Figure 8s with three bodies.

Richard Montgomery
Mathematics Dept. UCSC, Santa Cruz, CA 95064
USA, email: rmont@cats.ucsc.edu

July 3, 1999

Abstract: We prove the existence of a new type of **collision-free** periodic orbit for the planar three-body problem with the **standard Newtonian potential**, provided two of the masses are equal. In this orbit the unequal mass describes a figure eight around the two equal masses. The construction is non-perturbative. It is based on the direct method of calculus of variation, together with the symmetry of interchanging the two equal masses, and a detailed analysis of the behaviour of the action near collisions. Using these same methods we also construct a new collision-free periodic orbit which realizes the sequence of eclipses "earth-moon-sun-earth-moon-sun" in the case where all three masses are equal.

1 Introduction and Results

1.1 The result

We investigate collision-free periodic motions of three point masses, m_1, m_2, m_3, moving in the Euclidean plane according to the laws of Newtonian gravitational attraction. Draw the line between m_1 and m_2, and work in a moving frame in which this line is fixed. Also apply a time-dependent homothety to the motion so that these two masses lie at fixed points on this line. The remaining mass m_3 now moves about in this new moving coordinate system, avoiding m_2 and m_3. If the curve traced out by m_3 has the topological type of a figure 8 encircling m_1 and m_2, then we will say that m_3 draws a figure 8 around m_1 and m_2, or simply that the motion is a figure 8. See figure 1.

***** INSERT FIGURE 1

The masses are positive numbers, m_1, m_2, m_3. (By a standard abuse of notation, we use these same symbols m_1, m_2, m_3 to also stand for the positions of these masses.)

Theorem 1 *Suppose $m_1 = m_2$. Then, for any value of m_3 and any period $T > 0$, there is a collision-free periodic solution to Newton's equations in which*

m_3 draws a figure 8 orbit about m_1 and m_2. The angular momentum of this orbit is zero. Viewed in the moving frame described above, the figure 8 curve traced out by m_3 is symmetric with respect to both the reflections about the y-axis and about the x-axis, where the y-axis is the line joining m_1 to m_2.

1.2 Motivation and history.

The study of periodic orbits in the three body problem has a long history. Poincare instructed us to focus attention on the periodic orbits in a famous passage of his book ([16]). Most of the results in this arena are perturbational, perturbing away from cases where one or more of the masses is zero. Ours is not. It fits within the recent tradition of using variational methods to obtain periodic orbits. See [1], [2], for example. It is a direct outgrowth of the author's earlier paper [15].

The fundamental group of the configuration space for 3 points in the plane **without collision** is the colored braid group on three strands. The center of this group is generated by rotating the three points rigidly one full revolution. Divided by this center the fundamental group becomes that of the two-sphere minus three points. This two-sphere is realized as the space of **oriented similarity classes of triangles** which we henceforth call the *shape sphere*, and the three points to be deleted represent the three types of binary collisions: m_1 with m_2, m_1 with m_3, and m_2 with m_3. **See figure 2.** Any collision-free motion of the three bodies projects to a curve on this sphere which misses the three collision points, and if the motion is periodic modulo rotations then it projects to a closed loop. Consequently, such a motion represents a free homotopy class for the thrice-punctured sphere, or, what is the same, a conjugacy class for its fundamental group. **Question: is every free homotopy class for the sphere minus three points realized by some periodic (modulo rotation) solution to the planar three body problem?** This question was first asked, to my knowledge, by Wu-Yi Hsiang (private communication).

Free homotopy classes on the sphere minus collisions can be encoded by finite sequences of eclipses. Let "2" stand for any non-collision collinear configuration in which m_2 lies on the line segment joining m_1 and m_3. In astronomical terms, this is a configuration in which m_2 has eclipsed, or come between, m_1 and m_3. Similarly we have symbols 1 and 3 for these respective eclipses. Then the figure 8 drawn (see figures 1 and 2) is encoded by the word 1323, meaning first we have an eclipse of type 1, then 3, then 2, then 3, closing back up with the original eclipse 1. See figures A and B. In this way words of even length in the letters 1, 2, and 3 encode free homotopy classes. The words are subject to the rule of grammar "no stuttering" [1]. This means that the combinations 22 , 11, 33 are forbidden within a word. Moreover, since the curves on the sphere are periodic, the words should be viewed as cyclic words: 1323 = 3132 = 2313 = 3231, with these equivalences corresponding to changing the starting place of the curve on the sphere. The no stuttering rule applies to all cyclic permutation of the word. For example, 1231 is not an allowed word since it equals 1123.

We can now rephrase our above question. Subject to the above rules of grammar, **is every finite word in eclipses realized by an actual three-body motion?**

[1] We thank Mark Levi for this turn of phrase.

To date, the only words which are known to exist for all masses are the words 12, 13 and 23 and their powers. These correspond to tight binaries : two of the masses circle each other in a near-Keplerian orbit while the third body is far away and does not participate in the word or motion. Their existence has only been proved for sufficiently high total angular momenta, relative to the total energy. (See Moeckel [13].) We should add the existence of an orbit representing the "empy word" with no letters. This orbit is the Lagrange orbit in which the masses form an equilateral triangle which then rotates rigidly. (The empty word thus corresponds to the generator of the center of the colored braid group.)

Using Poincare's perturbation methods we can find a vast array of words in the **restricted** three-body problem ($m_3 = 0$), and these survive for m_3 very small. As far as we know, ours is the first nonperturbative existence result for periodic orbits which represent homotopy classes beyond the empty word, the tight binaries 12, 13, and 23, and their powers, 1212, 121212, etc.

1.3 A result for three equal masses.

We now have the language to state our other existence result.

Theorem 2 *Let T be any nonzero period and suppose that all three masses are equal: $m_1 = m_2 = m_3$. Then there is a periodic noncollision orbit representing the eclipse sequence* 123123. *SEE FIGURE 3. Each of the six collinearities of this orbit occurs at an Euler configuration, meaning that the eclipsing mass lies at the midpoint between the two masses which it eclipses. The angular momentum for the orbit is zero. The orbit is symmetric with respect to the composition of the following three maps: cyclic permutation $m_1 \to m_2 \to m_3$ of the triangle vertices, reflection of the instantaneous triangle, and translation of time by $1/6$th of the period.*

2 The methods.

We combine the following tools:

- 1) the direct method of the calculus of variations,
- 2) A knowledge of the geometry of the *reduced configuration space C*, by which we mean the space of oriented congruence classes of triangles,

  and
- 3) the reflectional symmetry corresponding to interchanging the two equal masses.

Tool 1 has been used in a great number of papers on the N-body problem ([1] and references therein), but to our knowledge has not resulted in any new collision-free solutions to the Newtonian N-body problem [2]. Tools 1 and 2 were used together in my paper [15], and also in Sbano [17]. See Hsiang([6], [20] for a closely related approach using the Jacobi metric.

2.1 Tool 1: The Direct method.

As is well-known, Newton's equations can be reformulated as the critical point equations for the classical action

$$A(\gamma) = \int_\gamma L dt,$$

where

$$L = K + U$$

is the Lagrangian which is the sum of the kinetic energy K plus the **minus** potential energy U. To be precise, suppose that $\gamma : [0,T] \to Q$ is a possible motion of the three masses, by which we mean a curve in $Q = \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$, and suppose that γ has no collisions. Then if γ is a critical point of A among all such possible motions which share γ's domain $[0,T]$ and its endpoints, then γ is indeed a solution to Newton's equations.

In ([15]) I applied the direct method by minimizing A over all loops in a fixed free homotopy class, for example, the class represented by the figure 8. It worked very well, **provided** the negative of the potential U satisfies the bound $U \geq c/r_{ij}^2$ whenever the interparticle distance r_{ij} is sufficiently small. This bound excludes finite action solutions with collision. However, it also excludes the case under present study, which is the case of most interest, the Newtonian case of $U = \Sigma m_i m_j / r_{ij}$. Here the central analytic difficulty with the direct method is binary collision, or near-collision with very small time intervals between successive eclipses. To circumvent this difficult, we instead pose:

Problem P: Suppose that $m_1 = m_2$. Consider the class of all curves which start at any collinear configuration realizing an eclipse of type 2 and ending in the collinear Eulerian configuration of type 3: ($r_{13} = r_{32} = \frac{1}{2} r_{12}$), and which take time $T/4$ to join these configurations. Find a curve in this class which minimizes the action among all curves in this class.

Proposition 1 *A solution to problem P exist. Any such solution is a smooth* **collision-free** *solution to Newton's equations. Its only collinear points are its endpoints.*

Our main theorem, theorem 1, follows directly from this proposition. Let c_1 be a solution to problem P. Using the operation of reflection in the plane of motion, together with the operation of interchanging m_1 and m_2, we can form

[2] after this paper was written, we learned of simultaneous works of Terracini et al, and of Chenciner, in which such new solutions are found, solutions different from ours

three new "reflected" copies of c_1. Altogther these four arcs concatenate to form a figure 8. The first variation equation shows that their derivatives match up at the concatenation points, thus resulting in a periodic solution to Newton's equation. See section 2.3 for details.

2.2 Tool 2: Geometry of the reduced configuration space.

In order to use and to prove proposition 1, we will need some knowledge of the geometry of the reduced oriented configuration space C. C is the space of oriented congruence classes of triangles. Oriented congruence is almost the same as usual congruence except that we do not allow reflections: a triangle of nonzero area and its reflection are not equivalent. The three side lengths r_{12}, r_{23}, r_{13} are good coordinates on C **except** near the collinear configurations – the triangles of zero area.

Formally speaking, C is the quotient space $Q/SE(2)$ obtained by dividing the usual planar three-body configuration space $Q = \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$ by the group $SE(2)$ of **orientation-preserving** isometries of the plane. $SE(2)$ is generated by rotations and translations of the plane and acts on Q by moving each of the three points simultaneously. In other words, think of the triangle formed by the three masses and move this triangle rigidly.

Points of Q will be written $q = (q_1, q_2, q_3)$ with q_i denoting the position of m_i in the inertial plane. As is standard, we first get rid of the translational part of the group action by setting the center of mass equal to zero:

$$\Sigma m_i q_i = 0,$$

thus defining a four-dimensional linear subspace $Q_0 \subset Q$. This subspace is invariant under the dynamics provided the total linear momentum $\Sigma m_i \dot{q}_i$ is zero, **which we henceforth assume**. Then $C = Q_0/SO(2)$ where $SO(2) \subset SE(2)$ consists of rotations about the the center of mass. Write

$$\pi : Q_0 \to C \quad \text{or } \pi : Q \to C$$

for the natural projections.

