JOURNAL OF DIFFERENTIAL EQUATIONS 26, 80–111 (1977)

A Global Existence Theorem for the Four-Body Problem of Newtonian Mechanics*

DONALD G. SAARI

Department of Mathematics, Northwestern University, Evanston, Illinois 60201

Received April 26, 1976

It is shown that for the planar or three-dimensional two-, three-, or four-body problems of celestial mechanics, almost all initial conditions (in the sense of Lebesgue measure and Baire category) lead to solutions which exist for all time.

1. INTRODUCTION AND NOTATION

The main purpose of this paper is to prove the following theorem.

THEOREM 1. *For almost all initial conditions (in the sense of Lebesgue measure and Baire category) in the two-, three-, and four-body problems of Newtonian mechanics, a unique solution exists for all time.*

This theorem holds for a p-dimensional physical space for $p \geqslant 2$, where we are mainly interested in the values $p = 2, 3$. Phase space can be either the full phase space or the reduced phase space obtained by fixing the center of mass at the origin of some inertial coordinate system. (Although we do not carry out the details, this result can be extended to energy manifolds and the algebraic varieties defined by constant angular momentum except for $n = 2$ and zero angular momentum.) Only the domain of existence needs to be established since solutions of the n-body problem are analytic, and hence unique, in their domain of existence.

Previously this type of global existence and uniqueness theorem was known to be true only for the two- and three-body problems. In the two-body problem it is known that the analytic singularities of the solutions are due to collisions between the particles. But it is also known that this system suffers a collision at some time if and only if the angular momentum of the system is zero. Thus the set of initial conditions leading to a collision coincides with the algebraic variety defined by a zero value for the angular momentum. Since, in phase space, this set is of Lebesgue measure zero and of first Baire category, the conclusion follows.

* This research was partially supported by an NSF grant.

ISSN 0022-0396

For $n \geqslant 3$, it cannot be expected, nor is it true, that the collision orbits are restricted to some fixed angular momentum surface. However, for the n-body problem it has been shown by Saari [7–9] that the set of initial conditions leading to collisions is of Lebesgue measure zero and of first Baire category. (Urenko [14] has established a similar statement in a lower-dimensional setting where the initial conditions are restricted to any angular momentum surface.) Thus in any situation where collisions constitute the only singularities of solutions, Theorem 1 follows immediately.

This is what happens for the three-body problem. Painlevé [5] showed for $n = 3$ that the singularities are caused by collisions. However, the situation changes for $n > 3$. Painlevé also showed that a necessary and sufficient condition for a singularity to occur at $t = t_0$ is for the minimum spacing between particles to approach zero as $t \to t_0$. Painlevé pointed out for $n > 3$ that this does not necessarily imply collision. While the existence of a noncollision singularity has not been established, their properties, if they do exist, have been studied. For example, it is known [13, 15] that the existence of a noncollision singularity is equivalent to the system becoming unbounded in physical space in a finite time. Indeed, the system must become unbounded at a "rapid" rate [11]. The existence of such a noncollision singularity is suggested by some recent work of Mather and McGehee [2] where they show, for the four-body problem, that binary collisions can accumulate in such a fashion that the system becomes unbounded in physical space in a finite time.

Therefore, in order to prove Theorem 1, we need to find an estimate on the size of the set of initial conditions which could potentially lead to a noncollision singularity.

THEOREM 2. *In the four-body problem, the set of initial conditions leading to a noncollision singularity lies in a set of Lebesgue measure zero and first Baire category.*

Theorem 1 now follows immediately from Theorem 2 and the fact that collisions are improbable.

Actually, all we need to show is that this set is of Lebesgue measure zero, as the Baire category statement follows immediately by continuity of solutions with respect to initial conditions. To see this, let $NC([n, n+1])$ be the set of initial conditions which suffer their first singularity in the time interval $[n, n+1]$, and it is a noncollision singularity. Then the set described in Theorem 2 is $\bigcup_{n=-\infty}^{\infty} NC([n, n+1])$. If this set were of second Baire category, then for some n, set $NC[n, n+1])$ is of second Baire category. This means that the closure of $NC([n, n+1])$ contains an open set U. Since both NC and the collision initiating set are of measure zero, there is some initial condition $p \in U$ such that its solution exists on the time interval $[0, n+1]$. But by continuity of solutions with respect to initial conditions, there exists an open neighborhood of initial conditions, V, such that their solutions exist in $[0, n+1]$ and $p \in V$. This contradicts the fact that p is in the closure of $NC([n, n+1])$, and it shows that the first Baire category statement follows from the measure zero statement.

Other results follow from the analysis, and they are stated in Section 9. One of the statements asserts the nonexistence of certain types of noncollision singularities for certain mass ratios of the masses. Another provides a lower bound on how fast the system becomes unbounded in finite time when a noncollision singularity is encountered.

The notation is standard. Assume the center of mass of the system is fixed at the origin of an inertial coordinate system. Let m_k, $\mathbf{r}_k$, and $\mathbf{v}_k$ denote, respectively, the mass, position vector, and velocity vector of the kth particle. (The same letter will be used to designate the magnitude of a vector, e.g., $v_k = |\mathbf{v}_k|$, $r_{kj} = |\mathbf{r}_k - \mathbf{r}_j|$.) If the gravitational constant is assumed to be unity, then the equations of motion are

$$m_k\ddot{\mathbf{r}}_k = \sum_{\substack{j=1\\ j\neq k}}^{4} (m_k m_j(\mathbf{r}_j - \mathbf{r}_k)/r_{kj}^3) = \partial U/\partial \mathbf{r}_k\,, \tag{1.1}$$

where

$$U = \sum_{1\leqslant k<j\leqslant 4} m_k m_j/r_{kj}\,.$$

The conservation of energy integral is

$$T = \tfrac{1}{2}\sum_{k=1} m_k v_k^2 = U + h \tag{1.2}$$

where h, the total energy, is a constant of integration. The angular momentum integral is

$$\sum_{k=1}^{4} m_k(\mathbf{r}_k \times \mathbf{v}_k) = \mathbf{c}, \tag{1.3}$$

where vector $\mathbf{c}$ is a constant of integration.

Define $2I = \sum_{k=1}^{4} m_k \mathbf{r}_k^2$. Since the center mass of the system is fixed at the origin, it follows immediately that

$$2MI = \sum_{1\leqslant k<j\leqslant 4} m_k m_j(\mathbf{r}_k - \mathbf{r}_j)^2,$$

where M is the total mass of the system.

The well-known Lagrange–Jacobi relationship which relates I to U is

$$\ddot{I} = U + 2h. \tag{1.4}$$

The importance of this relationship is partially derived from the fact that $I^{1/2}$ can be interpreted as a measure of the maximum spacing between particles, whereas U^{-1} can be viewed as a measure of the minimum spacing between particles; namely, if $R(t) = \max_{k \neq j} r_{kj}(t)$ and $r(t) = \min_{k \neq j} r_{kj}(t)$, then it can easily be shown that

$$m_0^2 R^2 \leqslant 2MI \leqslant (MR)^2/2,$$
$$m_0^2/r \leqslant U \leqslant M^2/2r, \tag{1.5}$$

where m_0 is the smallest mass.

In what follows, A and B will denote positive constants determined by the masses of the particles and other constants of the system. In general, they do not assume the same value with each usage. We shall be discussing several small intervals of time. In each case, the interval will be specified, and we shall use Δt to be the length of this particular interval. Also, if the interval is (t_1, t_2) then $\Delta f = f(t_2) - f(t_1)$. Finally, with the exception of t_0 which always will designate the time of a noncollision singularity, the subscripts on t may change meaning with each section.

2. Preliminary Estimates

The basic idea for the proof of Theorem 2 comes from [10] where essentially it was shown that if there is a noncollision singularity in the four-body problem at time $t = t_0$, then as $t \to t_0$ the commuting of the particles is restricted to a shrinking neighborhood of some fixed line in physical space passing through the origin. That is, it was shown that there exists an infinite sequence $\{t_\alpha\}$, $t_\alpha \to t_0$, such that at $t = t_\alpha$ three of the particles are extremely close together and the remaining particle is approximately distance R away. It turns out that this 3–1 configuration must break up, and some particles, either a singleton or a binary, must enter some small neighborhood of the fourth particle.

The idea behind the proof of Theorem 2 is to derive estimates on how fast these particles must approach this fixed line in physical space. Since this is a conservative system, it is measure preserving; thus these estimates lead to an upper bound on the measure of the initial conditions leading to this type of behavior. The basic idea behind the estimates is to exploit the fact that when some particle or binary is commuting between the remaining particles, its motion must be essentially a straight line. This is so since the acceleration is negligible at large distances. However, the velocities must be very large in order that all of this commuting can be done an infinite number of times in a bounded time interval. Furthermore, since the velocities are very large, some commuting particle or binary must pass very close to the target, or else nothing will return. (Close approaches effect the magnitude of the acceleration.) The actual estimates provide an estimate on the measure of the set of initial conditions leading to this behavior.

For the remainder of this paper, assume that a solution of the four-body problem suffers its first singularity at time $t = t_0$ and that it is a noncollision singularity. (This condition can be relaxed by "regularizing" binary collisions. But since collisions are already known to be improbable and since we are seeking an existence theorem, this complication is not needed here.)

LEMMA 2.1. (a) *As* $t \to t_0$, $r \to 0$ *and* $R \to \infty$.

(b) *If* $V(t) = \max_{k \neq j} v_{kj}(t)$, *then* $4m_0^2/M \leqslant \limsup rV^2 \leqslant M^3/m_0^2$ *as* $t \to t_0$.

(c) *For* $\epsilon > 0$, *there exists sequence* $\{t_\alpha\}$, $t_\alpha \to t_0$, *with the property that at* $t = t_\alpha$ *either three particles are within* ϵ *distance of each other and the remaining particle is at least* $R - \epsilon$ *units away, or there are two binaries no more than* ϵ *distance apart and the distance between the binaries is at least* $R - 2\epsilon$ *units. The first situation will be called a* 3–1 *configuration, while the second will be called a* 2–2 *configuration.*

(d) *There exists a sequence* $\{t_\alpha\}$, $t_\alpha \to t_0$, *such that at* $t = t_\alpha$ *the particles form a* 3–1 *configuration.*

(e) *For* $\delta > 0$ *and for each index* k *there exists index* j *and a sequence* $\{t_\alpha\}$, $t_\alpha \to t_0$, *such that* $r_{kj}(t_\alpha) < (t_0 - t_\alpha)^{1-\delta}$.

From part a, the system becomes unbounded.

From part d, the 3–1 configuration must be formed infinitely often. From part e, this configuration must break up infinitely often, and some particle or particles (at most two) must approach the singleton of the 3–1 configuration. Since the distances these particles must travel to reach this singleton approach infinity (part a), and the time span approaches zero ($t \to t_0$), the magnitudes of the velocities must approach infinity. By part b, these velocities can be no greater than a constant multiple of $r^{-1/2}$.

Proof of Lemma 2.1.

Part a. Painlevé [5] showed that a singularity at $t = t_0$ is equivalent to $r \to 0$ as $t \to t_0$. By [11, 13, 15], if the singularity is a noncollision singularity, then R must be unbounded. However, according to Eqs. (1.4) and (1.5), and since h is a constant,

$$\ddot{I} = U + 2h > A/r \to \infty,$$

so the graph of I is concave as $t \to t_0$. Since I is unbounded, $I \to \infty$. This, in turn, implies $R \to \infty$. Incidently, this proof also shows that I is monotonically increasing near t_0.

Part b. This follows from the conservation of energy integral,

$$h + \frac{M^2}{2r} = h + \frac{1}{r}\sum{}^* m_i m_j > h + U = T = \frac{1}{2M}\sum{}^* m_i m_j (\mathbf{v}_i - \mathbf{v}_j)^2$$

$$\geqslant \frac{1}{2M} m_i m_j V^2 \geqslant \frac{m_0^2}{2M} V^2.$$

The upper bound for lim sup rV^2 now follows. Again using the energy integral, we find that

$$h + \frac{m_0^2}{r} \leqslant h + U = T = \frac{1}{2M}\sum{}^* m_i m_j (\mathbf{v}_i - \mathbf{v}_j)^2 \leqslant \frac{MV^2}{4},$$

which establishes the inequality

$$\liminf rV^2 \geqslant 4m_0^2/M.$$

Part c. Let $\epsilon > 0$ be given. Since $r \to 0$ as $t \to t_0$, after some time $r < \epsilon/3$. We shall restrict attention to the open interval defined by this time and t_0. Since $R \to \infty$, assume that $R > 10\epsilon$. There are two cases to consider. The first is where different particles exchange infinitely often the role of defining distance r. The second case is where after some time some distance between particles agrees with $r(t)$.

