A New Solution to the Three-Body Problem

Richard Montgomery

We describe a new solution to the three-body problem (C. Moore [1993], Chenciner and Montgomery [2001]) and motivate its discovery. We also sketch its existence proof, which is based on the direct method of the calculus of variations. We begin with the statement of the N-body problem and some of its solutions.

Newton told us that two masses attract each other, the force of attraction being directed along the line joining them, proportional to the product of the masses, and inversely proportional to the square of the distance between them. If we have N masses, then the force on any one is the sum of the forces exerted on it by all the others. This gives us the nonlinear system of second-order differential equations

$$m_i \frac{d^2 x_i}{dt^2} = -\sum_{j \neq i} \frac{m_i m_j (x_i - x_j)}{r_{ij}^3}, \qquad i = 1, \dots, N, \tag{1}$$

m_i being the numerical value of the ith mass, $x_i(t) \in \mathbb{R}^d$ its position vector, and r_{ij} the distance between it and mass j. We are interested in the planar case $d = 2$. A *solution* to the N-body problem is then a solution $x(t) = (x_1(t), \dots, x_N(t))$ to these equations. We contrast this notion with that of "solving the N-body problem", which we suppose to mean finding an explicit expression for the general solution. Poincaré showed, in effect, that this is impossible for $N > 2$.

Richard Montgomery is professor of mathematics at the University of California, Santa Cruz. His e-mail address is rmont@math.ucsc.edu.

Newton solved the two-body problem. The difference vector $x = x_1 - x_2$ satisfies *Kepler's problem:*

$$\frac{d^2 x}{dt^2} = \frac{-kx}{|x|^3}, \tag{2}$$

all solutions of which are conics with one focus at the origin. The Kepler constant k is $m_1 + m_2$. Correspondingly, if we fix the center of mass of our two bodies to be the origin, then the two move along similar conic sections with one focus at this origin. The periodic two-body motions are ellipses. We refer to them as *Keplerian ellipses*. They include degenerate ellipses, sometimes called *elliptic collision-ejection orbits*, which are line segments with one endpoint at the origin. They represent collision solutions to the two-body problem.

It is impossible to describe all the solutions to the three-body problem. Following Poincaré, we focus on the periodic solutions $x_i(t) = x_i(t + T)$. Here T is called the period. The simplest periodic solutions for the three-body problem were discovered by Euler [1765] and by Lagrange [1772]. Built out of Keplerian ellipses, they are the only explicit solutions. To form the Lagrange solution, start by placing the three masses at the vertices x_1^0, x_2^0, x_3^0 of an equilateral triangle whose center of mass $m_1 x_1^0 + m_2 x_2^0 + m_3 x_3^0$ is the origin. Identify the plane of the triangle with the complex plane $\mathbb{C}$, so that $x_i^0 \in \mathbb{C}$. Take any solution $\lambda(t) \in \mathbb{C}$ to the planar Kepler problem (2) where the Kepler constant k is a certain rational expression in the three masses m_i. The Lagrange solutions are $x_i(t) = \lambda(t) x_i^0$. Each mass moves in an ellipse in such a way that the triangle formed by the three

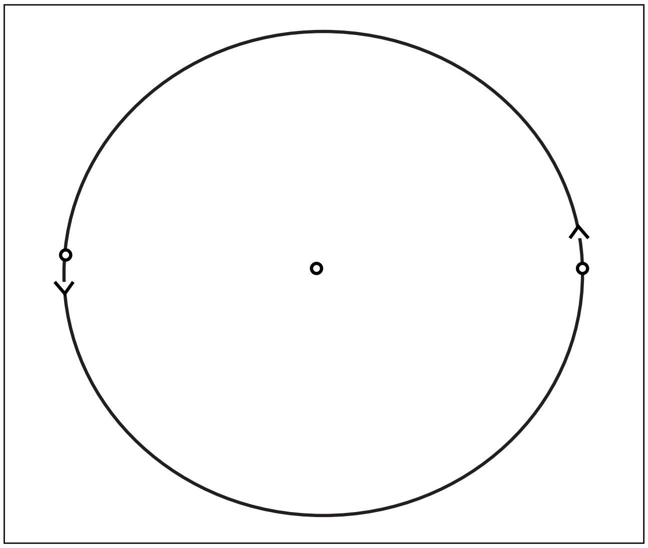

Figure 1. Euler's solution in the equal mass case.

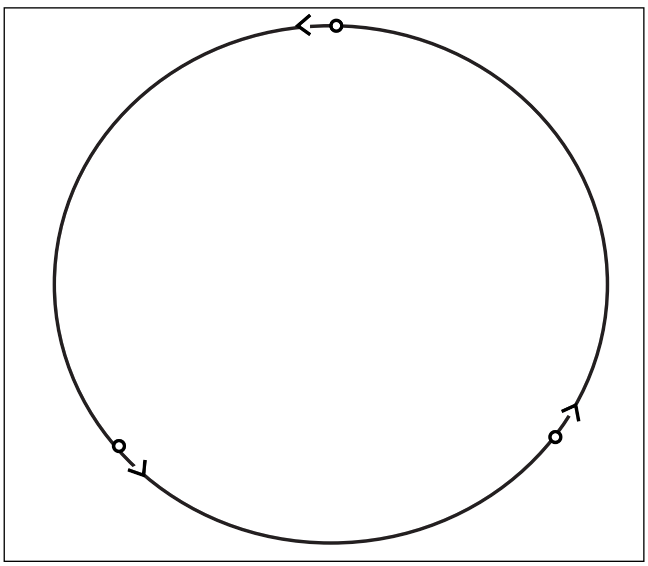

Figure 2. Lagrange's solution in the equal mass case.

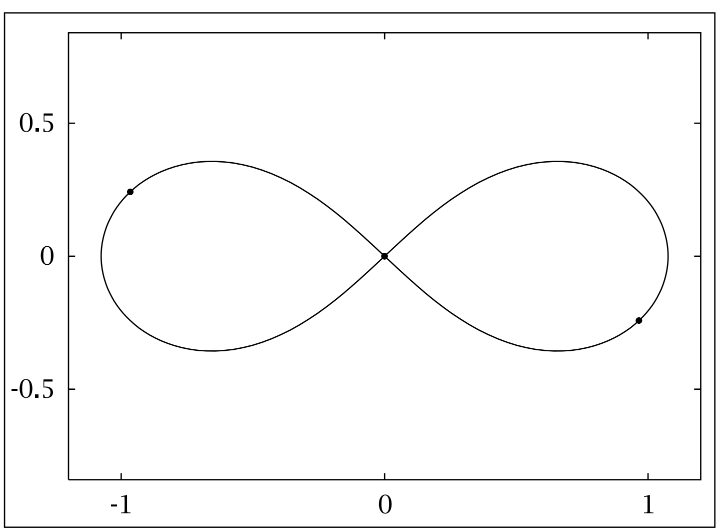

Figure 3. The figure-eight solution.

masses evolves by a composition of instantaneous dilations and rotations and hence is equilateral for all time. For the Euler solutions start by placing the three masses on the same line with their positions x_i^0 such that the ratios r_{ij}/r_{ik} of their distances are the roots of a certain polynomial whose coefficients depend on the masses. Again, take any solution $\lambda(t) \in \mathbb{C}$ to Kepler's equation (2) where the Kepler constant is a certain other rational expression in the masses m_i. The Euler solutions are $x_i(t) = \lambda(t)x_i^0$. At every instant the masses are collinear, and the ratios of their distances remain constant. There are three different families of Euler solutions, according to which mass remains between the other two. Together, the Euler and Lagrange solutions form the only solutions for which the similarity class of the triangle remains constant throughout the motion. Their beginning configurations x_i^0 are called *central configurations*.

