Four Open Questions for the N-Body Problem

Richard Montgomery

Contents

Dedication

This book is dedicated to the memory of Jerry Marsden for introducing me to geometric mechanics, to Chris Golé for insisting that I must introduce myself to Alains Albouy and Chenciner, and to my grandson Jude for his unending fascination and joy at being an intermediary in the interactions of elastic spherical objects with gravity and pavement.

1. Preface and book structure

Probably turn this section 1. into a 'Preface' -RM

We've been working on the N-body problem for more than 330 years. You might think we'd be done with it. But the problem remains very much alive. Substantial results have been achieved in the last decade. New open questions continue to arise. The purpose of this monograph is to illustrate the vibrancy of the classical N-body problem through four open questions.

Each question gets a chapter. There I explain what the question is asking and why it is important. I state known partial answers and take a shallow dive into methods employed in making what progress has been made so far.

My favorite case of the N-body problem is the zero angular momentum planar three-body problem. All of the questions but the first are open for this case. The third has been solved for the three-body problem when the angular momentum is ϵ, small but nonzero. The first is open for the planar 6-body problem.

This book was inspired by preparing for a talk at Vanderbilt after the colloquium chair emailed me the traditional warning research mathematicians seem to need before giving such a talk: "Remember, this is a general audience talk." I wrote this book with colleagues in number theory and algebraic geometry in mind, and with our stronger undergraduates in mind. I do not expect working knowledge of introductory physics. I do expect proficiency in linear algebra, and the understanding of what a differential equation is. I began with this introductory chapter to pose the N-body problem and some of the basic results surrounding it.

2. Introduction

Newton (Newton 1667) posited that any two masses in the universe attract each other by a gravitational force. He took this force to be proportional to $1/r^2$ where r is the distance between the masses. Supposing the universe to be populated by a finite number N of point masses subject only to their mutual gravitiational attractions and his laws of mechanics he formulated a set of differential equations (eq 5 below) which describe the motions of these N point masses. **The classical N-body problem is the study of this dynamical system.**

Newton solved his 2-body problem. Supposing these two masses to be the sun and a planet he derived Kepler's three laws of planetary motion and thereby gained himself a pre-eminent role in the history of science. See tour; figs 2.

More than two hundred years later Poincaré proved that a limiting case of the 3-body problem is unsolvable in a certain technical sense: it admits 'homoclinic tangles" and therefore is not " integrable by analytic functions". The effect of his proving unsolvability was analogous to that of Galois' proving the unsolvability of the general quintic. Rather than killing their subjects, they developed methods in establishing their impossibility results which opened up vast and previously unimagined vistas of research. For Galois these vistas included group theory, algebra, and number theory. For Poincaré the vistas were nonlinear qualitative dynamical systems (popularly referred to as "chaos theory") and its interaction with topology and analysis.

Part 1

Introduction

CHAPTER -1

A Tour of Solutions

ch: tour

The classical planar N-body problem is a system of differential equations (see equation 5) whose solutions are N parameterized curves in the plane which represent motions of planets or stars. We tour a few of the solutions, using N as a parameter. [1]

1. Two bodies

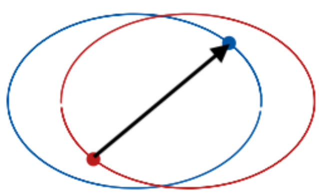


FIGURE 1. Equal mass two body problem

The Newtonian two-body problem reduces to an ODE known nowadays as Kepler's problem. The vector joining the two bodies moves on a Keplerian orbit. As a result each body moves on a Keplerian conic about their common center of mass. The center of mass moves along a straight line at constant speed. In the figure we show a solution to the equal mass two body problem. In the next subsection we go over the solutions to Kepler's problem.

FIGURE 2. Radial velocity dispersion

2. Kepler

[1] We often get a better sense of solutions through animations instead of stills, but this here is a book. Please see WEB PAGE? for animations.

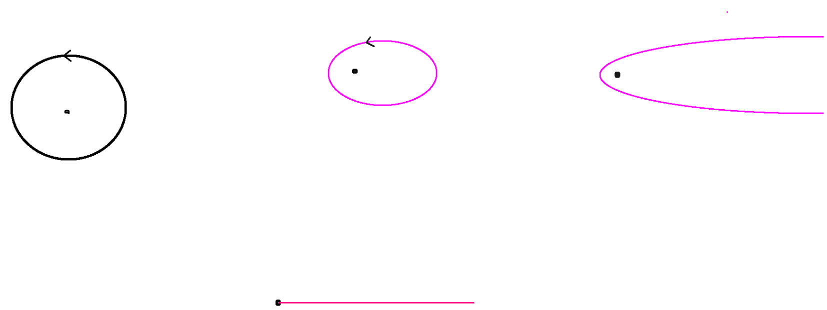


FIGURE 3. Solutions to Kepler's Problem

These days Kepler's problem refers to the planar ODE written here as equation (28). Kepler never wrote down ODEs. But Newton did and he argued successfully that this ODE governs the motion of a planet under the gravitational influence of an infinitely more massive sun. He proved that the solutions to Kepler's problem obey Kepler's three laws and in particular parameterize conic sections with one focus at the sun. The bounded collision-free Kepler solutions are all periodic and travel ellipses or circles. As the angular momentum degenerates to zero the ellipses degenerate to line segments colliding with the sun.

r and Lagrange solutions

3. Three bodies

3.1. The Central Solutions of Euler and Lagrange. There are exactly five solutions to the three-body problem for which we have closed-form expressions. Three were found by Euler in 1767. The remaining two were found by Lagrange ([**75**], [**74**]) in 1772. In their solutions all three bodies move along Kepler conics, in such a way that the "shape" of the triangle formed by the three bodies, viewed up to similarity, remains constant. The three conics making up a solution are scaled, rotated versions of the same Kepler conic and this can be any Kepler conic. Collectively, these five families are known as central solutions.

Tour: Lagrange

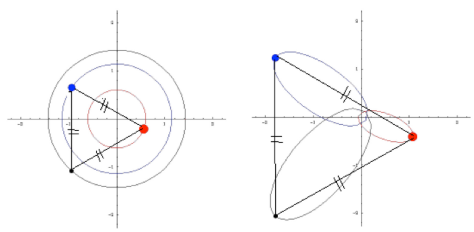


FIGURE 4. Lagrange's solution. Courtesy of an animation made by Rick Moeckel. See `http://www.scholarpedia.org/article/3-body_problem`

3.1.1. *Lagrange's solution.* For Lagrange's solution the three bodies forms an equilateral triangle at each instant. If we take our Kepler conic to be a circle then this equilateral triangle simply rotates rigidly about the center of mass of the triangle. If we take a general Keplerian conic then each body moves in a Keplerian conic about the center of mass, with the motion on the conics synchronized so that at each instant they form an equilateral triangle. Included in the family is the degenerate zero-angular momentum solution where the bodies are dropped from rest at an equilateral configuration, in which case each body moves on a line segment ending in simultaneous triple collision at the center of mass. The Lagrange families exists for all mass ratios. We count the Lagrange solutions as two families of solutions since we distinguish between 'right handed' and 'left handed' planar labelled triangles. See figure 5. The circular Lagrange solutions are linearly stable only when one of the two masses is much greater than the other two.

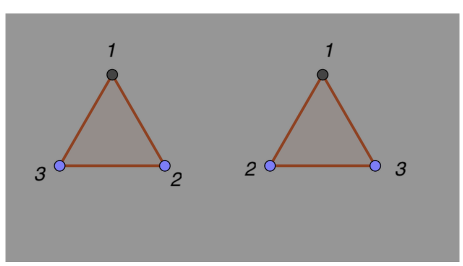


fig: trianglesRL

FIGURE 5. The two orientations of a labelled planar triangle

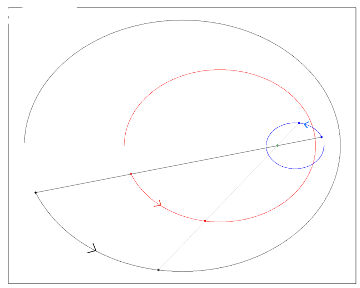


fig: Euler

FIGURE 6. Euler's Solution

3.1.2. *Euler's solutions.* For Euler's solutions the bodies are collinear at every instant, but the line they are on is typically spinning. They comprise three families labelled according to which of the three bodies sits in the middle during the motion. Throughout the motion the ratio between the distances of the bodies remains constants, the ratio being determined as the positive real solution to a fifth order polynomial whose coefficients depending on the masses. When the masses all equal then this middle mass must lie at at the midpoint of the segment formed by the other two. If we choose the circular solution to Kepler's equations then these two others simply rotate as the ends of a diameter of a circle about this center. See the left panel of figure 6. The Euler solutions are linearly unstable for all mass distributions.

3.1.3. *Prelude to Question 1.* The ansätz leading to the Euler and Lagrange solution works for more than three bodies. The resulting families of solutions keep the same shape over time and these shapes are called "central configurations". Question 1 asks if the number of central configurations is finite for all $N > 3$.

3.2. The circular restricted limit. The Euler and Lagrange solutions persist if we let one of the three masses tend to zero while the other two are fixed. If we take the circular Kepler element of the family, then these limiting solutions become special solutions to the *circular restricted planar three-body problem* which is the most studied of all three-body problems. In this limiting version, the two big masses, or "primaries" rotate in a circular Keplerian orbit. Newton's equations, written in coordinates rotating at the angular frequency of the primaries, reduces to a 2nd order autonomous ODE in the rotating plane for the remaining infinitesimal mass. The corresponding ODE in the rotating frame has 5 fixed points corresponding to these 5 limiting central configurations. These fixed points have named L1-L5 and are indicated in figure **??** Somehow Euler's name disappeared in their labelling.

fig: ccRestricted

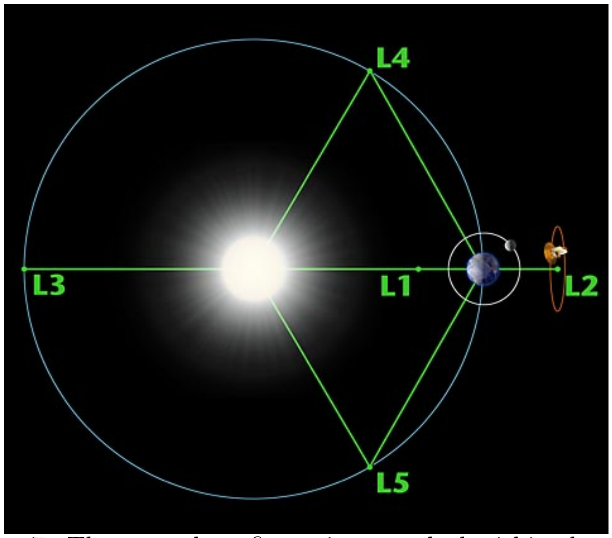


FIGURE 7. The central configurations marked within the circular restricted three body problem. Courtesy of NASA. `https://map.gsfc.nasa.gov/media/990528/990528b.jpg`

A large fraction of the existing work on the three-body problem concerns the circular restricted planar three-body problem. Figure 3.2 represents some of that work pre-dating electronic computers. I will return to restricted problems and perturbation theory at the end of the tour

will you? check? -RM

fig: Stromgren

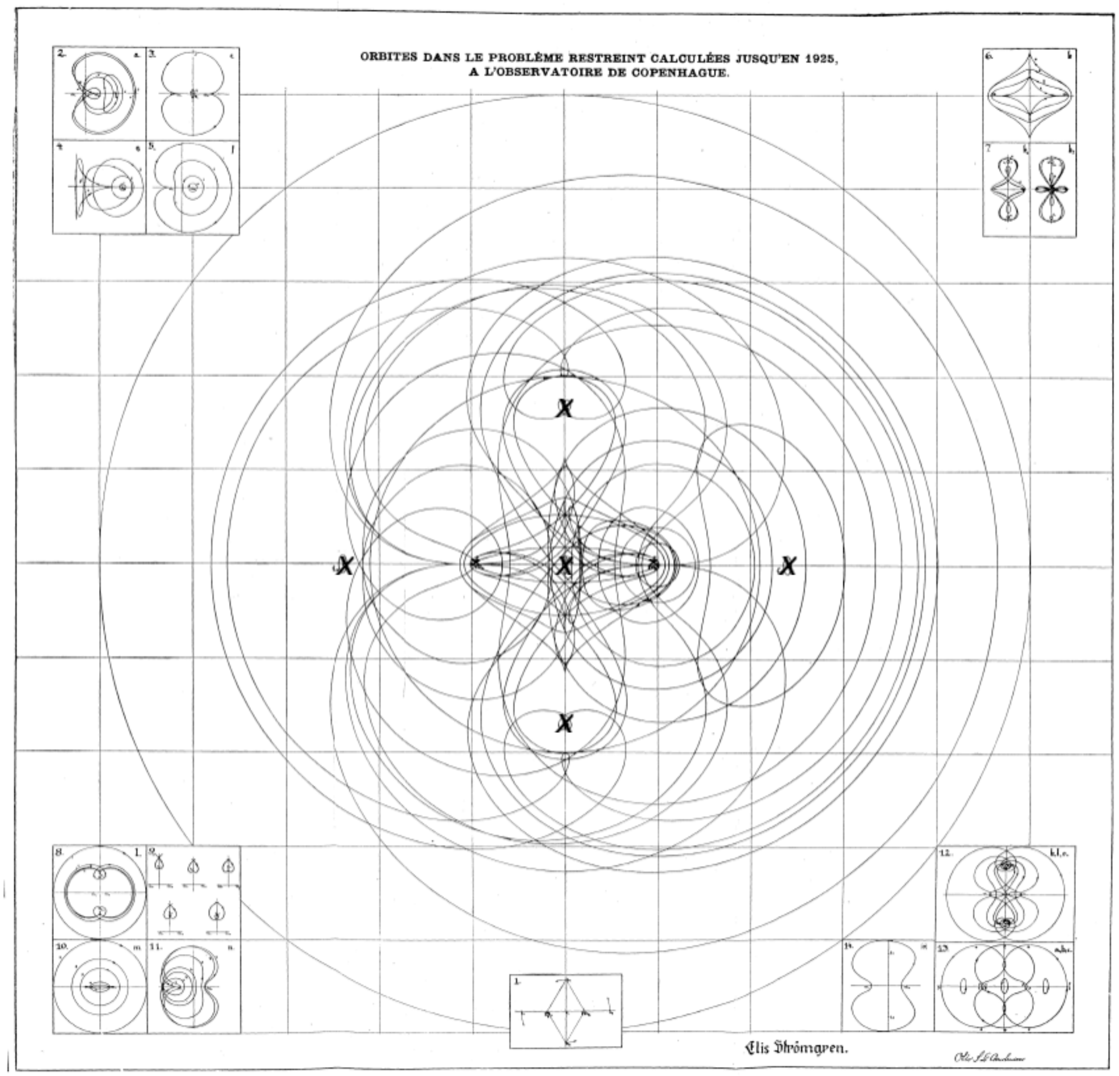


FIGURE 8. A few dozen periodic orbits for the planar restricted circular three-body problem in the case where the two primaries have equal masses, viewed in the rotating frame. The $*$s indicate the primaries and the xs the central configurations of Euler and Lagrange. Strömgren [**151**] found these in 1933 through numerical integration before digital computers were available.