In the first case there exists a sequence of times $\{t_\alpha\}$, $t_\alpha \to t_0$, and two mutual distances, say r_{12} and r_{ij}, such that $r_{12}(t_\alpha) = r_{ij}(t_\alpha) = r(t_\alpha)$. If $i, j \neq 1, 2$, then $r_{12}(t_\alpha) = r_{34}(t_\alpha) = r(t_\alpha) < \epsilon/3$. Using the triangle inequality we have that $R - \epsilon < r_{13}, r_{23}, r_{14}, r_{24} \leqslant R$. This is the 2–2 configuration.

If i equals 1 or 2, say 1, then j must be either 3 or 4, say 3. That is, $r_{12}(t_\alpha) = r_{13}(t_\alpha) = r(t_\alpha)$. We find by using the triangle inequality that $r_{23}(t_\alpha) \leqslant r_{13}(t_\alpha) + r_{12}(t_\alpha) < \epsilon$. Since $\mathbf{r}_4$ is the remaining particle and since $R \to \infty$, it follows that the distance from these particles to r_4 is at least $R - \epsilon$. This is the 3–1 configuration. This completes the first case.

In the second case, assume that after some time $r_{12}(t) = r(t)$. We shall now show there exists a sequence $\{t_\alpha\}$, $t_\alpha \to t_0$, such that at time $t = t_\alpha$ the distance from the center of mass m_1 and m_2 to some other particle is bounded above by ϵ. Since $r = r_{12}$, this statement establishes part d.

Assume the assertion is false. Let $\mathbf{C}_{12}$ denote the center of mass of particles $\mathbf{r}_1$ and $\mathbf{r}_2$. It follows from the equations of motion that

$$| M_{12}\ddot{\mathbf{C}}_{12} | = | m_1\ddot{\mathbf{r}}_1 + m_2\ddot{\mathbf{r}}_2 | = \left| \sum_{j=1}^{2} \sum_{i=3}^{4} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j)}{r_{ij}^3} \right| \leqslant A\epsilon^{-2},$$

where $M_{12} = m_1 + m_2$ and A is some positive constant. (Recall our usage of A and B.)

Since $\mathbf{C}_{12} = O(1)$, it follows by integrating this inequality twice and using the Cauchy criterion for the existence of a limit that $\mathbf{C}_{12}$ has a limit as $t \to t_0$. Since $r_{12} = r \to 0$, $\mathbf{r}_1$ and $\mathbf{r}_2$ approach this same limit.

Since $R \to \infty$ as $t \to t_0$, it follows from the above sentence that $\max(r_3, r_4) \to \infty$. However, since the center of mass of the system is fixed at the origin and since r_1 and r_2 approach a limit, it is clear that should either r_3 or r_4 become unbounded, then so must the other and the particles are going in essentially opposite directions.

In order for r_3 to become unbounded, some particle must approach particle 3 infinitely often in an arbitrarily close approach. If this were not the case, then $\ddot{\mathbf{r}}_3 = O(1)$, and r_3 would approach a limit as $t \to t_0$. According to our assumption that the assertion is false, it must be particle 4 which approaches particle 3 in arbitrarily close approaches. By center of mass considerations, whenever this occurs,

$$\mathbf{r}_3, \mathbf{r}_4 = -M_{12}\mathbf{C}_{12}/M_{34} + O(1).$$

Since this must happen infinitely often as $t \to t_0$, we are led to the implication $\liminf R < \infty$ as $t \to t_0$, which violates part a. This completes the proof.

Part d. From the above, we need only consider the case where infinitely often two distances, say r_{12} and r_{34}, exchange the role of defining $r(t)$. Furthermore, we can assume $r(t) < \epsilon/3$. Assume that statement d is false. This means that after some time the distance from either $\mathbf{r}_1$ or $\mathbf{r}_2$ to either $\mathbf{r}_3$ or $\mathbf{r}_4$ is bounded below by $\epsilon/2$. It is easy to see this must be the situation since there are only four particles, $r \to 0$ (so at all values of time at least one distance *must* be small), $R \to \infty$, and 3–1 configurations are not permitted by assumption.

Denoting the center of mass of particles m_1 and m_2 by $\mathbf{C}_{12}$, and of particles m_3 and m_4 by $\mathbf{C}_{34}$, it follows by an argument similar to the one used above that $\ddot{\mathbf{C}}_{12}, \ddot{\mathbf{C}}_{34} = O(1)$, and that they approach limits as $t \to t_0$. However, the assumption that infinitely often these distances trade the role of defining r, establishes the existence of t arbitrarily close to t_0 such that $r_{12}(t) = r_{34}(t) = r(t)$. This implies that $\mathbf{r}_1$ and $\mathbf{r}_2$ are close to $\mathbf{C}_{12}$, and $\mathbf{r}_3$ and $\mathbf{r}_4$ are close to $\mathbf{C}_{34}$, or that $\liminf R < \infty$ as $t \to t_0$. This contradiction proves part d of the lemma.

Part e. Assume false. Then there is some $0 < \delta$ and some index, say 4, such that $r_{34}(t) > (t_0 - t)^{1-\delta}$. This implies that $\ddot{\mathbf{r}}_4 = O((t_0 - t)^{-2(1-\delta)})$. By integrating both sides of this relationship twice, it follows that $\mathbf{r}_4 = O(1)$ as $t \to t_0$. This contradicts part d, establishing part e.

3. Commuting Particles

In this section we shall obtain some estimates on the behavior of the particles during the time period when they change from a 3–1 to either another 3–1 or a 2–2 configuration. The first lemma studies the behavior of the center of mass of the natural subsystems formed by these configurations. Extending the notation used in the preceding section, for set of indices S, define $M_S = \sum_{j\in S} m_j$, and $M_S\mathbf{C}_S = \sum_{j\in S} m_j\mathbf{r}_j$.

LEMMA 3.1. *If on the time interval* (t_1, t_2), $k \notin S$ *implies* $r_{kj} > AR(t)$ *for any* $j \in S$, *where* A *is some positive constant, then*

$$\dot{\mathbf{C}}_S(t) = \dot{\mathbf{C}}_S(t_1) + O((t_2 - t_1)/R_m{}^2) \tag{3.1}$$

and

$$\mathbf{C}_S(t) = \dot{\mathbf{C}}_S(t_1)(t - t_1) + \mathbf{C}_S(t_1) + O(((t_2 - t_1)/R_m)^2) \tag{3.2}$$

for $t \in (t_1, t_2)$. $R_m = \min_{[t_1,t_2]} R(t)$.

Proof. According to the definition of $\mathbf{C}_S$,

$$M_S\ddot{\mathbf{C}}_S = \sum_{j\in S} m_j\ddot{\mathbf{r}}_j = \sum_{\substack{k,j\in S\\ k\neq j}} \frac{m_j m_k(\mathbf{r}_k - \mathbf{r}_j)}{r_{kj}^3} + \sum_{\substack{j\in S\\ k\notin S}} \frac{m_j m_k(\mathbf{r}_k - \mathbf{r}_j)}{r_{kj}^3}.$$

The first double summation vanishes by the symmetry of the scalar terms and the antisymmetry of the vector terms with respect to subscripts. By hypothesis, $k \notin S$ implies $r_{kj}(t) > AR(t)$ for any $j \in S$, or

$$\ddot{\mathbf{C}}_S = O(R^{-2}(t)) = O(R_m^{-2}). \tag{3.3}$$

The conclusion follows by integrating both sides of this relationship.

Since the center of mass is at $\mathbf{0}$, since $R \to \infty$, and since three of the particles are within ϵ distance of each other infinitely often, it follows that *all* the r_i become unbounded. As shown in the previous section, the driving force which permits this unbounded motion in finite time is caused by the close approaches of particles when they commute to form the various 3–1 or 2–2 configurations. For the particles to do all this commuting infinitely often over distances which are becoming infinitely large and over time intervals which are becoming arbitrarily small, the velocities must approach infinity. From the conservation of energy integral and the definition of r, it follows that r approaches zero quite rapidly. The next lemma provides asymptotic bounds on these rates of expansion.

LEMMA 3.2. *After some value of time*

$$I \leqslant BV^2(t_0 - t)^2 \tag{3.4}$$

and

$$rR^2 \leqslant B(t_0 - t)^2; \tag{3.5}$$

where positive constant B *depends on the masses and exceeds* $(20M^3/m_0)^2$ *in value.*

Proof. Inequality (3.5) follows from inequality (3.4) by using part b of Lemma 2.1 and Eq. (1.5).

Assume the lemma is false; that is, assume the existence of a sequence $\{t_a\}$, $t_a \to t_0$, such that at time $t = t_\alpha$ inequality (3.4) is violated.

Let $1 > \epsilon > 0$. Define $t_{\alpha,1} = \inf\{t \mid t \geqslant t_\alpha$; at time t the particles form either a 3–1 or a 2–2 configuration$\}$. At time $t_{\alpha,1} \geqslant t_\alpha$, two clusters of particles are defined.

Define $t_{\alpha,2} = \sup\{t \mid t < t_\alpha$; at time t at least one particle from one of the clusters in the configuration resulting from the definition of $t_{\alpha,1}$ is within distance 1 of some particle from the other cluster$\}$.

Clearly $t_{\alpha,2} < t_{\alpha,1}$, and during the time interval $[t_{\alpha,2}, t_{\alpha,1}]$ the distances between particles in different clusters is bounded below by unity. The existence of values of $t_{\alpha,1}$ and $t_{\alpha,2}$ is immediate for t_α sufficiently close to t_0. Finally, it is clear we can assume sequence $\{t_\alpha\}$ is such that $t_{\alpha-1} \leqslant t_{\alpha,2}$. (If not, then replace $\{t_\alpha\}$ with a subsequence which does possess this property.)

Let S_1 and S_2 be the sets of indices defining the two clusters at time $t_{\alpha,1}$. Then by the definition of the involved terms

$$\tfrac{1}{2} \sum_{p=1}^{2} \sum_{j \in S_p} m_j(\mathbf{r}_j - \mathbf{C}_{S_p})^2 = I - \tfrac{1}{2} \sum M_{S_p} \mathbf{C}_{S_p}^2 = I - J. \tag{3.6}$$

By the definition of the 3–1 and 2–2 configurations, at time $t = t_{\alpha,1}$, we have

$$| \mathbf{r}_j - \mathbf{C}_{S_p} | \leqslant \epsilon, \qquad j \in S_p, \quad p = 1, 2.$$

This forces the left-hand side of Eq. (3.3) to be bounded above by $M\epsilon^2/2$. Consequently,

$$I(t_{\alpha,1}) - M\epsilon^2/2 \leqslant J(t_{\alpha,1}). \tag{3.7}$$

On the other hand, at time $t = t_{\alpha,2}$ two particles from some cluster, say $\mathbf{r}_1$ and $\mathbf{r}_2$ where $1, 2 \in S_1$, are at least $R/2$ units apart. From the inequality $2(a^2 + b^2) \geqslant (a - b)^2$, it follows that

$$m_1(\mathbf{r}_1 - \mathbf{C}_{S_1})^2 + m_2(\mathbf{r}_2 - \mathbf{C}_{S_1})^2 \geqslant m_0(\mathbf{r}_1 - \mathbf{r}_2)^2/2 \geqslant m_0 R^2(t)/8.$$

It now follows from inequality (1.5) that at time $t = t_{\alpha,2}$ the right-hand side of Eq. (3.6) is bounded below by $m_0 I(t_{\alpha,2})/4M$. Thus since I is monotonically increasing (proof of Lemma 2.1, part a), we find that

$$J(t_{\alpha,2}) \leqslant (1 - (m_0/4M))\, I(t_{\alpha,2}) < (1 - (m_0/4M))\, I(t_{\alpha,1}). \tag{3.8}$$

If I is sufficiently large (that is, t_α is sufficiently close to t_0), the combination of inequalities (3.7) and (3.8) implies that

$$M\left\{\sum_{p=1} (\mathbf{C}^2_{S_p}(t_{\alpha,1}) - \mathbf{C}^2_{S_p}(t_{\alpha,2})\right\}\Big/2$$
$$> J(t_{\alpha,1}) - J(t_{\alpha,2}) > I(t_{\alpha,1}) - (M\epsilon^2/2) - (1 - m_0/4M)I(t_{\alpha,1})$$
$$> m_0 I(t_{\alpha,1})/5M. \tag{3.9}$$

This string of inequalities forces one of the terms on the left-hand side, say for $p = 1$, to satisfy the inequality

$$\mathbf{C}^2_{S_1}(t_{\alpha,1}) - \mathbf{C}^2_{S_1}(t_{\alpha,2}) \geqslant m_0 I(t_{\alpha,1})/5M^2.$$

Since $m_0\mathbf{C}_1{}^2(t_{\alpha,1})/2 \leqslant I(t_{\alpha,1})$ (this follows from Eq. (3.6)) we have the inequality

$$|\,\mathbf{C}_{S_1}(t_{\alpha,1}) - \mathbf{C}_{S_1}(t_{\alpha,2})| \geqslant (m_0 I(t_{\alpha,1}))^{1/2}/10M^2. \tag{3.10}$$

We now demonstrate that should inequality (3.4) be repeatedly violated, this would lead to a contradiction of Eq. (3.10).