Most important to astronomy are Hill's periodic solutions, also called *tight binaries*. These model the earth-moon-sun system. Two masses are close to each other while the third remains far away. The two move in nearly circular orbits about their common center of mass. This center of mass and the third body in turn move in nearly circular orbits about the total center of mass. Like the Euler and Lagrange solutions, these Hill's solutions exist for all ratios of masses.

Perhaps next in order of complication is the orbit which is the subject of this paper, the figure eight. Unlike the earlier orbits, it is particular to the case when all three masses are equal. The three equal masses chase each other around the same figure-eight-shaped curve in the plane. The eight was discovered numerically by Chris Moore [1993]. Alain Chenciner and the author [2001] rediscovered it and proved its existence.

Description of the eight. The eight is a periodic solution $x = (x_1(t), x_2(t), x_3(t))$ to the equal-mass three-body problem. If T is the period, then $x_2(t) = x_1(t - T/3)$ and $x_3(t) = x_1(t - 2T/3)$. This says that the three bodies travel the same planar curve, phase shifted from each other by one-third of a period. This curve has the form of a figure eight. There is an eight orbit of any period T, according to a scaling symmetry of the equations (1) to be described below. Modulo this scaling symmetry and the other obvious symmetries of (1), the eight is unique according to all numerical investigations. Its unicity has not been proved.

The double point of the figure-eight curve is at the origin. This is also the center of mass. The eight curve has the reflectional symmetries of the x-y axes. Each of its two lobes is star-shaped (proved), indeed convex (unproved). The solution begins at $t = 0$ with mass 1 at the origin, forming the midpoint of masses 2 and 3. We call any such configuration an *Euler configuration* of type 1, since it is an initial configuration for Euler's solution described above when the masses are all equal. The set of Euler configurations with mass i forming the midpoint will be denoted by EUL_i. Every one-sixth of a period the eight solution returns to an Euler configuration, doing so in the order 132132 in a full period. (A different ordering occurs in Chenciner-Montgomery [2001].) At the times half-way between Euler configurations j and k, the triangle formed by the masses is isosceles, with $r_{ij} = r_{ik}$, ijk being a permutation of 123. We denote the set of all such isosceles configurations by $ISOSC_i$. Thus in time $T/12$ the curve x travels between EUL_1 and $ISOSC_2$. This is the key to constructing the eight.

Stability. Carles Simó [2000b] showed numerically that the figure eight is stable. This is surprising for two reasons. First, we know very few stable periodic orbits for the three-body problem and even fewer for the equal-mass case. Hill's solutions are always stable. The Lagrange orbits are only stable when one of the three masses is much greater than the other two. The Euler solutions are never stable. Second, the eight is found by minimizing action, a procedure which yields dynamically unstable orbits most of the time.

The stability of the eight is KAM stability. This is different from the standard stability of dynamical systems but is essentially the only kind of stability one can hope for in the N-body problem. When a periodic orbit is KAM stable, then the solutions through most initial conditions sufficiently near the orbit stay near it for all time. The density of these stable initial conditions approaches 1 as we tend to the orbit. Those solutions that do leave a neighborhood of the orbit do so extremely slowly (Nekhoroshev estimates), so slowly that their rate of escape cannot be detected by any power series in the distance from the orbit. If an orbit is KAM stable, then one can perturb its parameters, the mass ratios or angular momentum for the eight, and follow it to a nearby periodic KAM stable orbit having these perturbed parameters.

What stability means physically is that there is some chance that the eight might actually be seen in some stellar system. The domain of stability of the eight—the amount one can perturb the mass ratios in particular—is very small, so this chance is very small. Numerical experiments done by Douglas Heggie (2000) suggest that the probability of an eight is somewhere between one per galaxy and one per universe.

N-Body Generalities

The solutions described above have their center of mass $\sum m_i x_i / \sum m_i$ at the origin. This simplification is possible by virtue of the symmetries of Newton's equations (1). A *symmetry* is a transformation which takes solutions to solutions. The Galilean transformations $x_i \mapsto R(x_i) + c + tv$, $i = 1, \ldots, N$, are symmetries. Here R is an orthogonal transformation, which we will refer to as a "rotation"; c represents translation; and tv represents changing to a moving frame, moving with constant velocity v. Associated to symmetries are *conserved quantities*, which are functions of position and velocity that are constant along any solution. The total linear momentum vector $P = \sum m_i \dot{x}_i$ is the conserved quantity associated to translation. Its constancy lets us form the Galilean transformation with velocity $v = -P/\sum m_i$ which transforms us to a frame where the center of mass is constant.

The conserved quantity associated with rotation is the angular momentum $\sum m_i x_i \wedge \dot{x}_i$. Time translation $x_i(t) \mapsto x_i(t - t_0)$ is also a symmetry. Its conserved quantity is the total energy $H = K/2 - U$, where

$$K = \sum m_i \|\dot{x}_i\|^2$$

and

$$U = \sum_{i<j} m_i m_j / r_{ij}.$$

$K/2$ is called the kinetic energy, and $-U$ is called the potential energy. The positivity of U will be important momentarily.

By virtue of the homogeneity of the potential, (1) also enjoys the scaling symmetry $x_i(t) \mapsto \lambda x_i(\lambda^{-3/2} t)$. This implies that if we have a solution of one period, then we have similar solutions of any period. It also implies Kepler's third law, that the period of a Keplerian ellipse is proportional to its semi-major axis to the power $3/2$.

We will call a solution *bounded* if $c < r_{ij}(t) < C$ for all pairs i, j and all time T, where c and C are fixed positive constants. We call a solution *reduced periodic* of period T if $r_{ij}(t + T) = r_{ij}(t)$. Periodic implies reduced periodic implies bounded implies negative energy. For the two-body problem all these conditions are equivalent. In the three-body problem there are counterexamples to reversing any one of these implications.

For negative energies the N-body dynamics is believed to be a complicated mixture of chaotic and near-integrable (KAM) behavior. Very little is understood in general. For example, fix the energy and the angular momentum with the energy negative. Are the bounded orbits nowhere dense? Are the periodic orbits dense within the bounded orbits? These questions appear to be completely open. Poincaré pointed us towards the periodic orbits as being "the only breach in the fortress that is the three-body problem." We advocate looking for the simplest, most basic of these breaches using variational methods.