tour: eight

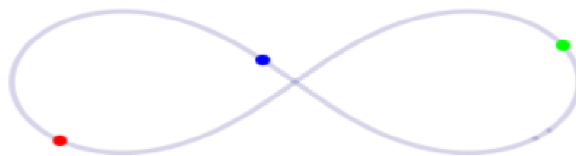


fig: fig8

FIGURE 9. Figure 8.

3.3. The eight. At the other extreme from having one of the three masses zero, take all three masses equal. In the figure eight solution three *equal* masses chase each other around a fixed figure eight shaped curve drawn (here laid on its side) on the plane. To generate the eight with a numerical integrator take inital positions as $(x_1, y_1) = (-0.97000436, 0.24308753), (x_2, y_2) = (-x_1, -y_1), (x_3, y_3) = (0, 0)$ with initial velocities $(\dot{x}_1, \dot{y}_1) = (\dot{x}_2, \dot{y}_2) = -\frac{1}{2}(\dot{x}_3, \dot{y}_3)$ where $(\dot{x}_3, \dot{y}_3) = (0.93240737, 0.86473146)$. Take all masses and G to be 1. I'm indebted to Carlés Simo for zeroing in on these initial conditions. (See [**144**].)

Unexpectedly, the eight is as stable as one can hope for in celestial mechanics: it is KAM stable. Question 2 of this book asks whether or not there are any periodic orbits that are Lyapunov stable which is the standard accepted meaning of "stable" in modern dynamics. See in particular section 2.10 for stability of the eight. Figure 3.3 depicts a perturbed eight, a solution resulting from near-eight initial conditions. We can imagine this one, which is periodic 'nestled between" two KAM 2-torii.

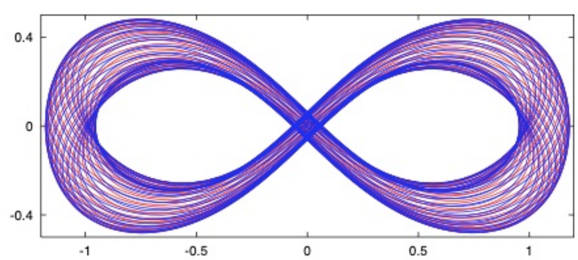


fig: near8

FIGURE 10. A near-eight. Courtesy of C. Simo. See [**144**] for more details.

3.4. Escape and scattering. Drop three bodies: let them go from rest. Or throw a binary system at a distant isolated third body. What can happen?

The first case was investigated for nearly a century in a long and convoluted series of papers for a specific case known as the 'Pythagorean three-body problem" For initial conditions place three masses at the vertices of a triangle whose edge lengths are in the ratio $3:4:5$. Choose the masses to have the same ratio, and place mass 5 opposite the edge of length 5, etcetera. See figure 3.4. Now drop the bodies: take all velocities zero, and numerically integrate. What happens? See figure 3.4

fig: Pythag0

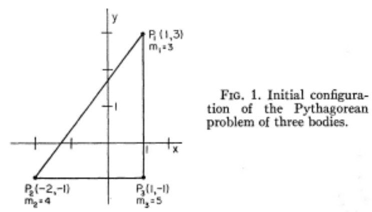


FIGURE 11. Initial conditions for the Pythagorean 3 body problem.Reprinted from Szebehely and Peters [].

fig: Burrau

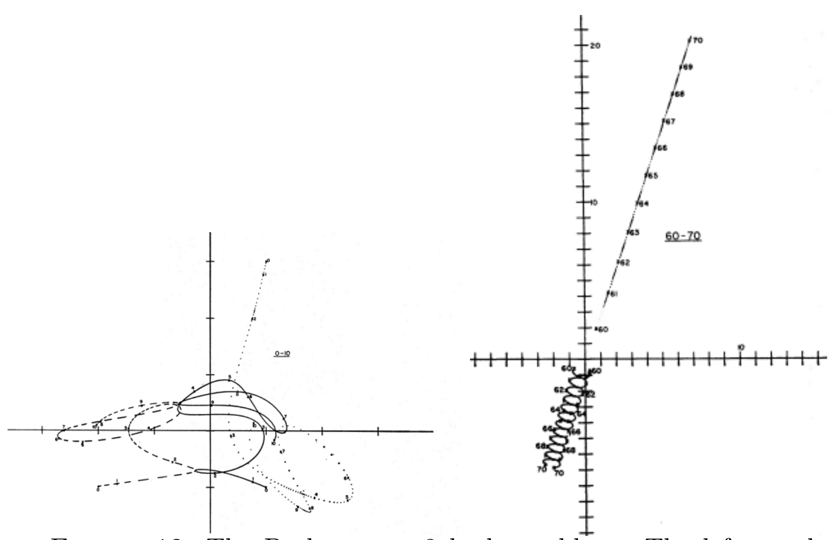


FIGURE 12. The Pythagorean 3 body problem. The left panel depicts the solution from being dropped at $t = 0$ to time $t = 10$ with one body's curve depicted as a solid line, one as a dashed curve and the other as a dotted curve. In the right panel the motion from $t = 50$ to 60 is depicted. We see that two of the masses form a tight binary and escape to infinity. Reprinted from Szebehely and Peters [**152**].

Hut, and collaborators threw binary stars at isolated stars in order to understand questions regarding the prevalence of binary stars in galaxies and their effects on galactic evolution. A sample solution is depicted in 3.4.

fig: interchange

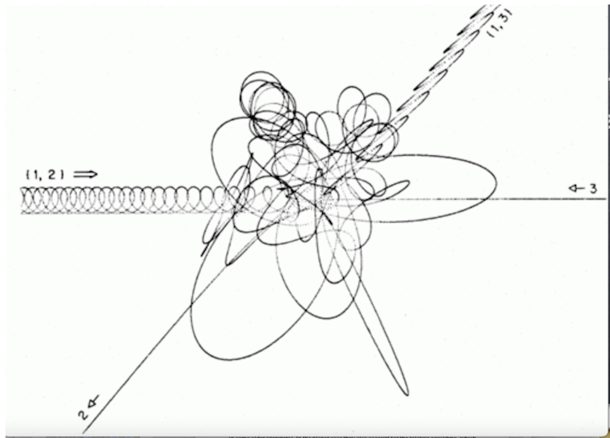


FIGURE 13. Chaotic interchange. Courtesy of P. Hut.

3.5. Schubart and Broucke-Henon. Place three bodies on the line and let them evolve according to gravity and they are typically forced to collide. Levi-Civita showed us how to analytically extend the N-body flow through isolated binary collisions and this allows us to discuss periodic solutions having such collisions. . In the year I was born Schubart [**137**] constructed such a solution for the collinear three-body problem depicted in figure 3.5. His orbit dominates the phase portrait for the negative energy collinear three-body problem. In 1974 Broucke [**16**] found many new orbits for the three-body problem and the last one he displayed has the form of the 3rd and 4th orbit in 15. Henon [**57**] used orbit continuation with the angular momentum acting as a bifurcation parameter, to follow Schubart's orbit through to Broucke's. continued Schubart's orbit into the planar problem, thus discovering what is now called the Broucke-Henon orbit 15.

appendix on– Levi-Civita [**?**] ? etc: Knauf? our paper? Heggie? -RM

figure of the Finlanders here? -RM

find ref -RM

discuss Broucke Henon vs Schubart in the braid chapter? -RM fig: Schubart

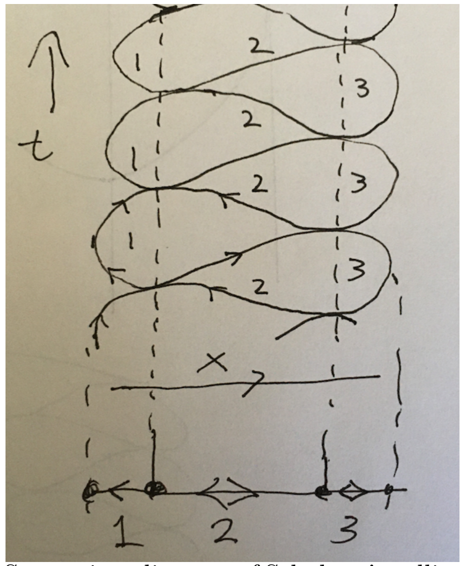


FIGURE 14. Space-time diagram of Schubart's collinear equal mass three-body solution.

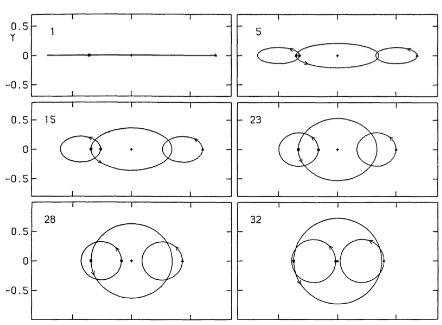


FIGURE 15. Broucke and Henon's continuation of the Schubart orbit. The orbits are shown in a rotating frame with respect to which the orbit is periodic. From [**57**]

fig: Broucke

3.6. A Bestiary. In 2015 Danya Rose [**131**] combined simultaneous regularization of all binary collisions, shape space thinking, and careful numerical integrations to catalogue 100s of equal mass zero angular momentum solutions to the three body problem. Figure 3.6 is an unstable orbit that Rose christened F1.2.7.

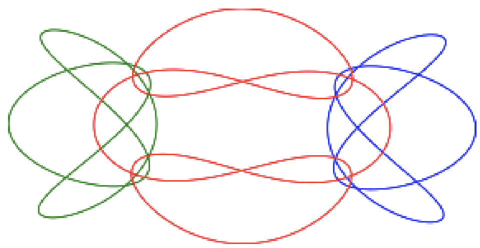


fig: bestiary

FIGURE 16. Another equal mass zero angular momentum solution, one of 100s found by Danya Rose.

In 2013 James Montaldi and Katrina Steckles [**106**] compiled an artistic bestiary of a different nature, based on equivariant homotopy ideas combined with the symmetry ideas that led to the figure eight. Below is a representative orbit of theirs whose existence has not yet been rigorously established.

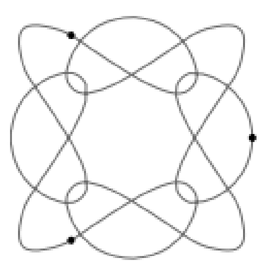


FIGURE 17. The three-body 'Celtic Knot' choreography denoted $D(3,4)$ by Montaldi and Steckles [**106**].

4. Four bodies

4.1. Central configurations. There are analogues of the Euler and Lagrange three-body solutions in which the four bodies keep the same shape throughout the motion. The shapes themselves are referred to as "central configurations". For the equal mass planar four-body problem there are precisely four possible shapes. See figure 18. When the vertices of these 4-gon shapes are labelled with the mass labels they yield 50 distinct labelled shapes modulo scaling and rotation, and hence 50 families of solutions, the analgues of the Euler and Lagrange three-body solutions. Counting the number of central configurations for $N \geq 4$ is the subject of question 1 of the book. The simplest four-body central solution to visualize is perhaps a rotating square at whose vertices are placed the four equal masses.