Following the ideas of the proof of Lemma 3.1, and using the fact that $r_{kj}(t) \geqslant 1$ for $t \in [t_{\alpha,2}, t_{\alpha,1}]$, $k \in S_1$, $j \in S_2$, we have that

$$|\,\ddot{\mathbf{C}}_{S_1}| \leqslant M,$$

or that

$$|\,\dot{\mathbf{C}}_{S_1}(t)| \leqslant |\,\dot{\mathbf{C}}_{S_1}(t_\alpha)| + M(t_0 - t_{\alpha-1}).$$

(Recall, $t_{\alpha-1} < t_{\alpha,2}$.)

The assumption that inequality (3.1) is violated at time $t = t_\alpha$ and the definition of $\dot{\mathbf{C}}_{S_1}$, combined with the above inequality shows that for all $t \in [t_{\alpha,2}, t_{\alpha,1}]$ the inequality

$$|\,\dot{\mathbf{C}}_{S_1}(t)| \leqslant M((I(t_{\alpha,1})/B(t_0 - t_{\alpha-1})^2)^{1/2} + (t_0 - t_{\alpha-1}))$$

holds.

According to this inequality, Eq. (3.10), and the mean value theorem, there is some $\xi \in [t_{\alpha,2}, t_{\alpha,1}]$ such that

$$\frac{(m_0 I(t_{\alpha,1}))^{1/2}}{10M^2} \leqslant |\,\mathbf{C}_1(t_{\alpha,1}) - \mathbf{C}_1(t_{\alpha,2})|$$
$$= |\,\dot{\mathbf{C}}_1(\xi)|\,(t_{\alpha,1} - t_{\alpha,2}) \leqslant \frac{MI^{1/2}(t_{\alpha,1})}{B^{1/2}} + M(t_0 - t_{\alpha-1})^2.$$

Since $M/B^{1/2} < m_0/20M^2$, the sought after contradiction can be obtained by choosing $t_{\alpha-1}$ sufficiently close to t_0. This completes the proof of the lemma.

4. Configurations

The next lemma provides a refined picture of the evolving 2–2 configurations. Essentially it states that the "lighter" particles do the commuting and if 2–2 configurations appear, then we can expect the lighter particle to approach the singleton and return to the binary. (Although stronger statements are possible, we shall not derive them since they detract from our main goal of establishing Theorem 2.)

LEMMA 4.1. *If a noncollision singularity exists, and 2–2 configurations occur infinitely often, then the values of the masses satisfy the relationship*

$$m_4 > m_3 \qquad \text{and} \qquad m_1 + m_2 > m_3. \tag{4.1}$$

Actually

$$(m_1 + m_2)/m_3 > (m_3 + m_4)/m_4. \tag{4.2}$$

If the 2–2 configuration occurs infinitely often, then it evolved from a 3–1 configuration, and it will evolve into a 3–1 configuration. Furthermore, the particle commuting to the singleton will return to the binary, forming again a 3–1 configuration, and it is the lighter of the two masses. The masses must satisfy Eq. (4.1) for some choice of the indices.

Proof. The only way a 2–2 configuration can arise is if it came from a 3–1 configuration. This means a particle commutes from a binary to the singleton. We shall show that this original binary must remain intact. This means that one particle from the newly formed binary (the singleton and the particle which approaches it) must commute back to the original binary.

Assume that particles 1 and 2 form the original binary; that is, the intact binary when the 3–1 configuration evolves into a 2–2 configuration. Let $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$, and define $h_{12} = \frac{1}{2}v_{12} - ((m_1 + m_2)/r_{12})$. The goal is to show that r_{12} must remain bounded above by a small positive constant, at least until some other particle enters some small neighborhood of this binary. We shall do this by showing in the time interval when the 2–2 configuration is being formed and until some particle returns to this binary, that $|h_{12}|$ is "large" but h_{12} is negative. This establishes the claim since $0 \leqslant ((m_1 + m_2)/r_{12}) + h_{12}$, or $r_{12} \leqslant (m_1 + m_2)/|h_{12}|$.

First we shall show that at some time h_{12} assumes a large negative value. According to the definition of the terms and the conservation of energy integral,

$$h = T - U = \frac{1}{2}\sum_{i=1}^{2} m_i(\dot{\mathbf{r}}_i - \dot{\mathbf{C}}_{12})^2 + \frac{1}{2} M_{12}\dot{\mathbf{C}}_{12}^2 + \sum_{j=3}^{4} m_j v_j^2 - U$$

$$= \frac{1}{2}\frac{m_1 m_2 v_{12}^2}{m_1 + m_2} + \frac{1}{2} M_{12}\dot{\mathbf{C}}_{12}^2 + \sum_{j=3}^{4} m_j v_j^2 - U$$

$$= \frac{m_1 + m_2}{m_1 m_2} h_{12} + \frac{1}{2} M_{12}\dot{\mathbf{C}}_{12}^2 + \sum_{j=3}^{4} m_j v_j^2$$

$$- \left[\frac{m_3 m_4}{r_{34}} + \frac{m_1 m_4}{r_{14}} + \frac{m_2 m_4}{r_{24}} + \frac{m_1 m_3}{r_{13}} + \frac{m_2 m_3}{r_{23}}\right].$$

Since one particle, say particle 3, is commuting to the singleton, at some time t_1 all the distances in the brackets are bounded below by unity. Using the fact that h is a constant, we have at time t_1 that

$$\left| \frac{m_1 + m_2}{m_1 m_2} h_{12} + \frac{1}{2} M_{12}\dot{\mathbf{C}}_{12}^2 + \sum_{j=3}^{4} m_j v_j^2 \right| < A \tag{4.3}$$

where according to our notation A is a constant depending on the masses and the value of h.

Equation (3.10), Lemma 3.10, and the mean value theorem imply that either $|\dot{\mathbf{C}}_{12}(t_1)|$ or $|\dot{\mathbf{C}}_{34}(t_1)|$ are bounded below by some multiple (depending on the masses) of $R(t_1)/(t_0 - t_1)$. Since $M_{12}\mathbf{C}_{12}(t) = -M_{34}\mathbf{C}_{34}(t_1)$ this is true for both of the terms. Substituting this estimate into Eq. (4.3) implies for t_1 sufficiently close to t_0, that

$$h_{12}(t_1) < -B(R(t_1)/(t_0 - t_1))^2. \tag{4.4}$$

This completes the proof of the assertion.

Before studying how h_{12} can vary, the equation for r_{12} will be derived. From the equations of motion we have

$$\ddot{\mathbf{r}}_{12} = \ddot{\mathbf{r}}_2 - \ddot{\mathbf{r}}_1 = -\frac{(m_1 + m_2)\,\mathbf{r}_{12}}{r_{12}^3} + m_3\mathbf{r}_3\left(\frac{1}{r_{12}^3} - \frac{1}{r_{31}^3}\right) + m_4\mathbf{r}_4\left(\frac{1}{r_{42}^3} - \frac{1}{r_{41}^3}\right)$$

$$+ m_3\left(\frac{\mathbf{r}_1}{r_{13}^3} - \frac{\mathbf{r}_2}{r_{32}^3}\right) + m_4\left(\frac{\mathbf{r}_1}{r_{14}^3} - \frac{\mathbf{r}_2}{r_{42}^3}\right). \tag{4.5}$$

Define $\boldsymbol{\rho}_j$ to be the vector from the center of mass of the binary formed by particles 1 and 2 to particle j. Then $r_{kj}^{-3} = \rho_j^{-3}(1 + O(r_{12}/\rho_j))$ for $r_{12} < \rho_j$ and $k = 1$ or 2.

Thus, for $r_{12} < \rho_j$, Eq. (4.3) can be expressed as

$$\ddot{\mathbf{r}}_{12} = -\frac{(m_1 + m_2)\,\mathbf{r}_{12}}{r_{12}^3} + O\left(\frac{r_{12}}{\rho_3^3}\right) + O\left(\frac{r_{12}}{\rho_4^3}\right). \tag{4.6}$$

We can now investigate changes in the value of h_{12}. Differentiating both sides of the defining equation and using Eq. (4.6) yields

$$\dot{h}_{12} = \dot{\mathbf{r}}_{12}\left(\ddot{\mathbf{r}}_{12} + \frac{M_{12}\mathbf{r}_{12}}{r_{12}^3}\right) = O\left(\frac{r_{12}\mid \dot{\mathbf{r}}_{12}\mid}{\rho_3{}^3}\right) + O\left(\frac{r_{12}\mid \dot{\mathbf{r}}_{12}\mid}{\rho_4{}^3}\right)$$
$$= O(r_{12}\mid \dot{r}_{12}\mid).$$

Since $2r_{12}^2 h_{12} = (r_{12}v_{12})^2 - 2M_{12}r_{12}$, in some neighborhood of t_1, $r_{12}v_{12} = O(r_{12}\mid h_{12}\mid^{1/2})$ or $|(d/dt)(h_{12})^{1/2}\mid = O(r_{12})$.

Therefore, until either particle 3 or 4 approaches within distance unity of some particle in the binary 1 and 2, or until $h_{12} = -BR(t_1)$, we have that

$$|(d/dt)(h_{12})^{1/2}\mid = O(r_{12}) = O(R^{-1}(t_1)).$$

However, function h_{12} *cannot* equal $BR(t_1)$ without some particle approaching the binary. To see this, note that the above relationship bounds the changes in $h^{1\ 2}$, and consequently the changes in $h_{12}^{1/2}$, namely,

$$\Delta(h_{12}^{1/2}) = O((t_0 - t_1)\, R^{-1}(t_1)).$$

For t_1 sufficiently close to t_0, this is too small of a change to allow a change in h_{12} from the value given in Eq. (4.4) to that of $-BR(t_1)$. Thus, the condition $R \to \infty$ and a center of mass argument, similar to that found in the proof of Lemma 2.1, implies that either particle 3 or particle 4 returns to a neighborhood of the binary.

Assume that particle 3 leaves the binary and approaches the singleton particle 4. We know from the above that one of these two particles must return to the binary. What we establish next is that it will be particle 3 and that $m_3 < m_4$.

From the equations of motion, if on some time interval of length Δt, $(t_1, t_1 + \Delta t)$, the distance from a given particle k to any other particle is bounded below by unity, then $\ddot{\mathbf{r}}_k = O(1)$, $\dot{\mathbf{r}}_k(t) = \dot{\mathbf{r}}_k(t_1) + O(\Delta t)$ and

$$\mathbf{r}_k(t) = \mathbf{r}_k(t_1) + \dot{\mathbf{r}}_k(t_1)(t - t_1) + O(\Delta t)^2). \tag{4.7}$$

That is, if Δt is small, then the motion of $\mathbf{r}_k(t)$ is restricted to a small neighborhood of a straight line. In our case $k = 3$, t_1 is the exit time when either $r_{13}(t_1)$ or $r_{23}(t_1)$ equals unity, and $t_1 + \Delta t$ is the approach time when r_{34} equals unity. During this time interval the motions of $\mathbf{C}_{34}$ and $\mathbf{C}_{12}$ are restricted to a small neighborhood of a line ($M_{12}\mathbf{C}_{12} = -M_{34}\mathbf{C}_{34}$, and Lemma 3.1). It turns out that these two lines can be chosen to agree with each other.