Action Principles

The construction of the eight is based on the principle of least action, which we now describe. With K and U as above, form the *Lagrangian*

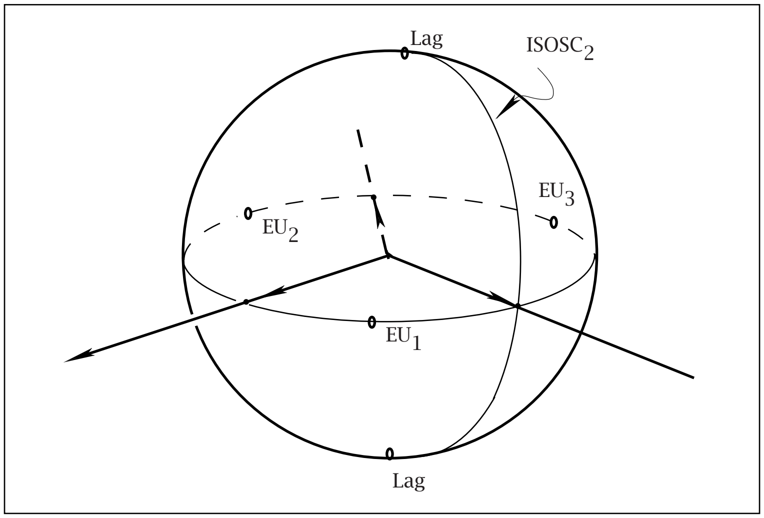

Figure 4. The shape sphere.

$$L(x, \dot{x}) = \frac{1}{2}K(\dot{x}) + U(x), \tag{3}$$

to be contrasted with the energy $H = K/2 - U$. The space in which x moves is called *configuration space*. For the planar N-body problem it is N copies of the Euclidean plane, if we allow collisions. We will, without loss of generality, restrict to the subspace Q of configuration space consisting of those configurations x whose center of mass is at the origin. The *action* of a path $x : [0, T] \to Q$ is

$$A(x) := \int_0^T L(x(t), \dot{x}(t))\, dt.$$

A curve $x(t)$ is called *collision-free* if it has no collisions, meaning that $r_{ij}(t)$ is never zero for any pair $i \neq j$. The *principle of least action* asserts that if $\mathcal{C}$ is an "appropriate class" of curves in Q, if x is a collision-free curve in $\mathcal{C}$, and if the derivative of the restriction of A to $\mathcal{C}$ is zero at x, then that path x is a solution to Newton's equations (1). An example of an appropriate class is the set of all curves joining two fixed submanifolds of Q in some specified time. To construct the figure eight, we take the starting submanifold to be EUL_1 and the ending manifold to be $ISOSC_2$.

Theorem 1. *(Chenciner and Montgomery [2001]) Fix a time $\bar{T}$. There is a collison-free curve x which minimizes the action among all curves joining EUL_1 to $ISOSC_2$ in the time $\bar{T}$. This curve comprises one-twelfth of the figure-eight solution.*

Our figure-eight solution $x(t)$ is assembled out of pieces congruent to the minimizer of the theorem. The assembly requires understanding the *shape space* for the three-body problem.

Shape Space

Because Newton's equations are invariant under isometries, we can push them down to the quotient space of configuration space by isometries. This quotient is the space of congruence classes of N-gons. It is more convenient to divide out instead by the group of orientation-preserving isometries, which means dividing by rotations, since we have fixed the center of mass. We call this quotient *shape space* and denote it by C.

In the case of the planar three-body problem, the shape space C is three-dimensional, by the side-side-side theorem from high-school geometry. It is homeomorphic to Euclidean three-space. Triple collision ($x_1 = x_2 = x_3$) is represented by the origin. It lies in the equatorial plane, the points of which represent the degenerate triangles whose masses are collinear. Reflection about this plane corresponds to replacing a triangle by its reflection. Distance from the origin in C is measured by $\sqrt{I}$, where $I = I(x)$ is the moment of inertia of a triangle whose center of mass is the origin: $I = \sum m_i |x_i|^2$. (Note that I is invariant under rotations, so it yields a well-defined function on C.)

Dilating a triangle by the factor λ induces the change $I \mapsto \lambda^2 I$. The sphere $I = 1$ in C is topologically a two-sphere. Its points represent oriented similarity classes of triangles. We call it the *shape sphere* and denote it by S. The equatorial plane intersects S in the equator whose points represent the similarity classes of collinear triangles. The three types of binary collisions—1 with 2, 2 with 3, and 3 with 1—are represented by three points on this equator.

The function U in the Lagrangian is invariant under isometries, hence descends to a function on shape space which we give the same name. This function is homogeneous of degree -1, so it can be expressed in the form $\tilde{U}/\sqrt{I}$ where $\tilde{U}$ is homogeneous of degree 0. We identify $\tilde{U}$ with the restriction of U to the shape sphere $I = 1$. The function $\tilde{U}$ blows up at the three binary collision points. It has five critical points: the two Lagrange configurations, one for each orientation of an equilateral triangle, and the three Euler configurations lying on the equator. The Euler points are saddle points for $\tilde{U}$ and the Lagrange points are its minima.

This picture of the shape sphere allows one to describe succinctly much of what is known about the planar three-body problem and suggests many directions and open problems. We refer the reader to the beautiful article of Moeckel [1988].

We have been viewing shape space C as simply a topological space. But it also has a metric which we call the *kinetic energy metric*, since it arises from the kinetic term K in the Lagrangian. K is the norm squared of the velocities for the kinetic energy inner product on the configuration space:

$$K = \|\dot{x}\|^2 := \sum m_i |\dot{x}_i|^2.$$

The associated norm squared of positions is the moment of inertia just introduced:

$$I = \|x\|^2 := \sum m_i |x_i|^2.$$

The group of orientation-preserving isometries of the plane acts isometrically on configuration space when we use this metric. Shape space, being the quotient of configuration space by this group, inherits a metric. The distance between two points in shape space is defined to be the distance in configuration space between the two corresponding orbits.

This metric on C is Riemannian away from the origin. The induced Riemannian metric on the sphere S is that of the standard Euclidean two-sphere of radius $1/2$. The act of dilating a triangle is a metric dilation on C. This information completely specifies the metric structure of C. It is called a cone over a two-sphere of radius one-half. The triple collision is the cone point, and the Riemannian metric is singular there.

For general masses, the binary collision points and central configurations are asymmetrically placed on the shape sphere, enjoying only reflectional symmetry about the equator. In our case of all equal masses, these eight points are placed as symmetrically as possible. The three Euler points and the three collison points are spaced out in equal intervals along the equator. The two Lagrange points are at the north and south poles. When we join the north pole to the six marked points on the equator by great circles, we obtain three meridians. Each meridian passes through the south pole and intersects the equator in a pair of antipodal marked points, one an Euler point and the other a binary collison. These meridians represent the manifolds of isosceles triangles $ISOSC_1$, $ISOSC_2$, and $ISOSC_3$ from above.

Reduced Action

If $x(t)$ is a curve in configuration space, then we will write $c(t)$ for its projection to the shape space C. If $x(t)$ is a solution to the three-body problem, then we will call $c(t)$ a *reduced solution*. We show how to pass back and forth between solutions and reduced solutions, provided the total angular momentum $J := \sum m_i x_i \wedge \dot{x}_i = 0$.

