a}_i \cdot \mathbf{a}_j = \frac{-1}{2m_i m_j} C_k = \frac{1}{2m_i m_j}(m_k I_k - m_i I_i - m_j I_j),$$

to replace all inner products in (161) by linear combinations of the $I_i's$. This procedure leads to the differential equations (55).

In the above differential equations the individual kinetic energies T_i are crucial terms, and their actual splitting (154) distinguishes the two cases of planary and non-planary 3-body motions $\Gamma(t)$. Of course, the actual case is also decided by the initial data $\Gamma(t_0), \dot{\Gamma}(t_0)$. However, it is not decided by $\Gamma(t_0)$ and the angular momentum vector, unless $\Gamma(t_0)$ is nondegenerate. Clearly, the statements of Theorem E1 and E2 must be modified if they should also cover the case where the initial configuration $\Gamma(t_0)$ is collinear.

First, observe that a planary three-body motion is characterized by having all normal kinetic energies $T_i^\eta = 0$, and then its moduli curve $\bar{\Gamma}(t)$ is a solution of the Ω-reduced equations (55) with $T_i = T_i^\tau$ given by the first summand in (154).

Conversely, let $\bar{\gamma}(t)$ be a moduli curve which is a solution of this ODE, for a given value of Ω. By Theorem B there is a unique lifting $\mathbf{X}(t)$, namely a virtual motion in the xy-plane, with angular momentum $\Omega\mathbf{k}$ and specified initial position $\mathbf{X}(t_0)$. From this knowledge we may calculate the initial velocity

$$\dot{\mathbf{X}}(t_0) = \boldsymbol{\omega}(t_0) \times \mathbf{X}(t_0) + \dot{\mathbf{X}}^h(t_0), \quad \boldsymbol{\omega}(t_0) = \frac{\Omega}{I(t_0)}\mathbf{k},$$

since the horizontal velocity $\dot{\mathbf{X}}^h(t_0)$ is determined by $\frac{d}{dt}\bar{\gamma}(t_0)$. On the other hand, Newton's equations (1) also has a unique solution $\Gamma(t)$ with the above initial conditions, namely $\Gamma(t_0) = \mathbf{X}(t_0)$ and $\dot{\Gamma}(t_0) = \dot{\mathbf{X}}(t_0)$, and clearly the moduli curve of $\Gamma(t)$ is a solution of the Ω-reduced ODE. By uniqueness of the lifting we conclude that $\Gamma(t) = \mathbf{X}(t)$ for all t, and this completes the proof of Theorem E1.

Next, we turn to the general case, described by Theorem E2. The kinetic energies T_i in the Ω-reduced ODE have the general form (154), and we claim they depend only on the moduli curve and the functions g_i. This clearly holds for the tangential summand T_i^τ, whose expression is even independent of g_1 and g_2. On the other hand, the normal summand is, say, for $i = 1$:

$$T_1^\eta = \frac{1}{2}m_1 \left|\boldsymbol{\omega}^\tau \times \mathbf{a}_1\right|^2 = \frac{1}{2}m_1 \left|(\frac{g_1}{\lambda_1}\mathbf{u}_1 + \frac{g_2}{\lambda_2}\mathbf{u}_2) \times (\cos\psi_1 \mathbf{u}_1 + \sin\psi_1 \mathbf{u}_2)\right|^2$$
$$= \frac{1}{2}m_1 (\frac{g_1}{\lambda_1}\sin\psi_1 - \frac{g_2}{\lambda_2}\cos\psi_1)^2, \tag{162}$$

where ψ_1 is the angle between $\mathbf{u}_1$ and $\mathbf{a}_1$, satisfying

$$\tan\psi_1 = \frac{\cos\varphi}{1+\sin\varphi}\tan\frac{\theta}{2}, \quad \text{cf. Theorem 18.} \tag{163}$$

In this formula θ is the longitude angle measured from the binary collision point $\hat{\mathbf{b}}_1$ and is increasing in the direction towards $\hat{\mathbf{b}}_2$. It follows (e.g. by symmetry) that one obtains the corresponding formula for T_2^η and T_3^η from (162) when ψ_1 is replaced by the corresponding angle ψ_2 and ψ_3, determined by the same formula (163) with θ measured from $\hat{\mathbf{b}}_2$ or $\hat{\mathbf{b}}_3$, respectively. This proves the above claim.

Problem 25. *It is an interesting task to simplify the expression for the energies T_i^η, and to express them in terms of coordinates I_j as in the case of T_i^τ. But we leave the topic here.*

Thus, in the general case our "reduced" ODE actually consists of the reduced Newton's equations (55) together with the Euler equations (53). The initial data will be a given nondegenerate configuration $\mathbf{X}(t_0)$ and angular momentum vector $\boldsymbol{\Omega}$, and as before, this determines the initial velocity $\dot{\mathbf{X}}(t_0)$ and hence the motion $\mathbf{X}(t)$ is unique (by Newton's equation (1)).

However, to avoid ambiguity in the initial value problem for the "reduced" ODE we must also specify the initial orientation (i.e. the normal vector $\mathbf{n}(t_0)$) which decides whether the shape curve starts out on the upper or lower hemisphere of M^*. With this choice the initial eigenframe $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{n}\}\,|_{t=t_0}$, and hence also the initial values $g_i(t_0)$, will be unique relative to a fixed convention, say, the angle ψ_1 between $\mathbf{u}_1$ and $\mathbf{a}_1$ is (initially) in the range (114). Now, the remaining part of the proof of Theorem E2 is similar to the proof of Theorem E1.

5.3. **Geometric reduction of the least action principles.** Recall from Section 1.2 the two classical least action principles with the action integral J_1 and J_2, respectively. The underlying geometric structure naturally associated to the former is *Riemannian* while that of the latter is *symplectic*. Thus the two types of least action principles are radically different in their basic geometric setting, although both of them characterize the same motion. Indeed, it is easy to verify that the Euler-Lagrange equations of both variational principles are equivalent to the Newton's equation of motion (1).

However, the classical approach to the three-body problem is mainly based on J_2, namely the least action principle of Hamilton and the Hamilton-Jacobi theory, in the framework of canonical transformations and symplectic geometry. On the other hand, the kinematic geometry of m-triangles is, on the other hand, more naturally associated with the least action principle of Euler-Lagrange-Jacobi and the action integral J_1, and it is in this geometric framework that we have established many basic results, such as the universal sphericality, the kinematic Gauss-Bonnet formula (and geometric phase), kinematic moving frames and the generalized Euler equations.

5.3.1. *Proof of Theorem F.* We consider virtual 3-body motions in the xy-plane, represented by motions $\delta(t)$ of oriented m-triangles with $\mathbf{k}$ as their common normal vector, and hence a (nondegenerate) m-triangle $\delta = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ is positively oriented if $\mathbf{a}_1 \times \mathbf{a}_2$ points in the direction of $\mathbf{k}$. The corresponding configuration space is $\mathbb{R}^4$, see (28), and $\mathbb{R}^4/SO(2) = \bar{M}$ is the full moduli space. The $SO(2)$-orbit $\bar{\delta}$ of an oriented m-triangle δ will be regarded both as a point in $\bar{M}$ and as a subset (congruence class) of $\mathbb{R}^4$. In the sequel, all motions in $\mathbb{R}^4$ are also assumed to have a constant angular momentum.

The set of differentiable (C^1-smooth) curves in $\bar{M}$ from $\bar{\delta}$ to $\bar{\delta}_1$ is denoted $\bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1}$, and similarly $\mathfrak{P}_{\delta,\bar{\delta}_1}$denotes the set of differentiable curves in $\mathbb{R}^4$ which start at δ and terminate at the orbit $\bar{\delta}_1$. For a fixed angular momentum Ω the kinetic energy

$$T = T_\Omega = \bar{T} + T^\omega = \bar{T} + \frac{\Omega^2}{2I},$$

the potential function U, the Lagrange function $L_\Omega = T_\Omega + U$ and total energy $h = T_\Omega - U$, are functions which are also defined at the level of $\bar{M}$. Therefore, for fixed value of Ω, h or time interval $[0, t_1]$, we may consider corresponding subsets of $\bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1}$

$$\bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1}(h), \ \ \bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1}([t_0, t_1]) \tag{164}$$

with the obvious meaning, and similarly subsets of $\mathfrak{P}_{\delta,\bar{\delta}_1}$

$$\mathfrak{P}_{\delta,\bar{\delta}_1}(\Omega), \;\; \mathfrak{P}_{\delta,\bar{\delta}_1}(\Omega, h), \;\; \mathfrak{P}_{\delta,\bar{\delta}_1}(\Omega, [t_0, t_1]). \tag{165}$$

Clearly, the solution curves of Newton's equation belong to sets of type (165).

We can also define the reduced action integrals

$$\bar{J}_{1,\Omega} = \int T_\Omega dt \;, \;\; \bar{J}_{2,\Omega} \; = \int L_\Omega dt \tag{166}$$

acting on moduli curves, and there is the following commutative diagram

$$\begin{tikzcd} \bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1} \arrow[dr, "\bar{J}_{i,\Omega}"'] & & \mathfrak{P}_{\delta,\bar{\delta}_1}(\Omega) \arrow[ll, "\pi_\Omega"'] \arrow[dl, "J_{i,\Omega}"] \\ & \mathbb{R} & \end{tikzcd} \tag{167}$$

where the map π_Ω takes a curve $\Gamma(t)$ to its moduli curve $\bar{\Gamma}(t)$ and $J_{i,\Omega}$ is the restriction of J_i to the subspace of curves with fixed angular momentum Ω. Moreover, for δ a nondegenerate m-triangle the map π_Ω is, in fact, a bijection due to the unique lifting property described by Theorem B.

Now, let us turn to the proof of Theorem F, which we restate as follows:

Theorem 26. *The solution curves of the planary Ω-reduced Newton's equation can be characterized as the extremal curves of $\bar{J}_{1,\Omega}$ (respectively $\bar{J}_{2,\Omega}$) restricted to the sets*

$$\bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1}(h), \;\; \textit{respectively}\; \bar{\mathfrak{P}}_{\bar{\delta},\bar{\delta}_1}([t_0, t_1])$$

of moduli curves, with fixed energy h or time interval $[t_0, t_1]$, respectively.

Proof. First of all, extremal curves of $\bar{J}_{1,\Omega}$ (respectively $\bar{J}_{2,\Omega}$) are solutions of the associated Euler-Lagrange equations, and one checks that these are second order ODE whose solution curves are (as usual) uniquely determined by their initial position and velocity.

On the other hand, the Ω-reduced Newton's equation is also a second order ODE whose solution curves are uniquely determined by their initial position and velocity. Therefore, to show that the Euler-Lagrange equations and the Ω-reduced Newton's equation have the same solutions it suffices to verify this locally. More precisely, it suffices to show that any small segment of a solution curve of the Ω-reduced Newton's equation is also an extremal curve of $\bar{J}_{1,\Omega}$ (respectively $\bar{J}_{2,\Omega}$).

Let $\bar{\Gamma}$ be a small segment from $\bar{\delta}$ to $\bar{\delta}_1$ of a solution curve of the Ω-reduced Newton's equation. We may assume the points are sufficiently close to ensure that $\bar{\Gamma}$ is the only segment (of a solution) linking them.

Choose δ in the orbit $\bar{\delta}$. Since the actions in (166) are always non-negative and the orbits $\bar{\delta}$ and $\bar{\delta}_1$ are sufficiently close, there exists a J_1-minimizing (respectively J_2-minimizing) curve Γ in $\mathbb{R}^4$ between δ and the orbit $\bar{\delta}_1$, say δ_1 is its end point. Then Γ is a solution of Newton's equation and it must, in fact, be the lifting of $\bar{\Gamma}$. Moreover, it is the unique curve in $\mathfrak{P}_{\delta,\bar{\delta}_1}(\Omega)$ with minimal action integral of J_1 (respectively J_2).

Consequently, Γ is, of course, also a small segment of an extremal curve of $J_{i,\Omega}$ and therefore by (167) its moduli curve $\bar{\Gamma}$ is a small segment of an extremal curve of $\bar{J}_{i,\Omega}$. ∎

Remark 27. *Theorem F does not extend to the case of general three-body motions. The reason is that the kinetic energy of a motion $\delta(t)$ does not depend only on the moduli curve and the angular momentum vector $\mathbf{\Omega}$, but also on the instantaneous configuration $\delta(t)$. Hence, one cannot proceed as above, since it is not clear what should be the appropriate action $\bar{J}_{i,\Omega}$ at the moduli space level.*

6. The Newtonian potential function

In this chapter our primary task is to analyze the Newtonian potential function (2) and its crucial dependence on the mass distribution. The function is naturally defined at moduli space level,

$$U = \sum_{i=1}^{3} \frac{\hat{m}_i}{s_i} = \sum_{i=1}^{3} \frac{\hat{m}_i^{3/2}}{\sqrt{(1-m_i)I - I_i}}, \tag{168}$$

where in each half-space $\bar{M}_\pm = \mathbb{R}^3_\pm$ the two triples (I_1, I_2, I_3) and (s_1, s_2, s_3), related by (64), are natural coordinate systems. Certainly, U has the simplest possible form when expressed by the mutual distances s_i. Even so, sometimes it is also convenient to use Euclidean coordinates or their associated spherical coordinates (I, φ, θ), where $I = \rho^2 = \sqrt{x^2 + y^2 + z^2}$, as explained in Section 4.5.

Let U^* be the restriction of U to the "unit" sphere $M^* = S^2 = (I = 1)$. As a function of the coordinates I_j (respectively s_j) U is homogeneous of degree $-\frac{1}{2}$ (respectively -1), namely

$$U(I_1, I_2, I_3) = \frac{1}{\rho} U^*(\frac{I_1}{I}, \frac{I_2}{I}, \frac{I_3}{I}) = \frac{1}{\rho} U^*(\delta^*) = \frac{1}{\rho} U^*(\varphi, \theta),$$

where $\delta^* \longleftrightarrow (\varphi, \theta)$ represents the shape of an m-triangle. For the sake of convenience, the formula for U^* in terms of spherical coordinates is

$$U^* = \sum_{i=1}^{3} \frac{\hat{m}_i^{3/2}(m_i^*)^{-1/2}}{\sqrt{1 - \sin\varphi\cos(\theta - \theta_i)}}, \tag{169}$$

where $\theta_1, \theta_2, \theta_3$ are the longitude angles of the binary collision points $\mathfrak{b}_{23}, \mathfrak{b}_{31}, \mathfrak{b}_{12}$, respectively. This follows by substituting the expression (123) with $\tilde{\theta}_i = \theta - \theta_i$ into (168). In particular, with the convention that $\theta_1 = 0$ made in Remark 15, we must use

$$(\theta_1, \theta_2, \theta_3) = (0, \beta_3, -\beta_2),$$

where the angles β_i, described in (131) - (133), measure the longitude differences between the points $\mathfrak{b}_{kl}$.

The analysis of U trivially reduces to that of the restriction $U^* = U\,|_{M^*}$. In fact, by symmetry it suffices to investigate U^* on the closed upper hemisphere, $0 \le \varphi \le \pi/2$. U^* has no maximum points since U^* tends to ∞ at the singular points $\mathfrak{b}_{kl}$ (which are poles of U^*). On the other hand, it is a classical result, dating (at least) back to Lagrange, that U^* has a unique minimum value at the shape of a regular triangle, namely (for $\rho = 1$)

$$s_1 = s_2 = s_3 = \frac{1}{\sqrt{\hat{m}}}, \quad \text{or} \quad I_i = 1 - m_i - \frac{\hat{m}_i}{\hat{m}}. \tag{170}$$

This defines a unique point $\mathfrak{p}_0^{\pm}$ on each hemisphere $M_{\pm}^*$ which we also refer to as the *physical center* (as opposed to the poles which are the *geometric center*). It is easy to prove the above statement using the coordinates s_i and Lagrange's multiplier method subject to the constraint $1 = I = \sum m_i^* s_i^2$, cf. (130). Another proof follows from Lemma 30 below. The spherical coordinates of the shape (170) is worked out in Section 8.8, cf. (322), (323).

6.1. **Vector algebra analysis of the Newtonian function.** The vector algebra representation of $\bar{M}$ in Section 4.5 is also a convenient setting for the local analysis of U^*, namely the function on the unit sphere of Euclidean 3-space defined by

$$U^*(\mathbf{p}) = \sum_{i=1}^{3} U_i^*(\mathbf{p}) = \sum_{i=1}^{3} \frac{k_i}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|}, \qquad k_i = \frac{2\hat{m}_i^{3/2}}{\sqrt{1 - m_i}}, \tag{171}$$

where

$$\left|\mathbf{p} - \hat{\mathbf{b}}_i\right| = \frac{k_i}{\hat{m}_i} s_i \qquad \text{(cf. (141))}$$

is the Euclidean distance from $\mathbf{p}$ to the binary collision point $\hat{\mathbf{b}}_i$.

Fix a point $\mathbf{p}$ on the sphere with $z > 0$, and consider nearby points $\mathbf{p}' = \mathbf{p} + \mathbf{x}$ on the sphere, that is, $\mathbf{x}$ is subject to the constraint

$$2\mathbf{p} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{x} = 0, \tag{172}$$

which in turn implies

$$\left|\mathbf{p}' - \hat{\mathbf{b}}_i\right|^2 = \left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^2 - 2\hat{\mathbf{b}}_i \cdot \mathbf{x}.$$

Hence, there is the following expansion at $\mathbf{p}$

$$U^*(\mathbf{p} + \mathbf{x}) = \sum_{i=1}^{3} \frac{U_i^*(\mathbf{p})}{(1 - z_i)^{1/2}} = \sum_{i=1}^{3} U_i^*(\mathbf{p}) \sum_{n=0}^{\infty} c_n z_i^n = \sum_{n=0}^{\infty} F_n(\mathbf{p}; \mathbf{x}), \tag{173}$$

where we use the notation

$$z_i = \frac{2\hat{\mathbf{b}}_i \cdot \mathbf{x}}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^2} = (\frac{m_i^* m_i}{\bar{m}})\frac{\hat{\mathbf{b}}_i \cdot \mathbf{x}}{s_i^2}$$

$$c_n = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n n!}, \quad c_0 = 1.$$

Remark 28. *It is easy to see that the convergence condition* $|z_i| < 1$ *for the series of* $U_i^*(\mathbf{p} + \mathbf{x})$ *in* (173) *is equivalent to the condition*

$$(\mathbf{p} + \mathbf{x}) \cdot \hat{\mathbf{b}}_i > 2\mathbf{p} \cdot \hat{\mathbf{b}}_i - 1.$$

Geometrically, this means that $\mathbf{p} + \mathbf{x}$ *belongs to the hemispherical cap* D_i *centered at* $\hat{\mathbf{b}}_i$ *which is cut out by the plane parallel to the tangent plane at* $\hat{\mathbf{b}}_i$ *and separated by the distance* $R_i = 2(1 - \mathbf{p} \cdot \hat{\mathbf{b}}_i)$. *In particular,* D_i *is the whole hemisphere if* $\mathbf{p} \cdot \hat{\mathbf{b}}_i \leq 0$*, and by Lemma 21, we know* $\mathbf{p} \cdot \hat{\mathbf{b}}_i \leq 0$ *holds for at least one* i. *The domain of convergence for the series of* U^* *is the "polygonal" region* $\cap D_i$.

The zero order term of the expansion (173) is, of course, $F_0 = U^*(\mathbf{p})$, and the first order term is

$$F_1(\mathbf{p}; \mathbf{x}) = (\frac{k_1 \hat{\mathbf{b}}_1}{\left|\mathbf{p} - \hat{\mathbf{b}}_1\right|^3} + \frac{k_2 \hat{\mathbf{b}}_2}{\left|\mathbf{p} - \hat{\mathbf{b}}_2\right|^3} + \frac{k_3 \hat{\mathbf{b}}_3}{\left|\mathbf{p} - \hat{\mathbf{b}}_3\right|^3}) \cdot \mathbf{x} \tag{174}$$

which is essentially the gradient of U^* at $\mathbf{p}$. To make this precise, consider the following function from $S^2(1)$ to the xy-plane

$$\mathbf{p} \to \mathbf{B}(\mathbf{p}) = \sum_{i=1}^{3} \frac{k_i \hat{\mathbf{b}}_i}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^3} \in \mathbb{R}^2. \tag{175}$$

Lemma 29. *The gradient vector of U^* at $\mathbf{p}$ is given by*

$$\nabla U^*(\mathbf{p}) = \mathbf{B}(\mathbf{p}) - (\mathbf{B}(\mathbf{p}) \cdot \mathbf{p})\mathbf{p}. \tag{176}$$

Proof. Since by (174)

$$\nabla U^*(\mathbf{p}) \cdot \mathbf{x} = F_1(\mathbf{p}; \mathbf{x}) = \mathbf{B}(\mathbf{p}) \cdot \mathbf{x}$$

and $\mathbf{x}$ is tangential to $\mathbf{p}$ in the limit as $\mathbf{x} \to 0$, it follows that

$$\nabla U^*(\mathbf{p}) \cdot \mathbf{t} = \mathbf{B}(\mathbf{p}) \cdot \mathbf{t}$$

holds for all tangent vectors $\mathbf{t}$. Hence, the tangent vector $\nabla U^*(\mathbf{p})$ is the orthogonal projection of $\mathbf{B}(\mathbf{p})$ in the direction of $\mathbf{p}$. ∎

Lemma 30. *The zero points of $\mathbf{B}$ are the two critical points of U^* outside the equator circle $z = 0$, namely the physical center $\hat{\mathbf{p}}_0^{\pm}$ (cf. (170)) on each hemisphere $z > 0$ or $z < 0$.*

Proof. By Lemma 21, $\mathbf{B}(\mathbf{p}) = 0$ if and only if for some constant $\lambda > 0$,

$$m_1^* \frac{\left|\mathbf{p} - \hat{\mathbf{b}}_1\right|^3}{k_1} = m_2^* \frac{\left|\mathbf{p} - \hat{\mathbf{b}}_2\right|^3}{k_2} = m_3^* \frac{\left|\mathbf{p} - \hat{\mathbf{b}}_3\right|^3}{k_3} = \lambda,$$

and by (141), this is equivalent to $s_1^3 = s_2^3 = s_3^3 = \frac{\lambda}{2}$, namely $s_1 = s_2 = s_3$. In particular, $\mathbf{p}$ is a point with $z \neq 0$, cf. (177c). On the other hand, for $z \neq 0$ it is easy to see that $\mathbf{B}(\mathbf{p}) = 0$ if and only if $\nabla U^*(\mathbf{p}) = 0$. ∎

The identity (176) also implies that the critical points of U^* on the unit circle $z = 0$ are the "eigenvectors" of $\mathbf{B}$, in the sense that $\mathbf{B}(\mathbf{p}) = \lambda\mathbf{p}$ for some λ, necessarily equal to $\mathbf{B}(\mathbf{p}) \cdot \mathbf{p} \neq 0$. Clearly, $\mathbf{B}$ has a pole at $\hat{\mathbf{b}}_i$ and $\mathbf{B}(\mathbf{p})/\left|\mathbf{B}(\mathbf{p})\right|$ tends to $\hat{\mathbf{b}}_i$ as $\mathbf{p}$ tends to $\hat{\mathbf{b}}_i$. A simple analysis of $\mathbf{B}$ will show there are exactly one solution $\mathbf{p}$ between each pair $\hat{\mathbf{b}}_j, \hat{\mathbf{b}}_k$ of poles. These are the so-called Euler points $\hat{\mathbf{e}}_i, i = 1, 2, 3$, and they are the saddle points of U^*on the 2-sphere. We omit the proof of this well known fact.

6.1.1. *Series expansion of U^* at its minimum point.* Henceforth, we shall focus attention on the expansion (173) at the physical center $\hat{\mathbf{p}}_0 = (\hat{x},\hat{y}, \hat{z}), \hat{z} > 0$, that is, $\hat{\mathbf{p}}_0$ is the minimum point of U^* on the hemisphere $z > 0$. The coordinates of $\hat{\mathbf{p}}_0$ are the following mass dependent constants

$$\hat{x} = \hat{\mathbf{p}}_0 \cdot \hat{\mathbf{b}}_1 = 1 - \frac{\bar{m}}{\hat{m} m_1^* m_1}, \tag{177a}$$

$$\hat{y} = \frac{\sqrt{\bar{m}}}{\hat{m}}\left(\frac{m_2 - m_3}{m_2 + m_3}\right), \tag{177b}$$

$$\hat{z} = \cos 2r_0 = 4\Delta_0 \sqrt{m_1 m_2 m_3} = \sqrt{3}\frac{\sqrt{\bar{m}}}{\hat{m}}, \tag{177c}$$

where Δ_0 is the area of the regular triangle with $I = 1$. These expressions follow from (111), (140) and (170). Note, for example, that $\hat{z}$ becomes arbitrarily small when some mass m_i tends to zero, and $\hat{z} = 1$, that is, $\hat{\mathbf{p}}_0$ is the north pole $\mathcal{N}$, precisely when the masses are equal. Concerning the sign of $\hat{y} = \pm\sqrt{1 - \hat{x}^2 - \hat{z}^2}$, we used (140) to check, for example, that $\hat{y} > 0$ if $m_2 > m_3$.

The constant term of the series (173) is the minimum value

$$F_0 = U^*(\hat{\mathbf{p}}_0) = \sum U_i^*(\hat{\mathbf{p}}_0) = \sum \sqrt{\hat{m}}\hat{m}_i = \hat{m}^{3/2}, \tag{178}$$

and $F_1 = 0$, of course. Moreover, by (170), $\hat{m}s_i^2 = 1$ holds for all i, and therefore the n-th order term can be written as

$$F_n(\hat{\mathbf{p}}_0; \mathbf{x}) = \left(\frac{c_n \hat{m}^{n+\frac{1}{2}}}{\bar{m}^{n-1}}\right) \sum_{i=1}^{3} \frac{1}{m_i}(m_i m_i^*)^n (\hat{\mathbf{b}}_i \cdot \mathbf{x})^n. \tag{179}$$

Now, let us turn to the local analysis of the series, namely U^* developed as a power series in suitable coordinates around $\hat{\mathbf{p}}_0$. To this end, consider a positively oriented orthonormal frame of the Euclidean 3-space of type

$$\{\mathbf{t}_1, \mathbf{t}_2, \hat{\mathbf{p}}_0\}, \quad \mathbf{t}_i \in \Pi \text{ for } i = 1, 2, \tag{180}$$

where Π is the tangent space of the sphere at $\hat{\mathbf{p}}_0$. By condition (172), the components of $\mathbf{x} = \mathbf{p} - \hat{\mathbf{p}}_0$ must satisfy

$$\mathbf{x} = \xi\mathbf{t}_1 + \eta\mathbf{t}_2 + \zeta\hat{\mathbf{p}}_0, \quad 2\zeta + \xi^2 + \eta^2 + \zeta^2 = 0, \tag{181}$$

and consequently the map

$$(\xi, \eta) \to (\xi, \eta, \zeta),$$

where

$$\zeta = -1 + \sqrt{1 - (\xi^2 + \eta^2)} = -\frac{1}{2}(\xi^2 + \eta^2) - \frac{1}{8}(\xi^2 + \eta^2)^2 - \dots, \tag{182}$$

is a parametrization of the region of the sphere lying above the plane through the origin and parallel to Π. Geometrically, the unit disk of Π is projected down to the sphere in the direction of the normal vector $\hat{\mathbf{p}}_0$.