To see this, note that at time t_1, the initial conditions establishing the lines approximating the motion of $\mathbf{C}_{123}$, $\mathbf{C}_{12}$, $\mathbf{C}_{34}$, and $\mathbf{r}_3$ are determined. Also, at $t = t_1$ there are some obvious inequalities, namely, $\mathbf{C}_{123}$, $\mathbf{C}_{12}$, and $\mathbf{r}_3$ are all within distance unity of each other. At time $t_1 + \Delta t$, vectors $\mathbf{r}_3$, $\mathbf{r}_4$, and

$\mathbf{C}_{34} = -M_{12}M_{34}^{-1}\mathbf{C}_{12}$ are within distance unity of each other. It is now an elementary task to establish the existence of a line L such that all the motion lies within distance unity of it. Also, since Δt is small, R is large, the directions of the velocities are also approximately given (except for orientation) by L.

During the binary encounter of particles 3 and 4, $\ddot{\mathbf{C}}_{34} = O(R^{-2})$, so a similar argument shows that line L serves as an approximation to the motion at least until some particle returns to a small neighborhood of binary 1, 2. (Of course, it may not define the velocity directions during a close encounter.)

Define the positive direction, or "right-hand side" of L by $\dot{\mathbf{C}}_{34}(t_1)$.

Equation (3.10), the inequality preceding it, and the fact that $M_{12}\mathbf{C}_{12} = -M_{34}\mathbf{C}_{34}$ show that $\dot{\mathbf{C}}_{34} > 0$. So particle 3 is coming from the binary on the left and going to the right.

Let ϵ be a small positive constant, and let t^* be the value of time when particle 3 is approaching particle 4 where $r_{34}(t^*) = M_{34}\epsilon^{-1}$. Let t^1 be the value of time when one of the two particles is returning to the binary where $r_{34}(t^1) = M_{34}\epsilon^{-1}$. Since $m_3\mathbf{v}_3 + m_4\mathbf{v}_4 = M_{34}\dot{\mathbf{C}}_{34}$ and since the velocity of the returning particle is directed to the left (when projected on L), the velocity of the remaining particle must be directed to the right. By Eq. (4.7), the velocity will remain essentially the same until some particle returns within distance unity of it.

This is a center of mass argument, and it holds independently of the particles involved. By assuming t^* is sufficiently close to t_0, it follows from Lemma 2.1 that at some time previous to t^* some other particle or binary left a neighborhood of particle 4. Therefore, the projection of $\mathbf{v}_4$ on L at $t = t^*$ is directed to the right.

This means,

$$m_3 v_3(t^*) + m_4 v_4(t^*) = m_{34}\,|\,\dot{\mathbf{C}}_{34}(t^*)| + O(1). \tag{4.8}$$

If at time t^1, particle 3 is returning to the binary, then

$$-m_3 v_3(t^1) + m_4 v_4(t^1) = M_{34}\,|\,\dot{\mathbf{C}}_{34}(t^*)| + O(1). \tag{4.9a}$$

If particle 4 is returning to the binary then

$$-m_4 v_4(t^1) + m_3 v_3(t^1) = M_{34}\,|\,\dot{\mathbf{C}}_{34}(t^*)| + O(1). \tag{4.9b}$$

Next we need an estimate on $v_{34}(t^1)$ and $v_{34}(t^*)$. This is obtained by studying h_{34}. So, define $h_{34} = \frac{1}{2}v_{34}^2 - (M_{34}/r_{34})$. Distance r_{34} is very large when particle 3 is halfway from the binary to particle 4. At this time h_{34} is given by $\frac{1}{2}v_{34}^2$ plus small error terms. Since v_3 and v_4 remain essentially constants (Eq. (4.7)) until a close approach, since $\mathbf{v}_4(t^*)$ is directed to the right, and since particle 3 must commute at least the distance $R(t^*)$, we have that

$$h_{34}(t^*) > Av_{34}^2 > A(R(t^*)/(t_0 - t^*))^2$$

for some positive constant A depending on masses. Using arguments similar to those following Eq. (4.6), it follows that $(d/dt)(h_{34}^{1/2}) = O(R(t^*)^{-5/2})$.

Consequently, $h_{34}^{1/2}$ changes very little in value during this time interval. This in turn implies

$$v_{34}(t^*) = v_{34}(t^1) + O(\epsilon).$$

At time t^*, both $\mathbf{v}_3$ and $\mathbf{v}_4$ are directed to the right. So, $v_{34}(t^*) = v_3(t^*) - v_4(t^*) + O(1)$. At time t^1, one particle is returning to the binary and its velocity is directed to the left; the other particle's velocity is directed to the right. This means $v_{34}(t^1) = v_3(t^1) + v_4(t^1) + O(1)$.

The combination of these two equations yields

$$v_3(t^*) = v_4(t^*) + v_3(t^1) + v_4(t^1) + O(1). \tag{4.10}$$

If particle 4 returns to the binary, then from Eqs. (4.8), (4.9b), and (4.10) we find that

$$m_3(v_3(t^*) - v_3(t^1)) = -m_4(v_4(t^*) + v_4(t^1)) + O(1)$$

or

$$m_3/m_4 = -1 + O(1/(v_4(t^*) + v_4(t^1))) < 0.$$

This contradiction means that particle 3 must return to the binary. According to Eqs. (4.8), (4.9a), (4.10), we have

$$m_3(v_3(t^*) + v_3(t^1)) = m_4(v_4(t^1) - v_4(t^*)) + O(1)$$

or

$$\frac{m_3}{m_4} = \frac{v_4(t^1) - v_4(t^*)}{v_4(t^1) + v_4(t^*) + v_3(t^1)} + O\left(\frac{1}{v_3(t^*) + v_3(t^1)}\right). \tag{4.11}$$

This means $m_3 < m_4$.

Notice that the above argument combined with Eqs. (3.9), (3.10) show that $R(t)$ is equal to a monotonically increasing function plus terms $O(r_{12})$ during this transit time. Eq. 4.2 follows from a similar center of mass argument using the fact that the singleton must catch the binary.

5. The 2–2 Configuration

Sharper estimates are needed on the behavior of particle 3 relative to particle 4 during the time interval when a 2–2 configuration is being formed. The basic idea is to approximate this behavior by a two-body Kepler motion. If $\mathbf{r}_{43} = \mathbf{r}_3 - \mathbf{r}_4$, then the goal is to estimate the normal component of $\mathbf{v}_{34}$ during the period of time that particle 3 is commuting to particle 4. The way we do this is to find a bound on the magnitude of $\mathbf{c}_{43} = \mathbf{r}_{43} \times \mathbf{v}_{43}$.

The equations of motion for $\mathbf{r}_{43}$ are

$$\ddot{\mathbf{r}}_{43} = \frac{-M_{34}\mathbf{r}_{43}}{r_{34}^3} + \left(\frac{m_1(\mathbf{r}_1 - \mathbf{r}_3)}{r_{13}^3} + \frac{m_2(\mathbf{r}_2 - \mathbf{r}_3)}{r_{23}^3}\right) - \left(\frac{m_1(\mathbf{r}_1 - \mathbf{r}_4)}{r_{14}^3} + \frac{m_2(\mathbf{r}_2 - \mathbf{r}_4)}{r_{24}^3}\right).$$

Since $r_{1j} = r_{2j} + O(r_{12})$, for $r_{1j}^{-3}(1 + O(r_{12}/r_{2j}))$ for $j = 3, 4$, the right-hand side of the above equation becomes

$$-\frac{M_{34}\mathbf{r}_{43}}{r_{43}^3} + \left(\frac{M_{12}(\mathbf{r}_2 - \mathbf{r}_3)}{r_{23}^3}\right) + \left(\frac{M_{12}(\mathbf{r}_2 - \mathbf{r}_4)}{r_{24}^3}\right) + O(r_{12}r_{23}^{-3}) + O(r_{12}r_{24}^{-3}).$$

But $\mathbf{r}_3 - \mathbf{r}_2 = \mathbf{r}_{23} = \mathbf{r}_{24} + \mathbf{r}_{43}$.

So

$$r_{23} = r_{24}\left(1 - 2\,\frac{\mathbf{r}_{24}\cdot\mathbf{r}_{43}}{r_{24}^2} + \left(\frac{r_{43}}{r_{24}}\right)^2\right)^{1/2},$$

or

$$r_{23}^{-3} = r_{24}^{-3}\left(1 + O\left(\left(\frac{\mathbf{r}_{24}\cdot\mathbf{r}_{43}}{r_{24}^2}\right) + \left(\frac{r_{43}}{r_{24}}\right)^2\right)\right).$$

Therefore,

$$\ddot{\mathbf{r}}_{43} = \frac{-M_{34}\mathbf{r}_{43}}{r_{43}^3} - \frac{M_{12}\mathbf{r}_{43}}{r_{24}^3} + \frac{\mathbf{r}_{42} + \mathbf{r}_{34}}{r_{24}^3}\,O\left(\left(\frac{\mathbf{r}_{24}\cdot\mathbf{r}_{43}}{r_{24}^2}\right) + \left(\frac{\mathbf{r}_{43}}{r_{24}}\right)^2\right). \quad (5.1)$$

Let t^* denote the time when particle 3 is halfway between the binary and particle 4, and let $t^* + \Delta t$ be the value of time when particle 3 is halfway on the return trip to the binary.

The angular momentum is given by

$$\mathbf{c}_{43} = \mathbf{r}_{43} \times \dot{\mathbf{r}}_{43}. \quad (5.2)$$

So, from Eq. (5.1) it follows that

$$\dot{\mathbf{c}}_{43} = \mathbf{r}_{43} \times \ddot{\mathbf{r}}_{43} = ((\mathbf{r}_{43} \times \mathbf{r}_{42})/r_{24}^3)\,O(\mathbf{r}_{43}/r_{24}) = O(R^{-1}(t^*)).$$

That is,

$$\Delta c_{43} = O(\Delta t/R(t^*)). \quad (5.3)$$

The next important element is the eccentricity, where we follow Pollard [6, Chap. 1] and define vector $\mathbf{e}_{43}$ as

$$M_{34}((\mathbf{r}_{43}/r_{43}) + \mathbf{e}_{43}) = \mathbf{v}_{43} \times \mathbf{c}_{43} . \tag{5.4}$$

It follows from [6, p. 35] that $\dot{e}_{43} = c_{43}O(1/R^2(t^*))$, or

$$\Delta e_{43} = c_{43}O(\Delta t/R^2(t^*)). \tag{5.5}$$

We now show that c_{43} is bounded above by constant D where

$$D = (10(|\cos 80°|^{-1} + 1)/M_{34})^{1/2}.$$

Assume this statement is false. It follows from Eq. (5.4) that

$$M_{34}(r_{43}(1 + e_{43}\cos f)) = \mathbf{r}_{43} \cdot \mathbf{v}_{43} \times \mathbf{c}_{43} = \mathbf{r}_{43} \times \mathbf{v}_{43} \cdot \mathbf{c}_{43} = c_{43}^2 ,$$

where f is the angle between vectors $\mathbf{e}_{43}$ and $\mathbf{r}_{43}$.

So,

$$r_{43} = c_{43}^2 M_{43}^{-1}/(1 + e_{43}\cos f). \tag{5.6}$$

Since c_{43} remains essentially a constant (Eq. (5.3) and the assumption c_{43} exceeds D), r_{43} attains its minimum value near the time when $1 + e_{34}\cos f$ attains its maximum value. But r_{34} must be less than unity at some time t', which implies that

$$e_{34}(t') > c_{43}^2 M_{34}^{-1} - 1.$$

Consequently, from Eq. (5.5), and by assuming t^* is close enough to t_0 so that $|\Delta e_{43}| < (5M_{34})^{-1}c_{43}$, we have

$$e_{43}(t) > c_{43}^2 M_{43}^{-1} - 1 - (5M_{34})^{-1} c_{43} > |1/\cos 80°|.$$

This in turn means that the domain of f is contained in $[-100°, 100°]$ (this follows from the condition $1 + e_{43}\cos f \geqslant 0$) which means that particle 3 cannot return to a neighborhood of the binary *provided* vector $\mathbf{e}_{43}/e_{43}$ remains essentially the same in the time interval $(t^*, t^* + \Delta t)$. But

$$\frac{d}{dt}\frac{\mathbf{e}_{43}}{e_{43}} = \frac{\dot{\mathbf{e}}_{43}e_{43}^2 - (\mathbf{e}_{43} \cdot \dot{\mathbf{e}}_{43})\,\mathbf{e}_{43}}{e_{43}^3} = \frac{(\mathbf{e}_{43} \times \dot{\mathbf{e}}_{43}) \times \mathbf{e}_{43}}{e_{43}^3} .$$

Therefore, $|(d/dt)(\mathbf{e}_{43}/e_{43})| \leqslant |\dot{\mathbf{e}}_{43}/e_{43}|$.