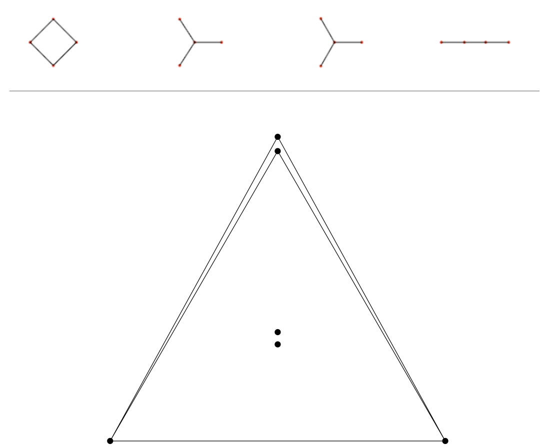


FIGURE 18. Four body central configurations for equal masses. Two of the four shapes on the left panel look the same. One of these two is an equilateral triangle with the fourth body placed at the center of mass. The other is isosceles but not equilateral. On the right panel we've blown up these two and super-imposed them after a rotation to show that they are actually different. Courtesy M. Hampton.

fig: 4bdyccs

4.2. Dancing quadrilaterals. A vast array of equal mass periodic orbits with high symmetry have been discovered in the last 25 years by combining symmetry and variational methods. Below are a few. See figure 19, **??**

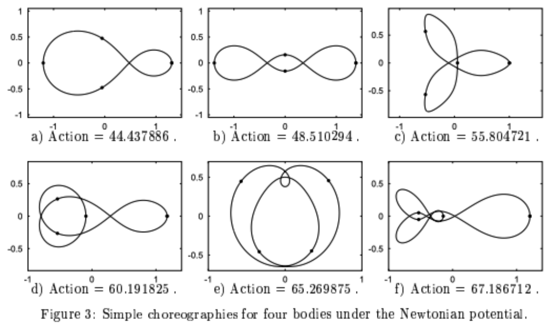


fig: 4bdychoreo

FIGURE 19. 4 body choreographies. Courtesy C. Simo.

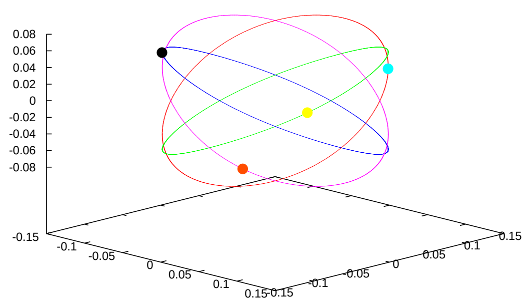


fig: hiphop

FIGURE 20. The hip-hop solution discovered by Chenciner and Venturelli [**?**] has 4 equal masses oscillating between the square and a tetrahedral configuration and enjoys the symmetry group $\mathbb{Z}_2 \times \mathbb{Z}_4$. Courtesy, Davide Ferrario.

4.3. Non collision-singularities. A solution has a singularity if there is a finite time t_* beyond which it cannot be continued as a solution. In collision solutions all bodies have limiting positions as t_* is approached, and at least two of them collide. In the three-body problem all singular solutions are collision solutions. Are there non-collision singularities for the $N > 3$ body problem? This old question was resolved by Gerver and Xia ($N = 5$). In 2022 Gerver et al **??** established existence of non-collision singularities for the 4 body, following up on earlier work by XXX. See figure 21.

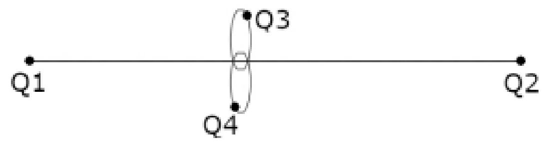


FIGURE 21. Cartoon of a non-collision singularity solutions with four bodies found by [**43**]. None of the bodies have finite limiting positions as $t \to t_*$, the singularity time. $Q_3(t)$ and $Q_4(t)$ accumulate onto the entire x axis with the distance between them going to zero, while $Q_1(t) \to (-\infty, 0)$ and $Q_2(t) \to (+\infty, 0)$. If $r_{ij}(t) = |Q_i(t) - Q_j(t)|$ is the distance between bodies i and j at time t then $\limsup_{t \to t_*} r_{ij}(t) = \infty$ and $\liminf_{t \to t_*} r_{ij}(t) = 0$ provided $i = 1$ or 2 and $j = 3$ or 4. Think $\frac{1}{|t|}|\sin(1/t)|$ with $t_* = 0$.

fig: noncollision

5. More.

A number of infinite families of choreographies have been established. Figure 23 is a sample. The chains with an odd number N of bodies travelling a chain of $N - 1$ loops has been established for all N, for example.

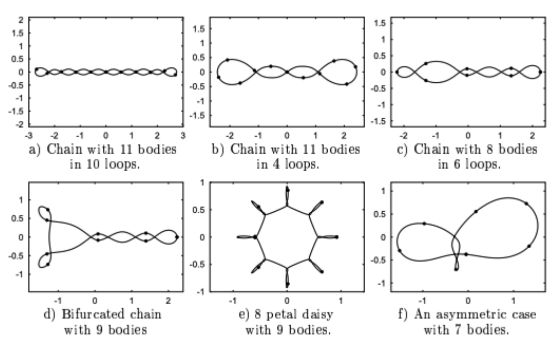


fig :nbdychoreos

FIGURE 22. A sampling of N body choreographies. Courtesy of C. Simo.

The variational and symmetry methods which established the existence of the figure eight solution have allowed researchers to discover and establish the existence of solutions having the symmetries of each of the Platonic solids. Below is depicted a 60-body solution enjoying dodecahedral symmetry, 60 being the order of the subgroup of symmetries of the dodecahedron which are orientation preserving.

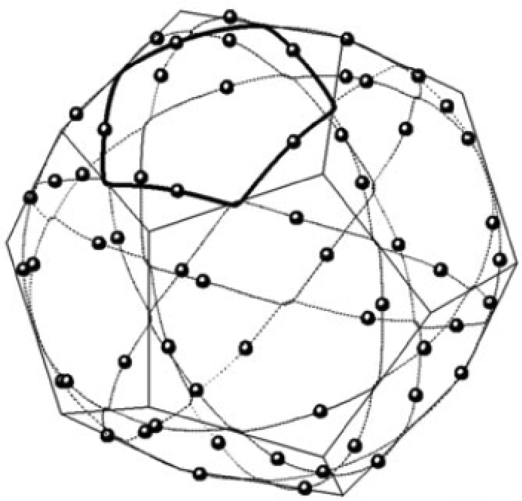


fig :nbdychoreos

FIGURE 23. A 60-body solution enjoying dodecahedral symmetry. Associated to each of 12 faces are 5 bodies moving in a choreography. Courtesy of Gronchi [**48**].

6. Perturbation theory and a return to the Solar system.

Rewrite this ** blah-blah-blah ** section. Make it shorter. Less philosophy and less personal. I think. -RM

A person has yet to look out into the night sky and find three stars moving in a figure eight orbit, let alone a constellation of 60 stars moving in the pattern of the dodecahedral orbit. The vast majority of orbits described by astronomy are Keplerian or slightly perturbed Keplerian orbits. Keplerian thinking, with a smattering of GR thrown in, continues to dominate astronomy as it appears that near Keplerian motions dominate the universe.

Our solar system's 8 planets move in nearly Keplerian orbits about the sun. Their motions themselves are almost always described in terms of deviations from purely Keplerian motion. [Figure XX Laskar]

The exo-planets were first discovered through a signature of two-body dynamics known as radial dispersion: a close-in Jupiter-mass planet makes its sun's orbit wobble enough that Doppler dispersion can detect it, leading to estimates of the period and planetary mass. Dark matter was originally discovered through discrepancies of stellar motions in galaxes from that of a model Keplerian motion as inferred by the guessing the density of the galaxy through visible stars.

The moon's orbital motion is the closest and perhaps most striking situation in our solar system in which non-Keplerian motion is visible to the naked idea on a time scale of decades. For the moon the strength of the Earth's gravitational force and the Sun's are comparable, which makes the Earth-Moon-Sun system a true three-body problem, not amenable to a naive perturbation theory.

Because of the enormous importance of perturbation theory and averaging to celestial mechanics, that is where the bulk of the theoretical work has gone, particularly before high-speed numerical computer simulations.

....

XXX

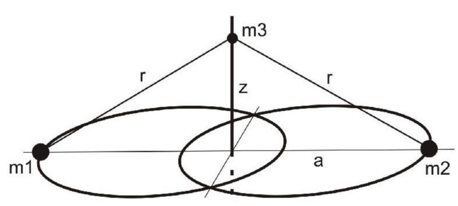
deviations from purely Keplerian behaviour are well described using perturbation theory. The mass of the heaviest planet, Jupiter, is about $1/1,000$ the mass of the sun, and Jupiter in turn is about 300 times more massive than the Earth. Trying to understand the effect of these perturbations led to deep and various developments in differential equations and analysis. vast majority of phenomenon observed XXX

As evidenced by the tour thus far, and underlined in this subsection, this book was written by a mathematician, not an astronomer. But the subject, and a great deal of mathematics developed over the millenia, owes its existence to astronomy (which in turn, heaven forbid, owes much of its existence to a belief in astrology), so, in this section we make a nod towards astronomy, and depict the more commonly described orbits.

Poincaré dedicated much of his famous three volume set [**?**] towards understanding the circular restricted planar three-body problem. In the process of investigating this problem he planted the seeds for the modern subjects of dynamical systems, and symplectic topology and made the discovery of what today is called "chaotic dynamics", notably "Hamiltonian chaos" by way of heteroclinic tangles.

There have been volumes and volumes written on the problem, so, except for this brief coverage here in the Tour, I will avoid the circular restricted three-body problem and all the other restricted three-body problems, almost like I had avoided crowded mask-free zones during the height of the pandemic while I was writing this book.

FIGURE ?? chaotic w tangle .. Simó ? where ?? -RM



[!ht]

FIGURE 24. The Sitnikov problem

find Sitnikov Ref -RM

The original Sitnikov problem is a special case of the **spatial** restricted three-body problem. In this version of the problem the primaries have equal masses but move on slightly elliptical Keplerian orbits. The infinitesimal mass moves on the line orthogonal to their plane of motion passing through their center of mass. A good chunk of the book of Moser [**86**] is dedicated to this problem. If the primaries move in circular orbits then the force on the infinitesimal is not time-dependent and the problem becomes a one-degree of freedom autonomous Hamiltonian system, hence integrable. Let them move in eccentric orbits and the problem becomes a non-autonomous time-periodic one-degree of freedom system. Moser uses the eccentricity of the ellipse as a perturbation parameter away from the integrable case of circular motion. He uses analysis at infinity and a study of Smale horseshoes to prove the following. Let time be counted by full revolutions of the primary once around their center of mass. For most negative energies, the infinitesimal mass will cross the plane of motion of the primaries over and over. Every time it crosses, mark down the integer number of years past since the last crossing. Then there is a (large) integer N_0 such that every infinite sequence $n_{-1}n_0n_1\ldots$ with $n_i \geq N_0$ is realized: there is some solution which ticks off this sequence in its travels.

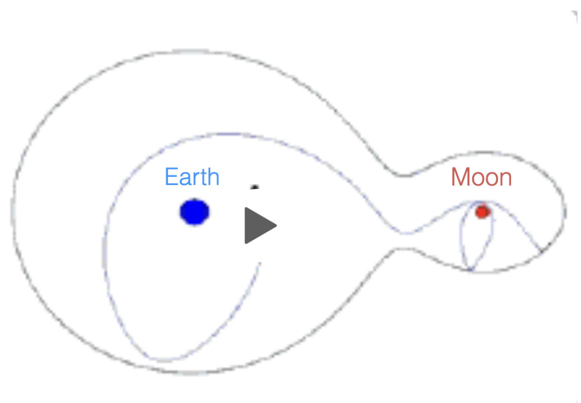


FIGURE 25. Transit orbit

6.1. Perturbation Theory. Much of the analytic work on the three-body problem over the 350 years since Newton has centered around the method of Perturbation theory. In the solar system the Sun's mass dominates that of the N planets. When cases like this, when one mass dominates all the others, and the others never come too close to hitting each other, then we can treat the problem as a small perturbation of N uncoupled Kepler problems. Similarly, when sending space craft from the earth to the moon, we can ignore the effect of the space craft on the earth and the moon. When the craft is very close to the earth we can treat its motion as a perturbation of Kepler motion.

I am under no obligation to slog through the restricted problem and hierarchical results. But should I?? -RM

In my N-body career I've avoided perturbation theory like the plague. I believed it hopeless to try to compete with its many illustrious practitioners: Lagrange, Laplace, Poincaré, Arnol'd, Meyer, etc. and have instead settled myself into some unexplored niches of celestial mechanics, notable variationally motivated crevices of the subject.

....

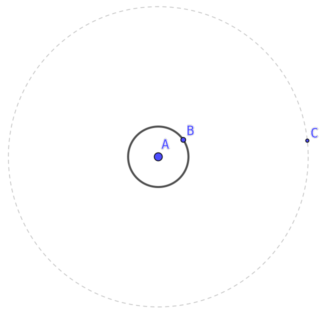


FIGURE 26. 'Solar system': hierarchical 3 body solution.

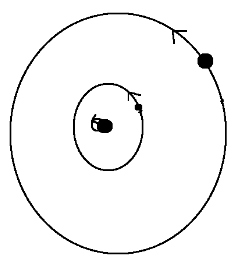


FIGURE 27. ??

TOUR NOTES:

The figure 8 orbit first appeared in the literature in the paper of C. Moore [**119**]. Chenciner and I rediscovered the eight several years later, not knowing Moore's work. (A referee told us about it!) Initial conditions for the eight are as follows. It used to irk Moore no end that we would get the credit for discovering this orbit and so we have tried to give Moore his credit where credit is due. Chenciner and I enjoyed our '15 minutes of fame' [**?**] thanks to this orbit being featured in a footnote on p. 199 of the English version of the science fiction novel [**?**], probably out as a movie by the time this book is out. Our work using variational methods to establish the existence of the eight led to a flurry of research on variationally establishing families of solutions . The eight features in chapters 2 and 3.

Schubart's orbit. Chenciner drew me this picture of the Schubart orbit in space-time.