The kinetic energy K of a motion $q(t) = (q_1(t), q_2(t), q_3(t)) \in Q_0$ is given by the standard expression

$$2K = \Sigma m_i ||\dot{q}_i(t)||^2.$$

It defines an inner product on Q. The kinetic energy splits into two orthogonal parts, one corresponding to shape changes, i.e. to motions in C, and the other corresponding to rotations:

$$2K(q, \dot{q}) = 2K_C(c, \dot{c}) + \frac{1}{R^2} J^2.$$

Here $c = \pi(q)$ is the projection of the motion to C. K_C is the kinetic energy corresponding to a metric on C which will be described shortly.

$$J = \Sigma m_i q_i \wedge \dot{q}_i$$

is the total angular momentum of the system. (For $v = (v_1, v_2), w = (w_1, w_2) \in \mathbb{R}^2$ we will write $v \wedge w = v_1 w_2 - v_2 w_1$.)

$$R^2 = \frac{1}{\Sigma m_i} \Sigma m_i m_j r_{ij}^2 = \Sigma m_i |q_i|^2,$$

is the total moment of inertial of the instantaneous triangle q and provides a measure of the overall size of the triangle.

Since J^2/R^2 is non-negative, it follows that any minimizer for problem P must have total angular momentum $J = 0$. Now the potential energy, or its negative

$$U = -V = \Sigma m_i m_j / r_{ij}$$

is a function on C. It follows that any minimizer q to problem P projects to form a minimizer $c = \pi \circ q$ for the problem of minimizing the integral of $K_C + U$ over curves on C, subject to endpoint conditions corresponding to those of problem P. Note that these endpoint conditions are invariant under the group of rigid motions. Conversely, if $c(t)$ is any solution to this minimization problem on C, let $q(t) \in Q_0$ be a curve with zero angular momentum which projects to c. Then $q(t)$ will be a minimizer for problem P. (The condition "angular momentum equals zero" defines a connection for the principal circle bundle $Q_0 \setminus \{0\} \to C \setminus \{0\}$. Consequently the set of all such "zero angular momentum lifts" forms a circle's worth of curves.) **In other words, problem P, and its analogue on C are equivalent via the projection** π.

As a topological space C is homeomorphic to $\mathbb{R}^3$. As a metric space it is not isometric to $\mathbb{R}^3$. The metric structure on C is defined by thinking of the kinetic energy $2K_C$ for C as a Riemannian metric on C (away from the triple collision). Using this metric, C becomes isometric to $C(S^2(1/2))$, the *cone over the sphere of radius one-half*. By the *cone* over a Riemannian manifold X, we mean the standard topological cone, $[0, \infty) \times X$ with $\{0\} \times X$ collapsed to a point, and with the Riemannian metric $dR^2 + R^2 d_X^s$ where R parameterizes the dilational ray $[0, \infty)$. Unless X is a sphere of radius 1, this metric is not smooth –i.e not Euclidean up to a quadratic error – at the cone point $R = 0$. In our case the cone point corresponds to triple collision. The sphere $S^2 = S^2(1/2)$ is the space of *oriented similarity classes* of triangles in the plane. Collinear configurations are represented by the cone $C(E)$ over the standard equator $E \subset S$. The three types of binary collisions are represented by three points $b_{12}, b_{13}, b_{23} \in E$, and consequently to three rays:

$$C(b_{ij}) := \{\pi(q) : q_i = q_j\} \subset C$$

intersecting at the origin, which is the cone point. The metric d^2s_C on C corresponding to the kinetic term K_C is given by

$$d^2s_C = dR^2 + R^2(\frac{1}{2}d\sigma)^2,$$

where $d\sigma^2$ is the standard metric on the two-sphere S^2 of radius 1 (in spherical coordinates $d\sigma^2 = d\phi^2 + sin^2\phi d\theta^2$).

These facts can be seen by first observing that Q_0 with the kinetic energy inner product $2K$ is simply a four-dimensional real inner product space, and that the circle action of $SO(2)$ on it is equivalent to the diagonal action of $SO(2)$ on $I\!R^2 \times I\!R^2 = \mathbf{C}^2$. Thus $Q_0 \to Q_0/SO(2)$ is the same as $\mathbf{C}^2 \to \mathbf{C}^2/S^1$. Finally, the restriction of this projection to $S^3 = \{R = 1\} \subset Q_0$ is the usual Hopf fibration $S^3 \to S^2$. The fact that the radius of our S^2 is one-half that of the S^3 comes about because any two antipodal points $(z_0, z_1) = (-z_0, -z_1) \in S^3$ project to the same point in S^2.

These facts regarding the metric structure of C are old. They can be found in various forms within Lemaitre [10], Deprit-Delie [5], Iwai [8], Montgomery [14], and Hsiang [6]. We also recommend the statistics book [18] for a description of this sphere valid in the case of equal masses. Moeckel [13] gives a particularly clear and useful picture of the sphere and its relation with the triple collision manifold.

The distance function d on C defined by this metric has the following direct description. Points of C are group orbits in Q. Q has a distance function, $d(q,p) = \sqrt{\Sigma m_i ||q_i - p_i||^2}$ induced by kinetic energy. **The distance between two points of C is the distance between their corresponding orbits in** Q, which is to say it is the minimum of all of the distances between the points on the corresponding orbits in Q. (Because the group acts by isometries, it is enough to take a single point on one of the orbit, and then minimize its distance to the other orbit.)

We will be using spherical coordinates (R, σ) on C, with $\sigma \in S$, $R \in [0, \infty)$. R is the coordinate described above. We also use $R\sigma$ instead of (R, σ) for the corresponding point of C. (This provides the homeomorphism of C with $I\!R^3$.) For $\Omega \subset S^2$ we use the notation

$$C(\Omega) = \{R\sigma : \sigma \in \Omega\} \subset C,$$

for the cone over Ω. We have already seen this notation, where we used $C(b_{ij})$ for the ray consisting of all binary collisions of m_i with m_j.

We have the following useful relationship between the Euclidean distance r_{ij} between masses m_i and m_j in the inertial plane and distances in C.

$$\sqrt{\mu}_{ij} r_{ij}(q) = dist_C(\pi(q), C(b_{ij})), \tag{1}$$

where

$$\mu_{ij} = m_i m_j/(m_i + m_j),$$

and where $dist_C$ is the usual distance between a point and a subset in a metric space C.

Let A_i denote the equatorial arc which represents the collinear configurations with m_i lying between m_j and m_k, with i, j, k any ordering of $1, 2, 3$. Assume that $m_1 = m_2$. Let $e_3 \in E \subset S^2$ denote the collinear configuration for which m_3 is midpoint between m_1 and m_2, and $C(e_3)$ the corresponding ray. (The "e" here is for Euler. This is is one of the three Eulerian configurations.) Then we can restate PROBLEM P: Minimize the action

$$A_C(c) = \int_0^{T/4} K_C(c, \dot{c}) + U(c)dt$$

among all arcs $c : [0, T/4] :\to C$ which satisfy the boundary conditions

$$c(0) \in C(A_2)$$

and

$$c(T/4) \in C(e_3).$$

2.3 Tool 3 and the proof of the theorem.

Any isometry $\sigma \to F(\sigma)$ of the shape sphere S^2 induces an isometry of $C = C(S^2)$ by sending $R\sigma$ to $RF(\sigma)$. (All isometries of C arise in this way.) A reflection τ about a great circle in is an isometry of S^2, and we will refer to the induced isometry of C as a reflection as well, and we will use the same symbol τ for it.

Two reflections are central to our proof. One is the symmetry τ_0 of reflecting about the collinear equator of S^2. This corresponds to reflecting triangles in the inertial plane. (The line of reflection in the plane does not matter, as the oriented equivalence class , or point of C will be the same.) The other reflection τ_3, corresponds to the interchange $m_1 \leftrightarrow m_2$ of vertices in the Euclidean plane. On S^2 it acts as reflection about the great circle $r_{13} = r_{23}$ which is the locus of points equidistant between b_{13} and b_{23}. This great circle is perpindicular to the equator E at the midpoint e_3 of arc $A_3 \subset E$.

Now any reflection is a symmetry of the reduced kinetic energy K_C, no matter what the masses are. The reflection τ_0 is always a symmetry of the potential energy as well, since it preserves the r_{ij}. **If $m_1 = m_2$ then τ_3 is also a symmetry of the potential energy.** It follows that in the equal mass case τ_0, τ_3, and $\tau_0 \circ \tau_3$ **take zero angular momentum solutions of Newton equations to other such solutions.**

The composition $\tau_0 \circ \tau_3$ is a half-twist about the ray e_3, both geometrically speaking, and in our Euclidean coordinates $R\sigma$. This means that it leaves the points of the ray $C(e_3)$ fixed, and when restricted to tangent planes orthogonal to $C(e_3)$ it is the operation $v \mapsto -v$ of 180 degree rotation.

Let c_1 be the solution of proposition 1. According to the first variation formula for the action, c_1 must be orthogonal to sector $C(A_2)$ at its starting point, and orthogonal to the ray $C(e_3)$ at its endpoint. Indeed the first variation for the action has boundary term $\langle \dot{c}, \delta c\rangle|_0^{T/4}$, where $\langle \cdot, \cdot \rangle$ is the Riemannian metric on C and δc denotes a variation of the curve c. In our case the variation δc is an arbitrary vectors field along c subject only to the constraint that it is tangent to the respective endpoint submanifolds $C(A_2)$ and $C(e_3)$ at $t = 0$ and $t = T/4$: i.e. $\delta c(0) \in TC(A_2)$ and $\delta c(T/4) \in TC(e_3)$. Now the boundary term of the first variation must vanish in order for c_1 to be a minimizer, and this boundary term is $\langle \dot{c}_1, \delta c\rangle|_0^{T/4}$, hence the orthogonality of c_1 to the endpoint sets. It follows that $(\tau_0 \circ \tau_3)_* \dot{c}_1(T/4) = -\dot{c}_1(T/4)$. Consequently the curve $c_2(t) = \tau_0 \circ \tau_3(c_1(T/2-t))$ satisfies $\dot{c}_2(T/4) = \dot{c}_1(T/4)$. Since c_2 is also a solution to Newton's equation, we see that continuing the solution c_1 past $t = T/4$ is the same as concatenating it with c_2. We now have a solution $\tilde{c} : [0, T/2] \to C$ which represents the eclipse sequence 1 3 2.

Similarly, since $\dot{c}_1(0)$ is perpindicular to E and since τ_0, acts by $v \to -v$ on vectors orthogonal to E , the process of dynamically continuing $\tilde{c}$ past $T/2$ is the same as concatenating it with the curve $\tau_0 \circ \tilde{c}(T-t), T/2 \leq t \leq T$. **The result is a closed solution curve in C representing the figure 8.** with symbol sequence 1323.