However, it follows from Eqs. (5.4) and (5.1) that

$$\begin{aligned} M_{34}\dot{\mathbf{e}}_{43} &= \ddot{\mathbf{r}}_{43} \times \mathbf{c}_{43} + \mathbf{v}_{43} \times \dot{\mathbf{c}}_{43} + \frac{M_{43}\mathbf{r}_{43}}{r_{43}^3} \times \mathbf{c}_{43} \\ &= \mathbf{v}_{43} \times \dot{\mathbf{c}}_{43} + O\left(\frac{c_{43}r_{43}}{r_{24}^3}\right) \\ &= \frac{\mathbf{v}_{43} \times (\mathbf{r}_{43} \times \mathbf{r}_{42})}{r_{24}^3} + O\left(\frac{c_{43}r_{43}}{r_{24}^3}\right) \\ &= O\left(\frac{r_{43}v_{43}}{r_{24}^2}\right) + O\left(\frac{c_{43}r_{43}}{r_{24}^3}\right) \\ &= O\left(\frac{r_{43}^{1/2}h_{43}^{1/2}}{r_{24}^2}\right) + O\left(\frac{c_{43}r_{43}}{r_{24}^3}\right). \end{aligned}$$

Thus,

$$\left| \frac{d}{dt}\frac{\mathbf{e}_{43}}{e_{43}} \right| = O\left(\frac{r_{43}^{1/2}h_{43}^{1/2}}{r_{24}^2}\right) + O\left(\frac{r_{43}}{c_{43}r_{24}^3}\right)$$

and

$$\Delta\frac{\mathbf{e}_{43}}{e_{43}} = O\left(\frac{r_{43}}{c_{43}r_{24}^3}\Delta t\right) + O\left(\int_{t^*}^{t_2=t^*+\Delta t} \frac{h_{43}^{1/2}}{r_{24}^{3/2}}\,dt\right).$$

The first term on the right can be made arbitrarily small by choosing t^* sufficiently close to t_0.

To eliminate the second term, notice that

$$h_{43}^{1/2} = O(v_3(t^*)) = O(\dot{C}_{43}).$$

A simple energy argument using $h_{43}^{1/2}$, and Eq. (4.7) and the fact that particle 3 must return to a neighborhood of the binary 1, 2, shows that if after close approach $r_{34} > 1$, then particle 3 returns to some neighborhood of the binary. This means $r_{24} > C_{43}$.

So,

$$\begin{aligned} \int_{t^*}^{t_2} \frac{h_{43}^{1/2}}{r_{24}^{3/2}}\,dt &= O\left(\int_{t^*}^{t_2} \frac{\dot{C}_{43}}{C_{43}^{3/2}}\,dt\right) \\ &= O\left(\left(\frac{1}{C_{43}(t^*)}\right)^{1/2} + \left(\frac{1}{C_{43}(t_2)}\right)^{1/2}\right) = O\left(\frac{1}{R^{1/2}(t^*)}\right). \end{aligned}$$

This can be made arbitrarily small by choosing t^* sufficiently close to t_0.

Consequently, $c_{43} \leqslant D$, and $\Delta e_{43} = O(\Delta t/R^2(t^*))$. That is, the value of e_{43}

cannot vary by much in this time interval. We now compute a bound for the value of e_{43}.

The motion is described by $r_{43} = c_{43}^2 M_{43}^{-1}/(1 + e_{43}\cos f)$, where $f = 0$ corresponds to the closest approach between the particles. At time t^*, $f = -\pi + \eta_1$ and at time $t^* + \Delta t$, $f = \pi - \eta_2$. This means that $1 + e_{43}\cos f$ does not vanish in this domain, or that

$$0 < 1 - e_{43}\cos\eta_i = 1 - e_{43}(1 - \eta_i^2 + O(\eta_i^4)).$$

That is,

$$0 \leqslant e_{43} \leqslant 1 + \eta_i^2 + O(\eta_i^4) = 1 + O((1/R(t^*))^2). \tag{5.7}$$

(The motion of $\mathbf{r}_3$ is close to a straight line, and it must approach the binary within distance unity. Thus $\eta_i = O(R^{-1}(t^*))$.)

Squaring both sides of Eq. (5.4) yields

$$v_{43}^2 c_{43}^2 = M_{43}^2(1 + 2\mathbf{e}_{43}\cdot(\mathbf{r}_{43}/r_{43}) + e_{43}^2) \leqslant A^2.$$

Therefore,

$$c_{43}^2 \leqslant A^2/v_{43}^2. \tag{5.8}$$

An immediate consequence is that at closest approach

$$r_{43} < A^2 M_{43}^{-1}/v_{43}^2.$$

Of greater interest to us is the fact that the normal component of $\mathbf{v}_{43}$, denoted by $\mathbf{v}_{43}^N$, is bounded by

$$|\mathbf{v}_{43}^N| \leqslant A/v_{43}r_{43}.$$

Consequently, in the period of time from t^* until $r_{43} = 1$,

$$|\mathbf{v}_{43}^N| = O(1/v_{43}(t^*)) \tag{5.9}$$

and the path of particle 3 can stray from the line determined by $\mathbf{r}_{43}(t^*)$ and $\mathbf{v}_{43}(t^*)$ by at most $O(\Delta t/v_{43}(t^*))$.

We now determine the behavior of the other particles. The angular momentum integral can be rewritten as

$$\mathbf{c} = \sum_{i=1}^{2} m_i(\mathbf{r}_i - \mathbf{C}_{12}) \times (\mathbf{v}_i - \dot{\mathbf{C}}_{12}) + \sum_{i=1}^{2} m_i(\mathbf{r}_i - \mathbf{C}_{34}) \times (\mathbf{v}_i - \dot{\mathbf{C}}_{34})$$
$$+ M_{12}\mathbf{C}_{12} \times \dot{\mathbf{C}}_{12} + M_{34}\mathbf{C}_{34} \times \dot{\mathbf{C}}_{34}.$$

The first summation is a multiple of $\mathbf{r}_{12} \times \mathbf{v}_{12}$, which is bounded above by

$O(r_{12}^{1/2})$ ($h_{12} < 0$ and $r_{12} = r$; see the proof of Lemma 4.1). The second summation is a multiple of $\mathbf{c}_{43}$. Defining $\boldsymbol{\lambda}$ as the vector from the center of mass of particles 1, 2 to the center of mass of particles 3, 4, $\boldsymbol{\lambda} = \mathbf{C}_{34} - \mathbf{C}_{12}$, and using the fact that $M_{12}\mathbf{C}_{12} = -M_{34}\mathbf{C}_{34}$, it follows that

$$\mathbf{c} = B\boldsymbol{\lambda} \times \dot{\boldsymbol{\lambda}} + O(1/v_{43}(t^*)). \tag{5.10}$$

This means for nonzero $\mathbf{c}$, that

$$|\boldsymbol{\lambda}^N| = Bc/\rho + O(1/v_{43}(t^*)), \tag{5.11}$$

where B is some constant (clearly not the same as above).

The next step is to relate $\boldsymbol{\lambda}(t^*)$ to the line defined by $\mathbf{r}_{43}(t^*)$ and $\mathbf{v}_{43}$. This will be done in Section 7.

We conclude this section by pointing out that since η_i is bounded above by a multiple of $R^{-1}(t^*)$, we have $c_{43}^2 = O(1/R^2(t^*)\, v_{43}^2(t^*))$.

6. The 3–1 Configuration

The dynamics of a 3–1 configuration need to be examined in order to understand the behavior of a particle returning to the binary after passing a singleton, or the behavior of 3–1 configurations evolving into a second 3–1 configuration by the ejection of a binary. In either case, we will assume the binary consists of particles 1 and 2, and particle 3 is either approaching or leaving binary 1, 2. Let $\mathbf{a}_i$ be the position of the ith particle, $i = 1, 2$, and $\boldsymbol{\rho}$ the position of particle 3 relative to the center of mass of the binary. Therefore $m_1\mathbf{a}_1 + m_2\mathbf{a}_2 = \mathbf{0}$ and $\mathbf{a}_2 - \mathbf{a}_1 = \mathbf{r}_{12}$.

The equations of motion for $\boldsymbol{\rho}$ are

$$\ddot{\boldsymbol{\rho}} = \frac{m_1(\mathbf{a}_1 - \boldsymbol{\rho})}{|\mathbf{a}_1 - \boldsymbol{\rho}|^3} + \frac{m_2(\mathbf{a}_2 - \boldsymbol{\rho})}{|\mathbf{a}_2 - \boldsymbol{\rho}|^3} + \frac{m_4(\mathbf{r}_4 - \mathbf{r}_3)}{|\mathbf{r}_4 - \mathbf{r}_3|^3}.$$

Since $(\boldsymbol{\rho} - \mathbf{a})^2 = \rho^2 - 2\mathbf{a}\cdot\boldsymbol{\rho} + a^2$, and since

$$|\boldsymbol{\rho} - \mathbf{a}|^{-3} = \rho^{-3}\left\{1 - \frac{3}{2}\left(\frac{-2\mathbf{a}\cdot\boldsymbol{\rho}}{\rho^2} + \left(\frac{a}{\rho}\right)^2\right) + O\left(\left(\frac{-2\mathbf{a}\cdot\boldsymbol{\rho}}{\rho^2} + \left(\frac{a}{\rho}\right)^2\right)^2\right)\right\}$$

we have

$$\ddot{\boldsymbol{\rho}} = -\frac{M_{12}\boldsymbol{\rho}}{\rho^3} + \frac{m_2 m_1^{-1}\mathbf{r}_{12}(\mathbf{r}_{12}\cdot\boldsymbol{\rho})}{\rho^5} + O\left(\frac{r^2}{\rho^5}\right) + O\left(\frac{1}{r_{43}^2}\right). \tag{6.1}$$

The main difficulty in the study of the dynamics of a 3–1 configuration is to determine a lower bound on ρ which will permit the motion of $\boldsymbol{\rho}$ to be approximated by a solution for the two-body problem. Even when we know it can be approximated in this fashion, we are mainly concerned with estimating how close the motion must be to a straight line. For our purposes it suffices to show that

$$\mathbf{c}_{123} = g_1 \mathbf{r} \times \dot{\mathbf{r}} + g_2 \boldsymbol{\rho} \times \dot{\boldsymbol{\rho}} = O(r_m^{1/2-\epsilon}), \tag{6.2}$$

where $g_1 = m_1 m_2/(m_1 + m_2)$, $g_2 = m_3(m_1 + m_2)/M_{123}$, r_m is some local maximum for r during the considered time interval, and $\epsilon > 0$. Notice that $\dot{\mathbf{c}}_{123} = O(\rho/R^2)$, so $\Delta \mathbf{c}_{123}$ can be made arbitrarily small.

By using an argument similar to the one preceding Eq. (4.3), and by using Eqs. (3.1) and (4.7) it follows that $h_4 = \frac{1}{2} m_4 v_4^2 - (m_1 m_4/r_{14})$ is positive, it differs by only a small amount from $\frac{1}{2} m_4 \mathbf{v}_4^2$, and it remains essentially a constant until another particle approaches particle 4. Thus

$$h_{123} = \tfrac{1}{2}(g_1 \dot{\mathbf{r}}^2 + g_2 \dot{\boldsymbol{\rho}}^2) - \sum_{1 \leqslant i < j \leqslant 3} (m_i m_j / r_{ij})$$

remains essentially equal to the constant $(-B v_4^2)$. A center of mass argument of the type found in Section 4 (Eqs. (4.8), (4.9)) demonstrates that $|\, \dot{\boldsymbol{\rho}} \,|$ is essentially a positive multiple of v_4, where the multiple is bounded by a positive constant depending on other constants of the system. (If it is a binary which is returning to particle 3, then a slight modification of the argument is needed.) The argument following Eq. (4.6) shows that $h_{12} = -A v_4^2$, and $\Delta h_{12}^{1/2} = O(r \Delta t)$ over any time interval such that particle 3 is bounded away from the binary by unity.