It is natural to choose $\mathbf{t}_1$ tangent to the meridian through $\hat{\mathbf{p}}_0$. Indeed, the intrinsic nature of this condition will lead to a series expansion whose coefficients are essentially symmetric functions of the masses m_i. The projection in the xy-plane of such an orthonormal basis $\{\mathbf{t}_1, \mathbf{t}_2\}$ of Π is (up to sign) given by

$$\mathbf{t}_1' = \frac{\hat{z}}{\sqrt{1-\hat{z}^2}}(\hat{x}, \hat{y}), \ \ \mathbf{t}_2' = \frac{1}{\sqrt{1-\hat{z}^2}}(-\hat{y}, \hat{x}) = \mathbf{t}_2. \tag{183}$$

In order to express the n-th term (179) of the U^*-series in terms of ξ, η, we proceed as follows. Expand the "variables" in (179)

$$\hat{\mathbf{b}}_i \cdot \mathbf{x} = a_i'\xi + b_i'\eta + c_i'\zeta, \ \ i = 1, 2, 3 \tag{184}$$

where the three triples (a_i', b_i', c_i') of coefficients are specific functions of the parameters m_i determined by

$$a_i' = \hat{\mathbf{b}}_i \cdot \mathbf{t}_1 = \hat{\mathbf{b}}_i \cdot \mathbf{t}_1', \ \ b_i' = \hat{\mathbf{b}}_i \cdot \mathbf{t}_2, \ \ c_i' = \hat{\mathbf{b}}_i \cdot \hat{\mathbf{p}}_0. \tag{185}$$

For example, by (139) and (183),

$$a_1' = \frac{\hat{x}\hat{z}}{\sqrt{1-\hat{z}^2}}, \ \ b_1' = \frac{-\hat{y}}{\sqrt{1-\hat{z}^2}} \ , \ \ c_1' = \hat{x},$$

where $\hat{x}, \hat{y}, \hat{z}$ are the known functions in (177a) - (177c), and the three triples in (185) permute covariantly with the parameters m_i.

In view of (179), it is slightly more convenient to replace (184) by

$$(m_i m_i^*)\hat{\mathbf{b}}_i \cdot \mathbf{x} = a_i\xi + b_i\eta + c_i\zeta, \ \ i = 1, 2, 3$$

and hence we calculate the modified coefficients, namely

$$\begin{aligned} a_i &= \frac{\sqrt{3\bar{m}}}{\sqrt{\hat{m}^2 - 3\bar{m}}}\left(m_i m_i^* - \frac{\bar{m}}{\hat{m}}\right), \\ b_i &= \frac{-\sqrt{\bar{m}}m_i}{2\sqrt{\hat{m}^2 - 3\bar{m}}}(m_{i+1} - m_{i+2}) \qquad (i \mod 3) \\ c_i &= m_i m_i^* - \frac{\bar{m}}{\hat{m}}. \end{aligned} \tag{186}$$

Thus, with the coordinate system (ξ, η), we finally arrive at the following *symmetrization* of the power series in (173):

$$\begin{aligned} U^*(\xi, \eta) &= \sum_{n=0}^{\infty} K_n \sum_{i=1}^{3} \frac{1}{m_i}(a_i\xi + b_i\eta + c_i\zeta)^n \\ &= \sum_{n=0}^{\infty} K_n \sum_{j+k=n} \binom{n}{j} A_{j,k}\xi^j\eta^k, \end{aligned} \tag{187}$$

where

$$K_n = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n n!} \frac{\hat{m}^{n+\frac{1}{2}}}{\bar{m}^{n-1}},$$

and $A_{j,k}$, k even, are symmetric functions of the masses m_i, whereas $A_{j,k}$ is alternating symmetric when k is odd.

Remark 31. *The Newton sums $S_k = \sum m_i^k$ are, of course, polynomials of $\hat{m}$ and $\bar{m}$, cf.* (15). *For example,*

$$S_4 = 1 - 4\hat{m} + 4\bar{m} + 2\hat{m}^2.$$

In terms of the S_k it is rather straightforward to obtain explicit expressions for the above symmetric functions $A_{j,k}$. For k odd, the alternating function $A_{j,k}$ is a product of a symmetric function and the basic alternating function

$$\begin{aligned}\mathfrak{A} &= (m_1 - m_2)(m_2 - m_3)(m_3 - m_1) \\ &= \sum_{i \bmod 3} m_i(1 - m_i)(m_{i+1} - m_{i+2}).\end{aligned} \tag{188}$$

6.1.2. *The quadratic term of U^*.* We shall work out explicitly the quadratic term of the function U^* expanded at the physical center $\hat{\mathbf{p}}_0$, namely the $n = 2$ term of (173) or (187)

$$F_2 = \frac{3\hat{m}^{5/2}}{8\bar{m}}(A\xi^2 + 2B\xi\eta + C\eta^2) = \kappa(\tilde{\lambda}_1\tilde{\xi}^2 + \tilde{\lambda}_2\tilde{\eta}^2), \tag{189}$$

where $(\tilde{\xi}, \tilde{\eta})$ is the coordinate system of a diagonalizing frame $\{\tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2\}$ of the tangent plane Π at $\hat{\mathbf{p}}_0$.

First, let us determine the coefficients $A = A_{2,0}, B = A_{1,1}, C = A_{0,2}$ in (187) as symmetric or alternating functions of the symbols m_i. By using (188), the identities

$$\sum m_i(1 - m_i)^2 = \hat{m} + 3\bar{m}, \quad \sum m_i(m_{i+1} - m_{i+2})^2 = \hat{m} - 9\bar{m}$$

and the expressions (186), we find

$$\begin{aligned}A &= \sum \frac{1}{m_i} a_i^2 = \frac{1}{4}\frac{3\bar{m}}{\hat{m}}\left(\frac{\hat{m}^2 + 3\hat{m}\bar{m} - 4\bar{m}}{\hat{m}^2 - 3\bar{m}}\right) \\ C &= \sum \frac{1}{m_i} b_i^2 = \frac{1}{4}\frac{\bar{m}(\hat{m} - 9\bar{m})}{\hat{m}^2 - 3\bar{m}} \\ B &= \sum \frac{1}{m_i} a_i b_i = \frac{1}{4}\left(\frac{-\sqrt{3}\bar{m}}{\hat{m}^2 - 3\bar{m}}\right)\mathfrak{A}.\end{aligned}$$

Hence, the eigenvalues of the quadratic in (189) are determined from the equations

$$\tilde{\lambda}_1 + \tilde{\lambda}_2 = A + C = \frac{\bar{m}}{\hat{m}}, \quad \tilde{\lambda}_1\tilde{\lambda}_2 = AC - B^2 = \frac{3}{4}\frac{\bar{m}^2}{\hat{m}},$$

which yield

$$\tilde{\lambda}_i = \frac{\bar{m}}{2\hat{m}}(1 \pm \sqrt{1 - 3\hat{m}}). \tag{190}$$

Consequently,

$$F_2 = \frac{3}{16}\hat{m}^{3/2}\left((1 \pm \mu)\,\tilde{\xi}^2 + (1 \mp \mu)\tilde{\eta}^2\right), \quad \mu = \sqrt{1 - 3\hat{m}}, \tag{191}$$

where $(\tilde{\xi}, \tilde{\eta})$ are coordinates with respect to the eigenvectors

$$\tilde{\mathbf{t}}_1 = \cos\tilde{\alpha}\ \mathbf{t}_1 + \sin\tilde{\alpha}\ \mathbf{t}_2, \quad \tilde{\mathbf{t}}_2 = -\sin\tilde{\alpha}\ \mathbf{t}_1 + \cos\tilde{\alpha}\ \mathbf{t}_2$$

obtained by rotating the intrinsic frame $\{\mathbf{t}_1, \mathbf{t}_2\}$ of the plane Π_0.

Similar to (117), the angle $\tilde{\alpha}$ is given by

$$\tan 2\tilde{\alpha} = \frac{2B}{A - C} = \frac{-\sqrt{3}\hat{m}(m_1 - m_2)(m_2 - m_3)(m_3 - m_1)}{\hat{m}^2 + 9\bar{m}\hat{m} - 6\bar{m}}. \tag{192}$$

Then it also follows from (119) that the largest eigenvalue $\tilde{\lambda}_1 \sim 1 + \mu$ corresponds to the vector $\tilde{\mathbf{t}}_1$ if and only if $\sin 2\tilde{\alpha}$ and B have the same sign.

Remark 32. *In the simplest case of uniform mass distribution, $m_i = 1/3$, the analysis of U^* is much simpler than in the general case. In this case, where the geometrical and physical center coincide, some of the above expressions such as* (192)*, are of indeterminate type. However, if $m_i = m_j \neq m_k$ then $\tan 2\tilde{\alpha} = 0$, and hence the frame $\{\mathbf{t}_1, \mathbf{t}_2\}$ is already diagonalizing.*

7. A geometric setting for the study of triple collisions

Recall the well known fact, stated by Weierstrass and proved by Sundman (cf. [14], [15]), that three-body motions leading to triple collision must have vanishing angular momentum, $\mathbf{\Omega} = 0$, and consequently they are also planary. Thus we shall focus attention on planary virtual motions $\mathbf{X}(t)$, namely curves in the configuration space $M_0 \simeq \mathbb{R}^4$, with zero angular momentum, and we continue to use the vector algebra representation (cf. Section 4.5) of the moduli space $\bar{M} = M_0/SO(2) = \mathbb{R}^3$, where the (equator) xy-plane $\bar{E} = \mathbb{R}^2$ represents congruence classes of eclipse (i.e. collinear) configurations.

The total kinetic energy $T = \bar{T}$ can be expressed as a positive definite quadratic differential form on $\bar{M} \setminus \{O\}$, namely the kinematic Riemannian metric

$$d\bar{s}^2 = 2Tdt^2 = d\rho^2 + \rho^2 d\sigma^2 = d\rho^2 + \frac{\rho^2}{4}ds^2\ , \ \text{cf. (138)}, \tag{193}$$

which describes $\bar{M}$ as the Riemannian cone over the shape space $M^* = S^2(1/2)$, and

$$ds^2 = 4d\sigma^2 = d\varphi^2 + (\sin^2\varphi)d\theta^2 \tag{194}$$

is the metric of the magnified sphere $S^2(1)$.

Following Jacobi, we introduce the following conformal modification of the metric (193) for each energy level h, namely

$$d\bar{s}_h^2 = (U+h)d\bar{s}^2, \tag{195}$$

which we refer to as the *physical metric*, and transform the action integral $\bar{J}_{1,0}$ of (166) into the arc-length integral in the Riemannian space

$$(\bar{M}_h, d\bar{s}_h^2): \ \bar{M}_h = \left\{\mathbf{p} \in \bar{M}; U(\mathbf{p}) + h \geq 0\right\}. \tag{196}$$

Consequently, the trajectories of three-body motions with total energy h are mapped to curves in $\bar{M}$ which are geodesics in the space $(\bar{M}_h, d\bar{s}_h^2)$. Notice that reflection in the equator plane $\bar{E} \subset \bar{M}$, that is, the transformation $\varphi \to \pi - \varphi$, restricts to an involutive isometry of the Riemannian space (196) with the eclipse subspace $\bar{E}_h = \bar{E} \cap \bar{M}_h$ as fixed point set, and hence this is a totally geodesic submanifold of $\bar{M}_h$. In particular, a geodesic curve in $(\bar{M}_h, d\bar{s}_h^2)$ is transversal to $\bar{E}_h$, unless it lies entirely in $\bar{E}_h$. Moreover, for a simple geometrical reason, a shortest geodesic in $(\bar{M}_h, d\bar{s}_h^2)$ between a point outside $\bar{E}_h$ and the origin O cannot have any intermediate intersection with $\bar{E}_h$.

Finally, we note that Newton's equation (1) has a 1-parameter group of space-time scaling symmetries $\{g_s\}$, where g_s sends a solution $\mathbf{X}(t)$ to a solution

$$\mathbf{Y}(t) = e^{2s/3}\mathbf{X}(e^{-s}t) \tag{197}$$

and changes the energy level from h to $e^{-2s/3}h$. Hence, all the Riemannian structures in (196) with energy h of the same sign are mutually homothetic, and consequently there are essentially only three distinct cases, namely when the total energy h is negative, zero or positive.

7.1. **Geodesic rays and distance estimates.** Clearly, for $h \geq 0$ the variety $\bar{M}_h$ is the whole moduli space $\bar{M}$, whereas for $h < 0$ it is the star-shaped union of all ray segments

$$\left[O, \frac{U^*(\mathbf{p})}{|h|}\mathbf{p}\right], \ \mathbf{p} \in M^* \tag{198}$$

from the origin O to the point where the ray through $\mathbf{p}$ intersects the boundary $\partial\bar{M}_h$, that is, the level surface $U = -h$. By definition, the physical metric (195) vanishes on the boundary, meaning that the distance between any two boundary points is zero.

For $h < 0$ the length of any ray segment (198) is

$$L_h(\mathbf{p}) = \int d\bar{s}_h = \int_0^{\frac{U^*(\mathbf{p})}{|h|}} \sqrt{\frac{U^*(\mathbf{p})}{\rho} + h}\; d\rho, \tag{199}$$

and therefore there is a unique pair of shortest length

$$L_h = L_h(\hat{\mathbf{p}}_0^{\pm}) = \int_0^{\frac{\mu_0}{|h|}} \sqrt{\frac{\mu_0}{\rho} + h}\; d\rho, \tag{200}$$

where the points $\hat{\mathbf{p}}_0^{\pm}$ on the hemispheres $z > 0$ and $z < 0$ represent the shape of a regular triangle and hence U^* has the minimal value

$$\mu_0 = U^*(\hat{\mathbf{p}}_0^{\pm}) = \hat{m}^{3/2}, \quad \text{cf. (178).}$$

The following is a useful fact in Riemannian geometry which follows from general analysis of the first variation of arc-length.

Lemma 33. *Let ds^2 and $d\tilde{s}^2$ be two Riemannian metrics on a given manifold such that*

$$d\tilde{s}^2 = f^2 ds^2,$$

where f is a smooth and positive function, that is, $d\tilde{s}^2$ is a conformal modification of ds^2. Let Γ be a C^2-smooth curve and let $\mathbf{n}$ denote a normal vector at a given point on Γ. Then the geodesic curvatures of Γ in (the direction of $\mathbf{n}$) with respect to the two metrics are related by

$$\tilde{\mathcal{K}}(\mathbf{n}) = \mathcal{K}(\mathbf{n}) - \frac{d}{d\mathbf{n}} \ln f.$$

We shall apply the lemma to the kinematic and physical metric, namely the metrics $d\bar{s}^2$ and $d\bar{s}_h^2$ on $\bar{M}$, cf. (195). Thus, a moduli curve $\bar{\Gamma}$ is a geodesic with respect to $d\bar{s}_h^2$ if and only if its geodesic curvature $\tilde{\mathcal{K}}(\mathbf{n})$ with respect to $d\bar{s}_h^2$ vanishes for all $\mathbf{n}$, or equivalently

$$\mathcal{K}(\mathbf{n}) = \frac{1}{2}\frac{d}{d\mathbf{n}} \ln(U + h), \tag{201}$$

where $\mathcal{K}(\mathbf{n})$ is the geodesic curvature in the normal direction $\mathbf{n}$, with respect to $d\bar{s}^2$.

The simplest type of 3-body motions are the *shape invariant* ones, that is, the shape curve is a single point on $M^* = S^2$ and hence the moduli curve is confined to a ray emanating from O in the cone $\bar{M}$. Rays are, of course, geodesics with respect to the kinematic metric $d\bar{s}^2$, consequently a ray through $\mathbf{p} \in S^2$ (or a ray segment (198) if $h < 0$) is also a geodesic of the metric $d\bar{s}_h^2$ if and only if the normal derivative vanishes in all directions $\mathbf{n}$ normal to the ray, that is,

$$\frac{d}{d\mathbf{n}}\ln(\frac{1}{\rho}U^* + h) = 0.$$

This condition is independent of the radial coordinate ρ, and for $\rho = 1$ the vectors $\mathbf{n}$ span the tangent plane of S^2 at the point $\mathbf{p}$. Consequently, the solutions are the five critical points $\mathbf{p}$ of U^*, namely the three saddle points (called Euler points) $\hat{\mathbf{e}}_i$ on the equator circle $\rho = 1$ in the xy-plane, and the pair $\hat{\mathbf{p}}_0^{\pm}$ of minimumspoints (also called Lagrange points). Thus, there are altogether exactly five geodesic rays (or ray segments) in $(\bar{M}_h, d\bar{s}_h^2)$.

Lemma 34. *For $h < 0$, the two ray segments $\left[O, \frac{\mu_0}{|h|}\hat{\mathbf{p}}_0^{\pm}\right]$ are the unique shortest geodesic curves in $(\bar{M}_h, d\bar{s}_h^2)$ linking a boundary point and the base point O (ignoring curve pieces of zero length along $\partial\bar{M}_h$). In particular, the distance from $\partial\bar{M}_h$ to O is the number L_h in* (200).

Proof. In $(\bar{M}, d\bar{s}^2)$, let B_h be the geodesic ball of radius $\frac{\mu_0}{|h|}$ centered at O. It lies inside $\bar{M}_h$ and touches $\partial\bar{M}_h$ at the two points $\frac{\mu_0}{|h|}\hat{\mathbf{p}}_0^{\pm}$. If $\bar{\Gamma}$ is any curve between O and a point $\mathbf{q}$ on $\partial\bar{M}_h$, let $\bar{\Gamma}_1$ be the portion of $\bar{\Gamma}$ between O and the first point $\mathbf{q}_1$ on $\partial\bar{M}_h$. From the calculation

$$L(\bar{\Gamma}) \geq L(\bar{\Gamma}_1) = \int \sqrt{h+U}d\bar{s} \geq \int \sqrt{h + \frac{\mu_0}{\rho}}d\bar{s} \geq \int_0^{\frac{\mu_0}{|h|}} \sqrt{h + \frac{\mu_0}{\rho}}d\rho = L_h$$

it is clear that the two ray segments of length L_h are, indeed, the shortest curves, and they are unique (modulo a portion along the boundary). ∎

Remark 35. *For $h \geq 0$, the geodesic rays through $\hat{\mathbf{p}}_0^{\pm}$ are still length minimizing, whereas for any h this fails for the three geodesic rays (or segments) in the eclipse plane $\bar{E}$.*

7.2. **Existence of triple collision motions with minimal action.** We shall combine the above differential geometric setting, Theorem F and Hilbert's direct method to study the *existence problem* of three-body motions leading to triple collision, starting from a given non-degenerate m-triangle and with *minimal* action integral (7), say. When this problem is pushed down to the level of $\bar{M}$, at a given energy level h, it can be reduced to the problem of existence of a shortest geodesic, with respect to the metric $d\bar{s}_h^2$, between a given point in $\bar{M}_h - \bar{E}$ and the base point O.

Remark 36. *For any energy h there are Newtonian motions, with the constant shape of a regular triangle or an Euler configuration, through which the m-triangle shrinks homothetically to a triple collision in finite time. To find the time parametrization of such a three-body motion, in fact, amounts to solve a two-body* (*or Kepler*) *problem, and this leads to the classical solutions found by Lagrange and Euler, see* [2], [7], [13]. *For these motions minimal action is achieved for the Lagrange motions* (*regular triangle*), *but not for the Euler motions* (*which are collinear*).

More generally, let us first consider the case $h < 0$, and define the variety

$$\bar{D}_h = \left\{ \mathbf{p} \in \bar{M}_h ; d(\mathbf{p}, O) \leq d(\mathbf{p}, \partial\bar{M}_h) + L_h \right\}, \tag{202}$$

where $d(\mathbf{p}, O)$ (respectively $d(\mathbf{p}, \partial\bar{M}_h)$) is the distance between O and $\mathbf{p}$ (respectively $\partial\bar{M}_h$) in $\bar{M}_h$ with the metric $d\bar{s}_h^2$. The interior D_h (respectively boundary $\partial\bar{D}_h$) of $\bar{D}_h$ is defined by strict inequality (respectively equality) in (202).

Remark 37. *The two surfaces $\partial\bar{M}_h$ and $\partial\bar{D}_h$ are interesting geometric objects in the study of triple collision orbits. They touch each other at the two points $\frac{\mu_0}{|h|}\hat{\mathbf{p}}_0^{\pm}$ on $\partial\bar{M}_h$ closest to the cone vertex O.*

We will prove the following existence result:

Theorem 38. (*cf.* [3], *Theorem* 5) *Starting from a given oriented m-triangle δ whose congruence class $\bar{\delta}$ belongs to $\bar{D}_h$, $h < 0$, there is a three-body motion with total energy h and minimal action integral which leads to a triple collision.*

Proof. As a consequence of Theorem F, the proof reduces to the existence of a curve with minimal length in $\bar{M}_h$ linking $\bar{\delta}$ to O. For $\bar{\delta}$ in D_h such a curve is necessarily a geodesic.

Assume first $\bar{\delta} \in D_h$, and let $\left\{\bar{\Gamma}_i\right\}$ be a sequence of curves in $\bar{M}_h$ between $\bar{\delta}$ and O whose lengths satisfy

$$L(\bar{\Gamma}_i) < d(\bar{\delta}, \partial\bar{M}_h) + L_h, \quad \lim_{i\to\infty} L(\bar{\Gamma}_i) = d(\bar{\delta}, O).$$

In particular, each $\bar{\Gamma}_i$ is disjoint from the boundary $\partial\bar{D}_h$.

Let us divide $\bar{\Gamma}_i$ into $m2^i$ segments of equal length and replace each segment by the unique shortest geodesic between its end points. Then it is quite straightforward to apply the direct method of Hilbert to find a suitable subsequence of $\left\{\bar{\Gamma}_i\right\}$ with a limiting curve $\bar{\Gamma}$, and this is necessarily a geodesic curve in $\bar{M}_h$ between $\bar{\delta}$ and O with minimal length $L(\bar{\Gamma}) = d(\bar{\delta}, O)$.

Assume next $\bar{\delta} \in \partial \bar{D}_h$, and let $\{\bar{\delta}_k\}$ be as sequence of points in D_h with $\bar{\delta}$ as its limit. Moreover, let $\{\bar{\Gamma}_k\}$ be a sequence of curves, where $\bar{\Gamma}_k$ is a shortest geodesic between $\bar{\delta}_k$ and O, that is, $L(\bar{\Gamma}_k) = d(\bar{\delta}_k, O)$. It follows that

$$\lim_{k\to\infty} d(\bar{\delta}_k, O) = d(\bar{\delta}, O) \tag{203}$$

and it is not difficult so see that there is a suitable subsequence of $\{\bar{\Gamma}_k\}$ with a limiting curve $\bar{\Gamma}$ whose length is the limit (203), and moreover, $\bar{\Gamma}$ is a geodesic between $\bar{\delta}$ and O. ∎

Finally, we consider the case $h \geq 0$, namely when $\bar{M}_h = \bar{M}$, and then we have the following analogue of the above theorem .

Theorem 39 (cf. [3], Theorem 5'). *Starting from a given oriented m-triangle δ, for a given energy level $h \geq 0$ there is always a three-body motion leading to triple collision and with minimal action integral.*

Proof. This is similar to the case $\bar{\delta} \in D_h$ of the previous proof, and the application of Hilbert's direct method will give the existence of the curve we seek. ∎

Remark 40. *The direct method of Hilbert can, of course, also be applied to study the existence problem of a geodesic curve $\bar{\Gamma}$ realizing the minimal distance between two given points $\bar{\delta}_1$ and $\bar{\delta}_2$ of $(\bar{M}_h, d\bar{s}_h^2)$. Then, by Theorem B, there are liftings Γ of $\bar{\Gamma}$ which are planary three-body motions (with specified angular momentum) starting from a given oriented m-triangle δ_1 belonging to the congruence class $\bar{\delta}_1$. However, the end point configuration δ_2 of Γ is already determined by $\bar{\Gamma}$ and δ_1, according to Theorem C2. Consequently, only three-body motions with specific relative positions of their initial and terminal m-triangles δ_i can have minimal action integrals. In the above two theorems there is no such relative position constraint since the triple collision configuration $\delta_2 = O$ consists of a single congruence class.*

7.3. **The uniqueness problem for triple collision motions with minimal action.** In view of the above existence theorems, it is natural to investigate the following uniqueness problem for motions starting from a given configuration at a sufficiently large energy level.

Problem 41. *To a given non-degenerate oriented m-triangle δ and energy level h above a lower bound, say $h \geq \mathfrak{h}(\bar{\delta})$, is there a unique three-body motion from δ leading to a triple collision with minimal action integral?*

This problem requires a considerable amount of in-depth analysis of the geodesic equation of $(\bar{M}_h, d\bar{s}_h^2)$, and here we shall leave it as an open problem.

However, to facilitate future analytical studies of the above problem and related problems we shall discuss a geometric reduction technique which reflects some useful feature of the Riemannian structure of $\bar{M} = C(M^*)$ as a cone over the subspace $M^* = (\rho = 1)$. In $\bar{M}$ the integral curves of the vector field $\frac{\partial}{\partial \rho}$ are the rays emanating from O, and they define the *radial* (i.e. a natural "vertical") direction at every point $\neq O$.

The above problem is, indeed, simple and has an optimal solution in the special case mentioned in Remark 36, namely for the shape invariant motions of Lagrange type. A 3-body motion is *shape invariant* if its moduli curve $\bar{\Gamma}$ is confined to a ray, that is, the associated shape curve Γ^* is a single point. In fact, if a Newtonian motion is shape invariant over some time interval of length > 0, then for all time Γ^* is a single point (necessarily a critical point of U^*). However, in view of Remark 36, the collinear solutions with the shape of an Euler point are not even action minimizing.