Gerver's super-eight. Joseph Gerver predicted this orbit during the question period at the conference where I presented our work on the figure eight orbit. Indeed, he asserted that if you drew any (k, ℓ) Lissajous figure and put $N = k + \ell$ masses on it, then there was a choreographic solution for the equal mass N body problem whose N bodies traced out a curve having the 'topological type' of that Lissajous figure. The figure eight is the case $(k, \ell) = (1, 2)$. The Gerver supereight has $(k, \ell) = (1, 3)$ and its existence was soon established numerically. A rigorous existence proof took some years to obtain. Gerver's assertion has largely been confirmed, although I could not point you to any precise statement or theorem.

Hut's scattering. In developing theories of galaxy formation and dynamics astrophysicists have tried to copy statistical mechanics. Crucial processes involved are the formation of binary pairs through capture in close three body encounters. Large numerical studies have been implemented to estimate probabilities of capture or of escape.

Sitnikov. The Sitnikov problem provides one of the easiest situations in which one can prove Hamiltonian chaos exists in the 3-body problem. See the book by Siegel and Moser, for example. The two primaries, 1 and 2, have equal mass and the 3rd is taken as infinitesimal, moving on the line through the center of mass of 1 and 2 and orthogonal to their plane of motion. The eccentricity of the primaries' conics acts as a perturbation parameter. The system is integrable if this parameter is zero. Call "one year" one period of the primaries. Every time the 3rd body crosses the primary plane, list the integer part of the time that has elapsed since the last crossing. The theorem asserts that there is an N_0 such that for any infinite sequence of integers $n_{-1} n_0 n_1 \ldots$, $n_i > N_0$, there is a solution realizing this sequence of crossing times.

CHAPTER 0

The Problem and its structures

ch: intro

1. The problem, its symmetries, and conservation laws.

1.1. Three. Imagine three point masses, alone in the universe, each exerting an attractive force on the other two. (See figure 1.) How do they move? Follow the Newtonian slogan 'force = mass times acceleration" to arrive at

{N3} (1)
$$\begin{aligned} m_1\ddot{q}_1 &= F_{21} + F_{31}, \\ m_2\ddot{q}_2 &= F_{12} + F_{32}, \\ m_3\ddot{q}_3 &= F_{23} + F_{13}, \end{aligned}$$

which is a coupled system of non-linear ordinary differential equations [ODEs] governing the motion. The bodies are modelled as points labelled $1, 2$ and 3. The ath body, $a = 1, 2, 3$, has mass m_a, position $q_a = q_a(t)$, and exerts force F_{ab} on body b. The positions depend on time: $q_a = q_a(t)$ where $t \in \mathbb{R}$ is Newtonian 'universal' time. The double dots over q_a denote acceleration : $\ddot{q}_a = d^2q_a/dt^2$.

In order to turn equations (1) into a self-contained ODE we need the forces. Newton, taking suggestions of Galileo, Hooke and Halley, settled on

$$F_{ba} = -\frac{Gm_am_b}{r_{ab}^2}\hat{q}_{ab}.$$

{N2} (2)

where

(3)
$$r_{ab} = |q_a - q_b|$$

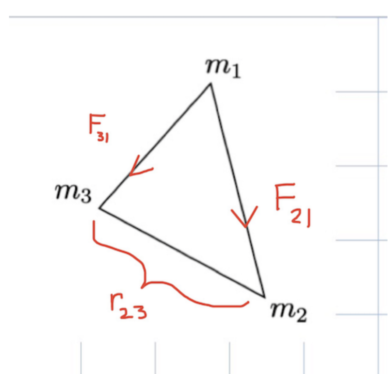


FIGURE 1. The three-body problem. The forces are along the edges of the triangle formed by the masses. fig:triangle

is the distance between bodies a and b, and

{rs} (4)
$$\hat{q}_{ab} = \frac{q_a - q_b}{r_{ab}}$$

is the unit vector pointing *to* body *a from body b*. This choice of force law between two bodies is determined by the conditions that it be attractive and directed along the line connecting the bodies, with a magnitude inversely proportional to the square of the distance between the bodies and directly proportional to their masses. The parameter G is the gravitational constant, a physical constant needed if the physical units on both sides of equation (1) are to agree. Being mathematicians we can and generally will set $G = 1$.

Where do the q_a live? In physics and astronomy one takes $q_a \in \mathbb{R}^3$. In this book we are primarily concerned with the planar three-body problem so that $q_a \in \mathbb{R}^2$. The planar three-body problem embeds as an sub-problem of the spatial one (see lemma 0.1 below and exercise 0.2), so can be understood as a special case of the spatial three-body problem. Most astronomers limit themselves for good practical reasons to dimensions $d = 3, 2$ and 1 for the dimension d of the universe $\mathbb{R}^d$ in which their stars and planets move. We are mathematicians at heart and can be free with our dimensions. At a crucial point in chapter 2 we will take $q_a \in \mathbb{R}^4$. You will see. It will be fun.

We depicted a number of solutions to the planar 3-body problem in the preceding Tour. The last three of the four Open Questions posed in this book are open for the planar 3-body problem. *The reader may want to turn to any one of these questions now.*

1.2. Any number. Any dimension. The reader can now write down the equations of motion for N point masses moving in Euclidean d-space:

{N} (5)
$$m_a \ddot{q}_a = F_a \qquad , a = 1, \ldots, N, \qquad F_a = \Sigma_{b \neq a} F_{ba}, \qquad q_a \in \mathbb{R}^d.$$

DEFINITION 0.1. *The N-body problem in d-space is the system of ODEs given by equations (5, 2) with $q_a \in \mathbb{R}^d$ and with $m_a > 0$ fixed parameters called masses. A solution to the N-body problem is a solution $q(t) = (q_1(t), \ldots, q_N(t)) \in (\mathbb{R}^d)^N$ to this system of ODEs.*

The N-body problem in d-space is not actually one problem, but rather a family of problems parameterized by the choice of masses m_a. Later, the choice of constant energy and angular momentum act as parameters. See section 0.5. Sometimes, notably in chapter three, it is useful to think of the force law itself as a parameter. See 1.5.

1.3. Flows and Phase space. Set

$$\mathbb{E} = \mathbb{E}(d, N) = (\mathbb{R}^d)^N = \mathbb{R}^d \times \mathbb{R}^d \times \ldots \mathbb{R}^d.$$

We call this space *configuration space* . We defined a solution to the N-body problem to be a curve in configuration space satisfying the ODEs (2 , 5). The letter $\mathbb{E}$ stands for "Euclidean". See equation (33) for the mass-dependent inner product on $\mathbb{E}$ which plays a central role..

A collision configuration is a point of $\mathbb{E}$ at which two masses collide: $q_a = q_b$ for some pair $a \neq b$. Equivalently, a collision occurs when $r_{ab} = 0$. Since forces blow up at collision configurations, we *expect* [1] that solutions are undefined beyond

include appendix on L-C? -RM

collisions.

collision locus

DEFINITION 0.2. *The collision locus* Δ *is the subset of configuration space*

$$\Delta = \{q = (q_1, \dots, q_N) \in \mathbb{E} : q_a = q_b \textit{ for some } a \neq b\} \subset \mathbb{E} \tag{6}$$

where two or more bodies collide.

Set

{collision-free}

$$\hat{\mathbb{E}} = \mathbb{E} \setminus \Delta. \tag{7}$$

and call it collision-free configuration space.

DEFINITION 0.3. *Phase space for the N-body problem in d-dimensions is the space*

{phase space}

$$\mathcal{P} = \mathcal{P}(d, N) = \hat{\mathbb{E}} \times \mathbb{E}. \tag{8}$$

We can rephrase the N-body problem in terms of a vector field on phase space. Introduce the N velocities $v_a = \dot{q}_a \in \mathbb{R}^d$ as additional variables, and combine them together into $v = (v_1, \dots, v_N) \in \mathbb{E}$. Then the equations defining the N-body problem take the first-order form

I will use a "$\mathcal{P}$" here instead of P because, later on, P gets used for linear momentum -RM

{NMO}

$$\dot{q} = v \tag{9}$$

{NM}

$$\mathbb{M}\dot{v} = F(q). \tag{10}$$

Here $F : \hat{\mathbb{E}} \to \mathbb{E}$ is the vector of forces, $F(q) = (F_1(q), \dots F_N(q))$ where F_a is the force on the ath body as in equation (5)and(2). The linear operator $\mathbb{M} : \mathbb{E} \to \mathbb{E}$ is a diagonal matrix in the standard basis with entries corresponding to the masses:

{M1}

$$\mathbb{M}(v_1, \dots, v_N) = (m_1 v_1, \dots, m_N v_N) \tag{11}$$

We call $\mathbb{M}$ the mass matrix. The mass matrix $\mathbb{M}$ is invertible since all the masses are positive, so we can rewrite these last equations as $\frac{d}{dt}(q, v) = X(q, v)$ where $X(q, v) = (v, \mathbb{M}^{-1}F(q))$, thus defining a vector field X on $\mathcal{P}$ whose corresponding ODE is equivalent to Newton's equations.

We write the flow of the Newtonian vector field X as

{eq: flow}

$$\Phi_t : \mathcal{P} \dashrightarrow \mathcal{P}, \qquad \Phi_t(q(0), v(0)) = (q(t), v(t)) \tag{12}$$

so that $q(t)$ is the solution to the N-body problem with initial conditions $(q(0), v(0)) \in P$ and $v(t) = \dot{q}(t)$. The broken arrow notation for the flow indicates that, for fixed time $t \neq 0$, the domain of the time t flow Φ_t is not all of $\mathcal{P}$. Some solutions fail to exist all the way up to time t due to collisions. In other words, the N-body flow is incomplete [2] By general theory, Φ_t is a one-parameter family of analytic diffeomorphisms. Φ_t also depends analytically on t and satisfies the semigroup property $\Phi_t \circ \Phi_s = \Phi_{t+s}$ on points $\zeta \in \mathcal{P}$ for which $t, s, t+s$ all lie on the interval of existence of the solution with initial condition ζ. It is worth restating this business about initial conditions in mundane terms. Initial conditions consist of N initial positions $q_a(0) \in \mathbb{R}^d$ and N initial velocities $v_a(0) = \dot{q}_a(0) \in \mathbb{R}^d$ at time $t = 0$. We put them all together to get the point $(q(0), v(0))$ in phase space.

find Saari ref -RM

REMARK 0.1. *about Levi-Civita .. .. yeah .. something ..*

Levi-civita -RM

[1] However, see the remark on the Levi-Civita transformation .. where?

[2] The set of inital conditions yielding incomplete solutions is non-empty but measure zero. NOPE! - Knauf. CITE SAARI. Appendix on incompleteness?

1.4. The planar problem sits in the spatial one. The spatial N-body problem contains the planar N-body problem.

3to2 LEMMA 0.1. *Consider* $\mathbb{R}^2 \subset \mathbb{R}^3$*. If initial conditions* $(q(0), v(0)) \in \mathcal{P}(3, N)$ *satisfy* $q_a(0), v_a(0) \in \mathbb{R}^2$ *then the corresponding planar solution* $\Phi_t(q(0), v(0))$ *is a curve in* $\mathcal{P}(2, N)$ *which solves the planar* N*-body problem with these same initial conditions.*

The lemma follows from the fact that under the assumptions of the lemma and equation (2) the forces $F_a(q) = \Sigma_b F_{ba}$, being linear combinations of the q_a, also lie in the plane $\mathbb{R}^2$. For an alternative proof of lemma 0.1 based on reflectional symmetry see exercise 0.2 below.

A similar lemma holds for the line $\mathbb{R} \subset \mathbb{R}^2$ so that the planar N-body problem contains within it the N-body problem on the line. More generally, this lemma holds for any d and $m < d$ by taking $\mathbb{R}^m \subset \mathbb{R}^d$. The m-dimenionsal N body problem embeds into the d-dimensional N-body problem as a sub-problem.

subsec: other forces

1.5. Other forces. The Newtonian two-body forces can be written $F_{ba} = \nabla_a \frac{G m_a m_b}{r_{ab}}$ where ∇_a means the gradient with respect to the vector variable $q_a \in \mathbb{R}^d$. Thus, if we put orthonormal coordinates on $\mathbb{R}^d$ so that q_a has components $q_{ai}, i = 1, \ldots, d$, and if $f : (\mathbb{R}^d)^N \to \mathbb{R}$ is a differentiable function then $(\nabla_a f)_i = \frac{\partial f}{\partial q_{ai}}$.

DEFINITION 0.4. *A general N-body problem is one for which the two-body forces* F_{ab} *of equation (5) are given by*

$$F_{ba} = \nabla_a f_{ab}(r_{ab}) \tag{13}$$

{N2ALT}

$$= f'(r_{ab}) \hat{q}_{ab} \tag{14}$$

for some collection of smooth functions $f_{ab} : (0, \infty) \to \mathbb{R}$*, subject to* $f_{ba} = f_{ab}$*,* $a, b = 1, 2 \ldots, N$*. The negatives of the functions* f_{ab} *are called the pair potentials.*

The N-body problem discussed earlier corresponds to $f_{ab}(r) = m_a m_b / r$. We typically require $f'_{ab}(r) \to 0$ as $r \to \infty$ and $|f'_{ab}(r)| \to \infty$ as $r \to 0$. The condition at infinity tells us forces fade away as particles diverge from each other, and suggest that eventually solutions tend to move along lines. The condition at collision $r = 0$ tell us we have singularities at collision.