We now have a closed curve on C which, by the above-mentioned principle, is the projection from Q_0 of a zero-angular momentum solution $q(t)$. A priori, q need not be closed, but rather only be closed modulo rotations: $q(T) = gq(0)$ for some rotation g. It follows from the "area rule" ([8], [14]) and symmetry that $q(t)$ is indeed closed. This area rule asserts that the rotation g is counterclockwise rotation by Θ radians, where Θ is equal to the signed spherical area enclosed by the spherical image $\sigma(t)$ of $c(t) = R(t)\sigma(t)$. Since the figure 8 curve c just constructed encloses zero area by symmetry, $\Theta = 0$, this rotation g is the identity. Consequently any of the lifted solution curves $q(t) \in Q_0$ which project to $c(t)$ are also closed.

2.4 The proof of theorem 2.

Replace problem P by the problem of minimizing the action among all arcs $c : [0, T/6] \to C$ which connect the ray $C(e_2)$ to the ray $C(e_3)$. Thus instead of minimizing to $C(e_3)$ starting from $C(A_2)$ as in problem P we insist that we start on the subset $C(e_2) \subset C(A_2)$. The analogue of Proposition 1 holds, with the result being collision-free solution arc joining the two rays, whose interior lies in the upper hemisphere. (See proposition 2 below.) In addition to τ_0 and τ_3 used in proving theorem 1, we will use the two other reflections, τ_1, which corresponds to the interchange $m_2 \leftrightarrow m_3$, and τ_2, which corresponds to the interchange $m_1 \leftrightarrow m_3$. Applying $\tau_0 \circ \tau_3$ to c_1 and shifting and reflecting the time as we did just above in the proof of theorem 1, we are able to continue c_1

on to a solution arc joining $C(e_3)$ to $C(e_1)$. Continue this process around the equatorial circle using $\tau_0 \circ \tau_1$, and then $\tau_0 \circ \tau_2$, we go around the circle in jumps of 2/3 of the way around, realizing the eclipse sequence 132. The derivatives match up by the same first variation argument of theorem 1 : the solution arcs intersect each ray $C(e_i)$ orthogonally, and $\tau_0 \circ \tau_i$ is a half-twist about this ray. The result is a non-collision solution curve $c : [0, T/2] \to C$, which is closed, but whose endpoint derivatives do not match up: indeed $\dot{c}(T/2) = -\dot{c}(0)$. To continue, either repeat the process again, or simply reflect c about the equator – the results are the same – a smooth closed solution curve with eclipse sequence $231231 = 123123$. SEE FIGURE 3. Its zero-angular momentum lift to Q_0 is closed for the same reasons that the figure 8 constructed in theorem 1 was closed.

3 Proof of proposition 1.

SKETCH:

Replace problem P by its closure, problem $\bar{P}$. *Problem* $\bar{P}$ is the problem of finding a minimizer for $A(\gamma)$ among all paths γ which satisfy

$$\gamma(0) \in C(\bar{A}_2) \ ; \gamma(T/4) \in C(e_3).$$

These endpoint conditions are the closures of the endpoint condition sets of problem P. Thus either endpoint may now be a triple collision, and the initial endpoint $\gamma(0)$ is allowed to be a type 12 or 23 binary collision initial conditions.

Upon closing problem P we can directly apply The Arzela-Ascoli theorem. The resulting minimizer may have, a priori, collisions. Most of our work lies in showing that it has no collisions, not even at its endpoints. Having no collisions at its endpoints means that it is a minimizer for the original problem P, thus proving the main proposition.

Doing all this is just as difficult as doing it for the general case in which the endpoint constraint sets $C(\bar{A}_2)$ and $C(e_3)$ are replaced by the cones over essentially arbitrary closed disjoint subsets of the sphere S^2.

Proposition 2 *Let α and β be closed disjoint subsets of the shape sphere S^2. Then, among all curves $\gamma : [a, b] \to C = C(S^2)$ satisfying $\gamma(a) \in C(\alpha)$, and $\gamma(b) \in C(\beta)$ there is at least one which minimizes the action $A(\gamma)$. If α and β are symmetric with respect to the reflection τ_0 about the equator of the shape sphere then every such minimizer is free of triple collisions on the whole interval $[a, b]$, free of binary collisions on the open interval (a, b), and is represented by a solution to the Newtonian three-body equations. If the binary collision points are not isolated points of α or of β then every such minimizer is free of collisions throughout the entire closed interval $[a, b]$.*

3.1 Existence.

Let γ_n be a minimizing sequence for the problem. This means that $A(\gamma_n) \to \inf_\gamma A(\gamma)$ where the infimum is taken over all curves satisfying the boundary conditions $\gamma(a) \in C(\alpha)$, $\gamma(b) \in C(\beta)$. The kinetic energy term of A, together with the positivity of the negative potential U and the Cauchy-Schwartz inequality provides the following standard bound:

$$d(\gamma_n(t), \gamma_n(s)) \leq \ell(\gamma_n([t,s]) = \int_t^s ||\dot{\gamma}_n(t)|dt \leq \sqrt{2A(\gamma_n)}\sqrt{|t-s|}, \qquad (2)$$

Since the $A(\gamma_n)$ tend to an infimum, we have $A(\gamma_n) \leq A^*$, for some uniform bound A^*, independent of n. This established equicontinuity of the γ_n.

The conical nature of the metric **and of the boundary conditions** implies the boundedness of the γ_n, which is to say that it prevents $R(\gamma_n(t_n)) \to \infty$. Indeed, as we will show momentarily, we have

$$R(\gamma_n(t_n)) sin(dist_{S^2}(\alpha, \beta)) \leq \ell(\gamma_n)$$

for any times t_n in the interval $[a, b]$. Since the distance occuring in the argument of the sine is positive and less than or equal to $\pi/2$, we see that $R(\gamma_n(t_n) \to \infty$ for some sequence of times t_n would imply that $\ell(\gamma_n) \to \infty$, and which would contradict the bound $A(\gamma_n) \leq A^*$, according to equation (1).

To prove the above inequality, note that the γ_n lies on the two-dimensional cone $C([\sigma_n])$ where $[\sigma_n] \subset S^2$ denotes the image of the spherical projection σ_n of $\gamma_n = R_n\sigma_n$. Since σ_n must connect α to β, the spherical length ϕ_n of $[\sigma_n]$ is greater than or equal to $dist_{S^2}(\alpha, \beta)$. The geometry of $C([\sigma_n])$ is that of a cone over the circle of circumference ϕ_n. This geometry is flat everywhere except at the cone point. Elementary planar geometry now shows that if γ is a curve lying on the cone over the circle of circumference ϕ, that if γ which has a point a distance R from the cone point, and if γ which sweeps out the full circumference ϕ then its length ℓ satisfies $\ell \geq Rsin(\phi)$ if $\phi < \pi/2$, and $\ell > R$ if $\phi \geq \pi/2$. This is the desired inequality, since $\phi_n \leq dist_{S^2}(\alpha, \beta)$.

We have seen that there is some constant C such that $R(\gamma_n(t) \leq C$ for the entire sequence. The γ_n thus form an equicontinuous bounded sequence of curves in the complete metric space C. The Arzela-Ascoli theorem now applies, yielding a convergent subsequence of the γ_n , converging uniformly to some γ. Since the convergence is uniform, γ also satisfies the requisite endpoint conditions, $\gamma(a) \in C(\alpha)$ and $\gamma(b) \in C(\beta)$. [Here is where we need that the endpoint conditions are closed.] By the Lebesque dominated convergence theorem $A(\gamma) \leq lim A(\gamma_n)$. So γ is a minimizer. If γ were collision-free, we would be done. The basic calculus of variations would imply that it is a solution to the Euler-Lagrange equations which are Newton's equations.

4 Getting rid of collisions.

THE OVERALL PLAN. It remains to show that the minimizer γ just constructed for proposition 2 is without collisions. The set of collision times is a closed set of measure zero; closed because γ is continuous, measure zero because $U = +\infty$ on the collisions and $A(\gamma) < \infty$. A priori, this set of collision times could be a Cantor set.

Our first task will be to show that the set of collision times has no cluster points. This implies that it is a finite set. Then we will remove these remaining finite number of collisions one-by-one by explicit action-decreasing perturbations.

4.1 Some tools.

We will need the following a priori properties of any minimizer γ . Note that the complement of the collision times is a countable union of disjoint open intervals.

- (i) Restricted to any collision-free interval, the minimizer satisfies the Euler-Lagrange equations. Hence we will call the restriction of a minimizer to such an arc a **solution arc**.
- (ii) In the neighborhood of a binary collison time the motion of the mass which is **not** participating in the collision is twice continuosly differentiable. The same holds for the Jacobi vector ξ, which is the vector joining this nonparticipating mass to the center of mass of the two colliding masses.
- (iii) The energy $H = \frac{1}{2}|\dot{\gamma}|^2 - U(\gamma)$ is constant almost everywhere.

Proof of (i). The minimizer is continous. Let $I = (a, b)$ be a collision-free open solution interval. If the Euler-Lagrange equations are not satisfied, then $dA(\gamma)(\delta\gamma) < 0$ for some variation $\delta\gamma$ whose support is contained within I. We can follow this variation with an actual perturbation $\gamma^\epsilon = \gamma + \epsilon\delta\gamma + O(\epsilon^2)$, thus decreasing the action while leaving γ's endpoints fixed. This contradicts minimality.

Proof of (ii). Although U, and consequently A, are not differentiable near binary collision, they are differentiable in the direction of perturbations in which only the non-participating mass is varied. Specifically, suppose that m_1 and m_2 collide at time t_c, while $r_{13}, r_{23} > 0$. As before, let $q = (q_1, q_2, q_3)$ denote the positions of the three planets in the Euclidean plane. Write $U = U(q_1, q_2, q_3) = \Sigma m_i m_j / r_{ij}$. U is differentiable with respect to q_3 in a neighborhood of any 12 binary collision configuration. Consider variations in which $q_1(t), q_2(t)$ are fixed, while $q_3(t)$ varies to $q_3(t) + \epsilon\delta q_3(t)$. The action functional is differentiable for such a perturbation. Consequently, the q_3-Euler-Lagrange equations are satisfied, for the same reasons as in (i). These equations are of the form $m_3 \frac{d^2}{dt^2} q_3 = -\frac{m_1 m_3}{r_{13}(t)^3}(q_3 - q_1) - \frac{m_2 m_3}{r_{23}(t)^3}(q_3 - q_2)$. The $r_{i3}(t)$ are continuous functions, bounded away from zero, and the q_i are continuous, thus q_3 has continuous second derivative.

The proof that the Jacobi vector ξ has continuous second derivative is the same, once we realize that the Lagrangian has the form $\frac{1}{2}(\mu|\dot{x}|^2 + \nu|\dot{\xi}|^2) + U(x, \xi)$ in the Jacobi variables, $x = q_1 - q_2$, $\xi = q_3 - \frac{1}{m_1 + m_2}(m_1 q_1 + m_2 q_2)$. (The constants μ, ν are "reduced masses" and depend only on the m_i.)