In general, let $\bar{\Gamma}$ be a smooth curve in $\bar{M} - \{O\}$. Since the associated shape curve Γ^* is the radial projection of $\bar{\Gamma}$ onto the transversal subspace M^*, Γ^* will be smooth as long as $\bar{\Gamma}$ is transversal to the radial direction, whereas a *cusp* may occur at points where this fails. Hence, in the long run the typical shape curves of 3-body motions are rather piecewise smooth, but still they can be parametrized by arc-length. Moreover, unless Γ^* is a single point, $\bar{\Gamma}$ may also be parametrized by the arc-length parameter of Γ^*.

Definition 42. *Let $\bar{\Gamma}$ be a curve in the moduli space $\bar{M}$ and let Γ^* be the associated shape curve. The cone consisting of all rays emanating from O and passing through points on $\bar{\Gamma}$ (or Γ^*) is called the* cone surface *of $\bar{\Gamma}$ (or Γ^*), and it is denoted either $C(\bar{\Gamma})$ or $C(\Gamma^*)$.*

We assume $\bar{\Gamma}$ (and hence also Γ^*) has a given orientation. Since the metric on $M^* = S^2(1/2)$ is denoted by $d\sigma^2$ (cf. e.g. 38), σ also denotes the arc-length parameter of Γ^*. For σ ranging over some interval $[\sigma_0, \sigma_1]$, the corresponding surface $C(\Gamma^*)$ is immersed in $(\bar{M}, d\bar{s}^2)$ with the induced kinematic metric

$$d\bar{s}^2|_{C(\Gamma^*)} = d\rho^2 + \rho^2 d\sigma^2; \quad \sigma_0 \leq \vartheta \leq \sigma_1, \tag{204}$$

and hence it is isometric to a flat Euclidean sector of angular width $\sigma_1 - \sigma_0$, with (ρ, σ) as polar coordinates centered at the origin O.

The moduli space $\bar{M} \approx \mathbb{R}^3$ has the standard (right handed) orientation and, in particular, the 2-sphere M^* has the induced orientation with $\frac{\partial}{\partial\rho}$ as positive normal vector field. The surface $C(\Gamma^*)$ is naturally oriented with the positive orthonormal frame

$$\left\{\frac{\partial}{\partial\rho}, \frac{1}{\rho}\frac{\partial}{\partial\sigma}\right\}, \tag{205}$$

and we choose its normal vector field $\boldsymbol{\nu}$ so that $\boldsymbol{\nu}$ followed by the frame (205) is a positive orthonormal frame in $\bar{M}$.

In (204) the curve Γ^* becomes the circular arc of radius $\rho = 1$, whereas the (original) moduli curve $\bar{\Gamma}$ appears as a radial deformation of Γ^*. When $\bar{s}$ in (204 is viewed as the arc-length parameter of

$$\bar{\Gamma} : \bar{s} \to (\rho(\bar{s}), \sigma(\bar{s})),$$

(204) becomes an identity along the curve.

The extrinsic geometry of $C(\Gamma^*) \subset \bar{M}$ is completely determined by the extrinsic geometry of $\Gamma^* \subset M^*$. Indeed, the *lines of curvature* are the two families of *coordinate curves,* namely the rays (σ constant) and the "circles" (ρ constant) in $C(\Gamma^*)$. The principal curvature of $C(\Gamma^*)$ at a point $\bar{\delta}$ is zero in the ray direction and is equal to $\mathcal{K}_g^*/\rho$ in the direction of $\frac{\partial}{\partial\sigma}$, where $\mathcal{K}_g^*$ is the geodesic curvature of Γ^* in M^* at the corresponding point δ^*.

Along the curve $\bar{\Gamma}$ we will also consider the positive orthonormal moving frame $\{\boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\nu}\}$, where $\boldsymbol{\tau}$ is the unit tangent vector in the (chosen) positive direction of $\bar{\Gamma}$, and hence the frame $\{\boldsymbol{\tau}, \boldsymbol{\eta}\}$ of $C(\Gamma^*)$ differs from the stationary frame (205) by a certain rotation angle α. Namely, we define the *(radial) inclination angle* α of $\bar{\Gamma}$ by writing

$$\boldsymbol{\tau} = \cos\alpha\frac{\partial}{\partial\rho} + \sin\alpha\frac{1}{\rho}\frac{\partial}{\partial\sigma}, \quad \boldsymbol{\eta} = -\sin\alpha\frac{\partial}{\partial\rho} + \cos\alpha\frac{1}{\rho}\frac{\partial}{\partial\sigma}, \tag{206}$$

$$\cos\alpha = \frac{d\rho}{d\bar{s}}, \quad \sin\alpha = \rho\frac{d\sigma}{d\bar{s}}\ , \quad \cot\alpha = \frac{1}{\rho}\frac{d\rho}{d\sigma} = \frac{d}{d\sigma}\ln\rho. \tag{207}$$

Briefly, α is the angle between the ray direction and the tangent direction, and $0 \le \alpha \le \pi$ since $\sin\alpha$ in (207) is not negative. The extreme values $\alpha = 0, \pi$ occur when $\bar{\Gamma}$ is not transversal to the radial direction, in which case α and Γ^* (as functions of σ or time) may encounter a singularity, namely Γ^* encounters a cusp.

Remark 43. *The angle $\alpha(\sigma)$ and the radial distance $\rho(\sigma)$ are mutually dependent according to* (207)*. For example, we have for $\rho(\sigma_0) \neq 0$*

$$\rho(\sigma) = \rho(\sigma_0)\exp\left(\int_{\sigma_0}^{\sigma}\cot\alpha(\sigma)d\sigma\right). \tag{208}$$

The geodesic condition for a curve $\bar{\Gamma}$ in $(\bar{M}_h, d\bar{s}_h^2)$ is evidently equivalent to two identities of type (201), namely for two linearly independent normal vectors $\mathbf{n}$ to $\bar{\Gamma}$. Thus, we shall consider the two cases

$$\begin{aligned} &(i)\ \mathbf{n} = \boldsymbol{\eta} : \text{ tangential to } C(\bar{\Gamma}) \\ &(ii)\ \mathbf{n} = \boldsymbol{\nu} : \text{ perpendicular to } C(\bar{\Gamma}). \end{aligned} \tag{209}$$

The first case amounts to the characterization of $\bar{\Gamma}$ as a geodesic in the (truncated) cone surface $C(\bar{\Gamma}) \cap \bar{M}_h$ with the metric $d\bar{s}_h^2$, as follows:

Lemma 44. *Let $u(\sigma)$ be the restriction of the potential function U^* along the shape curve $\Gamma^*(\sigma)$. Then the geodesic equation for the moduli curve $\bar{\Gamma}$ in the cone surface $C(\bar{\Gamma})$ with the metric*

$$d\bar{s}_h^2 = (\frac{1}{\rho}U^* + h)(d\rho^2 + \rho^2 d\sigma^2)$$

is equivalent to the equation

$$\frac{d\alpha}{d\bar{s}} + \frac{d\sigma}{d\bar{s}} = \frac{1}{2\rho}\left(\sin\alpha \frac{u(\sigma)}{u(\sigma) + h\rho} + \cos\alpha \frac{u'(\sigma)}{u(\sigma) + h\rho}\right). \tag{210}$$

Proof. Let $\boldsymbol{\eta}$ be the normal vector in (206). Then the geodesic condition is by (201)

$$\mathcal{K}(\boldsymbol{\eta}) = \frac{1}{2}\frac{d}{d\boldsymbol{\eta}}\ln(\frac{u(\sigma)}{\rho} + h), \tag{211}$$

where $\mathcal{K}(\boldsymbol{\eta})$ is the (geodesic) curvature of $\bar{\Gamma}$ in the Euclidean sector (204). However, in a Euclidean plane it is easy to see that $\mathcal{K}(\boldsymbol{\eta})$ can be expressed as $d\zeta/d\bar{s}$, where ζ is the angle between a fixed reference ray $\sigma = 0$ (say, the positive x-axis) and the tangent line, in fact, $\zeta = \alpha + \sigma$, see Figure 8. Finally, calculation of the normal derivative on the right side of the identity (211), using the orthonormal frame (205), leads to the formula (210). ∎

In the second case of (209) the geodesic condition is the identity (201) with $\mathbf{n}$ equal to the normal $\boldsymbol{\nu}$ of the surface. In this case $\mathcal{K}(\boldsymbol{\nu})$ equals the normal sectional curvature of $C(\bar{\Gamma})$ in the direction of $\bar{\Gamma}$, namely the value $\Pi(\boldsymbol{\tau}, \boldsymbol{\tau})$ of the second fundamental form. The latter has the frame (205) as eigenvectors, with eigenvalues 0 and $\mathcal{K}_g^*/\rho$ respectively, and hence by (206) and Euler's classical formula for the decomposition of normal geodesic curvature

$$\mathcal{K}_g^* \sin^2\alpha = \rho\mathcal{K}(\boldsymbol{\nu}) = \frac{\rho}{2}\frac{d}{d\boldsymbol{\nu}}\ln(\frac{1}{\rho}U^* + h). \tag{212}$$

As a summary we now state the following theorem, valid as long as the quantities involved are well defined.

Theorem 45 (cf. [3], Theorem 6)**.** *In the moduli space $\bar{M}$ with the kinematic Riemannian metric $d\bar{s}^2$, let $\bar{\Gamma}$ be the oriented moduli curve of a three-body motion with total energy h and vanishing angular momentum, and let Γ^* be the corresponding shape curve on the sphere $M^* = S^2(1/2)$ with unit tangent (respectively normal) vector $\boldsymbol{\tau}^*$ (respectively $\boldsymbol{\nu}^*$) so that $\{\boldsymbol{\tau}^*, \boldsymbol{\nu}^*\}$ is a positive frame on the sphere. Then $\bar{\Gamma} = (\rho, \Gamma^*)$ can be characterized as a solution of the following system of ODE*

$$(i) : \frac{d\alpha}{d\sigma} = -1 + \frac{1}{2}\frac{u(\sigma)}{u(\sigma) + h\rho}\left(1 + \cot\alpha \frac{d}{d\boldsymbol{\tau}^*}\ln(U^*)\right), \tag{213}$$

$$(ii) : \mathcal{K}_g^* \sin^2\alpha = \frac{1}{2}\frac{u(\sigma)}{u(\sigma) + h\rho}\ \frac{d}{d\boldsymbol{\nu}^*}\ln(U^*),$$

where $\mathcal{K}_g^*$ is the geodesic curvature of Γ^* in M^*, σ is the arc-length parameter of Γ^*, U^* is the restriction of U to M^* and $u(\sigma)$ is its further restriction along Γ^*, and $\alpha \in [0, \pi]$ is the angle between the (outgoing) ray direction and $\bar{\Gamma}$ in $\bar{M}$.

Proof. By using the expression for $\sin\alpha$ in (207), equation (210) can be stated as

$$\frac{d\alpha}{d\bar{s}} = (-1 + \frac{1}{2}\frac{u(\sigma)}{u(\sigma) + h\rho})\frac{d\sigma}{d\bar{s}} + \frac{1}{2\rho}\cos\alpha\frac{u'(\sigma)}{u(\sigma) + h\rho}.$$

When we replace the arc-length parameter $\bar{s}$ of $\bar{\Gamma}$ by σ, using a formula from (207), this equation reads

$$\frac{d\alpha}{d\sigma} = (-1 + \frac{1}{2}\frac{u(\sigma)}{u(\sigma) + h\rho}) + \frac{1}{2}\cot\alpha\frac{u'(\sigma)}{u(\sigma) + h\rho},$$

and by viewing $u'(\sigma)/u(\sigma)$ as the tangential derivative of $\ln(U^*)$ we obtain the first equation (213).

The second equation of (213) is merely a reformulation of (212), whose right side may be expressed as

$$\frac{1}{2}\frac{d}{d\boldsymbol{\nu}^*}\ln(\frac{1}{\rho}U^* + h) = \frac{2^{-1}}{u(\sigma) + h\rho}\frac{d}{d\boldsymbol{\nu}^*}U^*.$$

Here we use the fact that the normal vector $\boldsymbol{\nu}$ along $\bar{\Gamma}$ may be identified with the scaling of $\boldsymbol{\nu}^*$ by the factor $1/\rho$, that is, $\boldsymbol{\nu} = \boldsymbol{\nu}^*/\rho$, and moreover, differentiation in the direction of $\boldsymbol{\nu}$ commutes with the scaling. ∎

Remark 46. *The above system* (213) *is easily seen to be scaling invariant. Namely, when the size function ρ is multiplied by a fixed constant $k > 0$, the energy level changes as $h \to h/k$ and hence the product $h\rho$ stays invariant. In particular, since the energy level $h = 0$ is invariant with respect to scaling of solutions, the explicit dependence on ρ in* (213) *disappears in this case. Moreover, the angle* α*, geometrically interpreted in* (206) *as the inclination angle of the moduli curve* $\bar{\Gamma}$*, is a neat scaling invariant which together with* Γ^* *represents* $\bar{\Gamma}$ *uniquely up to scaling. In fact,* ρ *is generally obtained from* α *by "quadrature" along* Γ^*, cf. (208)*. This explains the following result.*

Corollary 47. *Let the pair* (α, Γ^*) *represent the moduli curve* $\bar{\Gamma}$ *of a three-body motion with vanishing angular momentum and vanishing total energy, where the shape curve* Γ^* *is not a single point and is viewed as a curve on the standard sphere* S^2 *of radius 1. Then* (α, Γ^*) *is a solution of the following system of ODE*

$$\begin{cases} \frac{d\alpha}{ds} = -\frac{1}{4} + \frac{1}{2}\cot\alpha \frac{d}{d\boldsymbol{\tau}^*}\ln(U^*) \\ \mathcal{K}_g^* \sin^2\alpha = \frac{1}{2}\frac{d}{d\boldsymbol{\nu}^*}\ln(U^*) \end{cases} \tag{214}$$

where $s = 2\sigma$ *is the arc-length parameter of* Γ^* *on* S^2 *and* $\mathcal{K}_g^*$ *is its geodesic curvature. Moreover, a solution* (α, Γ^*) *can only encounter a singularity* (*cusp*) *when* $\alpha = 0$ *or* π*, or when* Γ^* *reaches a collision point.*

We are particularly interested in applying the system (213) or (214) to the study of triple collision motions. A triple collision is simply expressed by the condition $\rho = 0$, but this singular event is not explicitly visible in (214) since the variable ρ is eliminated. However, the term $h\rho$ in (213) also disappears when $\rho \to 0$, so the two systems should behave "similarly" in the limit. Hence, the system (214) is likely to be significant also when $h \neq 0$.

One of the major results of Sundman and Siegel in their work on the local analysis of triple collisions prove the existence of both a *limiting shape*, necessarily a critical point of U^*, and a *limiting position*, cf. [14], [15], [11], [12]. The existence of a limiting position is the statement that the 3-body motion $\Gamma(t)$ has a "size normalized" limit,

$$\frac{\Gamma(t)}{|\Gamma(t)|} \to \delta, \tag{215}$$

at the configuration space level, and we shall express the statement concerning the limiting shape (due to Sundman) by saying the pair (α, Γ^*) approaches a specific pair $(\hat{\alpha}, \delta^*)$, namely

$$\hat{\alpha} \in \{0, \pi\} = \partial[0, \pi] \ , \ \ \delta^* \in \left\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3, \hat{\mathbf{p}}_0^+, \hat{\mathbf{p}}_0^-\right\} \subset S^2. \tag{216}$$

It is also known (cf. e.g. Siegel-Moser[13], p. 89) that an Euler point $\hat{\mathbf{e}}_i$ can only be the limiting shape of a triple collision motion confined to a fixed line. (However, $\hat{\mathbf{e}}_i$ may well be the limiting shape of a non-collinear motion as $t \to \pm\infty$).

The two "boundary" values of α in (216) actually distinguish between the two events *triple explosion* and *triple collision,* as follows: $\alpha = 0$ when $\bar{\Gamma}$ starts (or "explodes") out from the cone vertex O of $\bar{M}$, and $\lim \alpha = \pi$ when $\bar{\Gamma}$ is oriented towards O and terminates with a "total collapse". Anyhow, we are free to run a three-body motion in either directions, and the associated initial value problem for (213) or (214) is (a priori) of singular type in the above case since $\sin\alpha = 0$. In fact, a solution (α, Γ^*) of (214) may also encounter another type of singularity (called cusp) when $\sin\alpha = 0$, but with $\rho \neq 0$. In Chapter 8 these events and related problems will be further investigated in selected testing cases.

8. Case study of triple collision motions with zero energy

8.1. **The basic setting and statement of Theorem G.** Due to the simplicity of the system (214), the special case of vanishing total energy, $h = 0$, lends itself as the simplest testing case of three-body motions leading to a triple collision. We shall investigate this case more carefully, and for convenience, let us also restrict ourselves to the case of equal masses, $m_i = 1/3$, which largely simplifies the series expansions of the potential function and its derivatives.

We shall address the triple collision problem as an initial value problem, namely as a *triple explosion,* although we usually write "triple collision" motions. Let us first recall the so-called *Lagrange-Jacobi* equation which is the result of differentiating $I = \rho^2$ twice with respect to time t, using the homogeneity of U and conservation of energy $h = T - U$, namely in our case

$$\frac{d^2}{dt^2} I = 2(T + h) = 2T > 0. \tag{217}$$

It follows that I is a nonnegative convex function of time t, and starting from a triple collision (say, $I(0) = 0$) it is strictly increasing and tends to ∞ as $t \to \infty$.

Following the setup from Section 7.3, we seek a description of the moduli curves of triple collision motions, valid for some appropriate time interval $[0, t_1]$. For this purpose it is convenient to use the coordinates (ρ, φ, θ) in the cone $\bar{M} = C(M^*)$, where as before $\rho = \sqrt{I}$ and (φ, θ) are spherical coordinates on the unit sphere $S^2 = M^*$ centered at the north pole $\mathcal{N}$. In this setting a moduli curve $\bar{\Gamma}$ and the associated shape curve Γ^* have coordinate representations

$$\bar{\Gamma}(s) = (\rho(s), \varphi(s), \theta(s)), \ \ \Gamma^*(s) = (\varphi(s), \theta(s)), \tag{218}$$

where $s = s(t)$ is the arc-length parameter $s = s(t)$ of Γ^* and is an increasing function of time t. Here we must exclude, of course, the well understood shape invariant motions, namely the trivial case that Γ^* is a single point (in which case $\rho(t)$ is the solution of a 1-dimensional Kepler problem).

Thus, we seek a description of all those moduli curves $\bar{\Gamma}(s)$ emanating from a triple collision, at $s = 0$ say. According to Remark 46 it suffices to consider the class of $\bar{\Gamma}$ modulo scaling, represented by the pair (α, Γ^*) where $\alpha(s) \in [0, \pi]$ is the (radial) inclination angle of $\bar{\Gamma}$. In fact, recall from (208) that the size function ρ of $\bar{\Gamma}$ is recovered from (α, Γ^*) by the general quadrature formula

$$\rho(s) = \rho(s_0) e^{\frac{1}{2}\int_{s_0}^{s} \cot(\alpha) ds}, \quad \rho(s_0) \neq 0. \tag{219}$$

Remark 48. *Using the parameter s rather than time t is, of course, crucial for our geometric approach below. The relationship between s and t is follows from the kinematic metric, using e.g.* (204), (207), (219), (264)*. Namely, in the case of zero angular momentum there are the identities*

$$2(U+h)dt^2 = 2T dt^2 = d\bar{s}^2 = d\rho^2 + \frac{\rho^2}{4} ds^2 = (\cos^2\alpha) d\bar{s}^2 + \frac{\rho^2}{4} ds^2,$$

from which we deduce the relationship

$$dt = \frac{\rho(s)}{2^{3/2}\sin\alpha(s)\sqrt{u(s)/\rho(s) + h}} ds, \tag{220}$$

where $u(s) = U^*(\Gamma^*(s))$. Moreover, by switching over to t it is, in fact, not difficult to see that ρ (at any time t_0) can be determined solely from the time parametrized shape curve $\Gamma^*(t)$ and the normal derivative of U^* (near $t = t_0$).

Now, resuming the assumption $h = 0$, our approach is to determine the above pairs (α, Γ^*) by solving the system

$$ODE^* : \begin{cases} \frac{d\alpha}{ds} = -\frac{1}{4} + \frac{1}{2}\cot\alpha \frac{d}{d\boldsymbol{\tau}^*}\ln(U^*) \\ \mathcal{K}_g^* \sin^2\alpha = \frac{1}{2}\frac{d}{d\boldsymbol{\nu}^*}\ln(U^*) \\ 1 = (\frac{d\varphi}{ds})^2 + (\sin^2\varphi)(\frac{d\theta}{ds})^2 \end{cases} \tag{221}$$

as an appropriate initial value problem which represents a triple collision, see (242). The system (221) is a copy of (214) since the third equation merely expresses the constraint that Γ^* is a curve on the unit sphere. We will refer to the first and second equation of (221) as the *inclination* and *curvature* equation respectively. The first one relates the growth of the inclination angle α (of the moduli curve $\bar{\Gamma}$) with the tangential derivative of $\ln(U^*)$ along Γ^*, whereas the second one - which is of order two- relates α to the geodesic curvature of Γ^* and the normal derivative of $\ln(U^*)$ along Γ^*.

The following theorem summarizes the main result of this chapter. It describes the family $\mathfrak{S}(\hat{\mathbf{p}}_0)$ of all shape curves $\Gamma^*(s)$ representing triple collision motions, with the limiting shape of $\hat{\mathbf{p}}_0$ at the collision.

Theorem $\mathbf{G}_1$ *In the case of uniform mass distribution and zero total energy, consider the family $\mathfrak{S}(\hat{\mathbf{p}}_0)$ of arc-length parametrized shape curves $\Gamma^*(s), s \geq 0$, which emanate from the north pole $\hat{\mathbf{p}}_0 = \Gamma^*(0)$ of the 2-sphere S^2 and represent 3-body motions with a triple collision at $s = 0$. This family has the following properties:*

(i) *There is a unique curve $\Gamma^*_{\theta_0}$ for each initial longitude direction θ_0.*

(ii) *The family is invariant under the induced action of the dihedral isometry group $\mathfrak{D}_3$ of S^2 which fixes $\hat{\mathbf{p}}_0$ and permutes the three Euler points $\hat{\mathbf{e}}_i$. In particular, $\Gamma^*_{\theta_0+2\pi/3}$ is obtained from $\Gamma^*_{\theta_0}$ by rotating the sphere, $\theta \to \theta + 2\pi/3$.*

(iii) *Each curve stays within a sector of angular width $\pi/3$ and bounded by meridians representing the shape of isosceles triangles, at least until the first eclipse (i.e. crossing the equator circle).*

(iv) *Each curve extends analytically through $s = 0$ and $\Gamma^*_{\theta_0+\pi}(s) = \Gamma^*_{\theta_0}(-s)$, and it has no singularity before the first eclipse.*

(v) *For each θ_0 the associated inclination angle function $\alpha_{\theta_0}(s)$ is the unique analytic solution of* (221) *along $\Gamma^*_{\theta_0}$, with the singular initial condition $\alpha(0) = 0$, $\alpha'(0) > 0$. Moreover, $\alpha_{\theta_0+\pi}(s) = -\alpha_{\theta_0}(-s)$.*

(vi) *The sign of the curvature of the curves in $\mathfrak{S}(\hat{\mathbf{p}}_0)$ depends only on the sector, and in neighboring sectors the sign is opposite.*

Remark 49. *In the case of rectilinear three-body motions, the corresponding sets $\mathfrak{S}(\hat{\mathbf{e}}_i)$ are obviously "congruent", each consisting of the pair $\Gamma^*_\pm$ of arcs of the equator circle, in opposite directions and starting at the Euler point $\hat{\mathbf{e}}_i$. The associated inclination angle function $\alpha_\pm(s)$ is defined by a unique analytic function $\alpha(s)$ so that $\alpha_\pm(s) = \alpha(s)$ for $s \geq 0$, and $\alpha(-s) = -\alpha(s)$. We refer to Section* 8.4.

We also refer to Section 8.6.2 for more information concerning the geometric behavior of the curves in $\mathfrak{S}(\hat{\mathbf{p}}_0)$. Indeed, we are actually close to a stronger version of Theorem G_1, but the proof needs more elaboration, so we formulate the following conjecture as an open problem.

Conjecture 50. *The different triple collision shape curves* $\Gamma^*(s), s \geq 0$, *intersect the equator circle the first time at different points, and moreover, each point on the circle is reached by a unique curve. The curves do not intersect each other, except possibly after the first eclipse.*

Corollary 51. *Under the current hypothesis of uniform mass distribution and zero total energy, consider the "moduli space" consisting of all three-body motions in 3-space which start from a triple explosion at time* $t = 0$, *and moreover, the motion is neither rectilinear nor shape invariant (i.e. homographic). This space can be naturally identified with the manifold*

$$SO(3) \times SO(2) \times \mathbb{R}^+.$$

In particular, the "moduli space" for those triple collision three-body motions confined to a fixed plane is

$$O(2) \times SO(2) \times \mathbb{R}^+.$$

Indeed, starting with the space $\mathfrak{S}(\hat{\mathbf{p}}_0) \simeq SO(2)$ of shape curves described in Theorem G_1, $SO(2) \times \mathbb{R}^+$ is the space of their associated curves $\bar{\Gamma}(s)$ in the moduli space $\bar{M}$ since the size function $\rho(s)$ can be scaled by any positive number $\lambda \in \mathbb{R}^+$ without affecting the shape curve. Next, the possible liftings $\Gamma(s)$ of $\bar{\Gamma}(s)$ to three-body motions with $\Gamma(0) = 0$ are distinguished by the normalized limit δ in (215), which is an oriented, regular m-triangle Δ of unit size. We may identify the various positions of such an oriented triangle with the rotation group $SO(3)$ which measures its "deviation" from a fixed reference position. In terms of the fibration

$$O(2) \to SO(3) \to SO(3)/O(2) \simeq \mathbb{R}P^2$$

we can say that the projective plane $\mathbb{R}P^2$ represents the choices of 2-planes containing Δ (and the motion), whereas $O(2)$ represents the possible positions of an oriented regular triangle in a given plane. However, we mention that there is no global "field" of reference positions (or gauge) for all the planes, since this would imply the fibration is trivial, which is certainly not true.

8.2. **Analysis of the potential function for equal masses.** In this chapter we shall choose the zero meridian $\theta = 0$ for the polar coordinate system (φ, θ) of S^2 different from the convention in Remark 15. Namely, the three binary collision points $\hat{\mathbf{b}}_i, i = 1, 2, 3$, which are now equally spaced along on the equator circle $\varphi = \pi/2$, will have the longitude angles

$$\theta_1 = -\frac{\pi}{3}, \quad \theta_2 = \frac{\pi}{3}, \quad \theta_3 = \pi \qquad \text{(cf. Figure 9).} \tag{222}$$

Note, for example, the antipodal point of $\hat{\mathbf{b}}_i$ is the Euler point $\hat{\mathbf{e}}_i$, and now $\theta = 0$ at $\hat{\mathbf{e}}_3$. It is also convenient to use negative values of φ, with the usual interpretation so that (φ, θ) and $(-\varphi, \theta + \pi)$ is the same point on the 2-sphere. This is consistent with our trigonometric formulae for $U^*(\varphi, \theta)$ below, see (224) and (246).