Define $r^* = 10^{-2} M_{12} \,|\, h_{12} \,|^{-1}$, where $h_{12}(t)$ is evaluated at the time when $\rho = R/2$. It is clear from the representation for h_{12} and Lemma 2.1 that

$$r(t) \leqslant \tfrac{1}{2} A_1 r^* \tag{6.3}$$

and $|\, \dot{\boldsymbol{\rho}}(t)|^2 \geqslant A_2 r^*$ at least until $\rho(t) = 1$. Near the end of this section we shall establish the existence of several values of time t such that $r(t) \geqslant r^*$ where t is in the time interval defined by particle 3 commuting to the binary.

We shall show for $\epsilon > 0$ that as long as $\rho(t) \geqslant A_1 (r^*)^{1-\epsilon}$, then a solution for the two-body problem approximates the behavior of $\boldsymbol{\rho}(t)$. At the same time, we show that the dynamics of the 3–1 configuration demand that ρ violates this inequality. We obtain both results by assuming that throughout the total 3–1 configuration dynamics, this inequality is satisfied by $\rho(t)$. To accomplish these goals it will be shown that the elements of the orbit remain essentially constant as long as this inequality is satisfied. However, the techniques of the preceding sections are too crude to obtain these facts. What we need is a sharper estimate on the time intervals.

To this end, define t_{-1} as the first time since particle 3 left a neighborhood of particle 4 that $\rho(t) = A_1$. Let $t_1 = \min\{t \mid t > t_{-1}, \rho(t) = 2A_1\}$. The value t_1 exists since some particle must approach a neighborhood of particle 4. (*If* this is particle 3, then t_1 is the exit time.) Define k^* to be the first integer such that $2^{-k^*} < \epsilon$. Let

$$t_{-k} = \min\{t \mid t > t_{-1}, \rho(t) = A_1(r^*)^{1-2^{-k}}\}$$

and

$$t_k = \min\{t \mid t > t_{-k}, \rho(t) = 2A_1(r^*)^{1-2^{-k}}\}.$$

where $k = 1, 2,..., k^*$. Finally, define $t_s = \max\{t \mid t < t_{-1}, \rho(t) = A_1 + 1\}$.

Not all of these values of time need exist. However, for those that do, as long as (A1) $r(t) \leqslant A_1 r^*$, (A2) $|\dot{\boldsymbol{\rho}}(t)^2| \geqslant A_2 r^* 10^{-2}$, and (A3) the two-body approximation for $\boldsymbol{\rho}$ remains valid, then

$$\begin{aligned} t_k - t_{-k} \text{ and } t_{-(k+1)} - t_{-k} &= O(r^{*1-2^{-k}} r^{*1/2}) \\ &= O(r^{*(3/2)-2^{-k}}), \end{aligned} \tag{6.4}$$

and $t_1 - t_s = O(r^{*1/2})$.

Our approach will be the following. We shall assume that the assumptions A_1, A_2, A_3 remain valid until time t^*. By computing the elements of the orbit $\boldsymbol{\rho}$ and of the orbit for $\mathbf{r}_{12}$, it will follow that $t^* > t_{+1}$. (It is obvious that $t^* > t_{-1}$.)

We commence with an analysis of $\boldsymbol{\rho}$. Let $\mathbf{c}_3 = \boldsymbol{\rho} \times \dot{\boldsymbol{\rho}}$, and $M_{12}(\boldsymbol{\rho}\rho^{-1} + \mathbf{e}_3) = \dot{\boldsymbol{\rho}} \times \mathbf{c}_3$. It follows that

$$M_{12}^2(e_3{}^2 + 2\mathbf{e}_3 \cdot \boldsymbol{\rho}\rho^{-1} + 1) = |\dot{\boldsymbol{\rho}}|^2 c_3{}^2$$

and (6.5)

$$\rho = c_3{}^2 M_{12}^{-1}(1 + e_3 \cos f)^{-1}.$$

So, if e_3 and c_3 remain essentially constant, then the first of these equations provides an estimate for $|\dot{\boldsymbol{\rho}}|$, and the second allows an estimate on the closeness of the approach with the binary 1, 2.

Let $t_m = \min(t^*, t_1)$. Since $\dot{\mathbf{c}}_3 = \boldsymbol{\rho} \times \ddot{\boldsymbol{\rho}} = O(r^2\rho^{-3})$, it follows that on the time interval $[t_s, t_m]$,

$$\begin{aligned} \Delta \mathbf{c}_3 &= O(1) \sum_{k=1}^{k^*} \{r^2/(r^{*1-2^{-k}})^3\}\, r^{*(3/2)-2^{1-k}} \\ &= O(r^{*(1/2)+2^{-k^*}}). \end{aligned} \tag{6.6}$$

Since $\rho(1 + e_3) \geqslant c_3{}^2$ (from Eq. (6.5)), we have

$$\dot{e}_3 = c_3 O(r^2\rho^{-4}) = (1 + e_3)^{1/2} O(r^2\rho^{-7/2}),$$

or $(d/dt)(1 + e_3)^{1/2} = O(r^2\rho^{-7/2})$. Therefore,

$$\Delta(1 + e_3)^{1/2} = O(1) \sum_{k=1}^{k^*} \{r^2/r^{*1-2^{-k}})^{7/2}\} r^{*(3/2)-2^{1-k}}$$
$$= O(r^{*3(2)^{-(1+k^*)}}) \qquad \text{on the interval } [t_s, t_m]. \tag{6.7}$$

The next estimate will be for $\Delta(\mathbf{e}_3 e_3^{-1})$, where the equation $|(d/dt)(\mathbf{e}_3 e_3^{-1})| = O(|\dot{\mathbf{e}}_3| e_3^{-1})$ will be used. But

$$\begin{aligned} \dot{\mathbf{e}}_3 &= \dot{\boldsymbol{\rho}} \times \dot{\mathbf{c}}_3 + O(c_3 r^2 \rho^{-4}) \\ &= O(|\dot{\boldsymbol{\rho}}| r^2 \rho^{-3}) + O(r^2 \rho^{-7/2}(1 + e_3)^{1/2}) \\ &= O(r^{3/2}\rho^{-3}) + O(r^2\rho^{-7/2}(1 + e_3)^{1/2}). \end{aligned}$$

Since e_3 remains close to a constant value (Eq. (6.7)) and since ρ was arbitrarily large before it approached the binary ($R \to \infty$), it follows from the second of Eqs. (6.5) that $e_3 > 3/4$. Thus

$$\begin{aligned} (d/dt)(\mathbf{e}_3 e_3^{-1}) &= O(r^{3/2}\rho^{-3} e_3^{-1}) + O(r^2 \rho^{-7/2} e_3^{-1/2}) \\ &= O(r^{3/2}\rho^{-3}) + O(r^2\rho^{-7/2}), \end{aligned}$$

or

$$\Delta(\mathbf{e}_3 e_3^{-1}) = O(r^{*2^{-k^*}}) \qquad \text{during the time interval } [t_s, t_m]. \tag{6.8}$$

In summary, Eqs. (6.6), (6.8) show that the elements of the orbit for ρ remain essentially the same during passage of particle 3 from particle 4 to $\rho(t) = A_1(r^*)^{1-\epsilon}$, *provided* condition A_1 remains satisfied. The verification of this is the goal of the following perturbation analysis.

The equations of motion for $\mathbf{r} = \mathbf{r}_{12}$ are given in Eq. (4.6), and they are

$$\ddot{\mathbf{r}} = (-M_{12}\mathbf{r}/r^3) + O(r\rho^{-3}).$$

Defining $\mathbf{c}_{12} = \mathbf{r} \times \mathbf{v}$, we have that

$$\dot{\mathbf{c}}_{12} = \mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{r} \times O(r\rho^{-3}) = O(r^2\rho^{-3}),$$

from which it follows that

$$\Delta \mathbf{c}_{12} = O(r^{*(1/2)+2^{-k^*}}) \qquad \text{on the time interval } [t_s, t_m]. \tag{6.9}$$

Using Lemma 2.1, it follows that $c_{12} = O(r^{1/2})$. Therefore, $(d/dt)\, e_{12} = c_{12} O(r\rho^{-3}) = O(r^{3/2}\rho^{-3})$, or

$$\Delta e_{12} = O(r^{*2^{-k^*}}) \tag{6.10}$$

where $M_{12}(\mathbf{r}_{12} r^{-1} + \mathbf{e}_{12}) = \dot{\mathbf{r}}_{12} + \mathbf{c}_{12}$.

Defining $h_{12} = \frac{1}{2}\dot{\mathbf{r}}_{12}^2 - M_{12}r^{-1}$, we have

$$\dot{h}_{12} = \dot{\mathbf{r}}_{12} \cdot (\ddot{\mathbf{r}}_{12} + m_{12}\mathbf{r}_{12}r^{-3}) = O(rV\rho^{-3}) = O(r^{1/2}\rho^{-3}).$$

Thus, on the interval $[t_{-1}, t_m]$,

$$\Delta h_{12} = O(r^{*-1^{-2^{-k^*}}}) = O(r^{*-1}).$$

From the above, we have that $h_{12}(t_s) = -AV_4^2 = c/r^*$. This means that in the time interval $[t_s, t_m]$,

$$h_{12} = (1 + O(1))\, h_{12}(t_{-1}), \tag{6.11}$$

which in turn implies that

$$0 \leqslant \tfrac{1}{2}v^2 = h_{12} + (M_{12}/r), \tag{6.12}$$

or that $2r \leqslant M_{12}/|\, h_{12}(t_s)|$. Thus, condition A_1 is satisfied and it now follows that $t_m = t_1$. Furthermore, it follows that as long as $\rho(t) \geqslant A_1(r^*)^{1-\epsilon}$, the motion of $\mathbf{r}$ is closely approximated by that of a two-body elliptic motion. It follows from the two-body theory that the semimajor axis of this ellipse is $\frac{1}{2}M_{12} \mid h_{12} \mid^{-1}$, and the period of the motion is $2\pi(10)^3 2^{-3/2} M_{12}^{5/2}(r^*)^{3/2}$. Thus it follows for any time interval of length $A(r^*)^{3/2}$ intersecting the time interval whereby $\rho(t) \geqslant A(r^*)^{1-\epsilon}$ that there exist values of time whereby $r(t) > r^*$.

In summary, as long as $\rho \geqslant A_1(r^*)^{1-\epsilon}$, both the motion of $\boldsymbol{\rho}$ and $\mathbf{r}$ can be approximated by two-body motions. This, in turn, implies that should this inequality be satisfied throughout the whole 3–1 configuration encounter, then it is the particle (or particles) which left a neighborhood of particle 4 which returns to a neighborhood of it. By using arguments similar to those used in Section 5 and using Eq. (6.8), this provides an estimate for c_3^2 and e_3, namely, $c_3^2 = O(r^*)$, which in turn implies that $\rho = O(r^*)$ at close approach. This violates our basic assumption, thus either the two-body approximations do not suffice to handle the 3–1 encounter, or there exists some time such that

$$\rho = O((r^*)^{1-\epsilon}) \tag{6.13}$$

for any choice of $\epsilon > 0$. In either case, Eq. (6.13) is satisfied for some value of time, and at this time

$$\begin{aligned} \mathbf{c}_{123} = g_1\mathbf{r} \times \mathbf{v} + g_2\, \boldsymbol{\rho} \times \dot{\boldsymbol{\rho}} &= O(r^{1/2}) + O((r^*)^{1-\epsilon}\, v) \\ &= O((r^*)^{1/2-\epsilon}). \end{aligned} \tag{6.14}$$

It remains to show that this value for c_{123} persists in a time interval covering the flight of particle 3 from distance $\rho = R/2$ until some other particle is halfway on the return trip to particle 4. So, according to the definitions of the terms,

$$\mathbf{c} = \sum_{i=1}^{3} m_i(\mathbf{r}_i - \mathbf{C}_{123}) \times (\mathbf{v}_i - \dot{\mathbf{C}}_{123}) + M_{123}\mathbf{C}_{123} \times \dot{\mathbf{C}}_{123} + m_4\mathbf{r}_4 \times \mathbf{v}_4 \tag{6.15}$$
$$= A\mathbf{c}_{123} + B\mathbf{r}_4 \times \mathbf{v}_4 ,$$

where we have used the fact $M_{123}\mathbf{C}_{123} = -m_4\mathbf{r}_4$. Thus

$$A\dot{\mathbf{c}}_{123} = -B\frac{d}{dt}(\mathbf{r}_4 \times \mathbf{v}_4) = -B\mathbf{r}_4 \times \ddot{\mathbf{r}}_4 = -B\sum_{j=1}^{3}\frac{m_j\mathbf{r}_4 \times \mathbf{r}_j}{r_{4j}^3}.$$

In the period of time when the perimeter of the triangle defined by the three particles 1, 2, and 3 is $O(r^{*1-\epsilon})$ we have $\mathbf{r}_j = -A\mathbf{r}_4 + O(r^{*1-\epsilon})$, so $\mathbf{r}_4 \times \mathbf{r}_j = O(R^2 \sin\theta_{4_j}) = O(R^2 r^{*1-\epsilon}R^{-2})$, or $\mathbf{c}_{123} = O(r^{*1-\epsilon}R^{-3})$, where θ_{4_j} is the angle defined by $\mathbf{r}_4$ and $\mathbf{r}_j$. Thus in this period of time $\Delta\mathbf{c}_{123} = O(r^{*1-\epsilon})$.