Notation and Terminology. Above we used the following notation, which we will try to use consistently from now on. Whenever γ is a curve in C, then any zero-angular momentum curve in Q_0 whose projection to C is γ will be written $q(t) = (q_1(t), q_2(t), q_2(t))$. We may also say that q *represents* γ. Recall that if γ is a minimizer, or a solution to the Euler-Lagrange equations, then the same is true for q.

Proof of (iii). Vary the action with respect to changes of parameterization: $q_i^\tau(t) = q_i(\tau(t))$ where $\tau : [a, b] \to [a, b]$ is a smooth invertible time change. Write $t = t(\tau)$ for the inverse function, and $\lambda = dt/d\tau$, evaluated at τ. Write $V = -U$ for the potential, so that the Lagrangian is $K - V$, and the energy is $H = K + V$. We compute $L(q^\tau, \dot{q}^\tau)dt = [\frac{1}{\lambda}K - \lambda V]d\tau$. Consequently, the differential of the action with respect to parameterization change at $\lambda = 1$ (corresponding to $\tau = t$) is $\int \delta\lambda(t) H(t) dt$. Here $\delta\lambda(t)$ denotes the change in the differential of parameterization. This energy $H(t)$ is defined almost everywhere since the solution intervals of (i) have full measure. Remember the constraint $\int_a^b \lambda d\tau = \int dt = b - a$, and use the method of Lagrange multipliers to conclude that $\int(-H(t) + c)\delta\lambda(t) dt = 0$ for all possible parameterization changes. We conclude that $H(t) = c$ a.e.

A reflection principle. We will have occassion to use the *reflection principle* which asserts that if γ is a minimizer, then so is $\tau_0 \circ \gamma$, provided that the endpoint conditions for γ are invariant under τ_0. Here $\tau_0 : C \to C$ is the reflection about the collinear states $C(E)$. It holds because the Lagrangian on C is invariant under τ_0. (We already used this principle in concatenating solutions to proposition 1 in order to prove our theorems.)

4.2 Central configurations : topography of the potential.

We will need some understanding of the topography of U, and in particular of the important role played by the Lagrange configuration. The potential U is homogeneous of degree -1 with respect to dilations. It can be written in the form

$$U(R\sigma) = \frac{1}{R}\hat{U}(\sigma)$$

where $\hat{U}(\sigma)$ is a positive function on the sphere with poles at the binary collision points 12, 13, and 23.

The absolute minimum of $\hat{U}$ over the sphere occurs at the two Lagrange points. These two points represent the two oriented similarity classes of equilateral triangles. They are related to each other by the reflection τ_0. Either point will be refered to as σ_L.

$\hat{U}$ has five critical points in all. Besides the two Lagrange points there are the three *Euler points*, denoted e_1, e_2, e_3. These are saddle points for $\hat{U}$. They lie on the equator E of collinear states, with one for each of the three arcs A_i, which is to say one for each type of eclipse 1,2, or 3.

Together these five points form the *central configurations*, which are the critical points of $\hat{U}$. If one starts with a central configuration at rest, then its shape remains the same, while it shrinks toward triple collision. In our spherical coordinates, such solution curves have the form $R(t)\sigma_c$, where σ_c is one of the five central configurations. More generally, fix a realization $\{q_1(0), q_2(0), q_3(0)\} \subset \mathbb{R}^2$ of a given central configuration σ_c as a particular triangle in the plane, with size $R(0) = 1$. Identify the plane $\mathbb{R}^2$ with the complex numbers $\mathbf{C}$. Plug the ansatz $q_i(t) = z(t)q_i(0)$, $i = 1, 2, 3$ into Newton's equations. One finds that Newton's equation is satisfied if and only if the complex number $z(t)$ evolves as if it were a point in the z-plane subject to a Newtonian central force with potential $U(\sigma_c)/|z|$. The solution just described which shrinks to triple collision is of this form, with $z(t) = R(t)$ real, and with $\dot{z}(0) = 0$. For this motion the three bodies motion shrink to triple collison $z(t_c) = 0$ at some time, keeping their shape σ_* the same throughout. It is convenient to continue this path through t_c by having it retraces its path, arriving back at its maximal size of $z(0)$ at time $2t_c$. Such a continuation is natural, in that it can be made to fit analytically within the family of all solutions for the planar Keplerian problem motion through collision. We will call such an orbit a *elliptic collision rejection orbit* with shape σ_c.

4.3 Getting rid of triple collision cluster times

We show that there are at most two triple collision times. The main step here is:

Lemma 1 *Among all finite-action curves $c : [0, b] \to C$ beginning and ending at triple collision, the action is minimized by the Lagrange elliptic collision-rejection orbit with period b. There are two minimizers corresponding to the two Lagrange points $\sigma_L \in S^2$, and these are the only two minimizers.*

Proof: Let $R(t)\sigma(t)$, $0 \leq t \leq b$ be any competing curve, meaning that $R(0) = R(b) = 0$. Keep $R(t)$ the same, while replacing $\sigma(t)$ everywhere by one of the two Lagrange configurations, σ_L. This decreases both the kinetic energy K and the negative potential energy U. Indeed, before the replacement $K = \frac{1}{2}[\dot{R}^2 + \frac{1}{4}R(t)^2\|\dot{\sigma}(t)\|^2]$, and afterward it is $\frac{1}{2}\dot{R}^2$. And $\frac{1}{R}\hat{U}(\sigma(t)) \geq \frac{1}{R}\hat{U}(\sigma_L)$ with equality if and only if $\sigma(t) = \sigma_L$. It follows that $A(\gamma) \geq A(\gamma_L)$ with equality if and only if $R(t)\sigma(t) = R(t)\sigma_L$.

This reduces the minimization problem to the problem of minimizing the one-dimensional Keplerian action $\int_0^b \frac{1}{2}\dot{R}^2 + \frac{1}{R}\hat{U}(\sigma_L)$ over all scalar curves $R(t)$, $0 \leq t \leq b$ satisfying $R(0) = R(b) = 0$. The two-dimensional version was studied in detail by Gordon [7]. His results apply directly and imply that $R(t)$ must be the collinear collision-ejection Kepler orbit, as claimed.

Gordon's argument. For completeness, we recall Gordon's argument. For the Kepler problem, as for our problem, the complement of the collision times consists of a countable union of open intervals seperated from each other by the collision times. On each open interval the minimizer must satisfy the Euler-Lagrange equations, and it must tend to collision at the endpoints, i.e. the minimizer consists of an at most countable collection of solution arcs, attached at collisions.

For the one-dimensional Kepler problem there is only solution arc on a given interval, with collision at both endpoints. This is the collision-rejection solution arc defined on that interval. If the Kepler Lagrangian is $\frac{1}{2}\mu\dot{R}^2 + \frac{\alpha}{R}$ then the action of these period T solution arcs is computed to be $ct^{1/3}$, where $c = \frac{3}{2}(2\pi)^{2/3}(\mu\alpha^2)^{1/3}$. (This is also the action of any period t periodic solution for the two-dimensional Kepler problem.) In our case, $\mu = 1$ and $\alpha = \hat{U}(\sigma_L)$. Consequently the action of of our alleged minimizer is $c\Sigma(t_i)^{1/3}$ where the t_i are the lengths of the solution intervals. These open intervals are of full measure, so that $\Sigma t_i = b$. But the function $t \mapsto t^{1/3}$ is strictly concave, which implies that $\Sigma(t_i)^{1/3} \geq (\Sigma t_i)^{1/3}$ with equality if and only if all but one of the t_i's is zero. In other words, in order that it minimize the curve must consist of a single solution arc, in which case it is the collision-ejection orbit of period b.

Lemma 2 *If γ is a minimizer then it has at most two triple collison times.*

Proof: The set of triple collision times for γ is a closed subset of the interval. If it is empty, or consists of one point, we are done. Otherwise, it has a smallest point c, and a largest point d, and these are not equal. The arc $\gamma([c, d])$ joins triple collision to triple collision. This arc must be the Lagrange collision-ejection orbit over this interval. For if not, replace this section of γ with this Lagrange orbit. The resulting concatenated curve is still continuous and satisfies the correct boundary conditions, but its action is smaller than that of γ, according to the previous proposition. This would contradict the minimality of γ.

QED

4.4 Getting rid of binary collision cluster times.

We start the process of showing that the binary collision times have no cluster times. We will first need to dispense with the perverse possibility that a sequence of binary collisions converges to a triple collision. In other words, we want to dispose of the following scenario: $R(t) \to 0$ as $t \to t_c$ while $\hat{U}(\sigma(t_j)) = +\infty$ for times $t_j \to t_c$. We use a proof by contradiction. By translating time, we may assume that $t_c = 0$. (This shifts the domain of γ.) Suppose that there are binary collision times $t_j \to 0$. Set

$$m = lim_{t \to 0} \inf \hat{U}(\sigma(t)).$$

First, we discard the possibility that $m = +\infty$. Suppose $m = +\infty$, Then for any (large) value M for $\hat{U}$ there is an $\delta > 0$ such that $\hat{U}(\sigma(t)) \geq M$ over the interval $0 \leq t \leq \delta$. We may take δ to be the smallest time at which $\hat{U}(\sigma(t)) = M$. (Take M sufficiently large so this value is realized.) Now replace γ over the interval $[0, \delta]$ by $R(t)\sigma(\delta)$. The resulting curve still reaches triple collision at 0, but we have decreased both its kinetic energy, and its potential energy, since $U(R\sigma) = \frac{1}{R}\hat{U}(\sigma)$. Consequently the action of the curve has been decreased, contradicting minimality of γ.

Now suppose that m is finite. Then there are times $t_j \to 0$ with $\hat{U}(\sigma(t_j)) \to m$. Write

$$\Sigma_m = \{\sigma : \hat{U}(\sigma) = m\} \subset S^2 \; ; D_m = dist_{S^2}(\Sigma_\infty, \Sigma_m).$$

Here $\Sigma_\infty = \{b_{12}, b_{13}, b_{23}\} = \{\sigma \in S^2 : \hat{U}(\sigma) = \infty\}$ is the binary collision set. The distance D_m is positive. The arc γ consists of a countable collection of solution arcs I_j at whose endpoints the curve tends to binary collision. We may take the times t_j just introduced to be maximum points for $\hat{U}(\sigma(t))$ over their solution arcs I_j. By construction, we can find among the t_j one for which $\hat{U}(\sigma(t_j)$ is sufficiently close to m so as to guaranteee that

$$d(\sigma(t_j), \Sigma_m) < D_m/2 \text{ while } d(\sigma(t_j), \Sigma_\infty) > D_m/2.$$

Let $[t_{j-}, t_{j+}]$ be the solution interval I_j within which this maximum lies. Let η be the geodesic – an arc of a great circle on S^2 – which joins $\sigma(t_j)$ to Σ_m. Start η off at time t_j and parameterize it according to $|\dot{\eta}(t)| = |\dot{\sigma}(t)|$, $t > t_j$. Since the spherical length $\int_{t_j}^{t_j+} |\dot{\sigma}(t)| dt$ of σ over the interval $[t_j, t_{j+}]$ is greater than $D_m/2$ and since the length of η is $d(\sigma(t_j), \Sigma_m) < D_m/2$, the geodesic η will have reached Σ_m before we have run out of time in the interval I_j. Let t_* be the time when we hit Σ_m, so that $\hat{U}(\eta(t_*)) = m$. Replace our alleged minimizer $\gamma(t) = R(t)\sigma(t)$ by $\tilde{\gamma}(t) = R(t)\tilde{\sigma}(t)$ where

$$\tilde{\sigma}(t) = \begin{cases} \sigma(t), & \text{if } t \leq t_j \\ \eta(t), & \text{if } t_j \leq t \leq t_* \\ \eta(t_*), & \text{if } t \geq t_*. \end{cases}$$

Then the kinetic energy of this replacement curve is the same up to t_*, after which time it is zero, while the negative potential term $U(\tilde{\gamma})$ is the same as that of γ up to t_j and thereafter it is LESS than that of γ. Consequently the perturbed action $A(\tilde{\gamma})$ is less than the unperturbed action $A(\gamma)$, again contradicting minimality.