For convenience, let us normalize U^* by a constant factor to make its minimum value $U^*(\hat{\mathbf{p}}_0) = 1$. In fact, scaling of U^* has no effect on the system (221). Thus, for $\mathbf{p} \in S^2$

$$U^*(\mathbf{p}) = \frac{\sqrt{2}}{3} \sum_{i=1}^{3} \frac{1}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|} = \frac{1}{3} \sum_{i=1}^{3} \frac{1}{1 - z_i)^{1/2}}, \tag{223}$$

where (for any mass distribution, indeed)

$$\left|\mathbf{p} - \hat{\mathbf{b}}_i\right| = \sqrt{2}(1 - z_i)^{1/2}, \ \text{ with } z_i = \sin\varphi \cos(\theta - \theta_i), \tag{224}$$

is the usual Euclidean distance from $\mathbf{p}$ to the binary collision point $\hat{\mathbf{b}}_i$ and θ_i is the longitude angle of $\hat{\mathbf{b}}_i$, cf. (141), (169), (171), 222.

Consequences of the invariance of U^* with respect to permutation of the points $\hat{\mathbf{b}}_i$ will be analyzed and exploited later (cf. Section 8.3.1). At the algebraic level, however, the following symmetrization technique will facilitate the analysis of U^* and related series expansions. For each integer $k \geq 0$, define

$$S_k = \sum_{i=1}^{3} \cos^k(\theta - \theta_i)$$

and write

$$f(x) = \prod_{i=1}^{3} (x - \cos(\theta - \theta_i)) = x^3 - \frac{3}{4}x + \frac{1}{4}\cos(3\theta).$$

Logarithmic differentiation of $f(x)$ leads to the formal identity

$$\left(\sum_{k=0}^{\infty} \frac{S_k}{x^{k+1}} \right) f(x) = f'(x) = 3x^2 - \frac{3}{4}$$

from which we deduce the recursive formula

$$S_{k+3} = \frac{3}{4} S_{k+1} - \frac{1}{4}\cos(3\theta) S_k, \quad k \geq 0. \tag{225}$$

For convenience, the first few S_k are listed as follows:

$$S_0 = 3, \quad S_1 = 0, \quad S_2 = \frac{3}{2}, \quad S_3 = -\frac{3}{4}\cos(3\theta), \quad S_4 = \frac{9}{8}$$
$$S_5 = -\frac{15}{16}\cos(3\theta), \quad S_6 = \frac{27}{32} + \frac{3}{16}\cos^2(3\theta), \quad S_7 = -\frac{63}{64}\cos(3\theta).$$

It follows, for example, that S_k as a polynomial in $\cos(3\theta)$ has all its nonzero coefficients positive (respectively negative) when k is even (respectively k is odd). By expanding U^* as a sum of binomial series in the variables z_i and using the identity

$$\sum_{i=1}^{3} z_i^k = (\sin^k \varphi) S_k,$$

we arrive at the following trigonometric series

$$\begin{aligned} U^* = U^*(\varphi, \theta) &= \frac{1}{3}\sum_{k=0}^{\infty} \binom{-\frac{1}{2}}{k} (-1)^k (\sin^k \varphi) S_k \qquad (226) \\ &= 1 + \frac{3}{16}\sin^2 \varphi - \frac{5}{64}(\sin^3 \varphi)(\cos 3\theta) + \ldots + \,. \end{aligned}$$

For later use we also introduce the following functions on the sphere

$$F(\mathbf{p}) = \frac{\sqrt{2}}{3}\sum_{i=1}^{3} \frac{\sin(\theta - \theta_i)}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^3}, \quad G(\mathbf{p}) = \frac{\sqrt{2}}{3}\sum_{i=1}^{3} \frac{\cos(\theta - \theta_i)}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^3}. \qquad (227)$$

It follows that

$$\frac{\partial U^*}{\partial \theta} = -F \sin \varphi, \quad \frac{\partial U^*}{\partial \varphi} = G \cos \varphi. \qquad (228)$$

Remark 52. *It is easy to check that F vanishes precisely along the six meridians*

$$\Gamma_k^* : \theta = k\frac{\pi}{3}, \qquad 0 \le k \le 5 \qquad (229)$$

passing through either a binary collision point $\hat{\mathbf{b}}_i$ or an Euler point $\hat{\mathbf{e}}_i$ $(= -\hat{\mathbf{b}}_i)$. Hence, F changes its sign by crossing these meridians, but on the other hand, the function G is positive (except undefined at $\hat{\mathbf{p}}_0$). For example, F is negative for $0 < \theta < \pi/3$, and in this sector the gradient flow of U^* is depicted in Figure 9.

8.3. **Reduction, regularity and singularity.** Here we shall describe a finite group acting on moduli curves and, in particular, it is a symmetry group of the space of solutions $s \to (\alpha(s), \Gamma^*(s))$ of (221). Moreover, regularity and singularity aspects of the solutions we seek are also briefly discussed.

8.3.1. *Discrete symmetries and reduction.* In addition to time translation and space-time scaling symmetries which transform solutions of the general 3-body problem as in (197)), there is also an additional symmetry group of order 4 which we denote by

$$\mathfrak{D}_1 \times \mathbb{Z}_2 = \langle \bar{\sigma}, \bar{\tau} \rangle \simeq \mathbb{Z}_2 \times \mathbb{Z}_2. \tag{230}$$

The involution $\bar{\tau}$ represents reversal of time, $t \to -t$, and its induced action on oriented curves in the moduli space is expressed by

$$\bar{\tau} : (s, \alpha) \to (\tilde{s}, \tilde{\alpha}) = (-s, \pi - \alpha) \qquad \text{(reversal of direction)} \tag{231}$$

which takes a solution $(\alpha(s), \Gamma^*(s))$ of the system (221) on the interval $s_1 < s < s_2$ to the "reverse" solution in the opposite direction and defined on the interval $-s_2 < \tilde{s} < -s_1$(or any translation of this interval). The other involution $\bar{\sigma}$ is a purely geometric symmetry, arising from the reversal of orientation of m-triangles. At the moduli space level, the latter is the reflection of $\bar{M} \simeq \mathbb{R}^3$ in the (equator) xy-plane, that is, the map $\varphi \to \pi - \varphi$ in the coordinates (ρ, φ, θ).

On the other hand, under the present assumption of equal masses, there is also the (order 6) dihedral isometry group $\mathfrak{D}_3 \subset O(2)$ of S^2 which fixes the poles $\hat{\mathbf{p}}_0^{\pm}$ and permutes the Euler points $\hat{\mathbf{e}}_i$. It is a symmetry group of the 3-body problem since it leaves U^* invariant, cf. Section 8.2. The action is generated by the rotation $\theta \to \theta + 2\pi/3$ and the reflection $\theta \to -\theta$, and altogether we have a symmetry group of order 24,

$$\mathfrak{G} = \mathfrak{D}_6 \times \mathbb{Z}_2 = (\mathfrak{D}_3 \times \mathfrak{D}_1) \times \mathbb{Z}_2, \tag{232}$$

where we also regard $\mathfrak{D}_6 \subset O(3)$ as a dihedral isometry group of S^2 generated by reflections. Thus, the action of $\mathfrak{D}_6$ divides the sphere into 12 congruent spherical triangles called *chambers*, and we choose one of them to be our *fundamental chamber*, namely the following geodesic triangle on the upper hemisphere

$$\mathfrak{C}_0 = \left\{ (\varphi, \theta) \in S^2;\ 0 \le \varphi \le \frac{\pi}{2},\ 0 \le \theta \le \frac{\pi}{3} \right\} \tag{233}$$

(cf. Figure 9) with the vertices

$$\begin{aligned} \hat{\mathbf{p}}_0 = (\varphi = 0), \hat{\mathbf{e}}_3 = (\varphi = \pi/2, \theta = 0), \\ \hat{\mathbf{b}}_2 = (\varphi = \pi/2, \theta = \pi/3). \end{aligned} \tag{234}$$

In particular, the action of $\mathfrak{D}_6$ on solutions (α, Γ^*) of (221) reduces the study of solutions to the study of "solution segments" Γ^* inside $\mathfrak{C}_0$. In particular, we may restrict the study of triple collision solutions Γ^* to those emanating from the vertex $\hat{\mathbf{p}}_0$ or $\hat{\mathbf{e}}_3$ with initial direction leading into the chamber $\mathfrak{C}_0$. The natural first step of this program is to look for solutions whose curvature equation in (221) is trivially satisfied, and Section 8.4 is devoted to this preliminary study.

8.3.2. *Cusps and other singularities.* It is well known that a 3-body motion $\Gamma(t)$ can be "regularized" through a binary collision (cf. [13]), and the only "real" singularity must be a triple collision. Away from collisions the moduli curve $\bar{\Gamma}$ and the shape curve Γ^* are also analytic functions when we parametrize by time t or the arc-length $\bar{s}$ of $\bar{\Gamma}$. However, we also want to parametrize by the arc-length s of Γ^*, and then singularities may occur at specific instants where the time derivative $\dot{s}(t) \geq 0$ vanishes. Although this type of "singularity" is rather artificial, it has geometric significance which explains the possible cusps of the embedded curve Γ^* on the 2-sphere. These are also singularities of the system (221), and in this subsection we shall discuss them in some detail.

Since $\frac{d\bar{s}}{dt} > 0$ for all t, the event $\dot{s}(t_1) = 0$ is equivalent to the condition $\alpha(t_1) = 0$ or π. In this case, $\frac{d}{dt}\Gamma^*(t_1) = 0$ and we say $\mathbf{p} = \Gamma^*(t_1)$ is a *halting point* for the shape curve. Geometrically, this can be a singular point for Γ^*, namely it is the type of singularity that may occur when a regular curve $\bar{\psi}(t)$ in 3-space with vertical tangent at $t = t_1$ is projected to a curve $\psi^*(t)$ in the xy-plane, cf. also (218). On the other hand, when Γ^* is parametrized by s and $s_1 = s(t_1)$, the unit tangent vector $\frac{d}{ds}\Gamma^*(s_1)$ still exists (as a one-sided limit) at the halting point $\mathbf{p}$. Moreover, the (one-sided) geodesic curvature $\mathcal{K}_g^*$ of Γ^* will be bounded near $\mathbf{p}$. In fact, for a "thin" region $0 < \rho_1 \leq \rho \leq \rho_1 + \delta\rho$ of the moduli space $(\bar{M}, d\bar{s}^2)$ with the kinematic metric (193), the projection to the shape space (M^*, ds^2) may be viewed as a Riemannian submersion modulo an almost constant scaling. Restricting to the above region, the curvature of the moduli curve $\bar{\Gamma}$ is certainly bounded, and its image curve in M^* will also have bounded curvature.

Now, consider a pair $(\alpha(s), \Gamma^*(s))$ which is a solution of the system (221). The pair is *regular* on the interval (s_1, s_2) if the three functions $\alpha(s), \varphi(s), \theta(s)$ are analytic and $\alpha(s) \neq 0, \pi$, and a *singularity* is encountered at $s = s_i$ if we cannot extend the functions regularly beyond this point. In that case it follows from (221) that either $\sin\alpha(s_i) = 0$, in which case we call $\mathbf{p} = \Gamma^*(s_i)$ a *cusp*, or $\mathbf{p}$ is a binary collision point $\hat{\mathbf{b}}_j$ (in which case $\alpha(s_i) = \pi/2$ and $\alpha'(s_i) = \infty$). On the interval (s_1, s_2) the growth of $\alpha(s)$ is governed by the inclination angle equation (cf. (221))

$$\frac{d\alpha}{ds} = -\frac{1}{4} + \frac{1}{2}\cot\alpha(s)D(s) \tag{235}$$

whose dependence on the shape curve is solely through the tangential logarithmic derivative

$$D(s) = \nabla(\ln U^*) \cdot \boldsymbol{\tau}^* = \frac{u'(s)}{u(s)}, \quad u(s) = U^*(\Gamma^*(s)). \tag{236}$$

Assume there is a halting point at $s = s_i$, say $\alpha(s_i) = 0$ and write $\mathbf{p} = \Gamma^*(s_i)$. If $D(s_i) \neq 0$, then by (235) $\alpha'(s_i) = \pm\infty$ and $\mathbf{p}$ is a cusp. On the other hand, if $D(s_i) = 0$ and $\mathbf{p}$ is not a critical point of U^*, then the curvature (i.e. second) equation of (221), whose right side is nonzero at $s = s_i$, would force the geodesic curvature of Γ^* to become infinitely large towards $\mathbf{p}$. However, as observed above, such a behavior of the shape curve is not possible. Hence, $D(s_i) = 0$ is only possible when the halting point $\mathbf{p}$ of Γ^* is also a critical point of U^*.

Finally, assume the halting point $\mathbf{p}$ is also a critical point of U^*. For s close to s_i we have $\cot\alpha \sim 1/\alpha$, and the local behavior of $\alpha(s)$ near $s = s_i$ is largely governed by equation (235), from which we can show $\alpha'(s)$ is bounded in a neighborhood of s_i. Choose some s_0 so that $s_1 < s_0 < s_2$ and consider the two cases depending on whether s is approaching s_i from above or below:

$$\text{i) } s_i = s_1: \quad \int_{s_0}^{s_1} \cot\alpha(s)\, ds \approx \int_{s_0}^{s_1} \frac{ds}{\alpha(s)} = -\infty \tag{237}$$

$$\text{ii) } s_i = s_2: \quad \int_{s_0}^{s_2} \cot\alpha(s)\, ds = \infty. \tag{238}$$

In particular, in case i) an associated 3-body motion must necessarily encounter a triple collision at $s = s_1$ since formula (219) implies $\rho(s_1) = 0$, whereas in case ii) $\rho(s_2) = \infty$ and $\mathbf{p} = \boldsymbol{\Gamma}^*(s_2)$ is the limiting shape of the 3-body motion as $t \to \infty$. In contrast to this, for a cusp singularity with $0 < \rho(s_i) < \infty$, we have $\alpha(s) \to 0$ and $\alpha'(s) \to \pm\infty$ as $s \to s_i$, in such a way that the integrals (237) or (238) will converge. For example, this would be the case if $\alpha(s) \sim k(s - s_i)^p$ for some constant k and $p < 1$.

We claim, in fact, that $(\alpha(s), \Gamma^*(s))$ is regularizable at the above halting point $\mathbf{p} = \boldsymbol{\Gamma}^*(s_i)$ which is also a critical point of U^*, that is, as a solution of (221) the functions can be extended analytically beyond s_i. Indeed, once the derivative $\alpha'(s_i)$ exists, the power series developments at $s = s_i$ of the functions $\alpha(s), \varphi(s), \theta(s)$ are recursively determined from the system (221). The recursive scheme is worked out in detail in Section 8.4 for the case (237) with a triple collision at the north pole $\hat{\mathbf{p}}_0 = \boldsymbol{\Gamma}^*(s_1)$, and case ii) is similar.

The collinear type of triple collision is at an Euler point such as $\mathbf{p} = \hat{\mathbf{e}}_3$, in which case $\Gamma^*(s)$ moves along the equator. In particular, $u(s)$ in (236) has a minimum at $s = s_1$. In fact, approaching $\hat{\mathbf{e}}_3$ from other directions (not tangential to the equator) would force $\alpha'(s)$ to become complex valued. It is remarkable that $\alpha'(s_1) = a_0$ (respectively b_0) turns out to be the same constant for all possible triple collision curves emanating from $\hat{\mathbf{p}}_0^{\pm}$ (respectively $\hat{\mathbf{e}}_i$), see (250).

In order to describe analytically the regularization of the triple collision motions it is convenient to extend the domain of the angle α to negative values as well. Indeed, $-\alpha$ should be identified with $\pi - \alpha$, and therefore we introduce the *α-circle*

$$\frac{[0, \pi]}{(0 \sim \pi)} \simeq S^1 \subset \mathbb{C} : \alpha \to z_\alpha = e^{2i\alpha} \tag{239}$$

as the new domain for α. Then the continuous motion $z_\alpha(s)$ on the circle illustrates the qualitative behavior of the solution $(\alpha(s), \mathbf{\Gamma}^*(s))$ and hence also the moduli curve $\bar{\Gamma}(s)$ of an associated 3-body motion. For example, $\rho(s)$ is increasing (respectively decreasing) with s when z_α lies on the upper (respectively lower) semicircle. A halting point is characterized by $z_\alpha = 1$, and it represents either a cusp, a triple collision ($\rho \to 0$), or an escape ($\rho \to \infty$).

The only way $z_\alpha(s)$ enters the other semicircle is at $z_\alpha = 1$, but for $\rho \to 0$ or ∞, since at a cusp z_α "bounces back" on the same semicircle. Similarly, z_α reaches the value -1 at a binary collision point $\hat{\mathbf{b}}_i$, but again z_α "bounces back" (if the moduli curve is continued via regularization with ρ increasing).

Summary 53. *The solution* $(z_\alpha(s), \Gamma^*(s))$ *has a singularity at* $s = s_i$ *if either* i) *the pair takes the value* $(1, \mathbf{p})$ *where* $\mathbf{p}$ *does not belong to the set*

$$\{\pm\hat{\mathbf{p}}_0, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\} \cup \left\{\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \hat{\mathbf{b}}_3\right\}, \tag{240}$$

or ii) *it takes the value* $(-1, \hat{\mathbf{b}}_i), i > 0$. *The singularity is a cusp (respectively a binary collision) in the first (respectively second) case.*

8.3.3. *The initial value problem for triple collision solutions.* For later reference we introduce the set $\mathfrak{S}(\mathbf{p})$ of all shape curves $\Gamma^*(s), s \geq 0$, representing a triple collision motion with the limiting shape $\mathbf{p} = \Gamma^*(0)$ at the collision. Here $\mathbf{p}$ can be any of the five points of the first subset in (240). In fact, each $\Gamma^*(s)$ is associated with a unique (inclination angle) function $\alpha(s)$ so that the pair (α, Γ^*) is a solution of the system ODE^* in (221) with the additional and singular initial condition

$$\text{i)}\ \ \alpha(0) = 0, \qquad \text{ii)}\ \alpha'(0) \geq 0. \tag{241}$$

In particular, for a fixed $\Gamma^*(s)$, $\alpha(s)$ is the unique solution of (221) and (241). Thus we may as well consider the totality of pairs (α, Γ^*) and define

$$\mathfrak{S}(\mathbf{p}) = \{(\alpha(s), \Gamma^*(s));\ s \geq 0, \Gamma^*(0) = \mathbf{p},\ \alpha(0) = 0, \alpha'(0) \geq 0\} \tag{242}$$

as the solutions of a specific initial value problem for the system ODE^*, as indicated.

For the calculation of the sets (242) we may assume (by symmetry) the initial shape $\mathbf{p}$ belongs to the fundamental chamber $\mathfrak{C}_0$, namely $\mathbf{p}$ is either its vertex $\hat{\mathbf{e}}_3$ or $\hat{\mathbf{p}}_0$, see (234). Observe that the set (242) has the induced *symmetry group* $\mathfrak{G}_{\mathbf{p}}$ which is the "isotropy" subgroup at $\mathbf{p}$ of $\mathfrak{G}$ (cf. (232))

$$\mathfrak{G}_{\hat{\mathbf{e}}_3} = \{1, \bar{r}\} \times \{1, \bar{\tau}\} \simeq \mathfrak{D}_1 \times \mathbb{Z}_2, \quad \mathfrak{G}_{\hat{\mathbf{p}}_0} = \mathfrak{D}_3 \times \mathbb{Z}_2, \tag{243}$$

where $\bar{r} \in \mathfrak{D}_3$ is the reflection $\theta \to -\theta$.

Remark 54. *Contrary to the above, the initial value problem for (221) at a cusp or binary collision is not well defined since we would have* $\alpha'(s_i) = \pm\infty$. *Therefore, one cannot continue a solution* $(\alpha(s), \Gamma^*(s))$ *across the singularity using only the system* ODE^*. *To circumvent the problem one should, for example, turn to the moduli curve* $\bar{\Gamma}$ *itself, which is regular in any case. In Section* 8.4.2 *below we shall use Newton's equation of motion more directly to develop in time a specific solution through several cusps.*

8.4. **Isosceles and collinear triple collision motions.** In this section we take the opportunity to illustrate the above approach applied to the "simplest" type of triple collision motions apart from the shape invariant ones. We shall also supply with numerical calculations, for comparison reasons and illustration of examples only.

Namely, consider the possibility that the shape curve Γ^* of a triple collision motion is confined to a great circle on the sphere, that is, $\mathcal{K}_g^*(s)$ $= 0$ for all s. Then the curvature equation in (221) forces the normal derivative of U^* along Γ^* to vanish. Hence, by (228) and Remark 52, Γ^* "moves" either on the equator circle ($\varphi = \pi/2$) or on one of the six meridians (229) passing through an Euler point $\hat{\mathbf{e}}_i$ or a binary collision point $\hat{\mathbf{b}}_i$. These meridians represent the shape of isosceles triangles, so the associated 3-body motions are either of collinear type or isosceles triangle type.

By the symmetry reduction explained in Section 8.3.1 it suffices to consider three separate cases, namely Γ^* is (initially) confined to one of the boundary arcs of the fundamental chamber (233). We list the starting point (at the triple collision) and a choice of arc-length parameter s to be used (up to the first cusp or binary collision point) in each case:

$$\begin{aligned} &(1)\ \hat{\mathbf{p}}_0 : \theta = 0, \quad\ \ s = \varphi \geq 0, \\ &(2)\ \hat{\mathbf{e}}_3 : \varphi = \pi/2,\ s = \theta \geq 0, \\ &(3)\ \hat{\mathbf{p}}_0 : \theta = \pi/3,\ s = \varphi \geq 0. \end{aligned} \tag{244}$$

8.4.1. *The inclination angle α and its ODE.* We shall investigate the first equation of (221)

$$\frac{d\alpha}{ds} = -\frac{1}{4} + \frac{1}{2}\cot(\alpha)D_i(s)\ ,\ i = 1, 2, 3 \tag{245}$$

for each of the three cases (244), where $D_i(s)$ is calculated from the appropriate expression of $U^* = U^*(\varphi, \theta) = U^*(\varphi, \theta + 2\pi/3)$, namely (cf. (224))

$$\begin{aligned} U^*(\varphi, 0) = &\ \frac{1}{3}\Big(\frac{2}{\sqrt{1 - \frac{1}{2}\sin\varphi}} + \frac{1}{\sqrt{1 + \sin\varphi}}\Big), \\ U^*(\pi/2, \theta) = &\ \frac{1}{3}\Big(\frac{1}{\sqrt{1 - \cos(\theta - \frac{\pi}{3})}} + \frac{1}{\sqrt{1 - \cos(\theta + \frac{\pi}{3})}} + \frac{1}{\sqrt{1 + \cos\theta}}\Big) \\ U^*(\varphi, \pi/3) = &\ U^*(-\varphi, 0) = U^*(\varphi, \pi). \end{aligned} \tag{246}$$

As will be demonstrated below, there is a unique solution of the initial value problem (241).

The derivatives D_i are the following analytic functions expanded at the point $\hat{\mathbf{p}}_0$ in case (1) and (3), and $\hat{\mathbf{e}}_3$ in case (2):

$$\begin{aligned} D_1(\varphi) &= \frac{\frac{d}{d\varphi}U^*(\varphi, 0)}{U^*(\varphi, 0)} = \varphi\Big(\frac{3}{8} - \frac{15}{64}\varphi + \frac{23}{256}\varphi^2 - \ldots\Big), \\ D_2(\theta) &= \frac{\frac{d}{d\theta}U^*(\pi/2, \theta)}{U^*(\pi/2, \theta)} = \theta\Big(\frac{29}{20} + \frac{1801}{1200}\theta^2 + \frac{17569}{12000}\theta^4 + \ldots\Big), \\ D_3(\varphi) &= \frac{\frac{d}{d\varphi}U^*(\varphi, \pi/3)}{U^*(\varphi, \pi/3)} = \varphi\Big(\frac{3}{8} + \frac{15}{64}\varphi + \frac{23}{256}\varphi^2 + \ldots\Big). \end{aligned} \tag{247}$$

It follows that $D_1(\varphi) = -D_3(-\varphi)$ and hence case (3) of (244) can be subsumed under case (1) by using the range $\varphi < 0$ and moreover, $\alpha < 0$ interpreted in accordance with (239). In fact, if $\alpha_i(\varphi)$ is the solution of (245) for $i = 1, 3$, then $\alpha_3(\varphi) = -\alpha_1(-\varphi)$. Hence, we need only consider the cases (1) and (2) of (245).

To investigate the nature of the singularity $\alpha = 0$, let us first approximate the functions $D_1(\varphi)$, $D_2(\theta)$ and

$$\cot\alpha = \alpha^{-1} - \frac{1}{3}\alpha - \frac{1}{45}\alpha^3 + \ldots,$$

by their first term $\frac{3}{8}\varphi$, $\frac{29}{20}\theta$ and α^{-1} respectively. Then the initial value problem $\alpha(0) = 0$ for the simplified version of (245) has the two straight line solutions

$$\begin{aligned}
&(1): \alpha = a_0\varphi, \qquad a_0 = \frac{\pm\sqrt{13}-1}{8} \\
&(2): \alpha = b_0\theta, \qquad b_0 = \frac{\pm\frac{1}{5}\sqrt{1185}-1}{8}
\end{aligned} \tag{248}$$

found by solving a second order polynomial, namely

$$a_0 = -\frac{1}{4} + \frac{3}{16}\frac{1}{a_0}, \quad b_0 = -\frac{1}{4} + \frac{29}{40}\frac{1}{b_0}. \tag{249}$$

However, the extended initial value condition (241) demands α to be initially increasing and hence selects the positive solution in (248) as the leading coefficient for the two types of triple collision, namely

$$\begin{aligned}
&\text{Lagrange type: } a_0 = \frac{\sqrt{13}-1}{8} \approx 0.326, \\
&\text{Euler type: } \ b_0 = \frac{\frac{1}{5}\sqrt{1185}-1}{8} \approx 0.736.
\end{aligned} \tag{250}$$

Returning to the original equation (245) we can determine recursively the power series expansion of α,

$$\begin{aligned}
&(1) \quad \alpha = a_0\varphi(1 + a_1\varphi + a_2\varphi^2 + \ldots), \\
&(2) \quad \alpha = b_0\theta(1 + b_1\theta + b_2\theta^2 + \ldots)
\end{aligned} \tag{251}$$

which depends solely on the initial coefficients a_0, b_0 in (250). For convenience we list (with a few decimals only) the first terms of the expansions:

$$\begin{aligned}
&(1) \quad \alpha \approx \varphi(0.3257 - 0.0955\varphi + 0.0129\varphi^2 - 0.0233\varphi^3 + \ldots), \\
&(2) \quad \alpha = \theta(0.7356 + 0.1941\theta^2 + 0.0487\theta^4 + \ldots).
\end{aligned}$$

As indicated, in case (1) the series is alternating and in case (2) $\alpha(s)$ is an odd function since $b_1 = b_3 = b_5 = \ldots = 0$.