It will be useful to focus on a particular one-parameter family of general N-body problems whose pair potentials are homogeneous of degree $-\alpha$:

{homogeneous_force}

$$f_{ab}(r) = \frac{m_a m_b}{r^\alpha}. \tag{15}$$

so that $\alpha = 1$ corresponds to the original Newtonian gravity. For the limiting case $\alpha = 0$ we set

$$f_{ab}(r) = -m_a m_a \ln(r); \alpha = 0$$

which can be justified by noting that

$$\lim_{\alpha \to 0} \frac{1}{\alpha}(r^{-\alpha} - 1) = -\ln(r),$$

as is seen by differentiating the function $x \mapsto r^{-x}$ at $x = 0$.

DEFINITION 0.5. *When the pair potential has the form of equation (15) then we refer to the resulting generalized N-body problem as a power law N-body problem.*

When researchers refer to "the gravitational N-body problem in d dimensions" they often mean the power law N-body problem in $\mathbb{R}^d$ with power and dimension linked according to

$$\alpha = d - 2$$

This linking arises out of the fact that the fundamental solution to the Laplacian in $\mathbb{R}^d$ is a constant times $1/r^{d-2}$. For more on this perspective and fundamental physical principles that follow from it, see section 6.1 below. Despite compelling reasons for linking dimension and the power law exponent, in this book we will focus almost exclusively on the N-body problem as defined above, which is to say, the power law N-body problem with $\alpha = 1$.

All four Open Questions in this book make sense for the general N-body problems in the planar case of $d = 2$. The third open question, which only makes sense when $d = 2$, has been solved when $\alpha \geq 2$.

1.6. Symmetries. Galileo had argued, decades before Newton, that the laws of mechanics must be invariant under a transformation group acting on space-time $\mathbb{R}^d \times \mathbb{R}$ which now bears his name. Newton designed his equations to be invariant under the Galilean group. This Galilean group is generated by isometries of space, isometries of the time-line, and a special class of transformations called Galilean boosts which mix time with space called "Galilean boosts" [3].

The Galilean invariance of Newton's N-body equations implies that if $q(t) = (q_1(t), \ldots, q_N(t))$ solves the equations then, upon applying any Galilean transformation simultaneously to each of the component curves $q_a(t)$ in $\mathbb{R}^d$, the result is again a solution. The Galilean transformations are

homogeneous_force
$$q_a(t) \mapsto q_a(t) + c \text{ (space translation)} \tag{16}$$
$$q_a(t) \mapsto R(q_a(t)) \text{ (space rotation or reflection)} \tag{17}$$
$$q_a(t) \mapsto q_a(t - t_0) \text{ (time translation)} \tag{18}$$
$$q_a(t) \mapsto q_a(-t) \text{ (time reflection)} \tag{19}$$
Galilean_invariance
$$q_a(t) \mapsto q_a(t) + tv \text{ (boost).} \tag{20}$$

In the formulae $t \in \mathbb{R}$, $c, v \in \mathbb{R}^d$ while $R \in O(d)$, the group of orthogonal transformations of $\mathbb{R}^d$ with its standard inner product. Any general N-body problem also has the Galilean group as a symmetry group.

ex: 3to2

EXERCISE 0.1. *Check Galilean invariance of Newton's equations by direct computation, using equations (5, 2).*

ex: 3to2

EXERCISE 0.2. *Use the symmetry $R \in O(3)$ of reflection about the plane $\mathbb{R}^2 \subset \mathbb{R}^3$ to prove lemma 0.1.*

sec: conservation laws

1.7. Conservation Laws.

DEFINITION 0.6. *A conservation law is a smooth function on phase space which is constant along any solution $(q(t), v(t))$ to the N-body problem.*

The energy, linear momentum, and angular momentum are the basic conservation laws for the N-body problem. They can be obtained by applying an algorithm made explicit by Noether, an algorithm which associates a conservation law to each continuous symmetry of a "Lagrangian system". The N-body problem can be

[3]The boosts of special relativity also mix space with time

expressed as a Lagrangian system. The isometries of space and time form a continuous symmetry group - a Lie subgroup of the Galilean group. We further break up this subgroup into time translations, space translations, and space rotations. The associated conservation laws which follow from Noether are

{energy}

$$
\begin{aligned}
E(q,\dot{q}) &= K(\dot{q}) - U(q) \\
&= \frac{1}{2}\Sigma m_a|\dot{q}_a|^2 - \Sigma_{a<b} f_{ab}(r_{ab}) \\
&= \text{ energy (for time translations)}
\end{aligned} \tag{21}
$$

{eq:linear momentum}

$$P = \Sigma m_a \dot{q}_a = \text{ linear momentum (for space translations)} \tag{22}$$

{eq:angular momentum}

$$J = \Sigma m_a q_a \wedge \dot{q}_a = \text{ angular momentum (for space rotations) .} \tag{23}$$

We have split the energy E into kinetic ($K(\dot{q})$) and potential ($-U(q)$) energy. We have used U in place of the potential energy $-U$ because U is positive for the standard N-body problem:

{grav energy}

$$U = \Sigma \frac{m_a m_b}{r_{ab}} = -V \qquad \text{(standard gravity) .} \tag{24}$$

P and J are vector-valued conservation laws. In the formula for angular momentum '$\wedge$' denotes the wedge product on $\mathbb{R}^d$ which can be identified with the cross-product on $\mathbb{R}^3$ when $d = 3$. J takes values in $\Lambda^2\mathbb{R}^d$ while P takes values in $\mathbb{R}^d$.

These conservation laws were known well before Noether stated her theorem.

Exercise 0.3. *Show that energy, angular momentum and linear momentum are conservation laws for the general N-body problem by using the equations of motion, Newton's equations (5, 2) to compute the time derivatives of these functions along solutions to Newton's equations.*

One can have forces more general than the general N-body problem and still get conservation of linear and angular momentum.

Exercise 0.4. *Show that "equal and opposite reactions" implies conservation of linear momentum. In other words, prove that for any forces which satisfy $F_{ab} = -F_{ba}$ we have $dP/dt = 0$ along its solutions.*

Show that "force directed along lines connecting bodies" implies conservation of angular momentum. In other words, prove that for any forces which satisfy $F_{ab} = \lambda_{ab}(q_a - q_b)$ for some symmetric matrix valued function $\lambda_{ab} = \lambda_{ba}$ of q and $\dot{q}$ we have that $dJ/dt = 0$ along solutions. (Symmetry of λ_{ab} implies that $F_{ab} = -F_{ba}$.)

Below, in subsection 3.1 we will verify conservation of energy by putting Newton's equations in a particularly simple form.

1.7.1. *Planar angular momentum.* The expression for angular momentum when $d = 2$ deserves special attention. Let e_1, e_2 be the standard orthonormal basis for $\mathbb{R}^2$. Then

$$(xe_1 + ye_2) \wedge (\dot{x}e_1 + \dot{y}e_2) = (x\dot{y} - y\dot{x})e_1 \wedge e_2.$$

Now $e_1 \wedge e_2$ is a basis for the one-dimensional space $\Lambda^2\mathbb{R}^2 \cong \mathbb{R}$ so we should view J as a scalar. (Physicists would call it a pseudo-scalar.) Following this computation, we divide out by $e_1 \wedge e_2$ and simply set $(x,y) \wedge (\dot{x},\dot{y}) = x\dot{y} - y\dot{x}$.

In physics this expression $(x, y) \wedge (\dot{x}, \dot{y})$ is usually represented as $e_3 \cdot (xe_1 + ye_2) \times (\dot{x}e_1 + \dot{y}e_2)$. We will be representing the expression in complex variables: $(x, y) \wedge (\dot{x}, \dot{y}) = Im(\bar{z}\dot{z})$ where $z = x + iy, \bar{z} = x - iy, \dot{z} = \dot{x} + i\dot{y}$. We have identified $\mathbb{R}^2$ with $\mathbb{C}$ by sending $(x, y) \mapsto z = x + iy$. Using this convention we get

nar N body config space}

$$\mathbb{E}(2, N) = \mathbb{C}^N, \tag{25}$$

planar angular momentum}

$$J(q, \dot{q}) = \Sigma m_a Im(\bar{q}_a \dot{q}_a). \text{ the case } d = 2 \tag{26}$$

and the action of planar rotations on configuration space is given by : (equation 20)

{eq: planar rotations}

$$q_a \mapsto e^{i\theta} q_a, q_a \in \mathbb{C} \tag{27}$$

where we have identified the planar rotation group $SO(2)$ with the unit complex numbers $\mathbb{S}^1 \subset \mathbb{C}$.

put Kepler here ?; seperate section or subsection -RM

sec: Kepler

2. Kepler and the two-body problem.

good place to put? add a section on the circular restricted three body problem; mention La Mecanique Nouvel as being about this -RM

Newton's 2-body equations read

$$m_1 \ddot{q}_1 = F_{21}$$
$$m_2 \ddot{q}_2 = F_{12}$$

Add the two equations and use $F_{21} + F_{12} = 0$ to obtain $\ddot{Q}_{cm} = 0$ where $Q_{cm} = \frac{1}{m_1+m_2}(m_1 q_1 + m_2 q_2)$ is the center of mass of the two bodies. (Note that $\dot{Q}_{cm}$ is equivalent to conservation of linear momentum.) Subtract the two equations after dividing each by its mass to obtain the evolution equation

{Kepler}

$$\ddot{\vec{r}} = -\mu \frac{\vec{r}}{r^3}, r = |\vec{r}|, \mu = GM. \tag{28}$$

for the difference vector

$$\vec{r} = q_1 - q_2.$$

Here $M = m_1 + m_2$.

The ODE (28) is called "Kepler's problem" nowadays. Its solutions $\vec{r}(t)$ lies in a fixed plane within $\mathbb{R}^d$, namely the linear space spanned by $\vec{r}(0)$ and $\dot{\vec{r}}(0)$, a fact which can be established by verifying that the angular momentum

$$J := \vec{r} \wedge \dot{\vec{r}}$$

is constant along solutions. J is a bivector so defines a two- plane provided it is not zero. This two-plane is known as the invariable plane. In the usual spatial treatment, $d = 3$, $J \in \mathbb{R}^3$ under the identification $\Lambda^2\mathbb{R}^3 \cong \mathbb{R}^3$ and the invariable plane is the plane $J^\perp \subset \mathbb{R}^3$. Up to a mass-dependent constant, this J is the angular momentum described above (section 1.7), provided $Q_{cm} = 0$. See section 1.7 for more on angular momentum.

If $J \neq 0$, then the solution curve describes a conic section in the invariable plane defined by J. This conic section has one of its foci at the origin $\vec{r} = 0$ of the plane. This assertion contains Kepler's first law of planetary motion. Assuming that $\vec{r}(t)$ solves the Kepler probem we form the general solution to the gravitational two-body problem by adding constant multiples of $\vec{r}(t)$ to the motion of the center of mass $Q(t)$. Specifically, we have that the general solution is given by $q_1(t) = Q(t) + \frac{m_2}{M}\vec{r}(t); q_2(t) = Q(t) - \frac{m_1}{M}\vec{r}(t)$ where $Q(t) = tV + Q_0$ with V, Q_0 constant vectors in $\mathbb{R}^d$ and where $\vec{r}(t)$ is any solution to Kepler's problem. By this 'center of mass trick' we have reduced the study of the two-body problem to that of Kepler's problem. Kepler's problem itself is identical to the two-body problem upon imposing $Q_{cm} = 0$.

Assuming that $Q_{cm} = 0$, the conserved energy E is given by

$$\frac{E}{M} = \frac{1}{2}|\dot{\vec{r}}|^2 - \frac{\mu}{r} \tag{29}$$

$$= \frac{1}{2}(\dot{r}^2 + \frac{|J|^2}{r^2}) - \frac{\mu}{r} \tag{30}$$

Kepler's three laws, K1-3 below, summarize the most important features of the solutions to Kepler's problem. All three laws are clarified by using energy E and angular momentum J.

K1. The solutions lie on conics within the plane determined by the angular momentum. Each conic has one focus at the origin of the plane.

These conics are ellipses or circles when the energy E is negative, parabolas if $E = 0$, and hyperbolas if $E > 0$. See equation 21 regarding energy. In the Kepler problem the energy is given by $H = \frac{1}{2}|\dot{\vec{r}}\|^2 - \frac{\mu}{r}$. In the bounded, negative energy case, the energy H and size a of the orbit are related by

$$H = \frac{-\mu}{2a}$$

where 'size' means semi-major axis of the ellipse. In the case of a circle a is the radius. The conics degenerate to rays ($E \geq 0$) or line segments ($E < 0$) having one endpoint collision with the sun ($r = 0$) if and only if $J = 0$.

K2. "Equal angles in equal times". Let e_1, e_2 be an orthonormal basis for the invariable plane and introduce polar coordinates r, θ on the invariable plane by $\vec{r} = r\cos(\theta)e_1 + r\sin(\theta)e_2$. Then we compute that $J = r^2\dot{\theta}(e_1 \wedge e_2)$. Since J is constant we can rewrite this $r^2 d\theta = \pm|J|dt$. Integrating along an arc of a solution we find that the angle $\Delta\theta$ swept out by a fixed counterclockwise moving solution in time Δt is $\Delta\theta = |J|\Delta t$.

K3. Is a period-size relation:

{eq: Keplers3rd}

$$T^2 = \mu a^3. \tag{31}$$

where T is the period and a is the size, as per K1. Recall $a \sim -1/|H|$.

correct units -RM

A remarkable thing about Kepler's problem is that for fixed negative energy all its orbits are periodic with the same period. This fact and K3 play a particularly important role in question 3 on braids. In chapter 3 it will be important that the action of the periodic orbits also depends only on the energy, or, what is the same, the period. The Kepler problem appears in a central way in every open question on our list. One could say that the Kepler problem is to the N-body problem as planar Euclidean geometry is to differential geometry.