4.5 Isolating binary collisions

We finish the proof that the the set of binary collision times can have no cluster point. From the previous subsection, we know that the binary collisions along any minimizing path are bounded away from triple collision. Consequently the overall size R of a configuration at any binary collision time t_j is bounded away from zero and from infinity:

$$R_* \leq R(t_j) \leq R^* \;\text{ whenever } r_{ik}(t_j) = 0$$

where $R_* < R^*$ are positive numbers.

According to the conical geometry of C, the distance in C between two successive binary collisions $\gamma(t_j), \gamma(t_k)$ involving different masses, eg. 12 and 13, is subject to the bound

$$2R_* sin(\frac{1}{2}\theta_{23}) \leq d(\gamma(t_j), \gamma(t_k))$$

where $0 < \theta_{23} < \pi$ is the spherical distance between the binary collision points b_{12} and b_{13}, on the shape sphere S^2 . (See §3.1 for a similar argument.) Since $d(\gamma(t), \gamma(s)) \leq \sqrt{2A(\gamma)}\sqrt{|t-s|}$, (eq. (2)) this bounds the time $|t_j - t_k|$ between two such binary collisions away from zero.

Thus, if t_n is a sequence of binary collision times tending to a limit t_c , then from some point N on these binary collision $\gamma(t_n)$ must all be of the same type. Without loss of generality, we may assume that this type is 12, so that we have times $t_n \to t_c$ with $r_{12}(t_n) = 0$, while $R(t_n) \geq R_*$. We show that the existence of such a sequence contradicts minimality. In order to do this, we will need all three of the facts referred to at the beginning of this section: (i): on the open intervals I_j cut out by the collision times the minimizer must satisfy the Euler-Lagrange equations; (ii): the derivative of the Jacobi vector, and hence $|\dot{\xi}|^2$, is continuous at binary collisions, and so near t_c; (iii): the value of the energy H on different solution arcs $\gamma|_{I_j}$ is the same.

For the remainder of this subsection we will reserve the notation $O(1)$ to mean any function along the path γ which is bounded as $t \to t_c$. For example, it follows from the above discussion that $1/r_{13}$ and $1/r_{23}$ are both $O(1)$. To prove the impossibility of the sequence, it will suffice to prove the bound

$$\frac{d^2}{dt^2}\frac{1}{2}r_{12} = \frac{m_1 + m_2}{r_{12}} + O(1), \qquad (B)$$

uniformly on all of the solution arcs tending to t_c. To see why this suffices, assume the bound for the moment. The complement of the collision times is a union of open intervals (t_n, t_{n+1}) at whose endpoints we have collision: $r_{12}(t_n) = 0 = r_{12}(t_{n+1})$. In between, the function r_{12} achieves a maximum at some point t_n^*. The second derivative of r_{12} must be negative or zero at these points:

$$\frac{m_1 + m_2}{r_{12}} + O(1) \leq 0$$

which yields

$$k(m_1 + m_2) \leq r_{12}(t_n^*)$$

where $|O(1)| \leq 1/k$. Now recall (eq. (1)) the equality $d(P, C(b_{12})) = \sqrt{\mu}_{12} r_{12}$, relating r_{12} and the C-distance between a configuration and the nearest 12-type binary collision. (Here $\mu_{12} = \frac{m_1 m_2}{m_1 + m_2}$.) Using the distance bound $d(\gamma(t), \gamma(s)) \leq \sqrt{2A(\gamma)}\sqrt{s-t}$ of eq. (2) we obtain

$$k\sqrt{\mu}_{12}(m_1 + m_2) \leq \sqrt{2}A(\gamma)\sqrt{t_n^* - t_n}.$$

This shows that $|t_n^* - t_n|$ cannot go to zero, so that the time sequence t_n cannot be Cauchy, and consequently cannot converge to t_c.

We return to the proof of the bound (B). Levi-Civita [11] proved this bound in the case of a single solution arc. (See also Sundman [19].) We follow the lines of Wintner's treatment (see [21], p. 268) of the Levi-Civita bound We felt it is worth including the derivation, as opposed to simply quoting it, it not entirely obvious how to extend Levi-Civita's analysis to our situation of a a countable number of solution arcs with the same energy. Let q_1, q_2, q_3 be the vertices of the triangle in the plane and $q_{ij} = q_i - q_j$ as usual. Set

$$x = q_{12}.$$

The bound (B) will follow from the bound

$$\frac{1}{2}|\dot{x}|^2 - \frac{m_1 + m_2}{r} = O(1), \qquad (C)$$

where for convenience we set

$$r = r_{12} = |x|$$

for the rest of this subsection. To see why (B) follows from (C), note that

$$\frac{d^2}{dt^2}(\frac{1}{2}r^2) = \dot{x} \cdot \dot{x} + x \cdot \frac{d^2}{dt^2}x.$$

Newton's equations assert that

$$\frac{d^2}{dt^2}x = -(m_1 + m_2)\frac{x}{r^3} + f,$$

where

$$f = m_3(\frac{q_{31}}{r_{31}^3} - \frac{q_{32}}{r_{32}^3}).$$

(Wintner [21] , eq. (12), p. 260.) Since $1/r_{31}$ and $1/r_{32}$ are both $O(1)$ we have

$$|f| = O(1),$$

so that

$$f \cdot x = O(r).$$

Putting these results together, we obtain

$$\frac{d^2}{dt^2}(\frac{1}{2}r^2) = |\dot{x}|^2 - \frac{m_1 + m_2}{r} + O(r).$$

Now eq (C) implies that

$$|\dot{x}|^2 - \frac{m_1 + m_2}{r} = \frac{m_1 + m_2}{r} + O(1)$$

from which eq. (B) now follows.

It remains to derive the bound of eq (C), which is that $g(t) := \frac{1}{2}|\dot{x}|^2 - \frac{m_1+m_2}{r}$ is $O(1)$. The energy is

$$H = \mu g(t) + \frac{\nu}{2}|\dot{\xi}|^2 - \frac{m_1 m_3}{r_{13}} - \frac{m_2 m_3}{r_{23}},$$

where ξ is the Jacobi vector, as in fact (ii) above. According to that fact, $|\dot{\xi}|^2$ is bounded near t_c since ξ is C^2 near t_c. And the $1/r_{i3}$ are also $O(1)$. Fact (iii) above asserts that this energy H is constant a.e., and in particular is the **same** constant for all of the solution arcs. Call this constant h. We have proved $g(t) = \frac{1}{\mu}h + O(1) = O(1)$ near t_c, as desired.

4.6 Deleting Isolated Triple collisions

4.6.1 The Set-up.

We have shown that there are at most two triple collisions. We will show how to get rid of these remaining two.

Suppose that γ has an isolated triple collision at $t = t_c$. We will construct a perturbation γ_ϵ which agrees with γ except in a small neighborhood of t_c, has no collision in this neighborhood, still satisfies the endpoint conditions of problem $\bar{P}$, but for which

$$A(\gamma_\epsilon) < A(\gamma).$$

By a time translation we may assume that the collision time occurs at $t = 0$. This translation shifts the domain of γ. The collision time may be an interior point, or it may be either endpoint of the domain of γ. The arcs of γ on either side of 0 (or on one side if $t_c = 0$ is an endpoint) are solutions to Newton's equations (fact (i) above) which tend to triple collision at $t = 0$, and which have the same fixed energies $H = h$ (fact (iii) above). These collision solutions were investigated in some detail by Sundman, and later by other researchers, notably Siegel and McGehee. We will need some of their results.

4.6.2 Results needed from Sundman.

Let $q(t) = (q_1(t), q_2(t), q_3(t)) \in Q_0$ denote a solution to the three-body equations with a triple collision at time $t = 0$, and let $R(t)\sigma(t)$ be its projection to the shape space C. Sundman shows that :

$$R(t) = ct^{2/3} + O(t^{4/3}), c > 0$$

$$\dot{R} = \frac{2}{3}ct^{-1/3} + O(t^{1/3})$$

$$lim_{t\to 0}\sigma(t) := \sigma_0$$

and that the limiting shape σ_0 is one of the five central configurations σ_c discussed above. Moreover, this limiting shape is not just an abstract. The vertices of the triangle **in the inertial plane** settle down to a fixed triangle representing one of the σ_c after being appropriately rescaled:

$$lim_{t\to 0}\frac{1}{R(t)}q(t) = \hat{q}(0)$$

exists. Neccessarily $\pi(\hat{q}(0)) = \sigma_0$. Moreover

$$\frac{d}{dt}\hat{q} = O(1) \text{ as } t \to 0$$

where $\hat{q}(t) = q(t)/R(t)$.

4.6.3 The perturbation

Let $q(t), 0 \le t \le b$ be a solution arc with triple collision at $t = 0$. We will construct a perturbation of q supported in an arbitrarily small neighborhood of 0 which decreases the action. This will be our basic tool for getting rid of the triple collisions.

Fix a shape

$$\sigma^p \in Q = I\!R^2 \times I\!R^2 \times I\!R^2,$$

of unit size ($R(\sigma_p) = 1$) subject to the constraint:

$$\sigma^p_{ij} \cdot \hat{q}_{ij}(0) > 0, \text{ for } i \neq j. \quad (3)$$

Here $\sigma^p_{ij} = \sigma^p_i - \sigma^p_j$ denotes the vector connecting vertex m_i to vertex m_j for the perturbed triangle, $\hat{q}(t) = \frac{1}{R(t)}q(t)$ is the normalized solution, $\hat{q}(0)$ is the limit for these rescaled triangles, and $\hat{q}_{ij}(0)$ is the corresponding ij edge vector for $\hat{q}(0)$. Define a perturbation q^ϵ of q according to

$$\begin{aligned} q^\epsilon(t) &= q(t) + \epsilon f(t)\sigma_p \\ &:= q(t) + \epsilon\phi(t) \end{aligned}$$

Here $f(t) = f(t;\delta)$ is a strictly decreasing, nonnegative scalar function with $f(0) = 1$, $df/dt(0) < 0$, and with support $[0,\delta] \subset [0,b]$. The parameter δ is taken small enough so that the Sundman estimates hold on $[0,\delta]$. We also impose the conditions $|\dot{f}| \le c/\delta$ on $[0,\delta]$, while $\dot{f} < -c_2/\delta$ on $[0,\delta/2]$ for constants c, c_2 independent of δ.