Consider now the time period when ρ is between $3R/4$ and $A_1(r^*)^{1-\epsilon}$. It was shown above that a two-body approximation for the motion of $\boldsymbol{\rho}$ holds during this time span. At $t = t_{-k^*}$, $\mathbf{c}_3 = O(r^{*1/2-\epsilon})$, and the change in c_3 in the time interval $[t_s, t_{-k^*}]$ is at most $O((r^*)^{1/2+2^{-k}})$. Let $t_{m,-1}$ be the first time whereby $3R/4 \geqslant \rho \geqslant A_1 + 1$. During this time interval, $\dot{\mathbf{c}}_3 = O(r^2) + O(R^{-2}) = O(R^{-2})$. Consequently, $\Delta\mathbf{c}_3 = O(R^{-2}(t_s - t_{m,-1}))$. Since $(t_s - t_{m,-1}) = O(R/v_4)$, it follows from Eq. (6.1) that $\Delta\mathbf{c}_3 = O(r^{*1/2})$.

Some particle or particles must leave the 3-body configuration to approach particle 4. If it is either the binary 1, 2 or particle 3, then we continue our analysis with ρ. If it is some other particle or binary, then we define a new variable $\boldsymbol{\rho}^*$ as the position from the center of mass of the binary to the remaining particle. In any case, a definition for t_1 holds and a similar argument shows that $\boldsymbol{\rho}^* \times \dot{\boldsymbol{\rho}}^* = O(r^{*1/2-\epsilon})$, at least until $\rho^*(t) = 3R(t)/4$. (Since some particle must approach particle 4 which is escaping with velocity $\geqslant A(r^*)^{-1/2}$, and since $|\dot{\boldsymbol{\rho}}^*| \geqslant Ar^{*-1/2}$, it follows that this bound can be made sharper. We shall return to this point in a later section). In any case, we have that $\mathbf{c}_3 = O(r^{*1/2-\epsilon})$, whether we are describing ejection from or approach to a three-body configuration, i.e., a 3–1 configuration. Thus $\mathbf{c}_{123} = O(r^*)^{1/2-\epsilon})$ in the period whereby $3R(t)/4 \geqslant \rho(t)^{1-\epsilon}$, until $3R(t)/4 = \rho^*(t)$.

7. The Transition Orbit

We are now prepared to piece together the various estimates of the orbit to describe the motion of the particle or particles in transit from one configuration to another, whether it be from a 3–1 to a 2–2, or to another 3–1 configuration. Let $\boldsymbol{\rho}$ describe the motion of particle 3 from the binary 1, 2.

Since $\mathbf{c}_3 = O(r^{*1/2-\epsilon})$, the normal component of $\dot{\boldsymbol{\rho}}$, which is designated by $\dot{\boldsymbol{\rho}}^N$, can be computed:

$$\dot{\boldsymbol{\rho}}^N = O(r^{*1/2-\epsilon}\rho^{-1}). \tag{7.1}$$

This means for the time interval defined by $1 \leqslant \rho \leqslant 3R(t)/4$, we have

$$\Delta \boldsymbol{\rho}_N = O(r^{*1/2-\epsilon}\, \Delta t). \tag{7.2}$$

That is, $\boldsymbol{\rho}$ differs from the line defined by $\dot{\boldsymbol{\rho}}$ by at most $O(r^{*1/2-\epsilon}\, \Delta t)$. (Or, $\boldsymbol{\rho} \times \dot{\boldsymbol{\rho}} \mid \dot{\boldsymbol{\rho}} \mid^{-1} = O(r^{*1/2-\epsilon} \mid \dot{\boldsymbol{\rho}} \mid^{-1})$.

In the epoch commencing with $\rho(t) = R(t)/2$, the description of the motion is most easily achieved by introducing a new vector $-\boldsymbol{\lambda}$ which describes the behavior of the commuting particle(s) with respect to particle 4. If particle 3 is approaching particle 4, then let $\boldsymbol{\lambda}$ be the vector $\mathbf{r}_4 - \mathbf{r}_3$. Similarly, should it be the binary which is commuting to particle 4, then let $\boldsymbol{\lambda}$ be the vector from the center of mass of the binary to particle 4. In either case it follows from the developed theory in Sections 5 and 6 that $\boldsymbol{\lambda} \times \dot{\boldsymbol{\lambda}} = O(r^{*1/2-\epsilon})$ and

$$\dot{\boldsymbol{\lambda}}^N = O(r^{*1/2-\epsilon}\lambda^{-1}). \tag{7.3}$$

These estimates imply that the motion becomes extremely close to a straight line collinear behavior as $t \to t_0$. Indeed, the value of $\dot{\boldsymbol{\rho}}(t_1)$, where t_1 is such that $\rho(t_1) = 1$, defines the line. In fact it is easily seen that $\dot{\mathbf{r}}_4$ differs from this direction by at most $O(c/R) + O(r^{*1/2-\epsilon})$. We shall indicate why this is so for the case where particle 3 commutes to particle 4; the remaining possibility of the binary doing the traveling is handled in the same fashion.

Let t_2 be when $\rho(t_2) = (r^*)^{1-\epsilon}$ and t_3 be when $\lambda(t_3) = (r^*)^{1-\epsilon}$ where r^* is in terms of $\mid \dot{\rho}(t)\mid^{-2}$ for some time during the time period when $R(t)/2 \geqslant \rho > 1$. Now, since $\sum m_i\mathbf{r}_i = \mathbf{0}$, we have $\mathbf{C}_{123} = -m_4M_{123}^{-1}\mathbf{r}_4$. Using the estimate $\mathbf{C}_{12} = \mathbf{C}_{123} + O((r^*)^{1-\epsilon})$ at $t = t_2$, it follows that

$$\mathbf{C}_{12}(t_2) = -m_4\mathbf{r}_4(t_2) + O((r^*)^{1-\epsilon}). \tag{7.4}$$

Likewise, at time t_3,

$$\mathbf{C}_{12}(t_3) = -M_{34}M_{12}^{-1}\mathbf{r}_4 + O((r^*)^{1-\epsilon}).$$

But, for $t > t_1$, $\mathbf{C}_{12}^N = O(c/R)$. Also,

$$\mathbf{C}_{12} = \mathbf{C}_{123} - m_3M_{123}^{-1}\boldsymbol{\rho} = M_{123}^{-1}(-m_4\mathbf{r}_4 - m_3\boldsymbol{\rho}),$$

or

$$\dot{\mathbf{C}}_{12} = M_{123}^{-1}(-m_4\dot{r}_4 - m_3\dot{\boldsymbol{\rho}}).$$

Combining these estimates and the two-body approximation for $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$ shows that $\dot{\mathbf{c}}_{12}$ and $\dot{\mathbf{r}}_4$ differ from the line defined by $\dot{\boldsymbol{\rho}}(t_1)$ by at most $O(cR^{-1}) + O(r^{*1/2-\epsilon})$.

Should $\mathbf{c} = \mathbf{0}$, then in the adopted coordinate system, the velocity of $\mathbf{r}_4$ differs from this line by at most terms of the order $O(r^{*1/2-\epsilon})$. However, if $\mathbf{c} \neq \mathbf{0}$, this same error term applies only in the direction out of the invariable plane. Notice from Eq. (6.15) that $\mathbf{r}_4 \times \dot{\mathbf{r}}_4 = A\mathbf{c} + O(r^{*1/2-\epsilon})$. The above estimate for $\mathbf{v}_4$ shows that $\mathbf{v}_4^N \Delta t = O(cr^{*1/2})$. A standard vector analysis argument now shows that the components of $\mathbf{r}_4$ and $\mathbf{v}_4$ orthogonal to the plane defined by $\mathbf{c}$ and the center of mass are of order $O(r^{*1/2-\epsilon}R^{-1})$.

All of this means that $\mathbf{r}_4^N = O(cR^{-1}\Delta t) = O(r^{*1/2})$.

8. Completion of the Proof

Before completing the proof, the basic idea will be outlined. The goal is to use the estimates to determine the measure of the set of initial conditions which could lead to this motion. Fixing the center of mass at the origin, phase space is 18 dimensional. During the transition to a 3–1 configuration, the binary occupies $(rv)^3$ units of measure. At apicenter, $r > r^*$, so this can be expressed as $A(r^*)^3(r^*)^{-3/2}$. The velocity of the commuting particles, when they are still, say, 10^{10} units away, defines the line which approximates this motion. Taking all possible directions into account plus the velocity of $\dot{\boldsymbol{\rho}}$ yields a region of measure $|\dot{\boldsymbol{\rho}}|^3 = A(r^4)^{-3/2}$. The position vector fares much better having magnitude 10^{10} but differing from the line by no more than $A(r^*)^{1-\epsilon}$ units. Finally the velocity of the last particle differs from the line by almost BR^{-1}, and the position is bounded by AR, but differs from the line by $O(r^{*1/2})$. Thus, in total, the particles are restricted to a region of measure

$$A(r^*)^3(r^*)^{-3/2}(r^*)^{-3/2}10^{10}(r^*)^{2(1-\epsilon)}(r^*)^{1/2}R^{-2}Rr^* = A'(r^*)^{(5/2-\epsilon)}R^{-1},$$

a term which is small for small r^*.

The value of r^* changes with each exchange. In the proof which follows r^* is approximated by $2^{-2\alpha}$ for some choice of α. The value of R is approximated by 2^{δ}. The main difficulty is to select the partition points spaced sufficiently close together so that all of this motion, particularly the behavior of the binary, is captured at a partition point.

We now complete the proof of the theorem. Let $NC([n, n+1])$ denote the set of initial conditions for which the corresponding solutions satisfy the following constraints.

(i) The center of mass is fixed at $\mathbf{0}$ for all t.

(ii) Corresponding to $p \in NC([n, n+1])$ is a $t^* \in [n, n+1]$ with the property that the solution evolving from p exists on $[0, t^*)$, but it suffers a noncollision singularity at $t = t^*$.

We shall demonstrate that this set is of Lebesgue measure zero. Once this is shown, the proof of Theorem 2 follows since the initial conditions for a noncollision singularity lie in one of these sets for some integer n.

In proving this statement, the three-dimensional case will be handled first. Then the necessary modifications to prove the assertion for the planar case will be given. Furthermore, restrict attention to those initial conditions for which the center of mass is fixed at **0**.

Let α be a positive integer. Divide the time interval $[n - 1/2, n + 3/2]$ into $2^{\alpha+5}$ parts.

Starting at each of the resulting partition points, add $\nu 2^{20}$ addition partition points which are located distance $1/2\pi(2m_0)^{5/2}(5^3 2^{1/2}(2^{3\alpha})$ apart, where ν is some positive integer greater than $(M/m_0)^{5/2}$. This defines $1 + \nu 2^{20} 2^{\alpha+5} = \beta$ partition points $\{t_\gamma\}_{\gamma=1}^{\beta}$, $t_1 = n - 1/2$, $t_\beta = n + 3/2$.

For each t_γ define $D^\alpha(t_\gamma)$ to be the set of points in phase space satisfying the following constraints for some permutation of the indices 1, 2, 3, 4, given as (i, j, k, l).

(1) Viewing the point as an initial condition at time t_γ, the solution exists on the time interval $[0, t_\gamma]$ (or $[t_\gamma, 0]$, whichever applies. For simplicity assume $t_\gamma > 0$.)