The limiting case of angular momentum zero is worth focusing on for a moment, as it will come up several times, most notably in chapter 3. $J = 0$ if and only if the space spanned by the vectors $\vec{r}(0)$ and $\dot{\vec{r}}(0)$ is one-dimensional which in turn is true if and only if the solution is a solution to the one-dimensional Kepler problem, and thus travel along the ray joining $\vec{r}(0)$ to the origin $\vec{r} = 0$. Let us fix the two-plane, say $e_1 \wedge e_2$ and the energy at some negative value and let the angular momentum go to zero by setting $J = Le_1 \wedge e_2$ and letting $L \to 0$. In this way we get a family of ellipses crunching down to line segments. .....

.. something on eccentricity, limits here .. -RM

EXERCISE 0.5. *Figure out the units of the Kepler constant μ. Compare with K3.*

EXERCISE 0.6. *Consider the general 2-body problem as per subsection 1.5 where the negative of the pair potential is $-f(r)$ with $f_{21}(r) = -f_{12}(r) = f(r)$. Show that the relative motion vector $\vec{r}$ evolves according to $\ddot{\vec{r}} = f'(r)\frac{\vec{r}}{r}$. This ODE is refered to as the central force problem. It's potential function is $V(\vec{r}) = -f(r)$.*

check signs -RM

include effective potential; discussion of boundedness etc for $f = r^{-\alpha}$. -RM

sec: mass metric

3. Metric structure, boundedness, center-of-mass, scaling

3.1. Mass metric. The kinetic energy $K = \frac{1}{2}\Sigma m_a|\dot{q}_a|^2$ is half the quadratic form associated to mass-dependent inner product on configuration space $\mathbb{E} = \mathbb{R}^{dN}$. Thus

{KE}

$$K(\dot{q}) = \frac{1}{2}\langle\dot{q}, \dot{q}\rangle \tag{32}$$

where

{mass_innerproduct}

$$\langle q, v\rangle = \Sigma m_a q_a \cdot v_a \tag{33}$$

where $q_a \cdot v_a$ is the standard inner product on $\mathbb{R}^d$. The mass inner product can also be written as

{eq: intertwine}

$$\langle q, v\rangle = \langle q, \mathbb{M}v\rangle_1 \tag{34}$$

where $\mathbb{M} : \mathbb{E} \to \mathbb{E}$ is the mass metric defined in equation (11) above and where $\langle q, v\rangle_1 = \Sigma q_a \cdot v_a$ is the 'standard' inner product, that is to say, the mass inner product we get on $\mathbb{E} = \mathbb{R}^{dN}$ when we take all the masses equal to 1.

DEFINITION 0.7. *We call the above inner product (33) the mass inner product on configuration space.*

LEMMA 0.2. *The general N-body equations can be rewritten*

{N_Condensed}

$$\ddot{q} = \nabla U(q) \tag{35}$$

where ∇ is the gradient of $U = \Sigma f_{ab}(r_{ab})$ with respect to the mass inner product. We will call this equation the "condensed form" of Newton's equations .

PROOF. Equations (9, 36) show us that we can write the N-body problem as

{NM}

$$\mathbb{M}\ddot{q} = F(q) \tag{36}$$

where $\mathbb{M}$ is the mass matrix and $F : \hat{\mathbb{E}} \to \mathbb{E}$ is the vector of forces. Thus $\ddot{q} = \mathbb{M}^{-1}F(q)$. Now $F(q)_a = F_a(q) = \nabla_a U(q)$ where ∇_a is the usual gradient with respect to $q_a \in \mathbb{R}^d$. It follows that $F(q) = \nabla^{(1)}U$ where $\nabla^{(1)}$ is the gradient with respect to the standard inner product, which is to say, the mass metric in the case $m_1 = m_2 = \ldots = m_N = 1$. Recall, in general, the relation between the differential and the gradient of a function on an inner product space: $dU(q)(h) = \langle\nabla U(q), h\rangle$. The intertwining relation, equation 34 implies that our two gradient operators $\nabla^{(1)}$ and ∇ on $\mathbb{E}$ are related by raising and lowering of indices with respect to the mass metric: $\nabla^{(1)} = \mathbb{M}\nabla$ or $\nabla = \mathbb{M}^{-1}\nabla^{(1)}$, so that $\mathbb{M}^{-1}F(q) = \nabla U(q)$.

QED

Using the mass metric we have that the energy is given by

$$E = \frac{1}{2}\langle\dot{q}, \dot{q}\rangle - U(q).$$

EXERCISE 0.7. *Use the condensed form of the general N-body equations (35) to prove that energy is conserved.*

The mass metric gives geometric meaning to the vanishing of linear and angular momentum. Recall that "the infinitesimal generator" of a Lie group action on a manifold is the map from the group's Lie algebra to the space of vector fields on the manifold which we get by differentiating the action with respect to the group elements. The space of infinitesimal generators evalutated at a given point of the manifold spans the tangent space to the group orbit through that point.

EXERCISE 0.8. *Show that the space spanned by the infinitesimal generators (see equations 20) of the translation action of* $\mathbb{R}^d$ *on* $\mathbb{E}$ *consist of all vectors of the form* $(c, c, \dots, c), c \in \mathbb{R}^d$.

Show that space spanned by the infinitesimal generators for the rotation action of $SO(d)$ *on* $\mathbb{E}$ *(see equations 20) at* $q = (q_1, q_1, \dots, q_N)$ *consists of all vectors in* $\mathbb{E}$ *of the form* $(\xi(q_1), \xi(q_2), \dots, \xi(q_N))$ *for some* $\xi \in so(d)$. *Note that in case* $d = \mathbb{R}^3$ *we can write* $\xi(q_a) = \vec{\omega} \times q_a$ *for some angular velocity vector* $\vec{\omega} \in \mathbb{R}^3$.

EXERCISE 0.9. *Show that* $v = \dot{q}$ *is mass-metric perpindicular to the infinitesimal generators of translation if and only if the linear momentum* $P = \Sigma m_a \dot{q}_a$ *is zero.*

Show that $v = \dot{q}$ *is mass-metric perpindicular at* q *to the infinitesimal generators of rotation through* q *if and only if the angular momentum* $J(q, \dot{q})$ *is zero.*

We pause to recall:

def: rigid motion

DEFINITION 0.8. *The group* G *of rigid motions of a Euclidean space is the group of orientation preserving isometries of the plane. It consists of compositions of translations and rotations (no reflections).*

Thus linear and angular momentum are zero if and only if the solution curve is everywhere orthogonal to the orbits swept out by the group of rigid motions acting on configuration space.

put Saari velocity decomp here -RM

3.2. Getting rid of translations: Center-of-Mass. In arriving at the Kepler problem, (section 0.2) we separated the two-body problem into that of its center of mass, which is linear, and the relative motion which is described by Kepler's problem. The same trick works for any number of masses, except that the relative motion problem is no longer solvable in closed form when $N > 2$.

Write $Q_{cm} = \frac{1}{M} \Sigma m_a q_a$ for the center of mass of an N-body system where $M = \Sigma m_a$ is the total mass. Then $V := \dot{Q} = \frac{1}{M} P$ is constant by conservation of linear momentum so that $\ddot{Q} = 0$. The general solution of $\ddot{Q} = 0$ is a straight line so we have that the center of mass travels at uniform speed along a straight line in d-space:

$$Q(t) = Q(0) + tV.$$

Now take a solution $q_a(t)$ and look at it from the center of mass frame. By this we mean look at it relative to the inertial frame whose origin is $Q(t)$. This amounts to making the time dependent change of variables $q_a \mapsto \tilde{q}_a$ given by $q_a(t) = Q(t) + \tilde{q}_a(t)$. The inverse transformation

$$\tilde{q}_a(t) = q_a(t) - tV - Q(0)$$

is a Galilean boost followed by a translation so the new curve $\tilde{q}_a$ is again a solution, but now its center of mass zero is zero as its linear momementum.

We have shown that we can reduce the general N-body problem to the subproblem in which the center of mass is zero for all time. We will call this reduction the *center of mass trick*. It reduces the dimension of phase space by $2d$ dimension. We will say that a solution $\tilde{q}_a(t)$ to the zero-center-of mass subsystem is 'in center of mass frame" or "center of mass coordinates". Write

centered config space}

$$\mathbb{E}_0 = \{q \in \mathbb{E} : \Sigma m_a q_a = 0\} \cong \mathbb{R}^{d(N-1)} \tag{37}$$

for the center-of-mass zero subspace of configuration space. We have just argued that we may assume that $q(t) \in \mathbb{E}_0$ when working on the N-body problem. We will do so whenever it is more convenient or more illuminating than working on the full configuration space.

Set

$$\hat{\mathbb{E}}_0 = \mathbb{E}_0 \setminus \Delta.$$

and

: centered phase space}

$$\mathcal{P}_0 = \hat{\mathbb{E}}_0 \times \mathbb{E}_0 \tag{38}$$

$\mathcal{P}_0 \subset \mathcal{P}$ is invariant under the N-body flow. We have just shown that solving the general N-body problem is equivalent to solving it on the subspace $\mathcal{P}_0$ which we will call either "center-of-mass phase space" or just "centered phase space". Similarly we call $\mathbb{E}_0$ or "center-of-mass configuration space" or "centered" configuration space.

What we have just done is closely related to (but not quite the same) as what is called "reduction by translation" in symplectic geometry. We have gotten rid of d "translational degrees of freedom". The number of degrees of freedom of a mechanical system is defined to be the dimension of its configuration space, or half the dimension of its phase space. We have reduced the number of degrees of freedom by d in going from $\mathcal{P}$ to $\mathcal{P}_0$.

While we are on the center-of-mass subject it is worth recording an identity attributed to Lagrange.

LEMMA 0.3.

{eq: Lagrange identity}

$$\langle q - Q_{cm}, q - Q_{cm}\rangle = \frac{1}{\Sigma m_a}\Sigma_{a<b} m_a m_b r_{ab}^2, q \in \mathbb{E}_0 \tag{39}$$

PROOF. First assume that $q \in \mathbb{E}_0$ so that $Q_{cm} = 0$. Write Σ' for the sum over all pairs $a < b$ and Σ for the sum over *all pairs*. The sum in the lemma is a Σ'. By symmetry and the fact that $r_{aa} = 0$ we have, on the one hand, $\Sigma m_a m_b r_{ab}^2 = 2\Sigma' m_a m_b r_{ab}^2$, On the other hand,

$$\Sigma m_a m_b |q_a - q_b|^2 = \Sigma m_a m_b |q_a|^2 - 2\Sigma m_a m_b q_a \cdot q_b + 2\Sigma m_a m_b |q_b|^2.$$

The first and last term equal MI where

{eq: I}

$$I = \Sigma m_a |q_a|^2 = \langle q, q\rangle. \tag{40}$$

The second, cross term equals $\Sigma m_a q_a \cdot \Sigma m_b q_b$ which is zero since we have assumed $Q_{cm} = 0$ and $MQ_{cm} = \Sigma m_a q_a$. Thus $2\Sigma m_a m_b r_{ab}^2 = 2MI$.

To establish (39) in the general case observe that both sides of the equation are invariant under Galilean translation $q_a \mapsto q_a + c$.

QED

DEFINITION 0.9. *The quantity $I = I(q)$ of equation (40) is called the moment of inertia of q.*

Moment of inertia measures resistance to spinning. To understand how I arises in this context compute the kinetic energy associated to spinning a planar N-gon with constant angular velocity ω. Use the notation around equation (26) so that $\mathbb{E}(2,N) = \mathbb{C}^N$. In order to rotate the planar N-gon $q \in \mathbb{E}$ we form the curve $exp(i\omega t)q$. The curve's derivative is $\dot{q} = i\omega q$. It follows that

$$K(\dot{q}) = \frac{1}{2}\Sigma m_a \omega^2 |q_a|^2 = \frac{1}{2} I \omega^2$$

exhibiting I as a kind of mass associated to angular velocity ω. The rotational equivalent of mass is moment of inertia.

insert into next section, reduc by ang momentum; take out of this section (?) -RM

3.3. Jacobi vectors. Any two real inner product spaces of the same dimension are linearly isometric. It follows by a dimension count that

$$\mathbb{E}_0 \cong (\mathbb{R}^d)^{N-1},$$

where the squared inner product on $(\mathbb{R}^d)^{N-1}$ is such that if q corresponds to $(Q_1, \ldots, Q_{N-1})$ then $I(q)$ corresponds to $\Sigma |Q_a|^2$. The $Q_a \in \mathbb{R}^d$ are called *normalized Jacobi vectors* after Jacobi who described the now-standard method for constructing this isomorphism in the three-body problem. The Jacobi vectors are not unique! Any linear isometry $R \in O(d(N-1))$ of $(\mathbb{R}^d)^{N-1}$ will take one system of Jacobi vectors to another.

fig: Jacobi vectors

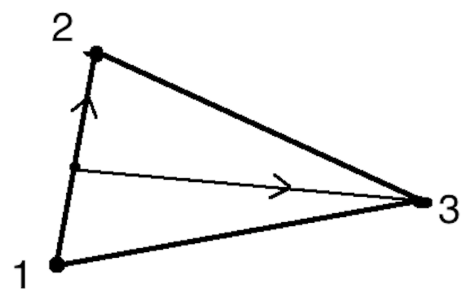


FIGURE 2. Jacobi vectors.