Let $K_C/2$ denote the kinetic energy associated to the metric on C. K_C is the function on the tangent bundle of C given by $K_C(c, \dot{c}) = \langle \dot{c}, \dot{c} \rangle_c$, where $\dot{c}$ represents a tangent vector to shape space at the particular shape $c \in C$ and $\langle \cdot, \cdot \rangle_c$ denotes the Riemannian metric at c. One proves that $K(\dot{x}) = K_C(c, \dot{c}) + |J(x, \dot{x})|^2/I$. It follows that $K \geq K_C$ with equality if and only if the angular momentum J is zero. Define the *reduced Lagrangian* to be $L_C(c, \dot{c}) = K_C(c, \dot{c})/2 + U(c)$, where $U(x) = U(c)$ is the negative potential, viewed as a function on C. The *reduced action* is then the function $A_C(c) = \int_c L_C \, dt$ on curves in C.

We relate the action principle to the reduced one. Write $\pi : Q \to C$ for the quotient map which associates to a triangle its shape, so that $c(t) = \pi(x(t))$. Let $M_1, M_2 \subset Q$ be two submanifolds invariant under the action of the group of rotations about the center of mass. Then $M_i = \pi^{-1}(E_i)$, $i = 1, 2$, for submanifolds $E_i \subset C$. Consider the problem of minimizing the action A among all paths in Q which connect M_1 to M_2 in time T. According to the above, $L(x, \dot{x}) \geq L_C(c, \dot{c})$ with equality if and only if the curve $x(t)$ has zero angular momentum. By applying a time-dependent family of rotations $g(t)$ to $x(t)$, we can obtain a new curve $g(t)x(t)$ whose angular momentum J is zero and whose shape curve is the same $c(t)$. This proves that any minimizer x for our original problem must satisfy $A(x) = A_C(c)$. Our original problem reduces to that of minimizing A_C among all curves in C joining E_1 to E_2 in time T.

We can reverse this procedure. Imagine that we have a minimizer c for our A_C-problem. Suppose that this curve is without triple collisions. Each curve in C without triple collision has a *lift* to Q, this being a curve x which projects to c and has angular momentum zero. The lift is unique up to rigid rotation. The lifts x of our minimizer c will minimize the original variational problem for curves on Q. We now apply this fact with $E_1 = EUL_1$ and $E_2 = ISOSC_2$.

Building the Eight Using the Discrete Symmetries

We build the figure-eight solution from the minimizing curve x of Theorem 1, beginning from its shape curve $c(t)$, $0 \leq t \leq T/12 = \bar{T}$.

The shortest curve in Euclidean three-space which joins a point to a plane is a line segment orthogonal to that plane. More generally, if a path minimizes the action over all paths which connect one submanifold to another in a given time interval, then this path must hit the endpoint manifolds orthogonally at the endpoint times. Upon applying this idea to our reduced action principle, we find that our shape curve $c(t)$ intersects EUL_0 orthogonally at $t = 0$ and $ISOSC_2$ orthogonally at $t = T/12$.

The equality of the masses implies that the reflections about the meridian planes $ISOSC_j$ in C are *symmetries*, meaning that they preserve the reduced Lagrangian and hence take reduced solutions to reduced solutions. Reflect our minimizing shape curve $c(t)$, $0 \leq t \leq T/12$, about $ISOSC_2$. The resulting reduced solution connects EUL_3 to $ISOSC_2$, hitting $ISOSC_2$ orthogonally with derivative opposite to c there. Thus if we concatenate c with its time-reversed reflection, the derivatives match up at $T/12$, and the result is a smooth reduced solution, still called c, now parameterized by the doubled interval $0 \le t \le T/6$. Continuing in this manner, applying reflections about meridian planes and the equator, which is a symmetry for all mass ratios, we march around the shape sphere, returning to where we started at time $6T/12$. The derivatives at $t = 0$ and $t = T/2$ are not equal, but instead are related by reflection about the equatorial plane. Continuing once more around the equator, we obtain a periodic reduced solution curve $c(t)$ in C of period T, made up of twelve pieces congruent to the initial shape curve coming from Theorem 1. It passes through the Euler configurations in the order 132132 at the specified times and the isosceles configurations at the intermediate times.

The lift $x(t) = (x_1(t), x_2(t), x_3(t)) \in Q$ of the shape curve just constructed is a solution to Newton (1). It remains to show that it is periodic of period T and has the form described. We do this by seeing how our reflections act on arcs of x. One might first guess that the symmetry $\sigma_3(x_1, x_2, x_3) = (x_2, x_1, x_3)$ of interchanging masses 1 and 2 induces the reflection about $ISOSC_3$. This is not true, since this interchange, unlike the reflection, reverses the orientations of triangles and so interchanges north and south hemispheres of the shape sphere. Instead, the interchange σ_3 induces the *half-twist* about EUL_3, by which we mean the isometry of C which is the identity on the ray EUL_3 and which rotates planes orthogonal to it by 180 degrees. The half-twist is the composition of the reflection about $ISOSC_3$ with the reflection about the equator and is the symmetry used to extend $c([0, T/6])$ past $T/6$. Now $x(T/6) = (x, -x, 0)$ with $x \neq 0$, thus $\sigma_3 x(T/6) \neq x(T/6)$, and so the concatenation of $x([0, T/6])$ with $\sigma_3(x(T/6 - t))$ yields a discontinuous curve, which is not right. To correct matters we instead concatenate x with $g \circ \sigma_3(T/6 - t)$, where g is an appropriate rigid rotation. (Any rotation g induces the identity on shape space.) The unique choice of g making the concatenation continuously differentiable at $T/6$ is rotation by 180 degrees, $gx_i = -x_i$. Since the arcs of x on either side of $T/6$ satisfy Newton (1), and since the derivatives now match, this concatenation is a solution over $0 \le t \le T/3$. The concatenation is achieved by substituting the time-reversed and translated time $T/6 + (T/6 - t) = T/3 - t$ for the old time. Hence our curve satisfies the functional equation $(x_1(t), x_2(t), x_3(t)) = (-x_2(T/3 - t), -x_1(T/3 - t), -x_3(T/3 - t))$ for $T/6 \le t \le T/3$.

The initial configuration $x(0)$ is in EUL_1 and so has the form $(0, a, -a)$ for some $a \in \mathbb{C}$. According to the functional equation, $x(T/3) = (-a, 0, a)$, or $x(T/3) = p(x(0))$ where p is the symmetry of permutation of masses: $p(x_1, x_2, x_3) = (x_3, x_1, x_2)$. The configurations $x(0)$ and $x(T/3)$ are identical, except that the masses have undergone the permutation (132). The same can be checked for the velocities. It follows that the continuation of the solution past time $t = T/3$ satisfies $x_2(T/3 + t) = x_1(t)$, $x_3(T/3 + t) = x_2(t)$, and $x_1(T/3 + t) = x_3(t)$. Armed with this functional equation, we can now continue x over the whole period T. The result is periodic of period T, since $p^3 = 1$. This also shows that all three masses trace out the same planar curve $q(t) = x_1(t)$ during the solution.