Proposition 3 (Perturbation Proposition) *For all ϵ, δ sufficiently small, $A(q^\epsilon) < A(q)$, for the perturbation defined above, provided that σ^p is subject to the inequality (3) above.*

Remark: Since the endpoint at $t = 0$ of q^ϵ is $\epsilon\sigma^p$, we have gotten rid of the triple collision and at the same time lowered the action.

Proof of Proposition 3.

The perturbed intermass distances are

$$(r_{ij}^\epsilon)^2 = r_{ij}^2 + \epsilon^2|\phi_{ij}|^2 + 2\epsilon\phi_{ij}\cdot q_{ij}$$

with $\phi_{ij} = \phi_i - \phi_j = f\sigma_{ij}^p$, and $r_{ij} := r_{ij}^0 = |q_{ij}|$ being the unperturbed distance. Now $\phi_{ij}\cdot q_{ij} = fR\sigma_{ij}^p\cdot\hat{q}_{ij}$ so that condition (3) together with the continuity of $\hat{q}$ at 0 implies that $r_{ij}^\epsilon > r_{ij}$ for δ sufficiently small. Consequently $U^\epsilon := U(q^\epsilon) < U = U(q)$ and so

$$\int U^\epsilon < \int U.$$

We now show that

$$\int K^\epsilon < \int K$$

where $K^\epsilon = \frac{1}{2}\langle\dot{q}^\epsilon,\dot{q}^\epsilon\rangle$, is the perturbed kinetic energy, and where we use $\langle q, v\rangle = \Sigma m_i q_i\cdot v_i$ denotes the kinetic energy inner product. The standard identity $\langle q, v\rangle = \frac{1}{\Sigma m_i}\Sigma m_i m_j q_{ij}\cdot v_{ij}$ shows that condition (3) implies that

$$\langle\sigma_p, \hat{q}(0)\rangle > 0, \tag{4}$$

which will be the key to the desired inequality.

In the remainder of the proof, "$\frac{d}{d\epsilon}$" will denote the derivative with respect to ϵ, taken at $\epsilon = 0$. We have

$$\frac{d}{d\epsilon}\frac{1}{2}\langle\dot{q}^\epsilon,\dot{q}^\epsilon\rangle = \langle\dot{\phi},\dot{q}\rangle.$$

And $\dot{\phi} = \dot{f}\sigma_p$. Also $\dot{q} = \dot{R}\hat{q} + R\frac{d}{dt}\hat{q}$ since $q = R\hat{q}$, where $\hat{q}$ is the normalized shape. Thus

$$\frac{d}{d\epsilon}K^\epsilon = \dot{f}\dot{R}\langle\sigma_p,\hat{q}(t)\rangle + \dot{f}R\langle\sigma_p,\frac{d}{dt}\hat{q}\rangle.$$

Now $\dot{f} = O(1/\delta)$, $R = O(t^{2/3})$, and $\dot{R} = O(t^{-1/3})$, so that the first term, $\dot{f}\dot{R}\langle\sigma_p, \hat{q}(t)\rangle$ dominates the integral. But $\dot{f} < 0$, $\dot{R} > 0$, and, according to condition (4) above, and the continuity of $\hat{q}$, we have $\langle\sigma_p, \hat{q}(t)\rangle > 0$ for t, and hence δ, sufficiently small. Consequently, the whole integrand is negative over the domain of variation, $[0, \delta]$, and the integrated kinetic energy has been decreased.

Let us be a bit more precise regarding this last argument. Let c stand for a positive constant, independent of ϵ, which can change from inequality to inequality in the rest of this paragraph. Then, for δ sufficiently small, and $t \leq \delta$, we have $\langle\sigma_p, \hat{q}(t)\rangle > c$, as well as $|\dot{f}| < c/\delta$, $\dot{R} > ct^{-1/3}$, and $R < ct^{2/3}$. Moreover $\dot{f}\dot{R} < 0$, with $\dot{f}\dot{R} < -\frac{c}{\delta}t^{-1/3}$ on a smaller interval, say $0 < t < \delta/2$. It follows that the change in the integral of the kinetic energy is given by

$$\begin{array}{rcl} \frac{d}{d\epsilon}\int K^\epsilon & = & \int_0^\delta \dot{f}\dot{R}\langle\sigma_p, \hat{q}(t)\rangle + O(\delta^{2/3}) \\ & < & -c\langle\sigma_p, \hat{q}(0)\rangle\delta^{-1/3} + O(\delta^{2/3}). \end{array}$$

Provided δ is small enough. The derivative of the kinetic energy is negative, so that the integral of the the kinetic energy has been decreased.

We have shown that $\int U^\epsilon$ and $\int K^\epsilon$ are each seperately less than their unperturbed counterparts, so the overall action has been decreased.

QED

We now use the perturbation proposition 3 to get rid of all triple collisions in our minimizer. If t_c is an interior triple collision time, then on either side of it are solution arcs, $q^-(t), t < t_c$, and $q^+(t), t > t_c$ with triple collision at t_c. If both have limiting shape $\hat{q}(t_c)$ equal to the Lagrange configuration, then we are done. For by rotating and possibly reflecting the arc of the alleged minimizer q on one side of the collision we can arrange that the limiting shape for both q^+ and q^- are the same (both being equilateral triangles in the plane). Apply the perturbation of the proposition simultaneously to both arcs $q^\pm$, taking the perturbed shape to be $\sigma^p = \hat{q}^+(t_c) = \hat{q}^-(t_c)$. Condition (3) is satisfied because $\sigma^p_{ij} = \hat{q}^\pm(t_c)$. The action of the perturbed curve $\hat{q}^\epsilon$ is less than that of q, and the perturbed curve has no triple collision at t_c. We have gotten rid of this collision.

If only one arc, say q^- is Lagrange, this trick will still work. For suppose the other arc is Eulerian. By appropriate rotation and reflection of one of the arcs we can arrange that the two limiting configurations are as in the figure below

***

Insert figure 4

****

By inspection, $\hat{q}^+_{ij}(t_c) \cdot \hat{q}^-_{ij}(t_c) > 0$ for this configuration. We can take σ^p to be either limit, say $q^-(t_c)$, and continue as above to get rid of this collision.

This same trick works if the triple collision is at one of the endpoints of the time interval, provided the limiting shape is equilateral. For example, suppose we tend toward Lagrangian triple collision at $t = T/4$, in our original problem, problem $\bar{P}$. Then we must take σ^p to be the Euler central configuration of type

132, otherwise the perturbation will violate the endpoint conditions of problem $\bar{P}$. But this works according to the previous paragraph. More generally, if $\pi(\sigma_p)$ lies in the same closed hemisphere of S^2 as our limiting Lagrage configuration σ_L then σ_p can be rotated so that condition (3) is satisfied. This can always be arranged by reflecting the solution arc. Recall that reflection keeps the action of a path the same. We need that the reflected arc satisfies the correct boundary conditions, here $c(a) \in C(\alpha)$. This is guaranteed by the assumption in proposition 3 that the endpoint sets α and β are symmetric with respect to reflection. **This is the only place where we use this assumption on the endpoint conditions.**

This disposes of the case where the limiting shape(s) $\hat{q}(t_c)$ are Lagrange. If t_c is interior and both endpoints are Euler but equal, or if we are in the endpoint case and the limiting shape corresponds to that endpoint's boundary condition (α or β) , then the tricks we just used to get rid of limiting Lagrange shapes carry through verbatim. It remains to deal with the case where t_c is interior, but the one-sided limits $\hat{q}_{\pm}(t_c)$ are **different** Eulerian configurations, or where t_c is one of the endpoints, but the limiting shape $\hat{q}(t_c)$ doe not match up with the boundary condition there. For example, in our original problem $\bar{P}$ this situation arises when $\pi(\hat{q}(t_c) = e_2$ at $t_c = T/4$, whereas according to the boundary conditions we must have $\pi(\sigma^p) = e_3$ at $T/4$.

We are left with trying to decrease the action in the case where the limiting shape $\hat{q}(t_c)$ is one euler configuration, but the desired perturbed shape, σ^p is a different Euler configuration, or more generally, any different shape. We proceed as follows.

Lemma 3 *Let $q(t), 0 \le t \le b$ be an action minimizer among all curves satisfying $q(0) =$ triple collision, $q(b) = q^1$, a fixed triangle. Then the limiting shape $\hat{q}(0)$ is a Lagrangian configuration.*

According to this lemma, we can replace any triple collision solution arc by one which tends to Lagrange, and thereby decrease the action. Now, proceed as before to get rid of this collision.

ASIDE. An alternative approach uses

Lemma 4 *The perturbation proposition still holds if condition (3) there is relaxed to $\sigma^p_{ij} \cdot \hat{q}_{ij}(0) = 0$ for all $i \neq j$.*

This approach is useful if the endpoint sets α, β are subsets of the equator, as is the case for our original problem $\bar{P}$. Suppose that σ^p and $\hat{q}_{ij}(0)$ are both Eulerian, or simply both collinear. Rotate the lines containing one of the states so that it becomes perpindicular to the line containing the other. The relaxed condition of this lemma holds, and we can lower the action as before, thus getting rid of this collision. The proof of lemma 4 follows the lines of the proof in the final section.

Proof of Lemma 3. Let γ be a minimizer for the problem of lemma 3. Collisions are isolated, so γ contains a solution arc which tends toward triple collision. The rescaled triangle $\hat{q}$ tends towards a central configuration. We suppose that this central configuration is one of the Eulerian configurations $\hat{q}^E$ and will show how to lower its action by a perturbation.

Let $R(t)\sigma(t)$ be the spherical decomposition of γ. We will keep $R(t)$ the same throughout our perturbaton.

To define the perturbation of σ recall that $\hat{q}^E$ is a saddle critical point of $\hat{U}$. According to the Morse lemma we can write

$$\hat{U} = c + x^2 - y^2$$

where x, y are Morse coordinates on the sphere near $\hat{q}^E$, and where $c = \hat{U}(\hat{q}^E)$. By the reflectional symmetry of $\hat{U}$, and the Z_2-equivariant version of the Morse lemma, we can take $\tau_0(x, y) = (x, -y)$ so that the y coordinate is a measure of the signed distance away from the equator, and so that the upper hemisphere is given (locally) by $y \geq 0$. According to the reflection principle above, we may assume that our unpertured shape curve $\sigma(t)$ lies in this hemisphere. This means that $y(t) \geq 0$ along our curve where $(x(t), y(t))$ are the coordinates of the unperturbed curve $\sigma(t)$.