For $N = 3$ write Q_0 for Q_{cm}. Then, if $q \in \mathbb{E}_0$ write

{Jac1}

$$Q_1 = q_2 - q_1, Q_2 = q_3 - (m_1 q_1 + m_2 q_2)/(m_1 + m_2). \tag{41}$$

See figure 3.3 for the meaning of the Jacobi vectors. Note Q_2 is the vector connecting the center of mass of 12 to mass 3. One then computes that

$$I(q) = M|Q_0|^2 + \mu_1 |Q_1|^2 + \mu_2 |Q_2|^2 \tag{42}$$

where

{Jac_constants}

$$\frac{1}{\mu_1} = \frac{1}{m_1} + \frac{1}{m_2} \qquad \frac{1}{\mu_2} = \frac{1}{m_1 + m_2} + \frac{1}{m_3}. \tag{43}$$

It follows that if $Q_0 := Q_{cm} = 0$ then $I(q) = \mu_1 |Q_1|^2 + \mu_2 |Q_2|^2$. The normalized Jacobi vectors are then $\sqrt{\mu_1} Q_1$ and $\sqrt{\mu_2} Q_2$.

An iterative binary tree algorithm for computing Jacobi vectors for any N is explained nicely in wiki. (See `https://en.wikipedia.org/wiki/Jacobi_coordinates` as of November 2021.) This algorithm yields the standard $N = 3$ Jacobi vectors just given. Alternatives to this binary tree algorithm are achieved by realizing that, as an inner product space, $\mathbb{E} = \mathbb{R}^d \otimes \mathbb{R}^N$ provided we give $\mathbb{R}^d$ its standard inner product while giving $\mathbb{R}^N$ the mass-induced inner product. Write $\epsilon = (1, \dots, 1) \in \mathbb{R}^N$ and observe that the generator of the translation space is $\mathbb{R}^d \otimes \epsilon$ so that $\mathbb{E}_0 = \mathbb{R}^d \otimes \epsilon^\perp$. Find an orthonormal basis $\epsilon_1, \epsilon_2, \dots, \epsilon_{N-1}$ for $\epsilon^\perp \subset \mathbb{R}^N$ by any method you like. You could, for example, fill out ϵ to a basis $(\epsilon, f_1, f_2, \dots, f_N)$ for $\mathbb{R}^N$ by any method and then apply Graham-Schmidt to this basis. Then any $q \in \mathbb{E}_0$ can be expanded uniquely as $\Sigma_{a=1}^{N-1} Q_a \otimes \epsilon_a$ with $Q_a \in \mathbb{R}^d$. The Q_a are then normalized Jacobi vectors. The basis corresponding to the original $N = 3$ (non-normalized) Jacobi vectors Q_1, Q_2 is $\epsilon_1 = (\frac{1}{m_1}, \frac{-1}{m_2}, 0), \epsilon_2 = (\frac{-1}{m_1+m_2}, \frac{-1}{m_1+m_2}, \frac{1}{m_3})$.

subsec: scaling

3.4. Scaling symmetry. The N-body problems with α- power law potential admit an additional space-time scaling symmetry

put as finals subsection of this section (?) -RM

{scaling}

$$q(t) \mapsto \lambda q(\lambda^{-\beta} t); \beta = \frac{\alpha}{2} + 1.. \tag{44}$$

which takes solutions to solutions and which arises from the homogeneity of U. This scaling transforms velocities according to $v \mapsto \lambda^{-\alpha/2} v$ and energy and angular momentum according to $E \mapsto \lambda^{-\alpha} E, J \mapsto \lambda^{1-\alpha/2} J$.

In the Newtonian case of $\alpha = 1$ we have $\beta = 3/2$. Write $q^\lambda(t) = \lambda q(\lambda^{-3/2} t)$ for the transformed solution of a solution $q(t)$. If $q(t)$ is periodic with period T and typical size a then $q^\lambda(t)$ is periodic with period $\lambda^{3/2} T$ and size λa. The quantity $a^3 T^{-2}$ remains invariant under scaling. This invariance yields Kepler's 3rd law for the case $N = 2$. See K3 and the associated equation (31) above.

Energy and angular momentum for the Newtonian case scale according to $E \mapsto \frac{1}{\lambda} E, J \mapsto \lambda^{1/2} J$ so that the quantity

{Dziobek}

$$Dz := E|J|^2 \tag{45}$$

is invariant under scaling. This quantity is called the Dziobek constant and is the basic bifurcation parameter for the planar N-body problem. For the planar three-body problem, it, along with the mass distribution $m_1 : m_2 : m_3$ are the only parameters.

really?? "DZIOB"? pretty ugly notation -RM

subsec:LagrangeJacobi

3.5. Boundedness, Relative Periodic Orbits and the Lagrange-Jacobi identity. An enormous body of work has gone into finding, counting, and classifying periodic orbits in dynamical systems generally and the N-body problem specifically. Poincaré's famous quote spurred us on, a quote we mangle by paraphrasing it as saying " periodic orbits are the only breach with which to enter the almost impenetrable fortress of the three body problem". Poincaré was actually referring to relative periodic orbits.

Definition 0.10. *We say that a curve $q : \mathbb{R} \to \mathbb{E}_0$ in the centered N-body configuration space is relative periodic if there exists a $T > 0$ and a $g_0 \in O(d)$ such that $q(T + t) = g_0 q(t)$. We will call T the period and g_0 the monodromy of the orbit.*

God has not painted Cartesian axes onto the fabric of the universe. This makes it difficult to impossible to observe honest periodic orbits out in the universe. Relative periodic orbits can be observed when mutual distances can be deduced from measurements.

The notion of a curve being *periodic* or *bounded* are standard notions. With these definitions in mind we have the following proposition.

PROPOSITION 0.1. *Let $q : \mathbb{R} \to \mathbb{E}_0$ be a solution to the general N-body problem defined for all time and viewed from center-of-mass frame. Let $r_{ab}(t) = |q_a(t) - q_b(t)|$ be the mutual distances along the curve. Then*

- *(i) the $r_{ab}(t)$ are bounded if and only if q is bounded*
- *(ii) the $r_{ab}(t)$ are periodic if and only if q is relative periodic.*

PROOF. For (i) we use equation (39) above: $\langle q, q \rangle = \frac{1}{M}\Sigma m_a m_b r_{ab}^2$ valid for points with center of mass zero.

One direction of (ii), that relative periodic curves have periodic mutual distance functions follows immediately from the definition of being relative periodic. For the other direction, use that two labelled N-gons in $\mathbb{R}^d$ have all of their mutual distances equal if and only if there is an element of $Iso(d)$ taking one N-gon to the other. If their centers of mass are zero then this g must lie in $O(d)$. It follows that $r_{ab}(T + h) = r_{ab}(h)$ implies that there is a smooth curve $g_h \in O(d)$ such that $q(T + h) = g_h q(h)$. I claim g_h is independent of h: $g_h = g_0$. For this purpose, observe that $g_0 q(t)$ is another solution to Newton's equations having the same initial *position* as $t \mapsto q(T + t)$. If the initial velocities $g_0 \dot{q}(0)$ and $\dot{q}(T)$ of these two curves agreed we would be done: they would be equal curves. To this purpose I will assume that $d = 2$ and leave the general case to the reader. Identify $\mathbb{E}$ with $\mathbb{C}^N$ as above so that $SO(2)$ acts by complex multiplication by unit complex numbers. See around equation (26). By doubling the period if necessary, we can assume that $g_0, g_h \in SO(2) = U(1)$ and write $g_h = exp(i(ah + O(h^2))g_0$ and differentiate $q(T + h) = g_h q(h)$ at $h = 0$ to get $\dot{q}(T) = iaq(T) + g_0\dot{q}(0) = (iag_0 q(0) + g_0 \dot{q}(0)$. Assume, for simplicity, that $g_0 \in SO(2)$ so that $g_0 = e^{i\theta_0}$. Now use the formula (26) for planar angular momentum to compute $J(q(T), \dot{q}(T)) = J(q(0), \dot{q}(0)) + aI(q(0)$ where I is the moment of inertia. By conservation of angular momentum we must have $a = 0$ so that $q(t + T) = g_0 q(t)$ as claimed. QED

A less well-known notion is that of a curve being *quasi-periodic*. The canonical example of a quasi-periodic curve is any integral curve for irrational flow on a torus: $\dot{\theta}_1 = \omega_1, \dot{\theta}_2 = \omega_2$ with ω_1, ω_2 constants and ω_1/ω_2 irrational. Every such curve is dense in the two torus coordinatized by θ_1, θ_2. More generally we have irrational flow on an n-torus. We will say that a smooth curve is quasi-periodic if it is the image of an integral curve for irrational flow on some torus under some smooth immersion of the torus. Quasi-periodic curves play a central role in the N-body problem and Hamiltonian mechanics due to the infamous KAM theorem described in chapter 2 on Stability. We have the following implications for the standard N-body problem

THEOREM 0.1. *The following implications hold for solutions to the standard N-body problem in the center-of-mass frame*

periodic $\Longrightarrow$ rel. periodic $\Longrightarrow$ quasi-periodic $\Longrightarrow$ bounded $\Longrightarrow$ negative energy

`thm: orbit types`

Contrast these implications with what we know about the Kepler problem, which is to say, the non-collision solutions to the two-body problem in the center of mass frame:

periodic $\Longleftrightarrow$ rel. periodic $\Longleftrightarrow$ quasi-periodic $\Longleftrightarrow$ bounded $\Longleftrightarrow$ negative energy

For $N > 2$ none of these reversed implications hold.

We can define relative periodicity, boundedness, etcetera in purely Galilean invariant terms, using only the mutual distances r_{ab}.

ef: distance properties

DEFINITION 0.11. *Suppose that $q(t)$ is a solution to the N-body problem Form its $\binom{N}{2}$ mutual distance functions $r_{ab}(t) = |q_a(t) - q_b(t)|$. Then the solution is*

- *relatively periodic if the $r_{ab}(t)$ are periodic functions of t*
- *quasi-periodic if the $r_{ab}(t)$ are quasi-periodic functions of t*
- *bounded if the $r_{ab}(t)$ are bounded functions of t*
- *unbounded if the $r_{ab}(t)$ are unbounded functions of t*
- *a relative equilibrium if the $r_{ab}(t)$ are constant functions of t*

PROOF OF THEOREM 0.1 All the implications of the theorem are obvious except for rel. periodic $\implies$ quasi-periodic and bounded $\implies$ negative energy .

We prove relative periodic $\implies$ quasi-periodic for the planar case at the end of the next section. See 5.0.1. We give a similar proof valid for $d > 2$ in the appendix on reduction.

Really? where?; maybe two sections down.. -RM

To establish bounded $\implies$ negative energy we differentiate I twice along a solution $q(t)$.

$$I(q) = \langle q, q \rangle$$

so

$$\dot{I} = 2\langle q, \dot{q} \rangle$$

and

$$\begin{aligned} \ddot{I} &= 2\langle \dot{q}, \dot{q} \rangle + 2\langle q, \ddot{q} \rangle \\ &= 4K(\dot{q}) + 2\langle q, \nabla U(q) \rangle \end{aligned} \tag{46}$$

This identity for general potentials U is called the "virial identity" . The gravitational potetial $-U$ is homogeneous of degree -1 so by one of Euler's many identities we have $\langle q, \nabla U(q) \rangle = -U(q)$. It follows that $\ddot{I} = 4K - 2U$. Since the energy $E = K - U$ this yields

{eq:LagJac}

$$\ddot{I} = 4E + 2U, \tag{47}$$

an identity known as the Lagrange-Jacobi identity . The implication "bounded implies negative energy" follows now from its contrapositive which is a corollary of the Lagrange-Jacobi identity.

lemma: LagrangeJacobi

LEMMA 0.4. *[Corollary to Lagrange-Jacobi] If a solution $q(t)$ to the gravitational N-body problem has energy $E \geq 0$ and is defined for all time then its motion is unbounded in both forward and negative time and it size $|q(t)| = \sqrt{I(q(t))}$ achieves a minimum value at a unique time.*

cor:LJ

PROOF OF LEMMA. $U > 0$ everywhere. If $E > 0$ then the Lagrange-Jacobi identity gives the strict convexity $\ddot{I} > 4E > 0$ proving that I is unbounded in both time directions with a unique minimum.