It remains to show that this planar curve $q(t)$ has the qualitative form of the eight. Using the proper choices of g in the realizations $g \circ \sigma_1$ of the half-twist about EUL_1 and of the equatorial reflection, we can check that q is odd in time t and that its image enjoys the symmetries of the x-y axes, where we take for the x-axis the symmetry axis of the isosceles triangle $x(T/12)$. The qualitative form of the eight now follows from the form taken by its first quarter, $q(t)$, $0 \le t \le T/4$. This quarter lies in the first quadrant, is star-shaped (proved), indeed convex (unproved). Star-shapedness is achieved by using minimality of c and the symmetries to show that the angular momentum $q \wedge \dot{q} = r^2 \dot{\theta}$ is positive, hence θ is monotone over this interval.

We now turn to the proof of Theorem 1, which is all that remains to demonstrate the existence of the figure-eight solution.

Hilbert's Direct Method; a Sketch of the Proof of Theorem 1

Consider the class $\mathcal{C}$ of all paths in C joining EUL_1 to $ISOSC_2$ in the time $\bar{T} = T/12$. Choose a sequence $x_n \in \mathcal{C}$ such that

$$\lim_{n \to \infty} A(x_n) = \inf_{x \in \mathcal{C}} A(x).$$

We call such a sequence a *minimizing sequence*. Try to extract a subsequence of the x_n which converges to some $x_* \in \mathcal{C}$. If we can show that x_* is collision-free, then we will be done, according to the principle of least action. This method of establishing the existence of critical points is *Hilbert's direct method* in the calculus of variations.

The existence of some limiting x_* follows quickly from the Arzelà-Ascoli theorem of real analysis. The crux of the matter is to show that x_* suffers no collisions. To do this we compute the infimum $A_{coll.}$ of the action A over all curves which suffer a collision during the specified time $\bar{T}$, regardless of whether or not these curves satisfy the endpoint conditions defining $\mathcal{C}$. Then we find a collision-free "test path" $x_{test} \in \mathcal{C}$ with $A(x_{test}) < A_{coll.}$. This will complete the proof, since $A(x_*) \le A(x_{test}) < A_{coll.}$, and so x_* must be collision-free.

To compute $A_{coll.}$ observe that we decrease the action if we replace a path $x(t) = (x_1(t), x_2(t), x_3(t))$ by a path in which one of the masses is positioned far from the other two and does not move. In other words, replace $x(t)$ by $\tilde{x}^{(n)}(t) = (x_1(t), x_2(t), x_3^{(n)})$ where $x_3^{(n)}$ is a fixed point of the plane satisfying

$$|x_3^{(n)} - x_1(t)| > |x_3(t) - x_1(t)|$$

and

$$|x_3^{(n)} - x_2(t)| > |x_3(t) - x_2(t)|.$$

In this way, a minimizing sequence $\tilde{x}^{(n)}$ realizing $A_{coll.}$ is obtained by letting two of the masses collide, while the third remains stationary and infinitely distant from them. This reduces the computation of $A_{coll.}$ to a two-body problem, one which had already been solved by Gordon [1977] and which will be described momentarily.

To construct our test path, note that the potential function $\tilde{U}$ takes on the same value $\tilde{U}(EUL)$ at all three of the Euler points $EUL_i \in S$ by symmetry. Consequently, its level curve $\tilde{U} = \tilde{U}(EUL)$ passes through each Euler point. These points are saddle points for $\tilde{U}$, and as we follow the level curve we alternate hemispheres, crossing the equator at the Euler points and nowhere else. In other words, the topology of this level curve vis-à-vis the collisions and Euler points is identical to that of the shape curve of the eight. Follow one-twelfth of this equipotential curve, namely, the upper branch connecting EUL_1 to $ISOSC_2$. Parameterize it at constant speed so that it arrives at $ISOSC_2$ in the desired time $\bar{T} = T/12$. Scale it so that it lies in the sphere $I = I_0$, thus obtaining a one-parameter family of test curves. Minimize the action with respect to the scaling parameter I_0. There will be a unique minimizing $I_0 = I_0(\bar{T})$, and this yields x_{test}. Its action depends only on the period $\bar{T}$ and the length of the starting equipotential curve on the sphere $S = \{I = 1\}$. We estimate this length with enough tolerance to guarantee that $A(x_{test}) < A_{coll.}$

Gordon and Kepler's Problem

We have just finished the sketch of our proof of the existence of the eight, except for the loose end concerning $A_{coll.}$. We now tie up this loose end by following Gordon's work on action minimization for the Kepler problem.

The action for the two-body problem is proportional to the Keplerian action $\int \frac{1}{2}|\dot{x}|^2 + k/|x|\, dt$ for the difference vector $x = x_1 - x_2$ in the plane. We insist on excluding the collision $x = 0$.

Theorem. *(Gordon [1977]) Consider the class of closed non-contractible curves of period T in the plane minus the origin. The infimum of the Keplerian action over this class is realized by any Keplerian ellipse with period T.*

The act of excluding collisions breaks up the space of closed curves in the plane into countably many components, indexed by the winding number $n \in \mathbb{Z}$ of the curve about the origin. (This index is the degree of the map $x/|x|$ from the circle into the circle.) The winding number of a single orbit of a Keplerian ellipse is either 1 or -1. Gordon asserts, among other things, that these are the only winding numbers which can be achieved by minimization. If we try to minimize the action over the winding number 2 component, then we will be led to a collision curve at the boundary between this component and the winding number +1 component.

All Keplerian ellipses of period T have the same action $cT^{1/3}$. (The constant c is $3(2\pi)^{2/3}k^{2/3}/2$.) This action is a function of the semimajor axis of the ellipse alone, as are both the energy and the period. Consequently we have a whole family of minimizers, all sharing the same infimal value of the action. At the boundary of this family lie the elliptic collision-ejection orbits representing the planet crashing into the sun and being expelled along the segment it came in on. These also realize the infimum of the action over our class despite the fact that they are not in the class but rather on its boundary.

These two linked phenomena—that of the direct method leading us to the boundary of a topological component and that of collision orbits sharing the infimal value of action—illustrate the main subtleties of using action principles for Newtonian N-body problems. Identical phenomena occur in the three-body problem. The solutions of Lagrange and Euler of a given period also occur in families parameterized by Kepler, all having the same action and including triple collision-ejection orbits on their boundary. Solutions corresponding to central configurations occur for all N, so this phenomenon persists in the Newtonian N-body problem for any N.

Still, the Keplerian ellipses do minimize, although they are degenerate minima. The crucial fact which allowed Gordon to prove this, which is to say, to exclude the possibility of more than one collision per period, is the concavity of the action $cT^{1/3}$ of a Keplerian orbit as a function of its period T. Suppose, for example, we have a minimizer with countably many collisions, with the times between successive collisions being $t_1, t_2, \ldots, t_j, \ldots$. By standard arguments from the calculus of variations, in between collisions the minimizer must be a solution to Kepler and hence must be a collision-ejection orbit. Its total action is thus $c\sum(t_i)^{1/3}$. But $\sum(t_i)^{1/3} \geq (\sum t_i)^{1/3} = T^{1/3}$ with equality if and only if there is but a single collision. This proves there is at most one collision. This is the argument which allows us to compute $A_{coll.}$.