The above series can be easily developed and used with high accuracy for small φ (or θ). However, we shall only use it to calculate an initial value of α for some small φ (or θ) and then solve equation (245) by standard numerical procedures, on the maximal interval bounded by the first singularity in each direction. Namely, in case (1) we find

$$\lim_{\varphi\to-\pi/2}\alpha(\varphi)=-\pi/2,\qquad \lim_{\varphi\to-\pi/2}\alpha'(\varphi)=\infty, \tag{252}$$
$$lim_{\varphi\to\varphi_1}\alpha(\varphi)=0,\quad \lim_{\varphi\to\varphi_1}\alpha'(\varphi)=-\infty,\qquad {}_1\approx 1.876\approx 107.5^\circ.$$

On the interval $[0,\varphi_1]$, $\alpha(\varphi)$ increases up to its maximum at $\varphi\approx 1.2$ and is thereafter decreasing, see Figure 10.

On the other hand, if we rotate by 180° the graph of $\alpha(\varphi)$ over the interval $[-\pi/2,0]$, then we obtain the graph of $\alpha(\varphi)$ on $[0,\pi/2]$ for case (3) of (244). Finally, the graph of $\alpha(\theta)$ in case (2) on the interval $[0,\pi/3]$ is quite similar to that of case (3), see Figure 11. Its graph over $[-\pi/3,\pi/3]$ is symmetric with respect to the origin.

As calculated above, the shape curve Γ^* along the meridian $\theta=0$ reaches the first cusp at $\varphi=\varphi_1$, which is beyond $\hat{\mathbf{e}}_3$. At the cusp the motion $\Gamma^*(s)$ changes its direction and continues northward and across $\hat{\mathbf{e}}_3$. See Section 8.4.2 and the time development of this motion.

Here is a brief summary of the analysis of the preliminary cases (244) and the solution sets (242) to which they belong:

- The set $\mathfrak{S}(\hat{\mathbf{e}}_3)$ has only two solutions $(\alpha_+,\Gamma^*_+),(\alpha_-,\Gamma^*_-)$, and they are equivalent modulo the isometric reflection $\bar{r}:\theta\to-\theta$ belonging to $\mathfrak{D}_3\cap\mathfrak{G}_{\hat{\mathbf{e}}_3}$. Let $\Gamma^*(\theta),\theta\in(-\pi/3,\pi/3)$, be the arc-length parametrization of the equator circle from $\hat{\mathbf{b}}_1$ to $\hat{\mathbf{b}}_2$. There is an analytic function $\alpha(\theta)$ so that the pair (α,Γ^*) is a solution of (221) and moreover,

$$\alpha_-(s)=-\alpha(-s)\text{ for }s=-\theta\ge 0;\ \ \alpha_+(s)=\alpha(s)\text{ for }s=\theta\ge 0;$$
$$\Gamma^*_-(s)=\Gamma^*(-s)\text{ for }s=-\theta\ge 0;\ \ \Gamma^*_+(s)=\Gamma^*(s)\text{ for }s=\theta\ge 0.$$

- The set $\mathfrak{S}(\hat{\mathbf{p}}_0)$ has a unique solution (α_0,Γ^*_0) and $(\alpha_{\pi/3},\Gamma^*_{\pi/3})$ in case (1) and (3) of (244), respectively. Consider also the solution

$$\bar{\mu}(\alpha_{\pi/3},\Gamma^*_{\pi/3})=(\alpha_\pi,\Gamma^*_\pi)\in\mathfrak{S}(\hat{\mathbf{p}}_0)$$

along the meridian $\theta=\pi$, obtained by applying the rotation $\bar{\mu}\in\mathfrak{D}_3:\theta\to\theta+2\pi/3$, as indicated, and let $\Gamma^*(\varphi),\varphi\in(-\pi/2,\pi/2)$, be the arc-length parametrization of the half-circle $(\hat{\mathbf{b}}_3\to\hat{\mathbf{p}}_0\to\hat{\mathbf{e}}_3)$. There is an analytic function $\alpha(\varphi)$ so that the pair (α,Γ^*) is a solution of (221) and moreover,

$$\alpha_\pi(s)=-\alpha(-s)\text{ for }s=-\varphi\ge 0;\ \ \alpha_0(s)=\alpha(s)\text{ for }s=\varphi\ge 0;$$
$$\Gamma^*_\pi(s)=\Gamma^*(-s)\text{ for }s=-\varphi\ge 0;\ \Gamma^*_0(s)=\Gamma^*(s)\text{ for }s=\varphi\ge 0.$$

8.4.2. *Time dependence and Newton's equation.* We shall choose case (1) of (244) and compare the above approach using the system (221) with the time parametrized motion $\Gamma(t)$ using Newton's equation (1). Namely, in the xy-plane we consider the motion of three point masses of mass 1/3, symmetric with respect to the y-axis and with position vectors

$$\mathbf{a}_1 = (-x, y), \mathbf{a}_2 = (x, y), \mathbf{a}_3 = (0, -2y).$$

Newton's equation (1) reads

$$\ddot{x} = -\frac{1}{12}\frac{|x|}{x^3} - \frac{1}{3}\frac{x}{(x^2+9y^2)^{3/2}}, \quad \ddot{y} = -\frac{y}{(x^2+9y^2)^{3/2}}. \tag{253}$$

As initial condition at time $t = 0$, assume $y = 0$ (i.e. Γ^* is at the Euler point $\hat{\mathbf{e}}_3$) and moment of inertia $I = \rho^2 = 1$. Moreover, let β denote the oriented angle from the positive x-axis to the initial velocity vector $\dot{\mathbf{a}}_2$, whose length is denoted v. Assuming the total energy $h = T - U$ vanishes, the initial condition now reads

$$(x, y)|_{t=0} = (\sqrt{\frac{3}{2}}, 0), \qquad (\dot{x}, \dot{y})|_{t=0} = v(\cos\beta, \sin\beta) \tag{254}$$
$$v = \sqrt{\frac{\frac{5}{6}\sqrt{\frac{2}{3}}}{1+2\sin^2\beta}}.$$

Finally, let us assume $0 < \beta < \pi$, which means the shape curve Γ^* is heading southwards from $\hat{\mathbf{e}}_3$ for small $t > 0$.

The above data specify a 1-parameter family (parametrized by β) of isosceles 3-body motions $\Gamma(t) = (x(t), y(t))$ with total energy $h = 0$, with normalized size $I = 1$ and collinear shape $\hat{\mathbf{e}}_3$ at time $t = 0$. We shall use the equation (253) to investigate the time dependence of the various geometric and kinematic quantities of the motion, such as I, T, φ, α, where

$$I = \frac{2}{3}x^2 + 2y^2, \quad \cos\varphi = \pm\frac{4}{3\sqrt{3}}\frac{\Delta}{I} = \frac{-4}{\sqrt{3}}\frac{xy}{\frac{2}{3}x^2+2y^2} \tag{255}$$
$$T = U = \frac{1}{9}(\frac{2}{|x|} + \frac{2}{\sqrt{x^2+9y^2}}) = \frac{1}{2}(\dot{\rho}^2 + \frac{\rho^2}{4}\dot{\varphi}^2).$$

By definition of the inclination angle α,

$$\cos\alpha = \frac{\frac{\partial}{\partial\rho}\cdot\frac{d}{dt}\bar{\Gamma}(t)}{\left|\frac{d}{dt}\bar{\Gamma}(t)\right|} = \frac{\dot{\rho}}{\sqrt{2T}} = \frac{\frac{2}{3}x\dot{x}+2y\dot{y}}{\sqrt{2IU}}, \tag{256}$$

and denoting the angle at $t = 0$ by α_0 we deduce

$$\cos\alpha_0 = \frac{3}{\sqrt{5}}(\frac{2}{3})^{1/4} v\cos\beta = \frac{\cos\beta}{\sqrt{1+2\sin^2\beta}}. \tag{257}$$

Thus, the correspondence $\beta \longleftrightarrow \alpha_0$ is a bijection of the interval $(0,\pi)$ such that $\pi - \beta$ corresponds to $\pi - \alpha_0$.

We claim there are exactly two values of β leading to a triple collision motion, namely the pair β_0 and $\pi - \beta_0$ for some $\beta_0 < \pi/2$. This choice of $\beta = \beta_0$ yields $\dot\rho(0) > 0$, and hence the triple collision occurred in the past, namely at some negative time $t_0 < 0$, with the shape curve Γ^* at the north pole $\hat{\mathbf{p}}_0$. Similarly, using $\beta = \pi - \beta_0$ the triple collision is reached at time $-t_0 > 0$ with Γ^* at the south pole.

The angle β_0 is calculated using the formula (257), where $\alpha_0 = \alpha(\pi/2)$ and $\alpha(\varphi)$ is the solution of the equation

$$\frac{d\alpha}{d\varphi} = -1/4 + \frac{\varphi}{2}\cot(\alpha)(\frac{3}{8} - \frac{15}{64}\varphi + \frac{23}{256}\varphi^2 - \ldots) \tag{258}$$

with initial condition $\alpha(0) = 0$, cf. case (1) of (245) and (247). The solution found by the approach in Section 8.4.1, is approximately

$$\alpha_0 = \alpha(\pi/2) \approx 0.18673\ldots, \qquad \beta_0 = 0.10865\ldots.$$

Thus, we know the initial data (254) corresponding to triple collision motions. As a test, by running the system (253) backwards in time one will find that the triple collision occurs approximately at $t_0 = -1.0228\ldots$

It is also interesting to follow the shape curve Γ^* of $\Gamma(t)$ for $t > 0$, for example, using the ratio $y(t)/x(t)$ or calculating $\varphi(t)$ directly from (255). The solution $\Gamma(t)$ can be continued in time t through the cusps since they are not singularities for Newton's equation, and only truncation errors or numerical instability may invalidate the calculation in the long run. At $t_1 \approx 10.4$, y/x is maximal and $\varphi(t_1) = \varphi_1$ (cf. (252)) is the colatitude of the first cusp - here $\alpha = 0$ and $\Gamma^*(t)$ turns northward. After passing $\hat{\mathbf{e}}_3$ there is a second cusp where Γ^* turns southward again and crosses $\hat{\mathbf{e}}_3$, but the next cusp is closer to $\hat{\mathbf{e}}_3$, and so on. Concerning the possible asymptotic behavior of $\Gamma^*(t)$, at triple collision or as $t \to \infty$, see also Section 8.6.3.

Finally, we consider the time dependent size function $\rho(t) = \sqrt{I(t)}$ and compare it with the integral formula (219). We have, by assumption, $\rho = 1$ at time $t = 0$, and Newton's equation (253) yields, for example,

$$\rho(-0.5) \approx 0.634726, \quad \varphi(-0.5) = \hat\varphi \approx 1.381793.$$

On the other hand, numerical integration of the solution $\alpha(\varphi)$ of (258) yields

$$\rho|_{\varphi=\hat{\varphi}} = e^{\frac{1}{2}\int_1^{\hat{\varphi}} \cot\alpha d\varphi} \approx 0.634725. \tag{259}$$

Alternatively, let us also evaluate this integral using the time parametrized function $\alpha(t)$ calculated by (256) and developed via Newton's equation. Thus we change the variable φ in (259) to t using (255), namely

$$d\varphi = \varphi'(t)dt = (\frac{d}{dt}\arccos(-\frac{4x(t)y(t)}{\sqrt{3}(\frac{2}{3}x(t)^2 + 2y(t)^2)}))dt,$$

and then numerical integration similar to the case (259) yields

$$\rho(-0.5) = e^{\frac{1}{2}\int_0^{-0.5}\cot(\alpha)\varphi'(t)dt} \approx 0.634726.$$

8.5. **Analytic uniqueness of triple collision motions .** In this section we turn to the full family $\mathfrak{S}(\hat{\mathbf{p}}_0)$ of triple collision solutions (α, Γ^*) of the system ODE^*, with Γ^* starting out from $\hat{\mathbf{p}}_0$. Due to the symmetry group $\mathfrak{D}_3$ we may assume the shape curve Γ^* enters the fundamental chamber $\mathfrak{C}_0$, which limits the initial direction θ_0 to the range $[0, \pi/3]$. Therefore, in terms of spherical coordinates (φ, θ) the appropriate and complete initial condition (242) now reads

$$\alpha(0) = 0,\ \alpha'(0) \geq 0;\ \varphi(0) = 0,\ \theta(0) = \theta_0,\ 0 \leq \theta_0 \leq \pi/3.$$

The border cases $\theta_0 = 0, \pi/3$ are the cases (1) and (3) of (244) already investigated in Section 8.4, and now it is natural to generalize the procedure used there to the whole range of angles θ_0. This time, however, the strength of the curvature equation of (221) must be fully utilized. At this point, we assume (tentatively) that the functions α, φ, θ have power series expansions at $\hat{\mathbf{p}}_0$, necessarily of type

$$\begin{aligned} \varphi &= s(c_0 + c_1 s + c_2 s^2 + c_3 s^3 + \ldots +) \\ \theta &= \theta_0 + s(d_0 + d_1 s + d_2 s^2 + \ldots +), \quad 0 \leq \theta_0 \leq \frac{\pi}{3}, \\ \alpha &= a_0 s(1 + a_1 s + a_2 s^2 + \ldots +). \end{aligned} \tag{260}$$

Recall from Section 8.1, we have excluded the trivial case of constant shape, and hence α does not vanish identically.To justify the notation in the third line, it will be demonstrated below that the leading term is $a_0 s$ with $a_0 \neq 0$.

By considering the leading coefficients of the series for φ and θ the third equation of ODE^* implies

$$c_0 = 1, \quad c_1 = 0, \quad c_2 = -\frac{1}{6}d_0^2 \leq 0 \tag{261}$$

and clearly

$$(\sin\varphi)\theta' = \varepsilon(1-\varphi'^2)^{1/2} = \varepsilon(-6c_2s^2+\ldots+)^{1/2}, \quad \varepsilon = \pm 1. \tag{262}$$

The calculation of a_0 in the expansion $\alpha = a_0 s + \ldots$ is really the same as in Section 8.4.1 and gives the same value (250) independent of θ_0. To see this, we write for clarity the first equation of ODE^* as

$$\alpha(2\alpha' + \frac{1}{2}) = (\alpha\cot\alpha)D(s), \tag{263}$$

where

$$\alpha\cot\alpha = 1 - \frac{1}{3}\alpha^2 - \frac{1}{45}\alpha^4 - \frac{2}{945}\alpha^6 - \frac{1}{4725}\alpha^8 + O(\alpha^{10}),$$

and the potential function (226) and its logarithmic derivative along Γ^* have the expansions

$$u(s) = U^*(\Gamma^*(s)) = \sum_{i=0}^{\infty} u_i s^i = 1 + \frac{3}{16}s^2 - \frac{5}{64}(\cos 3\theta_0)s^3 + \ldots, \tag{264}$$

$$\begin{aligned} D(s) &= \frac{d}{ds}\ln(u) = \frac{u'(s)}{u(s)} = \frac{1}{u}\left(\frac{\partial U^*}{\partial\varphi}\varphi' + \frac{\partial U^*}{\partial\theta}\theta'\right) \\ &= s\sum_{i=0}^{\infty}\mu_i s^i = s\left(\frac{3}{8} - \left(\frac{15}{64}\cos 3\theta_0\right)s + \ldots\right). \end{aligned} \tag{265}$$

In particular, the first order term in (265) is independent of θ_0 and the leading terms of (263) yield the single condition

$$4a_0^2 + a_0 - 3/4 = 0 \text{ , with positive root : } a_0 = \frac{\sqrt{13}-1}{8}. \tag{266}$$

The identity (263) also provides recursive relations for the calculation of $a_k, k > 0$, expressed in terms of the coefficients $\mu_i, i \le k$, see (277) below.

8.5.1. *The method of undetermined coefficients.* The proof of Theorem G_1, concerning the existence and uniqueness of the curves, is based upon formal power series substitution for the three functions (260) involved in ODE^*. This leads to a recursive procedure - the method of *undetermined coefficients* - which is consistent and determines successively the higher order coefficients in terms of c_2, d_0 and θ_0. By (261) c_2 is already determined by d_0, and at the final stage we shall find that d_0 is actually determined by θ_0. Consequently, the expansions in (260) are, indeed, determined by the initial angle θ_0 alone and hence θ_0 parametrizes the whole solution set $\mathfrak{S}(\hat{\mathbf{p}}_0)$.

On the 2-sphere there is the positive, orthonormal frame $\{\frac{\partial}{\partial\varphi}, \frac{1}{\sin\varphi}\frac{\partial}{\partial\theta}\}$ associated with the coordinates φ, θ. Along the oriented shape curve Γ^*

we also have the positive, orthonormal moving frame $\{\boldsymbol{\tau}^*, \boldsymbol{\nu}^*\}$, where $\boldsymbol{\tau}^*$ is the tangent vector. The latter frame differs from the stationary frame by a rotation angle β, namely in analogy with (206) - (207),

$$\boldsymbol{\tau}^* = \cos\beta\frac{\partial}{\partial\varphi} + \frac{\sin\beta}{\sin\varphi}\frac{\partial}{\partial\theta}, \quad \boldsymbol{\nu}^* = -\sin\beta\frac{\partial}{\partial\varphi} + \frac{\cos\beta}{\sin\varphi}\frac{\partial}{\partial\theta} \tag{267}$$

$$\cos\beta = \varphi', \quad \sin\beta = (\sin\varphi)\theta' = \varepsilon(1-\varphi'^2)^{1/2}, \text{ cf. (262).} \tag{268}$$

Now, we turn to the second equation of (221), namely the curvature equation written as an identity

$$LHS = RHS \tag{269}$$

between the left hand and right hand side

$$\begin{aligned} LHS &= (1-\cos 2\alpha)\mathcal{K}_g^* U^* \\ RHS &= U^*(\nabla U^* \cdot \boldsymbol{\nu}^*) = -\sin\beta\frac{\partial U^*}{\partial\varphi} + \frac{\cos\beta}{\sin\varphi}\frac{\partial U^*}{\partial\theta}. \end{aligned} \tag{270}$$

In LHS the geodesic curvature term decomposes as

$$\mathcal{K}_g^* = \frac{d\beta}{ds} + \cos\varphi\frac{d\theta}{ds} = -\varepsilon(1-\varphi'^2)^{-1/2}\varphi'' + (\cos\varphi)\theta', \tag{271}$$

where β' is calculated using (267). We have actually $\varepsilon = 1$, see the Remark below.

By substituting the first expression for $\sin\beta$ in (267) into RHS and comparing the leading terms of the expansions of LHS and RHS, it follows that either $c_2 \neq 0$ or all $c_i = 0$ for $i > 1$, and that $c_2 = 0$ implies $\sin 3\theta_0 = 0$. Thus, $c_2 = 0$ means $\theta_0 = 0$ or $\pi/3$, that is, the two meridian solutions of isosceles triangle type already discussed in Section 8.4.

Henceforth, we shall assume $c_2 \neq 0$. Comparison of the leading terms (of order 2) in (269) yields

$$4a_0^2 d_0 = -\frac{3}{8}d_0 + \frac{15}{64}\sin 3\theta_0$$

which combined with (261) gives

$$d_0 = \frac{15}{16}\frac{\sin 3\theta_0}{(16a_0^2+\frac{3}{2})}, \qquad c_2 = -\frac{75}{512}\frac{\sin^2 3\theta_0}{(16a_0^2+\frac{3}{2})^2} \tag{272}$$

with the approximate values

$$d_0 \approx 0.293\sin 3\theta_0, \quad c_2 \approx -0.014\sin^2 3\theta_0.$$

Remark 55. *In particular, in* (262) *we have* $\varepsilon = sgn(d_0) = 1$, *and the expressions in* (272) *are, in fact, valid in the whole closed chamber* $\mathfrak{C}_0$: $0 \leq \theta_0 \leq \pi/3$. *However,* $\varepsilon = \pm 1$ *actually changes sign across the border meridian of two neighboring chambers.*

In view of (267), (271) we also need the expansions

$$(1-\varphi'^2)^{1/2} = d_0 s(1 + b_1 s + b_2 s^2 + \ldots),$$
$$(1-\varphi'^2)^{-1/2} = \frac{1}{d_0 s}(1 + \bar{b}_1 s + \bar{b}_2 s^2 + \ldots),$$

where by writing $\tilde{c}_k = c_k/c_2$ for $k \geq 3$ we have by simple inspection

$$b_n = \frac{n+3}{6}\tilde{c}_{n+2} + B_n(\tilde{c}_3, \ldots, \tilde{c}_{n+1}), \;\; n \geq 1 \tag{273}$$
$$\bar{b}_n = -\frac{n+3}{6}\tilde{c}_{n+2} + \bar{B}_n(\tilde{c}_3, \ldots, \tilde{c}_{n+1}), \;\; n \geq 1$$

where B_n and $\bar{B}_n$ are polynomials and $B_1 = \bar{B}_1 = 0$.

For simplicity, let $P(y_1, y_2, \ldots)$ denote any (unspecified) polynomial in the variables y_i, except that $y_1 = \theta_0$ means it is polynomial in $\sin 3\theta_0$ and $\cos 3\theta_0$. Using the notation

$$\sin\varphi = s(1 + g_2 s^2 + g_3 s^3 + \ldots), \quad g_k = c_k + P(c_2, \ldots, c_{k-1}) \tag{274}$$
$$\cos\varphi = 1 - \frac{1}{2}s^2 + h_4 s^4 + h_5 s^5 + \ldots, \quad h_k = P(c_2, \ldots, c_{k-2})$$

we derive from (262) the following formula for d_n

$$(n+1)d_n = d_0 b_n - (d_0 g_n + 2d_1 g_{n-1} + \ldots + (n-1)d_{n-2}g_2), \;\; n \geq 1, \tag{275}$$

and from (271) we calculate the curvature expansion

$$\mathcal{K}_g^* = 2d_0(1 + k_1 s + k_2 s^2 + \ldots),$$

where

$$2d_0 k_n = \left(d_0\bar{b}_n + (n+1)d_n - (n+3)(n+2)\frac{c_{n+2}}{d_0}\right) \tag{276}$$
$$- \frac{1}{d_0}\sum_{k=3}^{n+1}(k+1)kc_k\bar{b}_{n+2-k} + \sum_{k=0}^{n-1}(k+1)d_k h_{n-k}.$$

Next, let us have a closer look at the coefficients of $u(s)$ and $u'(s)/u(s)$, using (226), (264), (265), and also at the recursive generation of the coefficients of $\alpha(s)$ using (263). It follows that they are of type

$$u_n = P(\theta_0, c_2, \ldots, c_{n-2}, d_0, \ldots, d_{n-4}), \quad n \geq 4$$
$$\mu_n = P(\theta_0, c_2, \ldots, c_n, d_0, \ldots, d_{n-2}), \quad n \geq 2 \tag{277}$$
$$a_n = P(a_1, a_2, \ldots, a_{n-1}, \mu_1, \ldots, \mu_n), \quad n \geq 1.$$

For example,

$$a_1 = \frac{2\mu_1}{12a_0^2 + a_0} = \frac{8\mu_1}{10 - \sqrt{13}} = -\frac{15/8}{10 - \sqrt{13}}\cos 3\theta_0.$$

For convenience, write

$$1 - \cos 2\alpha = 2a_0^2 s^2(1 + A_1 s + A_2 s^2 + \ldots), \tag{278}$$

where

$$A_1 = 2a_1, \quad A_2 = 2a_2 + a_1^2 - \frac{1}{3}a_0^2, \quad \ldots, \; A_n = P(a_1, a_2, \ldots, a_n),$$

and thus we arrive at the following presentation of LHS as a product of series

$$LHS = 4a_0^2 d_0 s^2(1 + A_1 s + \ldots)(1 + k_1 s + \ldots)(1 + u_1 s + \ldots). \tag{279}$$

From the structure of U^* as a trigonometric series (226) in the variables $\sin\varphi$ and $\cos 3\theta$, we can write

$$\frac{\partial U^*}{\partial \varphi} = (\sin\varphi\cos\varphi)R_1, \quad \frac{\partial U^*}{\partial \theta} = (\sin^3\varphi \sin 3\theta)R_2, \tag{280}$$

where

$$R_1 = \frac{3}{8} - \frac{15}{64}\sin\varphi\cos 3\theta + \ldots, \qquad R_2 = \frac{15}{64} + \frac{945}{4096}\sin^2\varphi + \ldots$$

are again polynomial series in the variables $\sin\varphi, \cos 3\theta$. Hence, by substituting the series of $\sin\varphi, \cos\varphi, \theta', \varphi', \cos 3\theta, \sin 3\theta$ into the expression

$$RHS = (\sin^2\varphi)\,(-\theta'\cos\varphi R_1 + \varphi'\sin 3\theta R_2) \tag{281}$$

equation (269) renders a recursive procedure for the calculation of c_n and d_{n-2}, $n \geq 2$, starting from c_2 and d_0 (272).

Lemma 56. *The coefficients c_m and d_{m-2} can be expressed as*

$$c_m = P(\theta_0)c_2, \qquad d_{m-2} = P(\theta_0)d_0; \quad m \geq 2 \tag{282}$$

where $P(\theta_0)$ denotes some polynomial of $\sin 3\theta_0$ *and* $\cos 3\theta_0$ (*generally different for each coefficient*).