Using these coordinates we define the perturbation by

$$\sigma^\epsilon(t) = (x(t), y(t) + \epsilon f(t)), \epsilon > 0.$$

Here f is a smooth non-negative, decreasing function with support on $[0, \delta_0]$ chosen so that

$$f(t) = 1, t \leq \delta_1$$

and

$$f(0) = 0, t \geq \delta_2$$

where δ_0 and δ_1 will be related to ϵ later. Then

$$\hat{U}^\epsilon = \hat{U}(\sigma^\epsilon) = c + x^2 - (y + \epsilon f)^2$$

so that

$$\hat{U}^\epsilon \leq U \text{ everywhere }.$$

Moreover

$$\hat{U}^\epsilon - \hat{U} \leq -\epsilon^2, \text{ for } 0 \leq t \leq \delta_1.$$

Also

$$||\dot{\sigma}^\epsilon||^2 = ||\dot{\sigma}||^2 + O(\epsilon/(\delta_1 - \delta_2)) + O([\epsilon/(\delta_1 - \delta_2)]^2)$$

Now take $\delta_1 = k_1\epsilon$, $\delta_2 = k_2\epsilon$ for fixed constants $k_1 < k_2$. The constants $k_1 = 1, k_2 = 2$ will do. Then the kinetic energy estimate becomes

$$||\dot{\sigma}^\epsilon||^2 - ||\dot{\sigma}||^2 = O(1)$$

It follows that the difference in actions is

$$A(\gamma^\epsilon) - A(\gamma) = \int \frac{1}{R^2}(\hat{U}^\epsilon - \hat{U})dt + \int \dot{R}^2(\|\dot{\sigma}^\epsilon\|^2 - \|\dot{\sigma}\|^2)dt.$$

According to Sundman's estimates for R and $\dot{R}$, and the above estimates on $(\hat{U}^\epsilon - U)$ and $\|\dot{\sigma}^\epsilon\|^2 - \|\dot{\sigma}\|^2$ we have

$$A(\gamma^\epsilon) - A(\gamma) < -c\int_0^{\delta_1} \frac{\epsilon^2}{t^{4/3}}dt + c\int_0^{\delta_2} t^{4/3}dt$$

for some constant c. Taking into account the linear relation between the δ_i and ϵ it follows that

$$A(\gamma^\epsilon) - A(\gamma) < -c_1\epsilon^{5/3} + c_2\epsilon^{7/3} + O(\epsilon^{8/3})$$

for some positive constants c_1, c_2. Taking ϵ sufficiently small, the first term dominates, showing that the action has been decreased.

We have just shown that we can always decrease the action to triple collision if the limiting shape is Eulerian. Since the limiting shape of a minimizing solution arc has to be either Eulerian or Lagrangian, it must be Lagrangian.

4.7 Deleting Isolated Binary Collisions

Suppose our alleged minimizer q has an isolated binary collision. By translating time we may assume that collision occurs at time $t = 0$. We perturb q as we did for isolated triple collisions:

$$q^\epsilon = q + \epsilon f \sigma^p$$

where f has the given shape:

INSERT FIGURE FOR f

In other words,

$$f = \begin{cases} 1, & \text{for } t \leq \delta_0 \\ 0 & \text{for } t \geq \delta_1 \end{cases}$$

$\dot{f} \leq 0$, and $|\dot{f}| \leq c/(\delta_1 - \delta_0)$. For negative t, extend f by reflection. Thus $f(-t) = f(t)$ and f is defined on both sides of the collision. We will take the cut-off parameters δ_0, δ_1 to be given by

$$\delta_0 = \epsilon^{3/2}$$

$$\delta_1 = \epsilon^{3/2} + \epsilon$$

The point $\sigma^p \in Q$ is a given unit non-collision "perturbed shape" with characteristics to be specified momentarily.

Solution arcs $q^{\pm}$ of our alleged minimizer q lie on either side of the collision time. Without loss of generality, we may suppose that m_1 and m_2 are the colliding masses. Results going back to Levi-Civita assert that:

$$r_{12}(t) = ct^{2/3} + O(t^{4/3}),$$

as $t \to 0$ and that the solution arcs, $q^{\pm}(t)$ are analytic functions of the variable $t^{1/3}$. (The two arcs $q^{\pm}$ may be different analytic functions.) By analyticity, the angular coordinate θ of the direction vector $q_{12} = q_1 - q_2$ for either arc satisfies $\theta^{\pm} = \theta_0^{\pm} + c_{\pm}t^{1/3} + O(t^{2/3})$ which implies that both one sided limits for the corresponding direction vector $\hat{q}_{12} = q_{12}/r_{12}$ are well-defined:

$$\hat{q}_{12}^{\pm} := lim_{t \to 0^{\pm}}[\frac{1}{r_{12}^{\pm}(t)}(q_1^{\pm}(t) - q_2^{\pm}(t))].$$

These limits need not be equal. The condition we impose on the perturbed shape is that

$$\sigma_{12}^p \cdot \hat{q}_{12}^+(0) \geq 0 \ \text{ and } \ \sigma_{12}^p \cdot \hat{q}_{12}^-(0) \geq 0.$$

This is always possible to do by taking σ_{12}^p to lie along the bisector of the angle defined by the $q_{12}^{\pm}(0)$. This inner product criterion implies the condition

$$\sigma_{12}^p \cdot q_{12}^{\pm} = (nonnegative) + O(t). \tag{5}$$

(This is stronger than the condition $\sigma_{12}^p \cdot q_{12}^{\pm} = O(t^{2/3})$ which is what we would have if we used the Levi-Civita decay rate alone with no condition on σ^p.)

We now show that the action A^ϵ of each solution arc $q^{\epsilon,\pm}$ is decreased through this perturbation. We will only present the case for the positive arc $q^+ = q(t)$, $t \geq 0$. The argument is the same for the negative arc. Let U^ϵ denote $U(q^\epsilon(t))$, and $U = U(q(t)$, with similar notation for K^ϵ , K, and A^ϵ, A. Then

$$\begin{array}{lclll} A - A^\epsilon & = & \int_0^{\delta_0}(U - U^\epsilon)dt & +\int_{\delta_0}^{\delta_1}(U - U^\epsilon)dt & +\int_{\delta_0}^{\delta_1}(K - K^\epsilon)dt \\ & = & I_0(\epsilon) & +I_1(\epsilon) & +I_2(\epsilon). \end{array}$$

(In obtaining the expression for $I_2(\epsilon)$, the integrated kinetic difference, we used the fact that $\dot{q} = \dot{q}^\epsilon$ except over the interval $[\delta_0, \delta_1]$.) We must show that the quantity $A - A^\epsilon$ is positive for small enough positive ϵ.

To begin our estimates, we argue that we can replace U by $m_1 m_2/r_{12}$, and U^ϵ by $m_1 m_2/r_{12}^\epsilon$,while at the same time replacing K by $\frac{1}{2}\mu|\dot{q}_{12}|^2$, and K^ϵ by $\frac{1}{2}\mu|\dot{q}_{12}^\epsilon|^2$. This replacements in the potentials U are legitimate because r_{13} and r_{23} are continuous and bounded away from zero, so that the terms in the potentials involving their reciprocals are $O(1)$. Similarly, for the kinetic term, if ξ is the Jacobi vector joining the 12 center of mass to m_3, then its time derivative is $O(1)$. (Fact (ii) of §4.1.) But $K = \frac{1}{2}(\mu|\dot{q}_{12}|^2) + \nu|\dot{\xi}|^2)$ so that $K = \frac{1}{2}\mu|\dot{q}_{12}|^2 + O(1)$ as $t \to 0$ so the error in K upon ignoring the $\dot{\xi}$ term is $O(1)$. Consequently, these replacements lead to an overall error in $A - A^\epsilon$ of size $O(\delta_1) = O(\epsilon)$. But as we will show below the dominant term of our three integrals is $I_0(\epsilon)$ and that it is of order $O(\sqrt{\epsilon})$, and consequently beats out these $O(\epsilon)$ errors.

4.8 Bounding I_0.

We estimate I_0. We compute

$$(r_{12}^{\epsilon})^2 = (r_{12})^2 + \epsilon^2 f^2 |\sigma_{12}^p|^2 + 2\epsilon f \sigma_{12}^p \cdot q_{12}^{\pm}.$$

Since $f = 1$ on $[0, \delta_0]$, and since $\sigma_{12}^p \cdot q_{12} = nonneg. + O(t)$ we have

$$(r_{12}^{\epsilon})^2 = (r_{12})^2 + \epsilon^2 |\sigma_{12}^p|^2 + 2\epsilon(nonneg. + Ct + ...)$$

on this interval. Recall

$$r_{12} = ct^{2/3} + O(t)$$

with $c > 0$. These suggest the substitution

$$t^{2/3} = \epsilon\tau$$

in the integral. Ignoring the terms r_{i3} as discussed in the previous paragraph, we have $Udt = m_1 m_2 \epsilon^{3/2} \tau^{1/2} d\tau / (c\epsilon\tau + c_2 \epsilon^{3/2} \tau^{3/2} + ...)$, while

$$U^{\epsilon} dt = \frac{m_1 m_2 \epsilon^{3/2} \tau^{1/2} d\tau}{(c^2 \epsilon^2 \tau^2 + \epsilon^2 |\sigma^p|^2 + \epsilon O(\epsilon^{3/2} \tau^{3/2}))}$$

where the error term $O(\epsilon^{3/2}\tau^{3/2})$ came from condition (5) above, the Levi-Civita expansion of r_{12} and our substitution $t = \epsilon^{3/2}\tau^{3/2}$. Consequently, on the interval in question, we have

$$(U - U^{\epsilon})dt = \sqrt{\epsilon} m_1 m_2 \tau^{1/2} d\tau [\frac{1}{c\tau + O(\epsilon^{1/2}\tau^{3/2})} - \frac{1}{\sqrt{c^2\tau^2 + |\sigma^p|^2 + O(\epsilon^{1/2}\tau^{3/2})}}].$$

This integrand is of the form $[\sqrt{\epsilon} f(\tau) + O(\epsilon)]d\tau$ where $f(\tau)$ is the **positive** function

$$f = m_1 m_2 \tau^{1/2} (\frac{1}{c\tau} - \frac{1}{\sqrt{c^2\tau^2 + |\sigma^p|^2}}).$$

Recalling that $\delta_0 = \epsilon^{3/2}$ and that $I_0 = \int_0^{\delta_0} (U - U^{\epsilon})dt$ we find that

$$I_0(\epsilon) = C\sqrt{\epsilon} + O(\epsilon),$$

with $C = \int_0^1 f(\tau) d\tau$ a **positive number**.