Poincaré Minimizes over a Homology Class

In the remainder of the article we describe other applications of the direct method to N-body problems, relating some of these to the eight's construction.

Poincaré [1896] was apparently the first to apply the direct method to find new periodic solutions to three-body equations. Poincaré was most interested in reduced periodic solutions, as described earlier. Specifically, he looked for solutions $x(t)$ with the property that there is some fixed rotation g and some time T for which

$$(R) \qquad x(t+T) = g(x(t)).$$

Equivalently, these are solutions x to (1) which are not necessarily periodic but whose shape curves are periodic.

Poincaré did not want collisions any more than we do. Shape space minus collisions is homotopic to the shape sphere minus three points. Its first homology group is $\mathbb{Z} \times \mathbb{Z}$. Poincaré fixed a homology class (K_2, K_3) as well as the overall rotation g in (R). Concretely he considered a rotating orthogonal frame with one axis being the edge $x_3 - x_2$ of the moving triangle. The integers K_3 and K_2 are the winding numbers of the nonzero vectors $x_3 - x_1$ and $x_2 - x_1$ about the origin, as viewed in this moving coordinate system. The rotation g is the overall rotation of the coordinate axes in one period. Together $(K_2, K_3; g)$ specify a a class of collision-free curves over which to apply the direct method.

Poincaré knew that the Newtonian problem admits finite action solutions. Indeed, all solutions ending in collision and defined over a bounded time interval have finite action. He thus faced the possibility of action-minimizing collision-free sequences pinching off to collision. This possibility blocked him from achieving reduced periodic solutions in the given class by the direct method. To circumvent this difficulty, Poincaré changed the potential and thus the differential equation (1), a trick which dozens, if not hundreds, of researchers have used since then without knowing that Poincaré had done this a hundred years earlier.

Replace the function U in the Lagrangian by one of the general form

$$U = \sum_{a<b} f_{ab}(r_{ab}),$$

where $f_{ab}(r) > 0$, $f_{ab}(0) = +\infty$, and $f_{ab}(\infty) = 0$. The resulting Euler-Lagrange equations define a new N-body problem whose differential equation is found by replacing the right-hand side of Newton's equations (1) by the gradient of U with respect to x_i. This N-body problem will be called *strong-force* if there exist positive constants δ and c such that the two-body potential functions f_{ab} making up U satisfy

$$f_{ab}(r) > \frac{c}{r^2} \text{ whenever } r < \delta.$$

The prime examples come from the *power laws*

$$f_{ab}(r) = \frac{k_{ab}}{r^\nu},$$

with $k_{ab} > 0$. These are strong-force if and only if $\nu \geq 2$.

Poincaré observed that when the potential is strong-force, the action of any curve tending to a collision diverges. Consequently, there are no finite action curves with collision. This let Poincaré prove the following, using the direct method.

Theorem. *(Poincaré [1896]) Consider the class of collision-free reduced periodic curves specified, as above, by the period T, the winding numbers (K_2, K_3), and the overall rotation g. If the potential is strong-force, then there is a solution to Newton's equations lying in this class and realizing the infimum of the action over this class.*

There is a small oversight in Poincaré's proof. For the statement to be correct we must exclude those classes in which at least one of the K_i is zero and g is the identity. To see what goes wrong without this constraint, consider the case $g = 1$, $K_2 = K_3 = 0$. The infimal action is zero. It is realized by a minimizing sequence x_n in which the $x_n(t)$ are constant, forming the vertices of an equilateral triangle (or any fixed noncollinear triangle $\Delta \in S$) of size $\sqrt{I} = n$. As $n \to \infty$, all three masses recede to infinity and there is no limiting curve.

If, on the other hand, $K_2 K_3 \neq 0$ or $g \neq 1$, then all three masses must move in order that the curve lie in the class $\mathcal{C}$. Indeed, the projection to the three-sphere $\{I = 1\}$ must be a curve whose length is bounded away from zero by a positive constant depending only on K_2, K_3, and g. It follows that if x_n is a sequence of curves realizing this class and if $I_n = I(x_n(t_n)) \to \infty$ for some times t_n, then the lengths of the x_n must tend to infinity, which implies that their action $A(x_n) \to \infty$. This excludes the possibility of points of a minimizing sequence wandering off to infinity in the plane.

Gordon [1977] formalized the above argument. He called a class $\mathcal{C}$ of collision-free curves *tied* to the collisions if whenever we have a sequence of curves $x_n \in \mathcal{C}$ with $\|x_n(t_n)\|^2 := I(x_n(t_n)) \to \infty$ for some sequence of times t_n, the actions $A(x_n)$ must also tend to infinity. A minimizing sequence extracted from a tied class cannot fail to converge by receding to infinity.

Instead of fixing a homology class of reduced periodic curves in shape space, let us fix a homology class of truly periodic curves in the usual configuration space minus collisions. Such a class is determined by the winding numbers n_1, n_2, n_3 of the three edges $x_i - x_j$. Gordon's tied condition is that at least two of these integers are nonzero. The direct method then yields a periodic collision-free solution with these winding numbers, provided the potential is a strong-force potential.

Venturelli [2001] transcended the strong-force restriction.

Theorem. *(Venturelli [2001]) Consider the class of all closed collision-free periodic curves of period T*

in the configuration space for the planar Newtonian three-body problem whose winding numbers n_i satisfy $n_1 n_2 n_3 \neq 0$. The infimum of the Newtonian action over this class is realized by any elliptic or circular Lagrange solution of this period.

The Lagrange solutions have winding numbers $1, 1, 1$ or $-1, -1, -1$. Their homology classes are the only classes realized by action-minimizing solutions. If we fix any other homology class with all nonzero winding numbers, such as $2, 1, 1$, and try to minimize the action over this class using the direct method, then we will be led out of that class to the Lagrange collision-ejection solution. This collision solution has the same action as the noncollision Lagrange solutions.

Note the close parallel with Gordon's theorem. Gordon's theorem is one of the main ingredients in Venturelli's proof. Another key ingredient is that the Lagrange configuration minimizes $\tilde{U}$.

Eclipse Sequences: Minimizing over a Homotopy Class

Instead of fixing the homology class of a reduced periodic curve, we now fix its free homotopy class. We explain how to encode such a class into an *eclipse sequence*. A loop in $C \setminus collisions$ radially projects to a loop in $S \setminus collisions$. Perturbing the loop slightly makes it intersect the equator of collinear triangles a finite number of times, each transversally. Each intersection represents an *eclipse*, a configuration in which the three masses are collinear, one lying between the other two. We label these eclipses 1, 2, or 3 according to which body is in the middle. Consequently, our curve yields a finite sequence, or "word", $abc\ldots$, a, b, and c being 1, 2, or 3 written down in the order they appear. We call such a word the *eclipse sequence* of the curve. For example, Hill's solution realizes the eclipse sequence 12 if 1 and 2 are the close bodies, while the figure eight has eclipse sequence 132132.