Proof. The first step is to compare terms of order 3 in (269), which renders the identity

$$4a_0^2 d_0(A_1 + k_1 + u_1) = -\frac{3}{4}d_1 + \frac{15}{64}d_0\cos 3\theta_0,$$

where

$$A_1 = 2a_1, \quad k_1 = \frac{c_3}{c_2}, \quad u_1 = 0, \quad d_1 = \frac{d_0 c_3}{3c_2}.$$

Consequently,

$$c_3 = \frac{15(16a_0+1)}{4(16a_0^2+1)(12a_0+1)}(\cos 3\theta_0)c_2, \tag{283}$$
$$d_1 = \frac{15(16a_0+1)}{12(16a_0^2+1)(12a_0+1)}(\cos 3\theta_0)d_0.$$

We proceed by induction and assume that (282) holds for m in the range $2 \le m \le n$. By (273) - (278), we infer that b_{m-2}, $\bar{b}_{m-2}$, k_{m-2}, g_m, h_m, u_m, μ_m and a_m are all of type $P(\theta_0)$ for $m \le n$. Furthermore,

$$nd_{n-1} = \left(\frac{n+2}{6c_2}c_{n+1} + P(\theta_0)\right) d_0, \tag{284}$$
$$k_{n-1} = \frac{(n+1)(n+2)}{12c_2}c_{n+1} + P(\theta_0).$$

Consider the terms of order $n+1$ in equation (269). By equating the coefficients of s^{n+1} in LHS and RHS we deduce

$$4a_0^2 d_0 k_{n-1} + P(\theta_0)d_0 = -\frac{3}{8}nd_{n-1} + P(\theta_0)d_0. \tag{285}$$

Here, k_{n-1} and d_{n-1} are the only coefficients depending on c_{n+1}, and by substituting their expressions from (284) into the identity (285), we deduce the identity

$$\left(\frac{4a_0^2(n+1)(n+2)}{12} + \frac{n+2}{16}\right) c_{n+1} = P(\theta_0)c_2,$$

and consequently,

$$c_{n+1} = P(\theta_0)c_2, \quad d_{n-1} = P(\theta_0)d_0.$$

This completes the induction step, and hence (282) holds for all $m \ge 2$. ∎

This settles the existence and uniqueness question for the series expansions (260), for each initial longitude angle θ_0. Their radius of convergence is certainly positive (e.g. by an inductive argument showing the coefficients are bounded), but we shall not try to estimate the radius here. Clearly, for $\theta_0 = k\pi/3$ the radius is at most $\pi/2$.

8.5.2. *Symmetries of the solution set* $\mathfrak{S}(\hat{\mathbf{p}}_0)$. In the previous subsection it was established that the triple collision solution set (242) for $\mathbf{p} = \hat{\mathbf{p}}_0$ is naturally parametrized by angles θ_0, namely

$$\mathfrak{S}(\hat{\mathbf{p}}_0) = \left\{(\alpha_{\theta_0}, \Gamma^*_{\theta_0}); 0 \le \theta_0 < 2\pi\right\} \tag{286}$$

is in 1-1 correspondence with points on a circle and therefore inherits "symmetries" of a circle. However, the actual symmetry group should act with orbits representing the various "species' or "congruence" classes of solutions, and moreover, knowledge of each class suffices to generate all solutions by a straightforward transformation procedure. We contend that $\mathfrak{G}_{\hat{\mathbf{p}}_0}$defined in (243) is, in fact, the appropriate group.

First of all, $\mathfrak{G}_{\hat{\mathbf{p}}_0}$ contains the group $\mathfrak{D}_3$ which acts on the 2-sphere and represents the purely geometric symmetries. On the other hand, we have also seen that each solution curve $\Gamma^* = \Gamma^*_{\theta_0}$ corresponds to three analytic functions $(\alpha(s), \varphi(s), \theta(s))$ in a neighborhood of $s = 0$, and for $s < 0$ these functions also describe a motion approaching a triple collision as $s \to 0^-$. Hence, by inverting its direction we should obtain a triple collision motion emanating with initial longitude angle $\theta_0 + \pi$, which by uniqueness must be the solution in (286) labeled by $\theta_0 + \pi$. Consequently, in agreement with the summary of Section 8.4.1, for each "antipodal" pair $(\alpha_{\theta_0}, \Gamma^*_{\theta_0}), (\alpha_{\theta_0+\pi}, \Gamma^*_{\theta_0+\pi})$ in (286) there is an analytic curve $\Gamma^*(s)$ passing through $\hat{\mathbf{p}}_0$ and an analytic function $\alpha(s)$ so that $(\alpha(s), \Gamma^*(s))$ is a solution of the system (221), and moreover,

$$\alpha_{\theta_0+\pi}(s) = -\alpha(-s) \text{ and } \alpha_{\theta_0}(s) = \alpha(s) \text{ for } s \geq 0,$$
$$\Gamma^*_{\theta_0+\pi}(s) = \Gamma^*(-s) \text{ and } \Gamma^*_{\theta_0}(s) = \Gamma^*(s) \text{ for } s \geq 0.$$

In particular, the inversion operator $\bar{\tau}$ applied to $(\alpha(s), \Gamma^*(s))$ induces an involution

$$\bar{\tau} : (\alpha_{\theta_0}, \Gamma^*_{\theta_0}) \to (\alpha_{\theta_0+\pi}, \Gamma^*_{\theta_0+\pi})$$

of the set $\mathfrak{S}(\hat{\mathbf{p}}_0)$ which commutes with the action of $\mathfrak{D}_3$, and together they generate the dihedral *symmetry group*

$$\mathfrak{G}_{\hat{\mathbf{p}}_0} = \mathfrak{D}_3 \times \{1, \bar{\tau}\} \ \simeq \mathfrak{D}_6 \tag{287}$$

which may be viewed as an "isotropy" subgroup of $\mathfrak{G}$ in (232). In effect, this divides the fundamental chamber $\mathfrak{C}_0$ (233) in two sectors of angular width $\pi/6$, say

$$\tilde{\mathfrak{C}}_0 : 0 \leq \theta_0 \leq \pi/6$$

is our *reduced* fundamental chamber. However, we remark that $\mathfrak{G}_{\hat{\mathbf{p}}_0}$ does not act on S^2, so $\bar{\mathfrak{C}}_0$ is not a fundamental domain in the geometric sense. But solutions starting out in this region suffice to generate the whole solution set (286) using analytic continuation.

The power series developments (260) of the three functions $\alpha(s)$, $\varphi(s)$, $\theta(s)$, where $\Gamma^*_{\theta_0}(s) = (\varphi(s), \theta(s))$ is the shape curve with initial direction θ_0 at the north pole, also exhibit a specific symmetry pattern which reflects the $\mathfrak{D}_6$-symmetry of their coefficients. For example, consider the "reflection" in $\mathfrak{G}_{\hat{\mathbf{p}}_0}$

$$\theta_0 \to \pi/3 - \theta_0$$

which divides $\mathfrak{C}_0$ into two reduced chambers, and write

$$X = \cos 3\theta_0, \quad Y = \sin 3\theta_0$$
$$a_k = \mathrm{A}_k(X,Y) \quad c_k = \mathrm{C}_k(X,Y)c_2, \quad d_k = \mathrm{D}_k(X,Y)d_0,$$

where A_k,C_k,D_k are polynomials of two variables (not unique, of course, since X, Y are algebraic dependent). Let a_k, c_k, d_k and $\bar{a}_k, \bar{c}_k, \bar{d}_k$ be the coefficients of the solutions $(\alpha, \varphi, \theta), (\bar{\alpha}, \bar{\varphi}, \bar{\theta})$ corresponding to initial angles θ_0 and $\pi/3 - \theta_0$, respectively. Then we have

$$\begin{cases} a_k = \bar{a}_k,\ c_k = \bar{c}_k,\ \ d_k = \bar{d}_k & \text{for } k \text{ even} \\ a_k = -\bar{a}_k, \quad c_k = -\bar{c}_k, \quad d_k = -\bar{d}_k & \text{for } k \text{ odd} \end{cases}$$

Equivalently, as functions of X the polynomials A_k,C_k,D_k are odd (respectively even) functions for k odd (respectively even). This is due to the fact that Y is invariant whereas X changes sign under the substitution $\theta_0 \to \pi/3 - \theta_0$.

8.5.3. *Symbolic manipulations and numerical calculation of power series.* The recursive scheme used in Section 8.5.1 will generate all higher order coefficients as polynomials of $Y = \sin 3\theta_0$ and $X = \cos 3\theta_0$, but the explicit calculations involve a substantial amount of symbolic manipulations. For example, various types of algebraic operations, together with composition, are applied to power series.

In principle, calculations involving elementary functions of power series, such as $\sin(\sum p_i x^i)$, can be reduced to symbolic manipulations on power series of the type

$$(p_0 + p_1 x + p_2 x^2 + \ldots)^n = \sum_{k=0}^{\infty} P_k^{(n)} x^k,$$

where the n-th *multinomial polynomial* $P_k^{(n)}$ records the n-partitions and associated multinomial coefficients which can be calculated recursively with some effort.

On the other hand, available computer software developed for symbolic computation have built-in procedures which effectively generate the intermediate power series expansions as well as recursive formulas. We have employed such symbolic software for the calculation[1] of a_k, c_k, d_k in (260), for small k, see also Section 8.7. For convenience, we list the first of them below (omitting the already known c_2, d_0, a_0), and we remark

[1]The symbolic and numerical calculations were performed in 1995 by Chee-Whye Chin, an undergraduate student at U.C. Berkeley, using *Mathematica* software.

that the exact (or symbolic) expressions are growing fast in complexity as k increases:

$$c_3 = c_2 \frac{15(10+\sqrt{13})}{116}(\cos 3\theta_0), \tag{288}$$

$$c_4 = c_2 \left(\frac{2040762505 + 136353812\sqrt{13}}{2891425280}, \right.$$

$$\left. + \frac{20984375(113+20\sqrt{13})}{2891425280}(\cos 6\theta_0) \right),$$

$$d_1 = d_0 \frac{5(10+\sqrt{13})}{116}(\cos 3\theta_0),$$

$$d_2 = d_0 \left(\frac{2(19733316 + 84347\sqrt{13})}{216856896} + \frac{302175(113+20\sqrt{13})}{216856896}(\cos 6\theta_0) \right),$$

$$a_1 = \frac{-5(10+\sqrt{13})}{232}(\cos 3\theta_0),$$

$$a_2 = \frac{3(28004 + 4175\sqrt{13})}{2745024} + \frac{-25(644+47\sqrt{13})}{2745024}(\cos 6\theta_0),$$

$$+ \frac{4350(9+\sqrt{13})}{2745024}\sqrt{113+20\sqrt{13}}(\sin^2 3\theta_0).$$

8.6. **Global behavior of the shape of triple collision motions.** First we shall investigate the curvature properties of the flow consisting of the gradient lines of the potential function U^* on the sphere $S^2(1)$. This information will be related to the curvature properties of the "flow" consisting of those curves Γ^* belonging to the set (286), that is, the triple collision shape curves emanating from the north pole $\hat{\mathbf{p}}_0$.

8.6.1. *Differential geometry of the gradient flow of U^*.* We start with the following elementary result about curves on the unit sphere S^2 in Euclidean 3-space.

Lemma 57. *Let $t \to \mathbf{p}(t)$ be a parametrized curve on S^2. Then its geodesic curvature is given by the following triple product*

$$K_g(t) = (\frac{dt}{ds})^3 \mathbf{p} \times \dot{\mathbf{p}} \cdot \ddot{\mathbf{p}} = \mathbf{p} \times \mathbf{p}' \cdot \mathbf{p}'', \tag{289}$$

where (as usual) s is arc-length, $\dot{\mathbf{p}} = \frac{d}{dt}\mathbf{p}$ and $\mathbf{p}' = \frac{d}{ds}\mathbf{p}$.

Proof. The unit tangent vector $\boldsymbol{\tau}^* = \mathbf{p}'$ points in the positive direction of the curve, and $\mathbf{p}''$ is the curvature vector in 3-space. With $\boldsymbol{\nu}^* = \mathbf{p} \times \boldsymbol{\tau}^*$ as the normal vector field along the curve, $\{\boldsymbol{\tau}^*, \boldsymbol{\nu}^*\}$ is a positively oriented frame of the sphere. By definition, the geodesic curvature vector in the sphere is the orthogonal projection of $\mathbf{p}''$ into the tangent plane, namely

$$\mathbf{p}'' - (\mathbf{p}'' \cdot \mathbf{p})\mathbf{p} = K_g \boldsymbol{\nu}^*,$$

where the coefficient K_g is the (scalar) geodesic curvature. Clearly, K_g equals the triple product in (289). ∎

Next, we turn to the gradient field ∇U^* on the sphere, whose integral curves will be referred to as the *gradient lines* of U^*. Their geodesic curvature will be denoted by K_g. Until further notice there is no restriction on the mass distribution, and we use the expressions (171), (175) and (176) for U^*, the vector function $\mathbf{B}(\mathbf{p})$ and the gradient ∇U^* field, respectively.

Lemma 58. *The geodesic curvature of the gradient line of U^* passing through $\mathbf{p}$ is given by the following triple product*

$$K_g(\mathbf{p}) = \frac{3}{|\nabla U^*(\mathbf{p})|^3} \mathbf{B} \times \mathbf{C} \cdot \mathbf{p}, \tag{290}$$

where $\mathbf{B} = \mathbf{B}(\mathbf{p})$ and

$$\mathbf{C} = \mathbf{C}(\mathbf{p}) = \sum_{i=1}^{3} \frac{k_i}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^5} \left[\mathbf{p} \times \mathbf{B} \cdot (\mathbf{p} \times \hat{\mathbf{b}}_i)\right] \hat{\mathbf{b}}_i. \tag{291}$$

Proof. Consider a gradient line parametrized by arc-length, $s \to \mathbf{p}(s)$. We can write

$$\nabla U^*(\mathbf{p}) = f(s)\mathbf{p}'(s),$$

where $f(s) = |\nabla U^*(\mathbf{p}(s))|$, and hence by (176)

$$\mathbf{p}' = \frac{\mathbf{B} - (\mathbf{B} \cdot \mathbf{p})\mathbf{p}}{|\nabla U^*|}.$$

On the other hand, $0 = \mathbf{p} \times \mathbf{p}' \cdot \nabla U^* = \mathbf{p} \times \mathbf{p}' \cdot \mathbf{B}$ and by differentiation

$$\mathbf{p} \times \mathbf{p}'' \cdot \mathbf{B} = -\mathbf{p} \times \mathbf{p}' \cdot \frac{d}{ds}\mathbf{B}. \tag{292}$$

Hence, by substituting the expression

$$\mathbf{B} = |\nabla U^*|\, \mathbf{p}' + (\mathbf{B} \cdot \mathbf{p})\mathbf{p}$$

into the left side of (292), we obtain

$$|\nabla U^*|\, \mathbf{p} \times \mathbf{p}' \cdot \mathbf{p}'' = \mathbf{p} \times \mathbf{p}' \cdot \frac{d}{ds}\mathbf{B} = \frac{\mathbf{p} \times \mathbf{B}}{|\nabla U^*|} \cdot \frac{d}{ds}\mathbf{B}. \tag{293}$$

Finally, we calculate from (175)

$$\frac{d}{ds}\mathbf{B} = 3\sum_{i=1}^{3} k_i \frac{(\hat{\mathbf{b}}_i \cdot \mathbf{p}')}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^5}\hat{\mathbf{b}}_i = \frac{3}{|\nabla U^*|}\sum k_i \frac{(\hat{\mathbf{b}}_i \times \mathbf{p}) \cdot (\mathbf{B} \times \mathbf{p})}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^5}\hat{\mathbf{b}}_i$$

and by substituting this into (293) it follows from Lemma 57

$$K_g = \mathbf{p} \times \mathbf{p}' \cdot \mathbf{p}'' = \frac{3}{|\nabla U^*|^3}\sum k_i \frac{(\mathbf{p} \times \mathbf{B}) \cdot (\mathbf{p} \times \hat{\mathbf{b}}_i)}{\left|\mathbf{p} - \hat{\mathbf{b}}_i\right|^5}(\mathbf{p} \times \mathbf{B} \cdot \hat{\mathbf{b}}_i).$$

This expression can be rewritten as (290). ∎

Problem 59. *Regarding K_g as a function on S^2, determine the curves defined by the condition $K_g(\mathbf{p}) = 0$.*

Remark 60. *Note that $\mathbf{B}(\mathbf{p})$ and $\mathbf{C}(\mathbf{p})$ are vectors in the xy-plane in the Euclidean model $\bar{M} = \mathbb{R}^3$, and the function K_g on the sphere S^2 : $x^2 + y^2 + z^2 = 1$ is undefined precisely at the critical points of U^*, namely for $z \geq 0$ these are the points $\hat{\mathbf{b}}_i, 0 \leq i \leq 3$, and the minimumspoint (physical center) $\hat{\mathbf{p}}_0^+$. From the triple product formula (290) it follows that $K_g(\mathbf{p})$ vanishes on the eclipse circle ($z = 0$), whereas for $z > 0$ $K_g(\mathbf{p})$ vanishes if and only if $\mathbf{B}(\mathbf{p})$ and $\mathbf{C}(\mathbf{p})$ are linearly dependent.*

Henceforth, we shall retain our assumption of uniform mass distribution, and a deeper understanding of the function K_g will be achieved. In fact, in this case the above problem has a simple solution, as explained at the end of this subsection.

By assumption,

$$m_i = \frac{1}{3}, \quad k_i = \frac{\sqrt{2}}{3}, \quad \hat{\mathbf{b}}_i \cdot \hat{\mathbf{b}}_j = -\frac{1}{2} \qquad (i, j \geq 1 \text{ and } i \neq j)$$

and we introduce the three distance functions

$$\delta_i = \delta_i(\mathbf{p}) = \left|\mathbf{p} - \hat{\mathbf{b}}_i\right| = \sqrt{2}\sqrt{1 - \mathbf{p} \cdot \hat{\mathbf{b}}_i}, \ i \geq 1, \quad \text{cf. Section 6.1} \tag{294}$$

which are algebraically related by

$$\delta_1^2 + \delta_2^2 + \delta_3^2 = 6, \tag{295}$$

due to the identity $\sum \mathbf{p} \cdot \hat{\mathbf{b}}_i = \mathbf{p} \cdot \sum \hat{\mathbf{b}}_i = 0$ and

$$\mathbf{p} \cdot \hat{\mathbf{b}}_i = 1 - \frac{1}{2}\delta_i^2. \tag{296}$$

Let us write

$$\mathbf{B}(\mathbf{p}) = \frac{\sqrt{2}}{3}\sum_{i=1}^{3} e_i\hat{\mathbf{b}}_i = \frac{\sqrt{2}}{3}\Big((e_1 - e_3)\hat{\mathbf{b}}_1 + (e_2 - e_3)\hat{\mathbf{b}}_2\Big), \tag{297}$$

$$\mathbf{C}(\mathbf{p}) = \frac{1}{9}\sum_{i=1}^{3} f_i\hat{\mathbf{b}}_i = \frac{1}{9}\Big((f_1 - f_3)\hat{\mathbf{b}}_1 + (f_2 - f_3)\hat{\mathbf{b}}_2\Big), \tag{298}$$

where

$$e_i = \frac{1}{\delta_i^3}, \quad f_i = \frac{3\sqrt{2}}{\delta_i^5}(\mathbf{p}\times\mathbf{B}(\mathbf{p}))\cdot(\mathbf{p}\times\hat{\mathbf{b}}_i). \tag{299}$$

To simplify our notation we denote products (monomials) of the functions δ_i by

$$\delta_{a,b,c} = \delta_1^a\delta_2^b\delta_3^c, \tag{300}$$

where a, b, c are nonnegative integers, and the *alternating* polynomial generated by the monomial (300) is

$$A_{a,b,c} = \begin{vmatrix} \delta_1^a & \delta_2^a & \delta_3^a \\ \delta_1^b & \delta_2^b & \delta_3^b \\ \delta_1^c & \delta_2^c & \delta_3^c \end{vmatrix} = \sum_{\sigma\in S_3} sgn(\sigma)\delta_{a,b,c}^{\sigma}, \tag{301}$$

where S_3 is the permutation group of $\{\delta_1, \delta_2, \delta_3\}$ acting on monomials in the obvious way. In particular, the *basic alternating* polynomial is

$$A = A_{0,1,2} = \begin{vmatrix} 1 & 1 & 1 \\ \delta_1 & \delta_2 & \delta_3 \\ \delta_1^2 & \delta_2^2 & \delta_3^2 \end{vmatrix} = (\delta_1 - \delta_2)(\delta_2 - \delta_3)(\delta_3 - \delta_1). \tag{302}$$

On the other hand, the *symmetric* function generated by $\delta_{a,b,c}$, where we may assume $a \geq b \geq c$, is the smallest S_3-invariant sum

$$S_{a,b,c} = \sum \delta_i^a\delta_j^b\delta_k^c = \sum_{\sigma\in S_3}\delta_{a,b,c}^{\sigma} \tag{303}$$

containing $\delta_{a,b,c}$. In particular, $S_{2,0,0} = 6$, by (295).

Note that $S_{a,b,c}$ is unchanged, whereas $A_{a,b,c}$ may change sign, when a, b, c are permuted. The alternating polynomials can be decomposed as a product of the basic alternating function (302) and a symmetric function, for example

$$\begin{aligned} A_{0,2,4} &= (S_{2,1,0} + 2S_{1,1,1})A, \\ A_{0,1,5} &= (S_{3,0,0} + S_{2,1,0} + S_{1,1,1})A, \\ A_{0,1,6} &= (S_{4,0,0} + S_{3,1,0} + S_{2,2,0} + S_{2,1,1})A, \\ A_{0,3,6} &= (S_{4,2,0} + S_{4,1,1} + 2S_{3,2,1} + 3S_{2,2,2})A. \end{aligned} \tag{304}$$

The induced action of S_3 on the triples $\{e_1, e_2, e_3\}$ and $\{f_1, f_2, f_3\}$ is covariant with the action on $\{\delta_1, \delta_2, \delta_3\}$. Certainly, the above coefficients e_i in (299) are simple rational functions of the $\delta_j' s$, and now we prove a similar statement for the $f_i' s$, as follows.

Lemma 61. *As a rational function of* $\delta_1, \delta_2, \delta_3$

$$f_1 = \frac{1}{\delta_{6,6,6}}(-3\delta_{1,3,6} - 3\delta_{1,6,3} + 2\delta_{0,6,6} + \delta_{3,6,3} + \delta_{3,3,6} + \delta_{1,5,6}$$
$$+\delta_{1,6,5} - \frac{1}{2}\delta_{2,6,6} - \frac{1}{2}\delta_{3,5,6} - \frac{1}{2}\delta_{3,6,5}),$$

and f_2 *(respectively* f_3*) is obtained from* f_1 *(respectively* f_2*) by cyclic permutation,* $a \to b \to c \to a$*, of the indices of each monomial* $\delta_{a,b,c}$*.*

Proof. Use the identities (296) and substitute the expression (297) for $\mathbf{B}$ into the formula (299) for f_1, namely

$$f_1 = \frac{3\sqrt{2}}{\delta_1^5}\left(\mathbf{B} \cdot \hat{\mathbf{b}}_1 - (\mathbf{p} \cdot \hat{\mathbf{b}}_1)(\mathbf{B} \cdot \mathbf{p})\right).$$

Then one obtains the above rational expression for f_1 by straightforward calculations, and by symmetry it is also clear that f_2 and f_3 are obtained from f_1 as claimed. ∎

Now, turning to the formula (290) for the curvature function K_g and inserting the expressions (297), (298), we write

$$\mathbf{B}(\mathbf{p}) \times \mathbf{C}(\mathbf{p}) \cdot \mathbf{k} = \frac{\sqrt{6}}{54}\tilde{A},$$

where $\mathbf{k}$ is the unit normal vector of the xy-plane and $\tilde{A}$ is, by definition, the function

$$\tilde{A}(\delta_1, \delta_2, \delta_3) = (e_1 f_2 - e_2 f_1) + (e_2 f_3 - e_3 f_2) + (e_3 f_1 - e_1 f_3). \tag{305}$$

When the products $e_i f_j$ are calculated using the above lemma, for example

$$e_1 f_2 = \frac{1}{\delta_{6,6,6}}(-3\delta_{3,1,3} - 3\delta_{0,1,6} + 2\delta_{3,0,6} + \delta_{3,3,3} + \delta_{3,1,5} + \delta_{0,3,6}$$
$$+ \delta_{2,1,6} - \frac{1}{2}\delta_{3,2,6} - \frac{1}{2}\delta_{3,3,5} - \frac{1}{2}\delta_{2,3,6}),$$

the expression (305) may be written as

$$\tilde{A} = -\frac{1}{\delta_{6,6,6}}(3A_{0,1,6} + A_{0,3,6} + A_{1,3,5} + A_{1,2,6}). \tag{306}$$

Finally, substitution of expressions from (304) into (306) yields

$$\tilde{A}(\delta_1,\delta_2,\delta_3) = -\tilde{S}(\delta_1,\delta_2,\delta_3)A,$$

where

$$\begin{aligned}\tilde{S}(\delta_1,\delta_2,\delta_3) = \frac{1}{(S_{1,1,1})^6}&[3(S_{4,0,0}+S_{3,1,0}+S_{2,2,0}+S_{2,1,1})+S_{4,2,0}+S_{4,1,1}\\ &+2S_{3,2,1}+3(S_{1,1,1})^2+2S_{1,1,1}S_{2,1,0}+3(S_{1,1,1})^2+S_{1,1,1}S_{3,0,0}].\end{aligned}$$

In summary, we have established the following proposition, where the factor $\tilde{S}$ of K_g is always positive!

Proposition 62. *The geodesic curvature function of the gradient lines of U^* on the unit sphere S^2 is given by the product*

$$K_g(\mathbf{p}) = -\frac{\sqrt{6}\ z\ \tilde{S}}{18\,|\nabla U^*(\mathbf{p})|^3}A \qquad \mathbf{p}=(x,y,z), \quad z\geq 0 \tag{307}$$

Observe that the $\mathfrak{D}_3$-chambers of the (upper) hemisphere of S^2 are defined by inequalities $\delta_i \leq \delta_j \leq \delta_k$, and therefore A, and hence K_g as well, has constant sign in each chamber. For example, the fundamental chamber (233) is given by

$$\mathfrak{C}_0 : \delta_2 \leq \delta_1 \leq \delta_3$$

and here $K_g \geq 0$. The meridians which are the walls of the $\mathfrak{D}_3$-chambers are defined by relations of type $\delta_i = \delta_j$; these are the zero set of the function A. Together with the equator circle they are the curves where K_g vanishes (or is undefined), and this also solves Problem 59 (in the case of uniform mass distribution).

It is easy to visualize the gradient flow on the 2-sphere. For example, in the interior of the spherical triangle $\mathfrak{C}_0$ the flow has the vertex $\hat{\mathbf{p}}_0$ as source and converges towards the vertex $\hat{\mathbf{b}}_2$, with positive curvature everywhere. The flow in $\mathfrak{C}_0$ is illustrated in Figure 9.