4.9 Bounding I_1

First, we claim that we can pick a positive constant c, arbitrarily small as $\epsilon \to 0$, such that

$$r_{12}^{\epsilon} \geq (1 - c\epsilon)^{1/2} r_{12}$$

holds on our interval $\delta_0 \leq t \leq \delta_1$. Indeed,

$$(r_{12}^{\epsilon})^2 = (r_{12})^2 + \text{ (nonneg.) } + 2\epsilon f \sigma^p \cdot q_{12}.$$

The last term is bounded by $C\epsilon t$ according to condition (5). On the other hand $r_{12}^2 = ct^{4/3} + O(t^{5/3})$ by Levi-Civita . What we require then is that $t^{4/3} >> \epsilon t$, which will be the case as long as $t^{1/3} >> \epsilon$. Use $\delta_0 = \epsilon^{3/2}$, and $t \geq \delta_0$ to conclude that $t^{1/3} \geq (\epsilon^{3/2})^{1/3} = \epsilon^{1/2} >> \epsilon$ over our interval.

Consequently,

$$\frac{1}{r_{12}} - \frac{1}{r_{12}^{\epsilon}}$$

is either nonnegative, or, if negative satisfies

$$\begin{array}{lll} |Neg\{\frac{1}{r_{12}} - \frac{1}{r_{12}^{\epsilon}}\}| & \leq & |\frac{1}{1} - \frac{1}{\sqrt{1-c\epsilon}}|\frac{1}{r_{12}} \\ & \leq & \frac{k\epsilon}{r_{12}} \\ \leq k_2\epsilon(\delta_0)^{-2/3}, & & \end{array}.$$

In the first line of this inequality "Neg " means the negative part of:

$$Neg\{x\} = min\{0, x\}.$$

In the second and consequent lines of the inequlality k stands for any constant, which can be taken independent of ϵ. Indeed, we will continue this tradition for k through the rest of this section. Thus in going from the second to the third line of the inequality we used the Taylor expansion $\frac{1}{\sqrt{1-c\epsilon}} - 1 = \frac{1}{2}c\epsilon + O(\epsilon^2)$ so that the constant k can be taken to be any number greater than line $\frac{1}{2}c$. And in the third line we again used $r_{12} \geq kt^{2/3}$ and $t \geq \delta_0$.

Finally we get

$$\begin{array}{lll} |Neg\{I_1\}| & \leq & \int_{\delta_0}^{\delta_1} k\epsilon\delta_0^{-2/3} dt \\ & = & k(\delta_1 - \delta_0)\epsilon\delta_0^{-2/3} \\ & = & k\epsilon \end{array}$$

where we used the defining relations between δ_0, δ_1 and ϵ. This proves the desired bound

$$I_1(\epsilon) = \text{nonnegative} + O(\epsilon).$$

4.10 Bounding I_2.

The kinetic difference $K - K^{\epsilon}$ is zero except along the interval $[\delta_0, \delta_1]$ where

$$\begin{array}{lll} K - K^{\epsilon} & = & -\epsilon \dot{f} \sigma^p \cdot \dot{q}_{12} \quad -\frac{1}{2}\epsilon^2(\dot{f})^2|\sigma^p|^2 \\ & = & K_1 \qquad\qquad\quad +K_2. \end{array}$$

This second term is easily bounded. As before, let k be an arbitrary positive constant which is allowed to change from inequality to inequality. Then $|\dot{f}| \leq k/(\delta_1 - \delta_0)$, so that

$$\begin{array}{rcl} \int_{\delta_0}^{\delta_1} |K_2| & \leq & \epsilon^2 \int_{\delta_0}^{\delta_1} k/(\delta_1 - \delta_0)^2 dt \\ & \leq & \epsilon^2 \frac{k}{\epsilon} = k\epsilon, \end{array}$$

where we used $\delta_1 - \delta_0 = \epsilon$. Thus

$$\int K_2 = O(\epsilon).$$

To bound the first term, $\int K_1$, will require more work. We will first need the bound:

$$\sigma^p \cdot \dot{q}_{12} = \text{ nonnegative} + O(1).$$

To prove this write q_{12} in polar coordinates: $q = r\hat{q}$, where $r = r_{12}$ and $\hat{q} = \hat{q}_{12}$, and $\hat{q} = (cos\theta, sin\theta)$ where θ is the angle which $\hat{q}$ makes with the inertial x-axis. It follows that

$$\begin{array}{rcl} \dot{q}_{12} & = & \dot{r}\hat{q} + r\dot{\theta}J(\hat{q}) \\ & = & \frac{\dot{r}}{r}q_{12} + r\dot{\theta}J\hat{q} \end{array}$$

where we have used the fact that $\frac{d\hat{q}}{d\theta} = J\hat{q}$ where J is ninety degree rotation. The Levi-Civita asymptotics $r = O(t^{2/3})$, $\dot{r} = O(t^{-1/3})$ imply that $\frac{\dot{r}}{r} = O(t^{-1})$, which combined with condition (5) yield $\frac{\dot{r}}{r}\sigma_{12}^p \cdot q_{12} = nonneg. + O(1)$. We also have the asymptotics $\dot{\theta} = O(t^{-1/3})$. which yields the bound $r\dot{\theta}\sigma_{12}^p \cdot Je = O(t^{1/3})$ for the second term in $\dot{q}_{12}$. Together these yield the desired bound above on $\sigma^p \cdot \dot{q}_{12}$.

Now $\dot{f} \leq 0$, and $|\dot{f}| \leq k/(\delta_1 - \delta_0) = k/\epsilon$, so that

$$0 \leq -\epsilon\dot{f} \leq k.$$

It follows that

$$-\epsilon\dot{f}\sigma^p \cdot \dot{q} = \text{ nonnegative} + O(1),$$

on the interval in question. Consequently

$$\begin{array}{rcl} |Neg\{\int_{\delta_0}^{\delta_1} K_1\}| & = & |Neg\{\int_{\delta_0}^{\delta_1} [-\epsilon\dot{f}\sigma^p \cdot \dot{q}]dt\} \\ & \leq & \int_{\delta_0}^{\delta_1} kdt \\ & = & k\epsilon. \end{array}$$

This proves that $\int_{\delta_0}^{\delta_1} K_1 \geq O(\epsilon)$. Combined with the same estimate for $\int K_2$ we now have

$$I_2(\epsilon) \geq O(\epsilon).$$

SUMMARY OF THE BINARY PERTURBATION ARGUMENT. The change in action $A - A_\epsilon$ is the sum of the three integrals I_0, I_1 and I_2, ignoring an $O(\epsilon)$ error. We have shown that the last two integrals are greater than or equal to $O(\epsilon)$. The first integral I_0 is greater than or equal to $C\sqrt{\epsilon}$ with C positive. Thus

$$A - A_\epsilon \geq C\sqrt{\epsilon} + O(\epsilon)$$

with $C > 0$. The action has been decreased.

5 Bibliography

References

[1] A. Ambrosetti and V. Zelati, **Periodic Solutions of Singular Lagrangian Systems**, Birkhäuser, 1993.

[2] A. Bahri and P. Rabinowitz, *Periodic solutions of Hamiltonian systems of 3-body type* , Annales Inst. Henri Poincaré, Analyse non linéaire, v. 8, n. 6, 561-649, (1991).

[3] G. Birkhoff, **Dynamical Systems** (esp. ch. 8, §11), American Mathematical Society, 1927.

[4] A. Chenciner and M. Desolneux, *Minima de l'integrale d'action et equilibres relatifs de n corps* C.R.A.S. Paris, S. I : Math, v. 326, 1209-1212, (1998). (erratum: v. 327, 193, (1998).)

[5] A. Deprit and A. Delie, *Régularisation du problème des trois corps*, Arch. for Rat. Mech., v. 12, 325-353, (1962). (see also: *Courbure de la métrique riemannienne du probème plan des trois corps* Bull. Cl. Sc. Acad. Belg. ,v. 48, n. 5, 652-658, (1962).)

[6] W-Y Hsiang, *Geometric Study of the Three-Body Problem I*, preprint report PAM-620, Center for Pure and Applied Math, University California, Berkeley, (1994) ; see also: *Kinematic geometry of mass-triangles and reduction of Schrdinger's equation of three-body systems to partial differential equations solely defined on triangular parameters*, Proc. Nat. Acad. Sci. U.S.A. , v. 94, no. 17, 8936–8938, (1997).

[7] W. Gordon, *A minimizing property of Keplerian orbits*, Am. J. Math. , v. 99,n. 5, 961-971, (1970).

[8] T. Iwai, *A geometric setting for classical molecular dynamics*, Ann. Inst. Henri Poincare, v. 47, n. 2, 199-219. (1987).

[9] N.S. Krylov, **Works on the Foundations of Statistical Mechanics**, Princeton Univ. Press, 1979.

[10] G. Lemaitre, *Coordonnées symétriques dans le problème des trois corps*, Bull. Cl. Sc. Acad. Belg. v. 38, n. 5, 582-592, 1218-1234, (1952).

[11] T. Levi-Civita, *Sur La Régularisation du Problème des Trois Corps*, Acta Math., v. 42, 99-144; (esp. p. 105, eq.(12)), (1921).

[12] P. Majer and S. Terracini, *Periodic solutions to some problems of n-body type*, Arch. Rat. Mech., v. 124 , no. 4, 381-404, (1993).

[13] R. Moeckel, *Some qualitative features of the three-body problem* in **Hamiltonian Dynamical Systems, proceedings, AMS Contemporary Math, Vol. 81,** ed. K. Meyer and C. Saari, 1-22, American Math. Society, (1988).

[14] R. Montgomery, *The geometric phase of the three-body problem*, Nonlinearity, v. 9, 1341-1360, (1996).

[15] R. Montgomery, *The N-body problem, the braid group, and action-minimizing periodic solutions* Nonlinearity, v. 11, 363-376, (1998).

[16] H. Poincaré, **Les Methodes Nouvelles de la Mécanique Céleste** v. 1, ch. 3, Paris, Gauthier-Villars, 1892; (see also D. Goroff's introduction to the English translation , p. I 42, in American Institute of Physics, History of Modern Physics and Astronomy series, v. 13, 1993.)

[17] L. Sbano, *The topology of the planar three body problem with zero total angular momentum and the existence of periodic orbits*, Nonlinearity, v. 11, n. 3, 641-658, (1998).

[18] Small, Christopher G. Small, **The statistical theory of shape,** Springer Series in Statistics. Springer-Verlag, New York, 1996.

[19] K. Sundman, *Mémoire sur Le Problème des Trois Corps*, Acta Math., v. 36, 105-179, (1912).

[20] M Szydlowski, *Curvature of gravitationally bound mechanical systems*, J. Math. Phys., v. 35, n. 4, (1994). (and references therein).

[21] A. Wintner, **The analytical foundations of celestial mechanics**, Princeton Univ. Press, 1941.