If two consecutive eclipses aa of the same type occur, then they can be deleted by pulling the arc between the two eclipses up into the opposite hemisphere. This deletion decreases the action, so it is consistent with the direct method. This process allows us to delete all "stutters" $\ldots aa \ldots$ occurring in our word. Since we are interested in periodic orbits, we should think of our eclipse sequence as being written out along the circumference of a circle, i.e., as a periodic word. For example, the class of the figure eight is $132132 = 321321 = 213213$. We insist that each such cyclic permutation be stutter-free. A cyclic word in the letters 1, 2, and 3 subject to this stutter-free rule will be called a *grammatically correct eclipse sequence*. Such an eclipse sequence uniquely specifies the free homotopy of the collision-free shape curve once

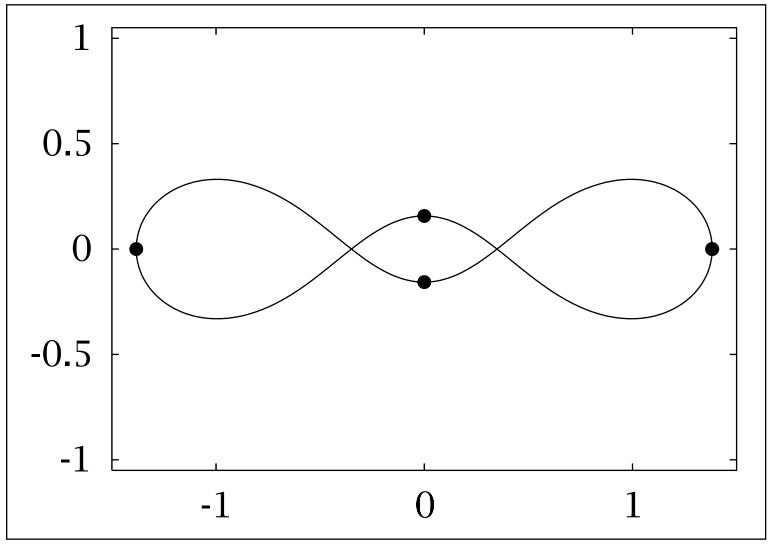

Figure 5. Gerver's solution.

we decorate the letters with superscripts + and − to indicate whether the corresponding ecliptic arc is approached from the northern or southern hemisphere. The pluses and minuses must alternate. A given eclipse sequence has exactly two such decorations, corresponding to a curve and its reflection. We have established a one-to-one correspondence between grammatically correct eclipse sequences and free homotopy classes in $C \setminus collisions$, up to reflection.

Open Problem. *(Wu-Yi Hsiang) Is every grammatically correct eclipse sequence realized by a solution to the three-body problem?*

Wu-Yi Hsiang posed this question in the spring of 1995. It helped lead to the construction of the figure-eight solution.

Cheating by assuming the strong-force condition and applying the direct method just as Poincaré did, we get an answer.

Theorem. *(Montgomery (1998)) Let $T > 0$ be given. For the three-body problem with a strong-force potential, any given eclipse sequence in which all three eclipse letters 1, 2, and 3 appear is realized by a reduced periodic solution of period T which minimizes the action over this eclipse sequence class.*

Gordon's tied condition is precisely that all three eclipse letters occur.

Poincaré, the inventor of the homotopy group, clearly had the tools to state and prove the above theorem. It is slightly surprising that he did not.

Summary. The set of periodic curves in C breaks up into infinitely many components when we exclude collisions. These components are indexed by eclipse sequences. The (reduced) action function creates walls between them. When we take the potential to be strong-force, these walls are

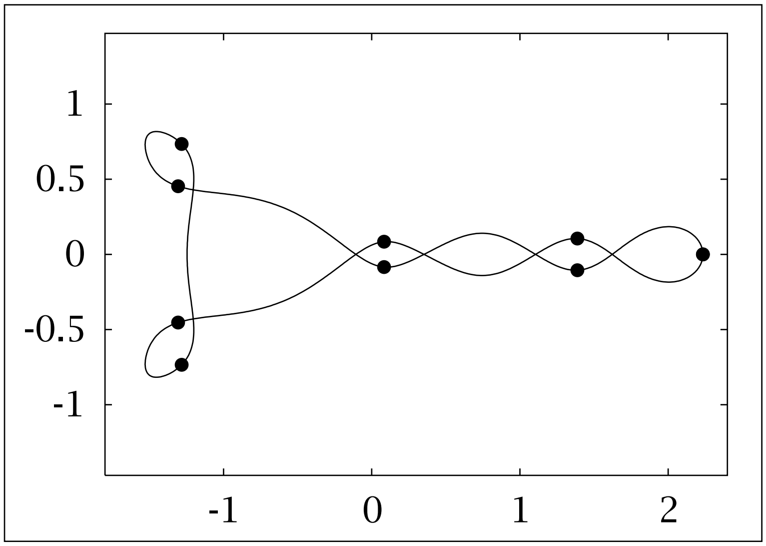

Figure 6. A choreography of nine bodies.

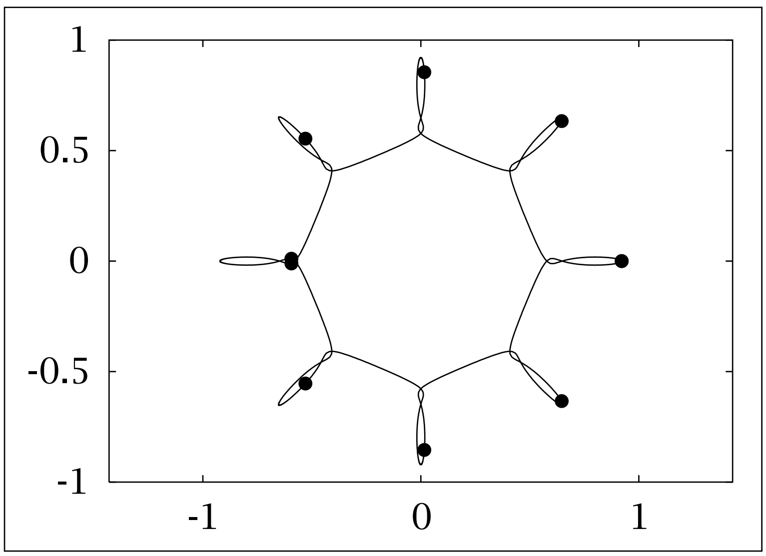

Figure 7. A floral arrangement with eleven bodies.

infinitely high, so we can minimize separately in each component. But as we lower the exponent ν of $f_{ab} = m_a m_b / r_{ab}^{\nu}$ coming from the power law potential above and so try to approach the case of real interest (the Newtonian case of $\nu = 1$), the walls descend in height, becoming finite in height as soon as $\nu < 2$. Then we can cross over from one component to the other. The walls are probably still steep and sharp in most places, but crossing is now possible. The Hilbert direct method might even lead us to "valleys" or "breaks" in these walls, forcing the limit x_* to have collisions. Because of this we can say little to nothing at present about whether or not the given eclipse sequence is realized by an action minimizer.