8.6.2. *Geometry of the triple collision shape curves.* It is possible to draw qualitative information about the family $\mathfrak{S}(\hat{\mathbf{p}}_0)$ of shape curves by relating it with the geometry of the gradient flow of U^*. By symmetry it suffices to consider those curves $\Gamma^*_{\theta_0}$ starting out in the chamber $\mathfrak{C}_0$, that is, $0 \leq \theta_0 \leq \frac{\pi}{3}$, and we observe that the boundary meridians $\theta_0 = 0$ and $\theta_0 = \pi/3$, emanating from the north pole $\hat{\mathbf{p}}_0$ towards the equator, are themselves both shape curves and gradient lines. So, the question is what one can say about the shape curves in the interior of $\mathfrak{C}_0$?

A rough description goes as follows. In $\mathfrak{C}_0$ there are three different "flows" of curves emanating from $\hat{\mathbf{p}}_0$, namely the shape curves, the gradient lines and the meridians (θ constant). At each point $\mathbf{p} = \Gamma^*(s_1), s_1 > 0$, the shape curve $\Gamma^*(s), s > s_1$, is "trapped" between the gradient line and the meridian through $\mathbf{p}$. Being positively curved one may imagine the shape curves arising by gradually bending the meridians towards the gradient lines by means of a "force" field directed eastward, see Figure 12.

To be more precise, we shall focus on five properties as stated below. For this purpose we introduce two angular functions $\beta(s), \gamma(s)$ as follows. Namely, β is the angle between the meridian and Γ^*, as defined in (267). It is the oriented angle from $\frac{\partial}{\partial\varphi}$ to the velocity vector $\boldsymbol{\tau}^* = \frac{d}{ds}\Gamma^*$, and γ is the angle from $\frac{\partial}{\partial\varphi}$ to the gradient vector

$$\begin{aligned}\nabla U^* &= \frac{\partial U^*}{\partial\varphi}\frac{\partial}{\partial\varphi} + \frac{1}{\sin^2\varphi}\frac{\partial U^*}{\partial\theta}\frac{\partial}{\partial\theta} \\ &= |\nabla U^*|\,(\cos\gamma\frac{\partial}{\partial\varphi} + \sin\gamma\frac{1}{\sin\varphi}\frac{\partial}{\partial\theta}).\end{aligned}$$

We also recall the role of the inclination angle $\alpha(s)$, which is not directly related to the geometry of the spherical curve Γ^* itself, but to the associated moduli curve $\bar{\Gamma}$. However, the curvature equation from (221) can now be stated as

$$2U^*\mathcal{K}_g^*\sin^2\alpha = |\nabla(U^*)|\cos(\frac{\pi}{2} + \beta - \gamma) \tag{308}$$

and hence relates all three angles with the curvature of Γ^*.

Now, we contend that the following properties are valid for $\Gamma^* = \Gamma^*_{\theta_0}, 0 < \theta_0 < \frac{\pi}{3}$, at least until Γ^* leaves $\mathfrak{C}_0$ the first time (but not necessarily later):

- (i) $0 < \alpha < \frac{\pi}{2}$, for $s > 0$. In particular, the curve $s \to \Gamma^*(s)$ has no cusp singularity for $s < \pi/2$.
- (ii) The spherical coordinates $\varphi(s), \theta(s)$ of $\Gamma^*(s)$ are strictly increasing functions of s.
- (iii) $0 < \beta \leq \gamma < \frac{\pi}{2}$, for $s > 0$.
- (iv) The geodesic curvature $\mathcal{K}_g^*$ of $\Gamma^*(s), s \geq 0$, is nonnegative.
- (v) Γ^* leaves the chamber $\mathfrak{C}_0$ by crossing its boundary arc $(\hat{\mathbf{e}}_3, \hat{\mathbf{b}}_2)$ on the equator circle.

In order to verify these statements one may proceed as follows. First, note that $0 < \gamma < \pi/2$ follows from the fact, due to (228) and Remark 52, that $\frac{\partial U^*}{\partial\varphi} > 0$ and $\frac{\partial U^*}{\partial\theta} > 0$ inside $\mathfrak{C}_0$, and moreover, the gradient lines emanate from $\hat{\mathbf{p}}_0$ with $\gamma = 0$ and approach the binary collision point $\hat{\mathbf{b}}_2$ with $\gamma = \pi/2$ in the limit. This also explains why property (v) follows from (iii), and using (267 and (308) we also readily deduce properties (ii) and (iv) from (iii). Thus, we are left with the statements (i) and (iii), and let us first establish property (iii) (using property (i) if necessary).

Observe that $\beta \geq 0$ by (267) and Remark 55. But $\beta = 0$ for some s would imply $\theta' = 0, \varphi' = 1$, and hence by (271) $\mathcal{K}_g^*(s)$ would vanish. However, with $\mathbf{p}$ in the interior of $\mathfrak{C}_0$ we also have $\frac{\partial U^*}{\partial \theta}(\mathbf{p}) \neq 0$ and then $\gamma > 0$ in the right side of the identity (308). Consequently, $\mathcal{K}_g^*(s) \neq 0$ and this contradiction shows $\beta > 0$ must hold. Again by (308), $\mathcal{K}_g^*$ is positive as long as $\beta < \gamma$, and this certainly holds for small s since

$$\mathcal{K}_g^*(0) = d_0 = \frac{15}{16}\frac{\sin 3\theta_0}{(16a_0^2 + \frac{3}{2})}.$$

We claim that $\beta \leq \gamma$ holds (at least) until $\Gamma^*(s)$ leaves the chamber. To see this, suppose we had $\beta = \gamma$ for $s = s_1$, $\beta < \gamma$ (respectively $\beta > \gamma$) for $s < s_1$ (respectively $s > s_1$)and s close to s_1. Then Γ^* would be tangent to the gradient line at $\mathbf{p} = \Gamma^*(s_1)$ and is (locally) lying on the "upper" side of it, hence $\mathcal{K}_g^*(s_1) \geq K_g(\mathbf{p}) > 0$. This contradicts the fact that $\mathcal{K}_g^*(s_1) = 0$, by (308).

Remark 63. *Property* (iii) *implies that U^* increases along the curve $\Gamma^*(s), s \geq 0$. In general, the event $\beta = \gamma$ means Γ^* is perpendicular to the level curve of U^*, and by* (308) *this can happen for two reasons, namely* i) *$\mathcal{K}_g^*$ vanishes or* ii) *Γ^* reaches a cusp. We can rule out the second case due to property* (i)*, but only up to the first crossing of the equator.*

Finally, we turn to property (i). From the relations

$$\frac{d\rho}{ds}\frac{ds}{dt} = \frac{d\rho}{dt} > 0, \quad \frac{d\rho}{ds} = \rho(s)\cot\alpha, \quad \text{cf. (219)},$$

we deduce $0 \leq \alpha < \pi/2$. In fact, $\alpha = \pi/2$ (and $ds/dt = \infty$) only at the vertex $\hat{\mathbf{b}}_2$, and our claim is that $\alpha = 0$ only holds at $\hat{\mathbf{p}}_0 = \Gamma^*(0)$.

It is certainly evident from the numerical analysis of the shape curves (cf. Table 1) that $\alpha > 0$ for $s > 0$, at least until Γ^* leaves $\mathfrak{C}_0$. Indeed, using numerical data and a continuity argument we can establish the uniform lower bound $\alpha \geq 0.18$ for $\pi/4 \leq \varphi \leq \pi/2$. However, we shall also explain an alternative and more qualitative approach to settle the problem.

To show $\alpha > 0$ holds, let assume the contrary and recall the geometric arguments in the setting in Section 7.3, where we regarded $\Gamma^*(\sigma)$,

$\sigma = s/2$, as a curve on the sphere $S^2(1/2)$ and $C(\Gamma^*) \subset \bar{M}$ denotes the cone surface with the induced Euclidean metric (204). The associated moduli curve $\bar{s} \to \bar{\Gamma}(\bar{s})$ lies in this surface, and assuming the first cusp occurs at $\sigma = \sigma_1$ we consider the Euclidean sector $0 \le \sigma \le \sigma_1$ bounded by the rays $\sigma = 0$ and $\sigma = \sigma_1$. By our assumptions, there is a bijective correspondence $[0, \bar{s}_1] \longleftrightarrow [0, \sigma_1]$ between the arc-length parametrizations of the moduli curve $\bar{\Gamma}$ and Γ^*, and $\alpha > 0$ for $0 < \bar{s} < \bar{s}_1$.

The curve $\bar{\Gamma}(\bar{s})$ starts out from the origin and its radial distance ρ is increasing. It has the positively oriented moving frame $\{\boldsymbol{\tau}, \boldsymbol{\eta}\}$ of (206), where $\boldsymbol{\tau}$ (respectively $\boldsymbol{\eta}$) is the unit tangent (respectively normal) vector. By (211) and a well known formula for the curvature of curves in the Euclidean plane, the curvature of $\bar{\Gamma}$ in the above sector can be expressed as

$$\frac{d\zeta}{d\bar{s}} = \frac{1}{2}\frac{d}{d\boldsymbol{\eta}}(\ln U), \qquad \zeta = \sigma + \alpha, \text{ cf. Figure 8.} \tag{309}$$

Towards the point $\bar{\Gamma}(\bar{s}_1)$ the curve $\bar{\Gamma}$ becomes tangential to the boundary ray $\sigma = \sigma_1$, that is, the angle α decreases to zero. Therefore $\frac{d\sigma}{d\bar{s}}$ vanishes, by (207), and hence

$$\frac{d\zeta}{d\bar{s}} \le 0 \ \text{ as } \bar{s} \to \bar{s}_1.$$

Moreover, the frame $\{\boldsymbol{\tau}, \boldsymbol{\eta}\}$ approaches $\left\{\frac{\partial}{\partial\rho}, \frac{1}{\rho_1}\frac{\partial}{\partial\sigma}\right\}$ as $\bar{s} \to \bar{s}_1$, so we also deduce

$$\frac{d}{d\boldsymbol{\eta}}(\ln U) \to \frac{u'(\sigma)}{u(\sigma)}|_{\sigma=\sigma_1} \le 0$$

and hence $u(\sigma) = U^*(\Gamma^*(\sigma))$ is decreasing at σ_1, or possibly $u'(\sigma_1) = 0$ and hence Γ^* is tangential to the level curve of U^* on the sphere. However, U^* is actually increasing towards the point $\Gamma^*(\sigma_1)$ by Remark 63 (which applies here since there is no cusp for $\sigma < \sigma_1$). This contradiction rules out any occurrence of cusps inside $\mathfrak{C}_0$.

8.6.3. *Final escape limiting behavior of the shape curves.* As an interesting example, recall the time parametrized meridian solution $\Gamma_0^*(t)$ from Section 8.4.2. The first cusp appears at $t_1 \approx 10.4$, with colatitude $\varphi_1 \approx 107.5°$. The next cusps occur roughly at times $t_2 \approx 435$, $t_3 \approx 162400$, $t_4 \approx 5{\cdot}10^7$, $t_5 \approx 1.24{\cdot}10^9$ (with due regard to numerical instability) and they appear to be approaching $\hat{\mathbf{e}}_3$ as a final limit of the shape curve. The behavior of the curve resembles a damped oscillation converging to its "stability" point $\hat{\mathbf{e}}_3$ as $t \to \infty$. Let us refer to this limiting behavior as *irregular*, namely the limit shape $\mathbf{p}$ is reached through converging cusps and $\alpha'(s)$ has no limit, as in the above example. Such a behavior at the final escape at infinity is, however, not necessarily related to the fact that the shape curve is a triple collision curve in the other direction.

Thus, one may consider more generally the limiting behavior of moduli curves $\bar{\Gamma}(t)$ of three-body motions as $t \to \infty$, assuming the limit shape $\mathbf{p}$ exists. In the irregular case, however, we do not claim that $\mathbf{p}$ is necessarily a central configuration (that is, a critical point of U^*), although this is rather likely. On the other hand, if $\mathbf{p}$ is a central configuration, we claim that it is an Euler points $\hat{\mathbf{e}}_i$, and moreover, Γ^* is not confined to the equator circle.

We define the final limiting behavior to be *regular* if the final shape $\mathbf{p} = \Gamma^*(s_2)$ is a central configuration, where $\alpha(s) > 0$ for $s_2 - \epsilon < s < s_2$ and $\alpha(s_2) = 0$. Moreover, Γ^* is confined to the equator circle (and hence the 3-body motion is collinear) if $\mathbf{p} = \hat{\mathbf{e}}_i$. As in the case of triple collisions, a useful tool in the study of such limiting behavior is again the system ODE* (221), whose solutions are pairs $(\alpha(s), \Gamma^*(s))$. Then the fact that $\rho \to \infty$ as $s \to s_2$ is expressed by the divergence of the integral (238), and moreover, the system (221) itself imposes the condition that $\mathbf{p}$ must be a critical point of U^*.

A closer study of the above regular solutions $(\alpha(s), \Gamma^*(s))$ near the final limit may proceed in the same way as we studied triple collision shape curves in Section 8.5. For convenience, let us translate the arc-length parameter, $s \to s - s_2$, and consider the power series expansions of $\alpha(s)$ and $\Gamma^*(s)$ at $s = 0$. The calculations are similar to the triple collision case worked out in Section 8.4.1 and 8.5.1, but this time $\alpha'(s)$ converges to the negative root of the polynomial in (249), namely

$$a_0 = -\frac{\sqrt{13}+1}{8} \approx -0.575\,69, \quad b_0 = -\frac{\frac{1}{5}\sqrt{1185}+1}{8} \approx -0.\,985\,6. \quad (310)$$

Thus, in the collinear (Euler) case with the limit shape $\Gamma^*(0) = \hat{\mathbf{e}}_3$, with $\Gamma^*(s)$ an arc-length parametrization of the equator circle near $\hat{\mathbf{e}}_3$, the differential equation (245) in the case $i = 2$ has a unique solution $\alpha(s)$ with $\alpha(0) = 0$, $\alpha'(0) = b_0$.

Next, for the final limit shape of Lagrange type we consider the following (singular) initial conditions

$$\Gamma^*(0) = \hat{\mathbf{p}}_0;\ \alpha(0) = 0, \alpha'(0) = a_0 \text{ (hence } \alpha(s) > 0 \text{ for } s < 0)$$

which define a family of analytic solutions $(\alpha(s), \Gamma^*(s))$ of the system (221). The calculations are similar to the triple collision case, with a_0

equal to the positive number in (250), worked out in Section 8.5.1. Therefore, we leave it to the reader to modify these calculations and perhaps establish the same kind of analytic uniqueness, namely that solutions are parametrized by the terminal angular (longitude) direction θ_0 of $\Gamma^*(s)$ at $\hat{\mathbf{p}}_0$, cf. (260). In particular, by reversing the direction of the curve segment $\{\Gamma^*(s), s \geq 0\}$ one obtains the shape curve with terminal direction $\theta_0 + \pi$.

8.6.4. *More about the asymptotic behavior at triple collision.* Finally, we turn to the asymptotic behavior, in terms of the time parameter t, of a triple collision taking place at $t = 0$. Let $\bar{\Gamma}(t) = (\rho(t), \Gamma^*(t))$ be the moduli curve of such a motion, with $\Gamma^*(0) = \hat{\mathbf{p}}_0$ or $\hat{\mathbf{e}}_i$, and write $\mu = U^*(\Gamma^*(0))$. Then it is a classical result, dating back to the work of Sundman and Siegel, that (for any energy level h and mass distribution)

$$\rho(t) \sim \kappa t^{2/3} \text{ as } t \to 0, \qquad \kappa = (\frac{9}{2}\mu)^{1/3} \tag{311}$$

and moreover, the total kinetic energy is asymptotically dominated by the "change of size", in the sense that

$$T(t) \sim T^\rho = \frac{1}{2}\dot{\rho}(t)^2 \sim \frac{2}{9}\kappa^2 t^{-2/3} \sim \frac{\mu}{\rho(t)}. \tag{312}$$

In particular, the residual kinetic energy

$$T^\sigma = \frac{1}{8}\rho(t)^2(\frac{ds}{dt})^2 = T - T^\rho \tag{313}$$

due to the "change of shape" must be of lower order in t than that of T^ρ, in the sense that $T^\sigma/T^\rho \to 0$. However, this does not exclude the possibility that $T^\sigma \to \infty$ as $t \to 0$. We claim, however, that $T^\sigma \to 0$, and we shall present the following (rather heuristic) argument for this.

Using the relationship (219) between $\rho(s)$ and $\alpha(s)$ and the series expansion of $\alpha(s)$ we obtain an expansion of type

$$\rho(s) = \rho_0 s^p(1 + r_1 s + r_2 s^2 + \ldots), \;\; p = 1/2\alpha'(0).$$

Hence, for suitable nonzero constants κ_0 and κ_s

$$s(t) \sim \kappa_0\rho(t)^{1/p} = \kappa_0\rho(t)^{2\alpha'(0)} \sim \kappa_s t^{4\alpha'(0)/3} = \kappa_s t^e,$$

where the exponent e depends on the two types of triple collision, namely by (250)

$$(i)\ \Gamma^*(0) = \hat{\mathbf{p}}_0 : e = \frac{4}{3}a_0 = \frac{1}{6}(\sqrt{13} - 1) \approx 0.434\,26 \tag{314}$$

$$(ii)\ \Gamma^*(0) = \hat{\mathbf{e}}_i : e = \frac{4}{3}b_0 = \frac{1}{6}(\frac{1}{5}\sqrt{1185} - 1) \approx 0.98079.$$

Now, as is the case of $\rho(t)$, let us assume differentiation of $s(t)$ also commutes with taking asymptotic limit, namely $\dot{s}(t) \sim e\kappa_s t^{e-1}$. Then by (313)

$$T^\sigma(t) \sim \frac{1}{8}(\kappa\kappa_s e)^2 t^{4/3+2(e-1)} = \kappa_\sigma t^\epsilon, \begin{cases} \text{Case (i)} : \epsilon \approx 0.201\,85 \\ \text{Case (ii)} : \epsilon \approx 1.294\,9 \end{cases} \tag{315}$$

8.7. **Numerical solutions of triple collision motions.** We shall describe a modified approach to provide numerical C^1-data for the 1-parameter family $\mathfrak{S}(\hat{\mathbf{p}}_0)$ of shape curves representing non-collinear triple collision motions, under the standard assumption of equal masses $m_i = 1/3$ and zero total energy. Recall that these curves $\Gamma^*(s)$ arise from solutions (α, Γ^*) of the system (221) of ordinary differential equations, where $\Gamma^* = \Gamma^*_{\theta_0}$ starts from the north pole $\hat{\mathbf{p}}_0$ on the 2-sphere with initial (longitude) direction θ_0, and the (radial inclination) angle $\alpha \geq 0$ has the initial value $\alpha = 0$. The basic idea is to obtain numerical data close to the initial point, by the analytical method, and use them as the initial data for the remaining integration by means of a Runge-Kutta method. With some more efforts we believe it is possible to settle Conjecture 50 by carefully combining numerical analysis and theory along these lines. The following numerical analysis serves at least to illustrate the geometry of those shape curves.

8.7.1. *Outline of a numerical approach.* As usual, (θ, φ) are the spherical coordinates of the 2-sphere $M^* \simeq S^2$, with $\varphi = 0$ at the initial point $\hat{\mathbf{p}}_0$. By symmetry (as explained earlier) we need only consider curves whose initial direction θ_0 lies in the interval $0 < \theta_0 < \pi/3$, and moreover, we may as well use φ to parametrize each curve since φ increases with the arc-length s. Thus, let

$$\Gamma^*_{\theta_0} = \left\{ (\varphi, \theta_{\theta_0}(\varphi)), \varphi \geq 0; \lim_{\varphi \to 0} \theta_{\theta_0}(\varphi) = \theta_0 \right\}$$

denote the shape curve with initial angle θ_0, and let $\alpha_{\theta_0}(\varphi)$ denote the corresponding angle α as a function of φ. We shall compute the values of θ_{θ_0}, $\frac{d}{d\varphi}\theta_{\theta_0}$ and α_{θ_0} for $0 \leq \varphi \leq \pi/2$ and $0 \leq \theta_0 \leq \pi/3$.

Elimination of the arc-length parameter s in the system (221) is achieved by using the third equation to rewrite the first two equations with φ as the independent variable. In standard (explicit) form the new system reads:

$$\begin{aligned}\alpha' &= -\frac{1}{4}\sqrt{1+(\sin^2\varphi)\theta'^2} + \frac{\cot\alpha}{2U^*}\left(\frac{\partial U^*}{\partial\varphi} + \frac{\partial U^*}{\partial\theta}\theta'\right)\\ \theta'' &= \frac{1}{2U^*}\left[\frac{\partial U^*}{\partial\theta}\csc^2\alpha\csc^2\varphi - \left(4U^*\cot\varphi + \frac{\partial U^*}{\partial\varphi}\csc^2\alpha\right)\theta'\right.\\ &\quad \left. + \left(\frac{\partial U^*}{\partial\theta}\csc^2\alpha\right)\theta'^2 - \left(U^*\sin 2\varphi + \frac{\partial U^*}{\partial\varphi}\csc^2\alpha\sin^2\varphi\right)\theta'^3\right],\end{aligned} \tag{316}$$

where $\alpha', \theta', \theta''$ means differentiation with respect to φ.

We consider the power series expansions at $\varphi = 0$:

$$\begin{aligned}\theta_{\theta_0}(\varphi) &= \theta_0 + \varphi(f_0 + f_1\varphi + f_2\varphi^2 + \ldots)\\ \alpha_{\theta_0}(\varphi) &= \varphi(g_0 + g_1\varphi + g_2\varphi^2 + \ldots)\end{aligned} \tag{317}$$

whose coefficients can be calculated recursively as functions of θ_0 by the method of undetermined coefficients. This is similar to the calculation of the expansions in (260) using the system ODE^*. Let

$$\theta_{\theta_0}^{[n]}(\varphi),\ \ \alpha_{\theta_0}^{[n]}(\varphi),\ \frac{d}{d\varphi}\theta_{\theta_0}^{[n]}(\varphi) \tag{318}$$

be the polynomials in φ of degree $n+1$ (respectively n for the third polynomial), where the first two are obtained by substituting the calculated expressions for f_i, g_i (as functions of θ_0), $0 \le i \le n$, into (317) and truncating higher order terms. The last polynomial is the derivative of $\theta_{\theta_0}^{[n]}(\varphi)$. Then for sufficiently small φ the "true" functions $\theta_{\theta_0}(\varphi)$, $\alpha_{\theta_0}(\varphi)$ and $\frac{d}{d\varphi}\theta_{\theta_0}(\varphi) = \theta'_{\theta_0}(\varphi)$ will be closely approximated by the polynomials (318), and the approximations can be made arbitrarily accurate by taking n sufficiently large, i.e. by computing enough coefficients f_i, g_i.

Given a value of θ_0 with $0 < \theta_0 < \pi/3$, we fix a small φ_0 and compute the polynomials (318), to be regarded as approximations of $\theta_{\theta_0}(\varphi)$, $\alpha_{\theta_0}(\varphi)$ and $\theta'_{\theta_0}(\varphi)$ on the interval $0 \le \varphi \le \varphi_0$. In particular, the values of the three polynomials at $\varphi = \varphi_0$ will serve as (approximate) initial values for the functions α, θ and θ', whose further development on the interval $\varphi_0 \le \varphi \le \pi/2$ is governed by the system (316). This allows us to obtain numerical solutions for θ and α on this interval, using any of the standard iterative methods. Pieced together, these data furnish us with the C^1-data of the triple collision shape curves $\Gamma^*_{\theta_0}$ within the interval $0 \le \varphi \le \pi/2$.

8.7.2. *C^1-data for a selection of triple collision motions.* As in Section 8.5.3 we perform symbolic computations to calculate successively the coefficients f_i and g_i of the expansions (317). As before, these are trigonometric polynomials of $3\theta_0$, namely polynomials of $\sin 3\theta_0$ and $\cos 3\theta_0$, and we have calculated the exact expressions for $i \leq 9$. Beyond that they tend to be rather untractable in their exact form. The exact expressions for the first few coefficients are listed below for the sake of reference:

$$\begin{aligned}
f_0 &= \frac{5(10+\sqrt{13})}{232}\sin 3\theta_0, \quad f_1 = \frac{25(113+20\sqrt{13})}{53824}\sin 6\theta_0\\
f_2 &= \frac{15(544702206+58374421\sqrt{13})}{201243199488}\sin 3\theta_0\\
&+\frac{5875625(1390+313\sqrt{13})}{201243199488}\sin 3\theta_0\cos 6\theta_0\\
f_3 &= \frac{2(1909168577687-51730231240\sqrt{13})}{1318947929444352}\sin 6\theta_0\\
&+\frac{1422740625(17969+4520\sqrt{13})}{1318947929444352}\sin 6\theta_0\cos 6\theta_0
\end{aligned} \tag{319}$$

$$\begin{aligned}
g_1 &= \frac{-15(1+3\sqrt{13})}{1856}\cos 3\theta_0,\\
g_2 &= \frac{543509+352091\sqrt{13}}{29280256}+\frac{-8925(49+31\sqrt{13})}{29280256}\cos 6\theta_0\\
g_3 &= \frac{10(253093537+56946315\sqrt{13})}{134162132992}\cos 3\theta_0\\
&+\frac{-3525375(893+359\sqrt{13})}{134162132992}\cos 3\theta_0\cos 6\theta_0
\end{aligned} \tag{320}$$

The coefficient g_0 equals the constant a_0 in (266) and hence is omitted here. To illustrate the (decreasing) magnitude of the coefficients of the trigonometric polynomials f_i, g_i we list a few approximate expressions

$$\begin{aligned}
f_1 &\approx (8.6)10^{-2}\sin 6\theta_0 \quad g_1 \approx -(9.6)10^{-2}\cos 3\theta_0,\\
f_3 &\approx (2.6)10^{-3}\sin 6\theta_0 + (3.7)10^{-2}\sin 6\theta_0\cos 6\theta_0,\\
g_3 &\approx (3.1)10^{-2}\cos 3\theta_0 - (5.7)10^{-2}\cos 3\theta_0\cos 6\theta_0.
\end{aligned}$$

In this way one obtains the approximating polynomials (318) for $n = 9$, say. Hence, to obtain approximate numerical data for the solutions $\Gamma^*_{\theta_0}$ of the system (316), we have chosen $\varphi_0 = 0.05$ and $\theta_0 = k\pi/300, k = 0, 1, \ldots, 100$, and we have computed the numerical solutions using the Runge-Kutta method. These 101 solution curves outline the general behavior of the shape curve $\Gamma^*_{\theta_0}$ of triple collision motions parametrized by the initial longitude angle θ_0. We refer to Table 1, Table 2 and Table 3 which list the calculated values of $\alpha_{\theta_0}(\varphi)$, $\theta_{\theta_0}(\varphi)$ and $\frac{d}{d\varphi}\theta_{\theta_0}(\varphi)$ for $\theta_0 = k\pi/30, k = 0, 1, .., 10$, and for 6 different values of φ. All angles are measured in radians.