(2) For some choice of the indices, say k and l, $|\mathbf{r}_{kl}(t_\gamma)| \leqslant M^3 m_0^{-2} 2^{-2\alpha+10}$ and $|\dot{\mathbf{r}}_{kl}(t_\gamma)| < 2^{\alpha+10}$.

(3) Let $\boldsymbol{\rho}$ be the vector from the center of mass of the binary defined by particles k and l to particle j. Vector $\dot{\boldsymbol{\rho}}(t_\gamma)$ and the center of mass of the binary defines a line θ. We require

(a) $\rho(t_\gamma) \leqslant 10^{10}$; the normal component of $\boldsymbol{\rho}(t_\gamma)$ with respect to θ is bounded by $10^5(2^{-2\alpha})^{1-0.001}$;

(b) $|\dot{\boldsymbol{\rho}}(t_\gamma)| \leqslant 2^\alpha$;

(c) for the remaining index i, $v_i \leqslant 2^\alpha$, and the normal component of $\dot{\mathbf{r}}_i$ with respect to θ is bounded by unity. Furthermore $r_i \leqslant 2^\alpha$, but the normal component of $\mathbf{r}_i$ with respect to θ is bounded by $(2^{-\alpha})^{1-0.001}$.

From 1 and the continuity of solutions with respect to initial conditions, it follows that set $D^\alpha(t_\gamma)$ is measurable and that it has 18-dimensional measure $E'(2^{-\alpha})^{4-0.006}$ where E' is some positive constant depending on the masses.

Let $D^\alpha[n - 1/2, n + 3/2]$ be the set of initial conditions such that the solution will be in $\bigcup_{\gamma=1}^{\beta} D^\alpha(t_\gamma)$. By condition 1, $D^\alpha[n - 1/2, n + 3/2]$ is measurable. Since the system is measure preserving, the measure of $D^\alpha[n - 1/2, n + 3/2]$ is bounded by $\beta E'(2^{-\alpha})^{4-0.006} = E''(2^{-\alpha})^{3-0.006}$.

Let $D[n - 1/2, n + 3/2] = \limsup D^\alpha[n - 1/2, n + 3/2]$. Since $D[n - 1/2,$

$n + 3/2] \subset \bigcup_{\alpha=N}^{\infty} D^{\alpha}[n - 1/2, n + 3/2]$, the measure of this set is bounded above by

$$E'' \sum_{\alpha=N}^{\infty} (2^{-(3-0.006)})^{\alpha} \leqslant 2E'' \, 2^{-(3-0.006)N}$$

for arbitrary $N > 1$. Thus, the measure of $D[n - 1/2, n + 3/2]$ is equal to zero.

By construction of sets $D^{\alpha}(t_{\gamma})$, it is clear that $p \in NC([n, n + 1]$ implies that $p \in D^{\alpha}[n - 1/2, n + 3/2]$ for an infinite number of choices of α. Indeed, whenever a 3–1 configuration is being either formed or desolved, and this must occur infinitely often, there is some choice of α such that $|\dot{\rho}(t)| \leqslant 2^{\alpha}$. This in turn defines a bound for the term r^* in terms of $O(2^{-2\alpha})$, and all the estimates in condition three hold during the transition from, or to, a 3–1 configuration. A partition point can be found during this transition period since the partition points t_{γ} are close enough together so that ρ can commute a distance of less than unity between successive points. The only problem is to satisfy the velocity constraint in condition 2. But since the orbit of $\mathbf{r}_{kl}$ is elliptical, since the second type of partition points divide the period of this orbit into 10^3 equal parts, it follows that a partition point can be selected during the transition of particle j such that r_{jk} is near a local maximum, and r_{jk} is bounded below by some constant multiple of $2^{-2\alpha}$. That the velocity constraint in condition 2 holds at such a partition point now follows from Lemma 2.1.

All this shows that $NC([n, n + 1]) \subset D[n - 1/2, n + 3/2]$, which implies it is of measure zero. This completes the proof for the three-dimensional problem.

To handle the planar problem, we make the following modifications. Divide the interval $[n - 1/2, n + 3/12]$ into the β partition points as described above.

For each partition point and positive integers E, $\delta \leqslant \alpha$ define $D^{\alpha,\delta}(t_{\gamma}, E)$ to be the set of points in phase space which satisfy the following constraints.

(1) Constraint 1 is the same except that we require, in addition, that the angular momentum of the system is bounded in magnitude by E.

(2) There are no changes in constraints 2 to 3b. Constraint 3c is changed to require the normal component of $\dot{\mathbf{r}}_i$ with respect to θ to be bounded by $10^{20}E(M/m_0)^3 2^{-\delta}$ and $r_i \leqslant 2^{\delta}$. The remaining constraints are the same.

This set is clearly measurable, and its measure is bounded by $E'(2^{-\alpha})^{2-0.002}$.

Set $D^{\alpha,\delta}([n - 1/2, n + 3/2], E)$ is defined to be the set of initial conditions leading to $\bigcup D^{\alpha,\delta}(t_{\gamma}, E)$. This set is measurable and its measure is bounded by $E''(2^{-\alpha})^{1-0.002}$. Define set $D([n - 1/2, n + 3/2], E)$ to be the

$$\limsup_{\alpha\to\infty} \left\{ \bigcup_{\delta=1}^{\alpha} D^{\alpha,\delta}([n - 1/2, n + 3/2], E) \right\}.$$

For any choice of N, the measure of this new set is bounded above by

$$\sum_{\alpha=N}^{\infty} \sum_{\delta=1}^{\alpha} E''(2^{-\alpha})^{1-0.002} = E'' \sum_{\alpha=N}^{\infty} \alpha(2^{-1+0.002})^{\alpha} < 10^2 N(2^{-1+0.002})^N.$$

Since this holds for any N, the measure of $D([n - 1/2, n + 3/2], E)$ is equal to zero.

Let $D([n - 1/2, n + 3/2]) = \bigcup_{E=1}^{\infty} D([n - 1/2, n + 3/2], E)$. Clearly, the measure of $D([n - 1/2, n + 3/2])$ is zero.

The argument that $NC([n, n + 1]) \subset D([n - 1/2, n + 3/2])$ is similar. The only changes are due to the 2^{δ} term which corresponds to bounds in terms of R. (Since $R < |\dot{\rho}|$, we have $\delta \leqslant \alpha$.)

The proof for initial conditions restricted to a fixed angular momentum surface is similar. While we do not supply any details, we shall outline some of the differences. There are two approaches. One is to reduce the order of the system of differential equations by eliminating the node, as done in [14]; and then using the above estimates in the new variables. A second approach would be to compute the invariant measure on these angular momentum surfaces. (See, for example, [1, pp. 101–103]).

If $\mathbf{c} \neq \mathbf{0}$, a reduction in the order of the terms results from the fact that θ can be selected to be on the invariable plane, that is, the plane orthogonal to $\mathbf{c}$. This reduction in the degrees of freedom of θ helps when computing what corresponds to $D^{\alpha,\delta}(t_{\gamma}, E)$. Here the estimates on how fast the motion is approaching the invariable plane help out. If $\mathbf{c} = \mathbf{0}$, the extra degree of freedom in the choice of θ, which enlarges the size of $D^{\alpha,\delta}(t_{\gamma}, E)$, is compensated for by a sharper estimate for the normal component of $\mathbf{r}_4$ with respect to θ.

9. Comments

In this section we use some of the lemmas to point out some additional properties of noncollision singularities. The first follows from Lemma 4.1.

THEOREM 3. *There exist ratios of the masses for which certain types of noncollision singularities cannot occur. In particular, unless one mass is much smaller than the others, the 2–2 configuration cannot occur infinitely often.*

It is a simple center of mass argument to show that certain mass ratios are inconsistent with this type of motion. Sharper mass estimates result by determining how fast the commuting particle(s) must travel to overtake their target.

While this fact follows from a center of mass argument, reasonably strong results should follow from a careful velocity analysis. (For example, see

Eq. (4.11)). Although this is clearly a problem of interest, we have not investigated it beyond the above comments.

A combination of the crude arguments of Section 2 with the analysis used in the Mather–McGehee [2] existence theorem, shows that *all* collinear noncollision singularities (actually, limit points of binary collisions) are caused by a mechanism similar to the one they employed. Namely, if there is a collinear noncollision singularity, then the initial condition leading to this behavior does not belong in any triple-collision manifold, but it is in the closure of some triple-collision manifold. (Indeed, the same holds for motion such that $R/t \to \infty$; see [3, 12].) This means there exist values of masses such that some triple collision manifold is immersed in phase space, while there are other mass ratios such that these manifolds are embedded.

One might hope that a similar argument employing "almost triple collisions" could be used to establish the existence of noncollision singularities in the four-body problem. This leads to problems in the values of the angular momentum. First of all, in order that the particles "almost" suffer a triple collision, it is necessary that $c_{123}^2 h_{123}$ be sufficiently small, for otherwise one particle will be bounded away from the binary (see [4]). Secondly, a perturbation analysis shows that $\mathbf{c}_3$ and $\mathbf{c}_{12}$ must decrease in value with an infinite number of the 3–1 configuration passes. Since c_{123} remains essentially a constant during a pass, this is only possible should one orbit, say $\mathbf{r}$, be direct and the other, $\boldsymbol{\rho}$, be retrograde, or should the elliptical orbit be "nearly" collisions. The first possibility also satisfies the condition concerning c^2h.

As it was mentioned earlier, noncollision singularities are equivalent to unbounded motion in finite time. Lemma 2.1 provides a bound on how fast the system must become unbounded in physical space.

THEOREM 4. *For the four-body problem, if there is a noncollision singularity at* $t = t_0$, *then for any positive constant* α,

$$I(t) \geqslant \ln^{\alpha}(t_0 - t)^{-1} \quad \text{as} \quad t \to t_0 .$$

A similar statement can be found for arbitrary n.

Proof. From Lemma 2.1, we have $r^*R^2 = O(t - t_0)^2$. Using the second part of inequalities (1.5) and Eq. (1.4), it follows that $\ddot{I} \geqslant A(t - t_0)^{-2}$, or that $I > A \ln (t_0 - t)^{-1}$. Using the first of inequalities (1.5), and Lemma 2.1, this estimate yields $\ddot{I} \geqslant A \ln (t_0 - t)^{-1}(t - t_0)^{-2}$. Continual integration and substitution yields the stated result.

REFERENCES

1. R. ABRAHAM, "Foundations of Mechanics," Benjamin, New York, 1967.
2. J. MATHER AND R. MCGEHEE, *in* "Proceedings of Bat. Conference, Seattle, 1974."

3. C. Marchal and D. G. Saari, On the final evolution of the n-body problem, *J. Differential Equations* **20** (1976), 150–186.
4. C. Marchal and D. G. Saari, Hill Regions for the general three-body problem, *Celestial Mechanics* **12** (1975), 115–129.
5. P. Painlevé, "Lecons sur la théorie analytique des équations différentielles," Hermann, Paris, 1897.
6. H. Pollard, "A Mathematical Introduction to Celestial Mechanics," Prentice–Hall, Englewood Cliffs, N.J., 1967.
7. D. G. Saari, Improbability of collisions in Newtonian gravitational systems, *Trans. Amer. Math. Soc.* **162** (1971), 267–271.
8. D. G. Saari, Improbability of collisions in Newtonian gravitational systems, II, *Trans. Amer. Math. Soc.* **181** (1973), 351–368.
9. D. G. Saari, Collisions are of first category, *Proc. Amer. Math. Soc.* **47** (1975), 442–445.
10. D. G. Saari, Singularities of Newtonian gravitational systems, *in* "Dynamical Systems" (M. M. Peixoto, Ed.), pp. 479–487, Academic Press, New York, 1973.
11. D. G. Saari, Singularities and collisions of Newtonian gravitational systems, *Arch. Rational Mech. Math.* **49** (1973), 311–320.
12. D. G. Saari, Expanding gravitational systems, *Trans. Amer. Math. Soc.* **156** (1971), 219–240.
13. H. Sperling, On the real singularities of the n-body problem, *J. Reine Angew. Math.* **245** (1970), 15–40.
14. J. Urenko, Improbability of collisions in a newtonian gravitational system of specified angular momentum, *SIAM J. Appl. Math.*, to appear.
15. H. von Zeipel, Sur les singularités du problème des n corps, *Ark. Mat. Astron. Fys.* **4** (1908).