The situation is worse. The Lagrange collision-ejection orbit x_L is a kind of "universal" low pass allowing us to cross from any component to any other. Indeed, it lies in the boundary of every component and has very small action, hence is a very attractive place to head. See the discussion of Venturelli's theorem and also Montgomery [2000]. It could be that the direct method applied to any component will always lead to x_L in the Newtonian case. If so, the direct method would yield no solutions beyond Lagrange and thus no eclipse sequences beyond the empty sequence, which is the sequence for Lagrange.

We were able to get around these difficulties in constructing the figure eight because of our assumption of equal masses, with its consequent discrete symmetries. In effect, these allowed us to replace the specification of the eclipse sequence 132132 with the endpoint conditions appearing in Theorem 1.

New N-Body Solutions

We ask N equal Newtonian masses to dance around a fixed curve. The eight is such a solution, as is the circular Lagrange solution. For each N we can obtain such a solution where the curve is a circle by placing the N points at the vertices of a regular N-gon inscribed in the circle and then rotating this N-gon at the proper frequency. Additional orbits like this were found numerically by Davies et al. (1983) and G. Hoynant (1999) in space and by Moore [1993] in the plane. Right after our rediscovery of the eight, Gerver conjectured the existence of solutions of this type for all N, with the number of distinct topological types of solutions increasing rapidly with N. He numerically found one of his conjectured solutions when $N = 4$. In Gerver's solution the configuration formed by the four bodies is a parallelogram at every instant, and the curve they move on is a "super-eight", a figure eight with an extra twist. Simó [2000a] numerically verified Gerver's conjecture, finding hundreds of new equal mass planar N-body solutions. See also Chenciner et al. [2001]. Of all these solutions, apparently the only stable one is the eight. And, except for the eight, we lack rigorous existence proofs for any of these new Newtonian N-body solutions.

To make N equal bodies perform a desired dance, begin with the circle $S^1 = \mathbb{R}/T\mathbb{Z}$ of circumference T. The cyclic group $\mathbb{Z}_N$ of order N acts on this circle with its generator ω acting by $\omega(t) = t + T/N$, which is to say by rotation by $2\pi/N$. It acts on $\mathbb{C}^N$ by $\omega(x_1, x_2, \ldots, x_N) = (x_N, x_1, x_2, \ldots, x_{N-1})$. Then $\mathbb{Z}_N$ acts on the space of all loops $x : S^1 \to \mathbb{C}^N$ by $(\omega x)(t) = \omega(x(\omega^{-1}(t)))$. A fixed point of this action on loops is a map $x : S^1 \to \mathbb{C}^N$ satisfying $x_{j+1}(t) = x_1(t - jT/N)$. We call such a map a *choreography*. In a choreography all N masses travel along the same closed planar curve $q(t) = x_1(t)$, staggered in phase from each other by T/N.

This action of $\mathbb{Z}_N$ on loops x leaves the N-body action $A(x)$ invariant when the masses are all equal. It follows from a general principle, which Palais calls the principle of symmetric criticality, that if we restrict A to the fixed points of the action—the choreographies—and find a collision-free choreography which is a critical point for this restricted A, then this choreography will be a solution to the N-body problem. Excluding collisions breaks up the set of choreographies into countably many different components, which we call *choreography classes*. By the argument which Poincaré used above, if the potential is strong-force, then there is a solution realizing each choreography class.

Simó has numerically implemented this equivariant action-minimization procedure for the Newtonian problem. He discretizes the gradient flow of the action, starting with an initial guess for a choreography. The resulting iteration scheme stays within the space of (discrete) choreographies and will either converge to a choreography with collisions or to a collision-free choreography. (In the strong-force case the first possibility is excluded.)

In this way Simó has produced a huge number of new solutions to the equal mass Newtonian N-body problem. Apparently all choreographies are unstable except the original figure eight for $N = 3$. Simó's results suggest that for each N only finitely many choreography classes are realized by local action minimizers. For $N = 3$ it seems only two choreography classes are realized: that of the Lagrange solution and that of the eight. What topological criterion selects the classes which are realized by a Newtonian choreography? An answer to this question might lead to an existence proof for this horde of new orbits and perhaps a better understanding of the planar N-body problem.

Acknowledgments

I thank Alain Chenciner for numerous e-mail conversations and critiques, Carles Simó for most of the pictures here, Robert MacKay and Phil Holmes for alerting us (via A.C.) to Poincaré's variational work [1896] and to Chris Moore's [1993] discovery of the eight, Wu-Yi Hsiang for starting me on these problems, and finally Gil Bor for his strong criticism of an early version of this article.

Bibliography

[1] A. Chenciner and R. Montgomery [2000], A remarkable periodic solution of the three-body problem in the case of equal masses, *Ann. of Math.*, November.

[2] A. Chenciner, J. Gerver, R. Montgomery, and C. Simó [2001], Simple choreographic motions of N-bodies, a preliminary study, to appear.

[3] Leonard Euler [1767], *De motu rectilineo trium corporum se mutuo attrahentium,* Novi commentarii academiæ scientarum Petropolitanæ 11, pp. 144–151 in Oeuvres, Seria Secunda tome XXV Commentationes Astronomicæ (page 286).

[4] W. B. Gordon [1977], A minimizing property of Keplerian orbits, *Amer. J. Math.* **99**, 961–971.

[5] Joseph-Louis Lagrange [1772], *Essai sur le Problème des Trois Corps,* Prix de l'Académie Royale des Sciences de Paris, tome IX, in vol. 6 of Oeuvres (page 292).

[6] R. Moeckel [1988], Some qualitative features of the three-body problem, Contemp. Math., vol. 81, Amer. Math. Soc., Providence, RI, pp. 1–21.

[7] C. Moore [1993], Braids in classical gravity, *Phys. Rev. Lett.* **70**, 3675–3679.

[8] H. Poincaré [1896], Sur les solutions périodiques et le principe de moindre action, *C.R.A.S. Paris* **123**, 915–918.

[9] C. Simó [2000a], New families of solutions in N-body problems, to appear in the *Proceedings of the ECM 2000* (Barcelona, July 10–14); see `http://www.iec.es/3ecm/`.

[10] _____ [2000b], Dynamical properties of the eight solution of the three-body problem, to appear in the *Proceedings of the Chicago Conference Dedicated to Don Saari* (December 15–19, 1999).

[11] A. Venturelli [2001], Une caractérisation variationnelle des solutions de Lagrange du Problème plan des trois corps, to appear in *C.R.A.S. Paris.*

About the Cover

The cover shows several frames of an animation constructed by Carles Simó—redrawn slightly—that exhibits the periodic orbital motion of six objects of equal mass on a single orbit. This is what is called in Richard Montgomery's article a "simple choreography". The original animation is number 42 in a spectacular series of animations by Simó which run in `gnuplot`. The whole series can be obtained from `http://www.maia.ub.es/dsg/nbody.html`.

As Montgomery explains, almost all of these choreographies are known at the moment only through computer calculation, primarily the work of Carles Simó. Producing them is a difficult but interesting process. To see what is involved one should look at various papers of Simó. Many recent ones are available as preprints at `http://www.maia.ub.es/dsg/`.

Simó is currently working on finding rigorous computer-assisted proofs of the existence of choreographies.

—*Bill Casselman* (`covers@ams.org`)