Table 1 : Inclination angle $\alpha_{\theta_0}(\varphi)$, for $\theta_0 = k\pi/30$

$k \setminus \varphi$	$\frac{\pi}{4}$	$\frac{3\pi}{8}$	$\frac{15\pi}{32}$	$\frac{63\pi}{128}$	$\frac{255\pi}{512}$	$\frac{1023\pi}{2048}$
0	.1947	.2292	.2072	.1926	.1883	.1871
1	.2132	.3470	.5938	.7243	.7705	.7835
2	.2464	.4554	.7917	.9694	1.0379	1.0586
3	.2742	.5170	.8895	1.0914	1.1747	1.2014
4	.2943	.5542	.9465	1.1643	1.2597	1.2922
5	.3083	.5780	.9827	1.2121	1.3179	1.3566
6	.3181	.5938	1.0067	1.2247	1.3599	1.4054
7	.3248	.6043	1.0226	1.267?	1.3904	1.4433
8	.3292	.6110	1.0328	1.2816	1.4116	1.4723
9	.3317	.6147	1.0385	1.2898	1.4243	1.4917
10	.3325	.6159	1.0403	1.2925	1.4285	1.4989

Table 2 : Longitude angle $\theta_{\theta_0}(\varphi)$, for $\theta_0 = k\pi/30$

$k \setminus \varphi$	$\frac{\pi}{4}$	$\frac{3\pi}{8}$	$\frac{15\pi}{32}$	$\frac{63\pi}{128}$	$\frac{255\pi}{512}$	$\frac{1023\pi}{2048}$
0	0	0	0	0	0	0
1	.23558	.40978	.66916	.77595	.80954	.81866
2	.42158	.61693	.83930	.92403	.95180	.95968
3	.55924	.73215	.90859	.97447	.99697	1.0037
4	.66448	.80830	.94747	.99907	1.0173	1.0231
5	.74914	.86481	.97349	1.0137	1.0283	1.0333
6	.82059	.91025	.99302	1.0236	1.0350	1.0391
7	.88344	.94901	1.0089	1.0310	1.0394	1.0426
8	.94075	.98370	1.0226	1.0370	1.0425	1.0448
9	.99475	1.0160	1.0352	1.0423	1.0450	1.0462
10	1.0472	1.0472	1.0472	1.0472	1.0472	1.0472

Table 3 : Values of $\frac{d}{d\varphi}\theta_{\theta_0}(\varphi)$, for $\theta_0 = k\pi/30$

$k \setminus \varphi$	$\frac{\pi}{4}$	$\frac{3\pi}{8}$	$\frac{15\pi}{32}$	$\frac{63\pi}{128}$	$\frac{255\pi}{512}$	$\frac{1023\pi}{2048}$
0	0	0	0	0	0	0
1	.29933	.63143	1.2517	1.7182	1.9466	2.0210
2	.40322	.60768	.99427	1.3887	1.6602	1.7708
3	.39183	.50135	.76959	1.0998	1.3930	1.5427
4	.34216	.40251	.60029	.87390	1.1710	1.3623
5	.28289	.31737	.46599	.68801	.97142	1.2023
6	.22295	.24306	.35361	.52754	.78004	1.0406
7	.16476	.17645	.25526	.38365	.58941	.85626
8	.10857	.11502	.16580	.25043	.39595	.62716
9	.05389	.05675	.08166	.12369	.19908	.33781
10	0	0	0	0	0	0

8.8. **An outlook on the general case.** Finally, let us briefly consider the more general case of non-equal masses and/or non-vanishing total energy h. Namely, by Theorem 45, the system (221) must be replaced by

$$\begin{cases} \frac{d\alpha}{ds} = -\frac{1}{2} + \frac{1}{4}\frac{u(s)}{u(s)+h\rho}\left(1 + 2\cot\alpha \frac{d}{d\boldsymbol{\tau}^*}\ln(U^*)\right) \\ \mathcal{K}_g^* \sin^2\alpha = \frac{1}{2}\frac{u(s)}{u(s)+h\rho}\ \frac{d}{d\boldsymbol{\nu}^*}\ln(U^*) \\ 1 = (\frac{d\varphi}{ds})^2 + (\sin^2\varphi)(\frac{d\theta}{ds})^2 \end{cases} \tag{321}$$

where $u(s) = U^*(\Gamma^*(s))$ and $\Gamma^*(s)$ is the arc-length parametrized shape curve (218) on the unit sphere $S^2(1)$. Now the size function ρ in the moduli space $\bar{M}$ appears explicitly, so the system involves the two "auxiliary" functions $\alpha(s), \rho(s)$ which are still related by (207) and (208). Their initial value at triple collision is $\alpha(0) = \rho(0) = 0$, and the shape curve Γ^* emanates from the physical center $\hat{\mathbf{p}}_0 = \Gamma^*(0)$, namely the minimums-point of U^* on the northern hemisphere. Due to the space-time scaling symmetries of the Newtonian equation (cf. Chapter 7), for $h \neq 0$ there are essentially only two cases, $h > 0$ and $h < 0$, and our case $h = 0$ may be viewed as the limiting case between negative and positive energies. However, for $h \neq 0$ we may scale and assume $h = \pm 1$.

We point out the open problem of finding the appropriate version of Theorem G (or G_1) in the two cases $h = \pm 1$ or when the masses m_i are unequal. For this purpose, it is natural to try first the following two special cases.

- $h = \pm 1$ and equal masses (i.e., $m_i = 1/3$). Many results in Section 8.5 still apply and there are symmetries as before, e.g. it suffices to consider shape curves whose angular direction at the north pole is in the range $0 \leq \theta_0 \leq \pi/3$. However, the initial value problem is "essentially" singular, in the sense that the solutions of (321) are singular at $s = 0$. For example, if we assume a series expansion

$$\alpha = a_0 s^q (1 + a_1 s + ..), \quad \rho = \rho_0 s^p (1 + r_1 s + ..)$$

  then we would have $q = 1$ and ρ would have the leading exponent $p = 1/2a_0$. Moreover, from the first equation of (321) it follows that a_0 has the value from (266), but this equation also tells us that $\Gamma^*(s)$ is singular at $s = 0$.
- $h = 0$ and the masses are not equal. The system (321) is the same as (221). Note that the mass distribution $\{m_i\}$ affects the system (321) solely via the potential function U^*, but the trigonometric series development of U^*, similar to that in Section 8.2, remains to be done for non-equal masses. We also seek a convenient coordinate system on the sphere near the point $\hat{\mathbf{p}}_0$. In Section 6.6.1 we actually worked out a series expansion of U^* centered at $\hat{\mathbf{p}}_0$, involving coefficients which are symmetric functions of the masses, but here we rather need its spherical polar coordinate version.

On the other hand, in Chapter 6 there are expressions for U^* in terms of spherical polar coordinates (φ, θ) centered at the north pole $\mathcal{N}$. Therefore, one approach is to find the coordinates $(\hat{\varphi}, \hat{\theta})$ of $\hat{\mathbf{p}}_0$ in this coordinate system and then use spherical trigonometry to determine the transformation from (φ, θ) to spherical polar coordinates $(\bar{\varphi}, \bar{\theta})$ centered at $\hat{\mathbf{p}}_0$. For this purpose, we recall

$$\cos \hat{\varphi} = \frac{\sqrt{3}\sqrt{\bar{m}}}{\hat{m}}, \quad \text{cf. (177c)} \tag{322}$$

and the value of $\hat{\theta}$ (which depends on the choice of zero meridian, $\theta = 0$) can be determined from the longitude differences $\hat{\omega}_i = \pm(\hat{\theta} - \theta_i)$, with $0 \leq \hat{\omega}_i \leq \pi$, $i = 1, 2, 3$, where θ_i denotes the longitude angle of the binary collision points $\hat{\mathbf{b}}_i$. To this end, consider the spherical triangle with vertices $\mathcal{N}$, $\hat{\mathbf{p}}_0$ and $\hat{\mathbf{b}}_i$ and note that $\hat{\omega}_i$ is the angle at the vertex $\mathcal{N}$. The arc opposite to $\mathcal{N}$, connecting $\hat{\mathbf{p}}_0$ and $\hat{\mathbf{b}}_i$, has length

$$2\sigma_i = 2 \arccos \sqrt{\frac{I_i}{1 - m_i}} = 2 \arccos \sqrt{1 - \frac{\hat{m}_i}{(1 - m_i)\hat{m}}},$$

cf. (127), (170), and the two other sides have length $\pi/2$ and $\hat{\varphi}$. Hence, by the spherical cosine law we deduce the formula

$$\cos\hat{\omega}_i = \frac{\cos(2\sigma_i)}{\sin\hat{\varphi}} = \frac{1 - \frac{2\hat{m}_i}{(1-m_i)\hat{m}}}{\sqrt{1 - \frac{3\bar{m}}{\hat{m}^2}}}. \tag{323}$$

Knowing the expansion of U^* as a trigonometric series in terms of the angles $\bar{\varphi}, \bar{\theta}$, we believe the initial value problem at $s = 0$ for the system (221) can be investigated in the same way as we did for equal masses. But this time we expect regularity issues to depend crucially on the mass distribution.

FIGURES

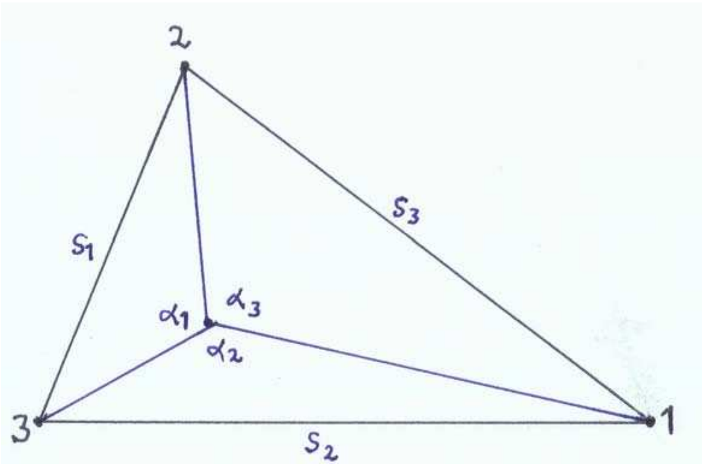

Figure 1: An m-triangle

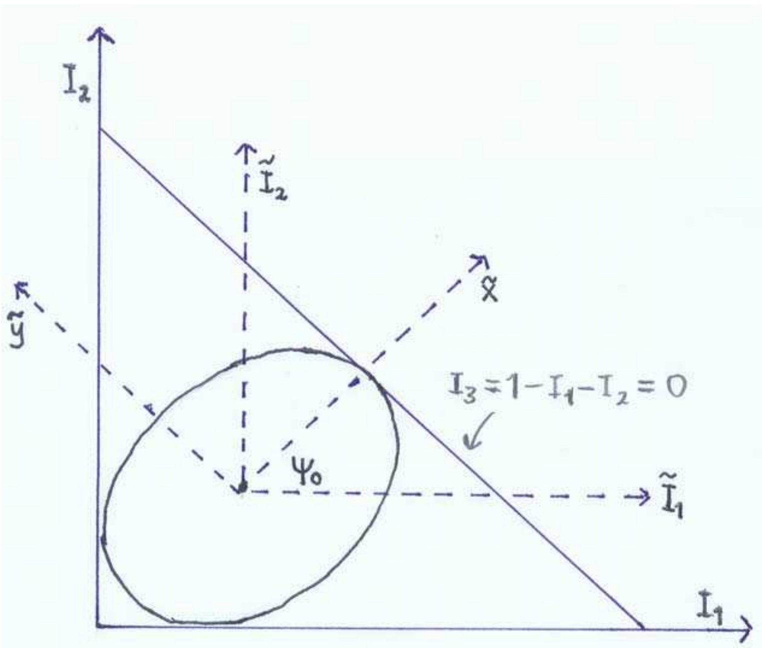

Figure 2

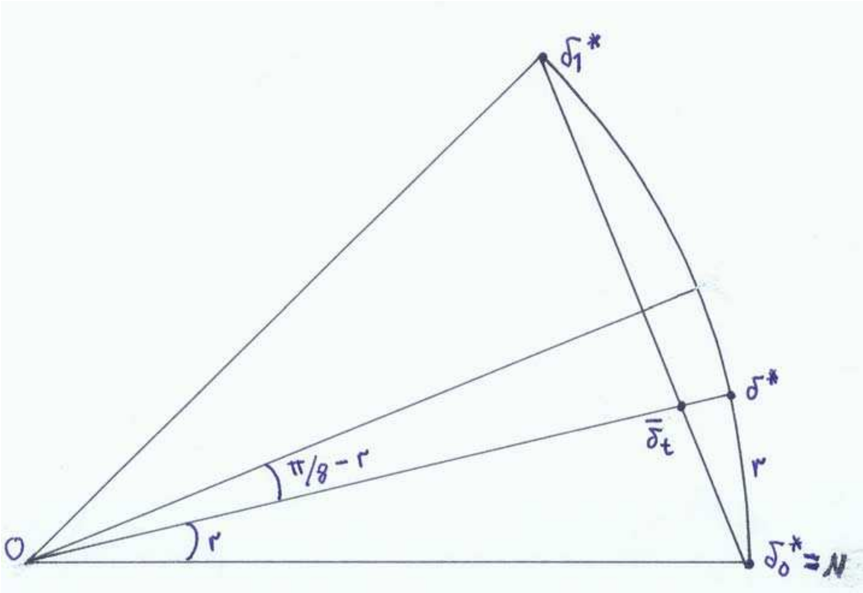

Figure 3

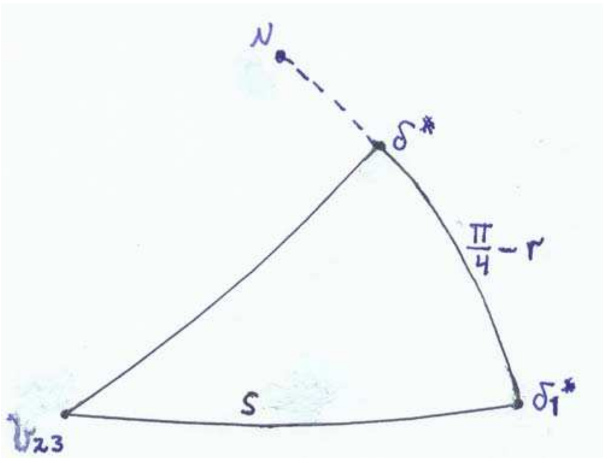

Figure 4

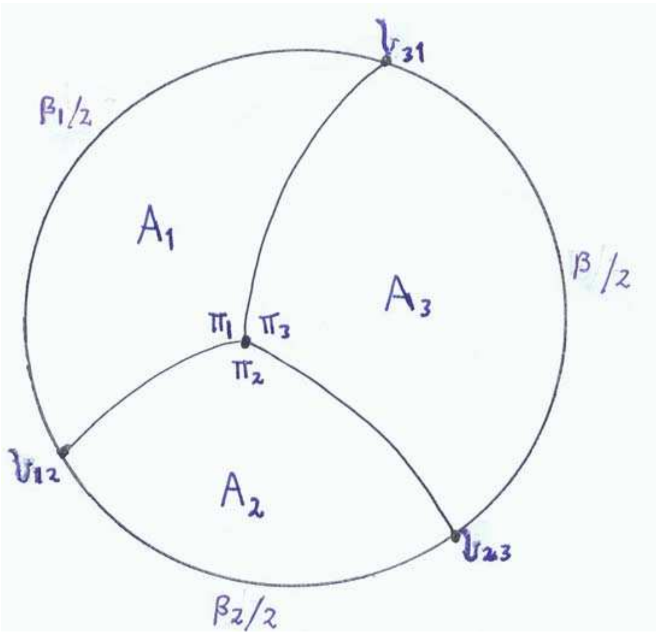

Figure 5 : Shape invariants

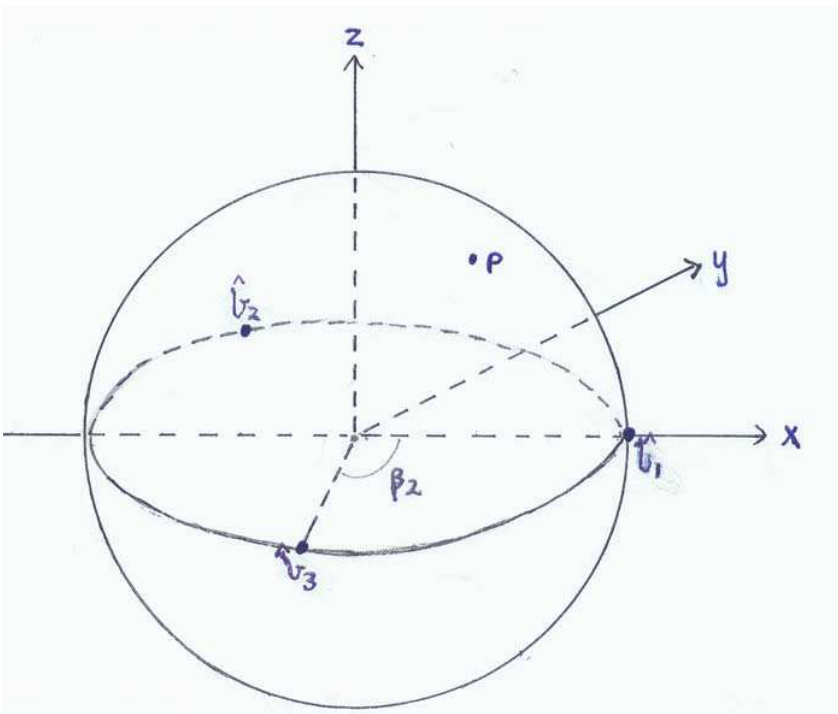

Figure 6 : 2-sphere as Euclidean model of the shape space

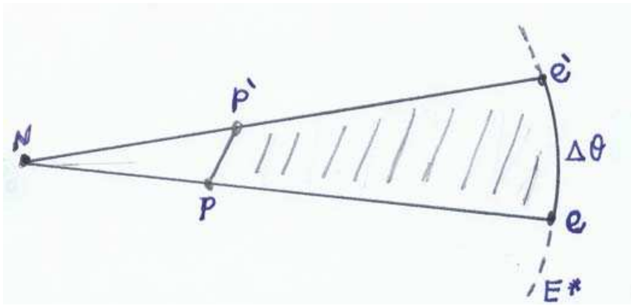

Figure 7

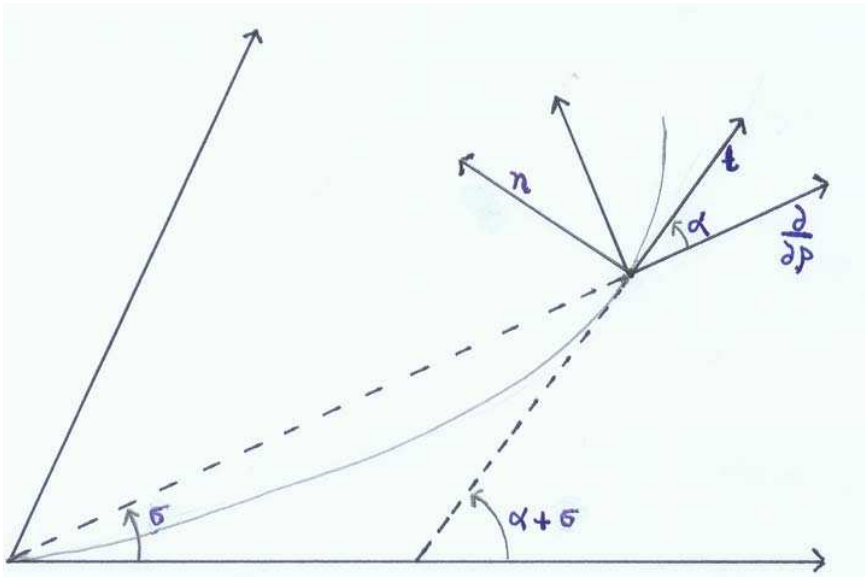

Figure 8 : Moduli curve in the cone surface $C(\Gamma^*)$

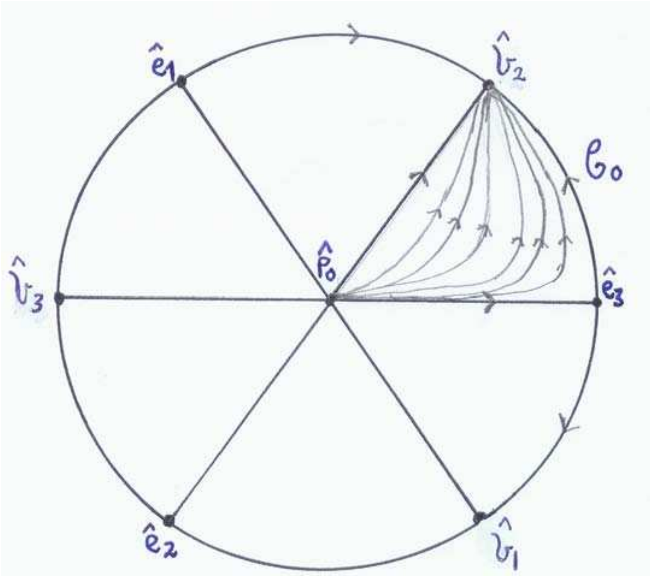

Figure 9 : Gradient flow in the basic chamber $\mathcal{C}_0$

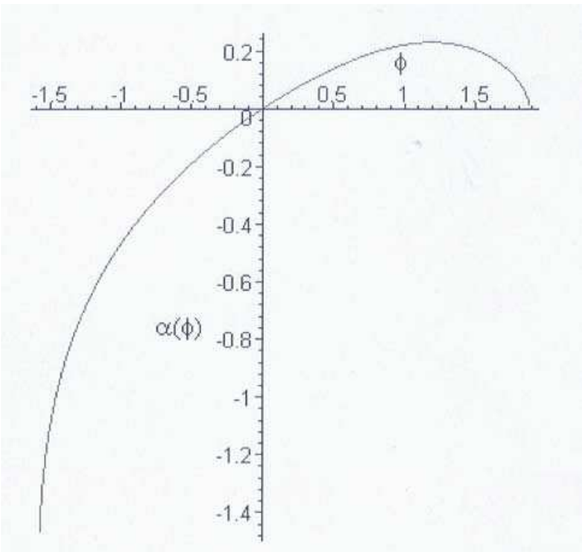

Figure 10 : Inclination angle, isosceles case

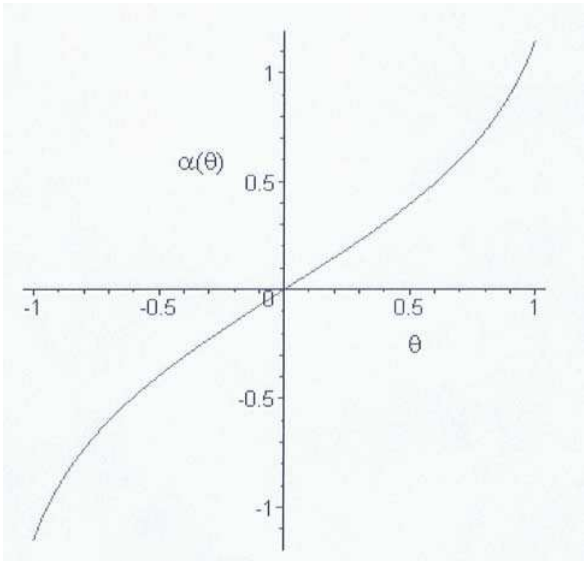

Figure 11 : Inclination angle, collinear case

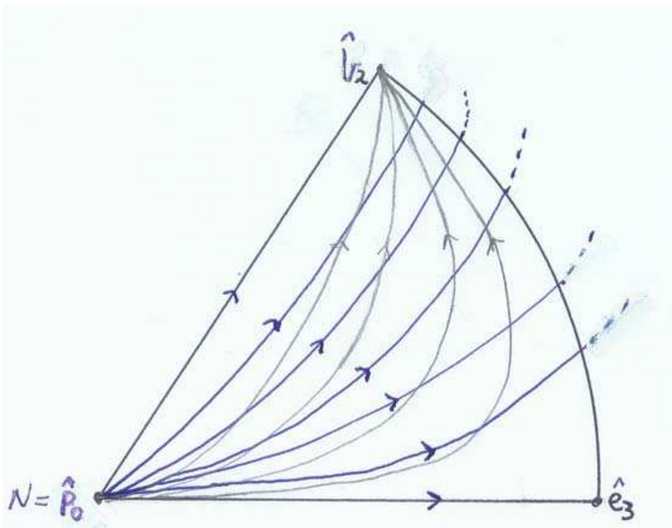

Figure 12 : Triple collision shape curves (dark) in $\mathcal{C}_0$

References

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics;* Graduate texts in Mathematics **60**, Springer-Verlag 1978.

[2] L. Euler, *De motu rectilineo trium corporum se mutuo attahentium*, Novi Comm. Acad. Sci. Imp. Petrop. **11** (1767), 144-151.

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

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

[5] S. Kobayashi, K. Nomizu, *Foundations of Differential Geometry, Volume I*; Interscience Publishers 1963.

[6] C. G. J. Jacobi, *Vorlesungen über Dynamik;* Berlin, A. Clebsch, ed., 1866.

[7] J. L. Lagrange, *Essai sur le problème des trois corps*; Ouvres, vol. 6.

[8] J. Lützen, *Interactions between mechanics and differential geometry in the 19th century;* Archive for History of Exact Sciences **49** (1995), 1 -72.

[9] C. Marchal, *The Three-body Problem;* Elsevier 1990, New York.

[10] H. Poincaré, *Sur les solutions périodiques et le principe de moindre action;* Comptes rendus de l'Aacadémie des Sciences, t.123, p. 915-918, 1896.

[11] C.L. Siegel, *Der Dreierstoss;* Ann. of Math. **42** (1941), 127-168.

[12] C.L. Siegel. *Lectures on the Singularities of the Three-body Problem;* Tata Institute of Fundamental Research Lectures on Mathematics, no. 42.

[13] C.L. Siegel, J.K. Moser, *Lectures on Celestial Mechanics;* Die Grundlehren der mathematischen Wissenschaften , vol. 187, Springer-Verlag 1971.

[14] K.F. Sundman, *Recherches sur le problème des trois corps,* Acta Soc. Sci. Fennicae **34** (1907), no.6, 144-151.

[15] K.F. Sundman, *Mémoire sur le problème des trois corps;* Acta Math. **36** (1912), 105-179.

Department of mathematics, University of California, Berkeley, USA

Department of mathematical sciences, Norwegian University of Science and Technology, Trondheim, Norway

E-mail address: eldars@math.ntnu.no

Received October 10, 2006