If $E = 0$ we just get $\ddot{I} > 0$ and a bit more work is required. We still get that I is a strictly convex non-negative function of time defined for all time. Such a function either limits monotonically to a finite value at one end of the line or the other, (like e^{-t}) or is unbounded in both directions with a unique minimum in between. We must eliminate the possibility that $I(t)$ tends to a finite value I_* at $t = \pm\infty$. In order to do so we introduce a normalized version of U.

f: normalized potential

DEFINITION 0.12. *The normalized potential* $\tilde{U}$ *is the function*

: normalized potential}

$$\tilde{U} = \sqrt{I}U. \tag{48}$$

In other words, $\tilde{U}$ is the function U, the negative of the Newtonian potential, made homogeneous of degree 0 and hence scale-invariant by multiplying U by $r = \sqrt{I}$. Identify the sphere $\mathbb{S} \subset \mathbb{E}$ as the set of rays through the origin of $\mathbb{E}$. so that $\tilde{U}$ becomes a function on the sphere with poles at the collision locus. It is positive and continuous (as a function with values in $\mathbb{R} \cup \{\infty\}$) and since the sphere is compact there is a positive constant c such that $\tilde{U} > c$ everywhere. Since $U = \tilde{U}/\sqrt{I}$ it follows that if $I(t) \to I_*$ as $t \to +\infty$ then $U > c/\sqrt{I_*} > 0$ in some neighborhood of $t = +\infty$, so that by the Lagrange-Jacobi identity $\ddot{I} \geq 2c/\sqrt{I_*}$ near $t = +\infty$ This inequality is inconsistent with $I(t) \to I_*$, a contradiction implying that the finite limit I_* cannot exist. A similar argument holds as $t \to -\infty$. QED

EXERCISE 0.10. *For power law potentials with exponent* α *derive the variant of the Lagrange-Jacobi identity*

{eq:LagJac_alpha}

$$\ddot{I} = 4E + (4 - 2\alpha)U. \tag{49}$$

EXERCISE 0.11. *Use the Lagrange-Jacobi identity to show that if* $q(t)$ *is a periodic solution for the standard N-body problem then its energy* E *satisfies*

$$E = -\frac{1}{2} < U >$$

where $< U > := \frac{1}{T}\int_0^T U(q(t))dt$ *is the average of the negative of the potential energy over one period* T *of the orbit.*

EXERCISE 0.12. *In the case of the power law potential in the limit* $\alpha = 0$ *appropriate to 'real' planar gravity, we have* $U = -\Sigma m_a m_b \log(r_{ab})$. *Show that regardless of energy every solution is bounded over its entire interval of existence.*

some words on strong force -RM

some words on instability, stability and lower bounds on energy? -RM

4. The Shape Sphere for three bodies.

word 'reduction' not used in body of section. take out, or use somewhere -RM

The three masses of the three-body problem form the vertices of a moving triangle. Galilean invariance (see figure 1) suggests the triangle's motion depends only on its congruence class. The three side lengths r_{12}, r_{23}, r_{31} fix this congruence class so we expect there to be a 2nd order differential equation in sidelengths which is equivalent to the equations defining the planar three-body problem. Such an equation, depending parametrically on the value J of the angular momentum, exists *away from collinearity*. We can avoid the problems at collinearity by working with *oriented* congruence classes rather than congruence classes of triangles.

DEFINITION 0.13. *Two configurations (points of* $\mathbb{C}^3$*) are 'oriented congruent' if a rigid motion (see definition 0.8) takes one to the other.*

DEFINITION 0.14 (Shape Space). *Shape space for the planar three-body problem is the space of oriented congruence classes of triangles, endowed with the quotient metric as inherited from the mass metric on* $\mathbb{C}^3$.

When we use oriented congruence instead of congruence then triangles related to each other by a single reflection are considered to be distinct, or have different 'shapes'.

The oriented congruence class of a triangle $(q_1, q_2, q_3) \in \mathbb{C}^3$ is specified by fixing its *signed* area A in addition to its three side lengths. The signed area is given by

$$(q_2 - q_1) \wedge (q_3 - q_1) = 2Ae_1 \wedge e_2.$$

Shape space is three-dimensional so there is a relation between these four numbers. That relation is given by Heron's formula:

$$A^2 = p(p - r_{12})(p - r_{23})(p - r_{31}), \qquad \text{where } p = \frac{1}{2}(r_{12} + r_{23} + r_{31}).$$

thm: shape space

THEOREM 0.2. *(See figure ??.) Shape space is homeomorphic to $\mathbb{R}^3$, with its origin corresponding to triple collision. The quotient map $\pi : \mathbb{C}^3 \to \mathbb{R}^3$ from configuration space to shape space is real quadratic homogeneous. The linear coordinates (w_1, w_2, w_3) on $\mathbb{R}^3$ can be chosen so that*

$$w_1^2 + w_2^2 + w_3^2 = (\frac{1}{2}I)^2$$

where I is the moment of inertia as defined above, and so that

$$w_3 = (\sqrt{\frac{m_1 m_2 m_3}{m_1 + m_2 + m_3}})A$$

where A is the signed area of the corresponding triangle q with $\pi(q) = (w_1, w_2, w_3)$.

A formula for the w_i in terms of the vertices q_a of a located triangle can be found in the appendix. From this formula one sees that reflections act on shape space by $(w_1, w_2, w_3) \mapsto (w_1, w_2, -w_3)$. Consequently, the 'unoriented shape space' which is to say, the space of usual congruence classes of triangles in the plane, is diffeomorphic to the closed half space $w_3 \geq 0$ within shape space $\mathbb{R}^3$.

okay? then needs to be written, placed, referenced -RM

The relation between objects in shape space and configuration space $\mathbb{C}^3$ is usually best understood through the medium of the *shape sphere*.

DEFINITION 0.15. *The shape sphere is the unit sphere $w_1^2 + w_2^2 + w_3^2 = 1/4$ in shape space.*

We will now show that shape sphere realizes the space of oriented similarity classes of triangles.

def: oriented similar

DEFINITION 0.16. *Two planar triangles $q, q' \in \mathbb{C}^3$ are 'oriented similar' if there is an orientation-preserving similarity -a rigid motion composed with a scaling – which takes q to q'.*

In forming the space of oriented similarity classes of triangles we do not allow triple collision triangles to count as viable triangles. Triple collisions are excluced for the same reason that we exclude the origin of a vector space when we projectivize the vector space to form the corresponding projective space. Triple collision triangles are all of the form $q = (z, z, z) = z\mathbf{1}$ where $\mathbf{1} = (\mathbf{1}, \mathbf{1}, \mathbf{1})$.

PROPOSITION 0.2. *The shape sphere is diffeomorphic to the space of oriented similarity classes of non-triple collision triangles, which is to say, it is the quotient space of $\mathbb{C}^3 \setminus \mathbb{C}\mathbf{1}$ by the group of orientation-preserving similarities.*

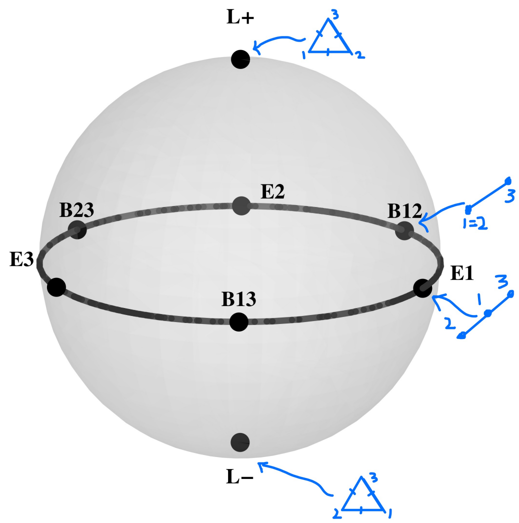


FIGURE 3. The shape sphere. The points marked $B12, B23, B31$ represent the three binary collisions. The points marked $E1, E2, E3$ represent the three Euler configurations.

fig:shape sphere

PROOF. Scaling a triangle $q \in \mathbb{C}^3$ by $\lambda > 0$ sends $q \mapsto \lambda q$ and hence sends the corresponding shape $\mathbf{w} \in \mathbb{R}^3$ to $\lambda^2 \mathbf{w}$. It follows that the oriented similarity classes of triangles are in bijective correspondence with the rays through the origin (triple collision!) of shape space. As with any real vector space, this set of rays is in turn in bijective correspondence with the sphere about the origin. We could use any non-zero radius, relative to any norm to define this sphere. We use the sphere $w_1^2 + w_2^2 + w_3^2 = 1/4$ for this sphere and call it $\mathbb{S}^2$. The metric rationale behind the $1/4$ and this choice will be given later.

We have drawn the shape sphere in figure 3. Its equator $w_3 = 0$ represents the collinear configurations. The three binary collision points $r_{12} = 0, r_{23} = 0, r_{31} = 0$ are three special points which lie on the equator of the shape sphere. Euler's three collinear configurations are three other special points, and they are interleaved with the collision points. Finally we have Lagrange's two equilateral triangles, a 'right handed one' and a 'left handed one' which we may think of as the North and South pole of the shape sphere.

The mass inner product induces a metric on shape space to be described momentarily. In this metric, the distance of a triangle to the origin (triple collision) is

$$r = \sqrt{I}.$$

We will use $(r, s) \in [0, \infty) \times \mathbb{S}^2$ as spherical coordinates on $\mathbb{R}^3$, so that

$$\mathbf{w} = \|\mathbf{w}\| s, r^2 = \sqrt{2}\|\mathbf{w}\|$$

U, the negative of the potential, is invariant under isometries of the plane, so induces a function $U = U(w_1, w_2, w_3)$ on shape space. Being homogeneous of degree

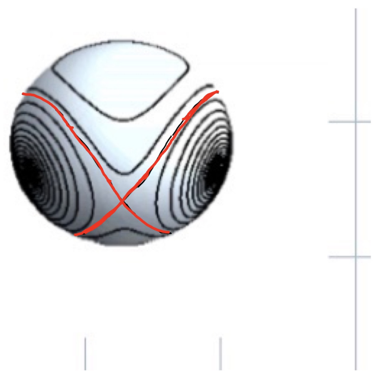


FIGURE 4. Contour plot for the shape potential. The absolute minimum occurs at the Lagrange points at the north and south poles. The binary collisions are poles where the potential tends to infinity. The red curve is a saddle curve which passes through all three Euler points if the three masses are equal.

fig: shapePotential

−1 with respect to scaling we have, in spherical coordinates

$$U(r,s) = \frac{1}{r}\tilde{U}(s), s \in \mathbb{S}^2.$$

where

$$\tilde{U} : \mathbb{S}^2 \to \mathbb{R},$$

the homogenized version of U will be called the shape potential. Our eight marked points are special points for the three-body problem precisely because they are critical points of $\tilde{U}$. The binary collision points are the poles of $\tilde{U}$, the three points where it achieves its maximum of ∞. The Lagrange points are the absolute minimum points of $\tilde{U}$ and the Euler points are its saddle points. See figure **??**.

4.0.1. *Shape Metric.* Shape space inherits a metric from the Euclidean structure (mass metric) of configuration space. The group of rigid motions of the plane acts on configuration space by isometries. The quotient of any metric space by a 'nice' [4] group of isometries inherits a metric. The distance on the quotient is defined by declaring that the distance between two points of the quotient is the distance

[4]'nice' means all orbits are closed

between their corresponding orbits in the original metric space. Shape space is such a metric quotient.

The shape space metric is Riemannian away from the origin. Its associated kinetic energy will be denoted by $K_{sh} = K_{sh}(w, \dot{w})$. If we know the angular momentum J of a curve $q(t)$ in configuration space, then its kinetic energy $K = \frac{1}{2}\langle \dot{q}, \dot{q} \rangle$ is given by

$$K = K_{sh}(w, \dot{w}) + \frac{1}{2}\frac{J^2}{r^2}, \qquad w(t) = \pi(q(t)).$$

It follows that the energy associated to $q(t)$ is

{eq: energy decomp}

$$E = K_{sh} + \frac{1}{2}\frac{J^2}{I} - \frac{1}{r}\tilde{U}(s) \tag{50}$$

The explicit expression for the shape kinetic energy is

$$K_{sh}(w, \dot{w}) = \frac{1}{2}\frac{\|\dot{w}\|^2}{4\|w\|}$$

where $\|w\|^2 = w_1^2 + w_2^2 + w_3^2$ is the standard Euclidean length squared. Said in other terms, the corresponding Riemannian metric is

$$ds^2_{shape} = \frac{\|dw\|^2}{4\|w\|}.$$

which appendix? - or maybe the section on Natural Mech systems w symmetry? -RM

Let ∇ be the corresponding Levi-Civita connection. In the appendix we show that, at fixed total angular momentum J the 3-body equations, projected to shape space, become :

{eq: Wong 3bdy}

$$\nabla_{\dot{w}}\dot{w} = -\nabla\left(V + \frac{1}{2}\frac{J^2}{I}\right) - \frac{2J}{\|w\|^2}\mathbf{w} \times \dot{\mathbf{w}} \tag{51}$$

We call these the *reduced equations* after the notion of "symplectic reduction". See equation (108) of our first appendix and the paragraph following that equation for a derivation of these reduced equations. The reduced equations have been most useful to date in the case $J = 0$ where they become:

$$\nabla_{\dot{w}}\dot{w} = \nabla U.$$

4.1. The case of equal masses and its subproblems. When the masses are all equal then their permutations induce isometries of $\mathbb{C}^3$ and hence of shape space. For example, to interchange 1 and 2 is the map $(q_1, q_2, q_3) \mapsto (q_2, q_1, q_3)$. Combined with reflection $(q_1, q_2, q_3) \mapsto (\bar{q}_1, \bar{q}_2, \bar{q}_3)$ these isometries generate the action of the 12-element dihedral group D_6 of a hexagon on shape space. This action is best seen through its induced action on the shape sphere. There, the group action is generated by reflections about 4 great circles, the three isosceles great circles $r_{ab} = r_{ac}$ and the collinear equator $w_3 = 0$.

These four circles correspond to four planes through the origin in shape space. Solutions tangent at any non-collision instant to any one of the planes remains within that plane. In this way we can see that the equal mass zero angular momentum planar three body problem contains within it four subproblems. These are the three isosceles subproblems and the collinear three-body subproblem. These subproblems are two-degree of freedom problems so are simpler to analyze and understand than the full problem.

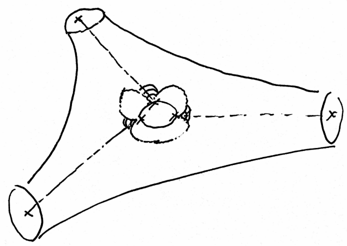


FIGURE 5. The Hill region in shape space for values $DZIOB = HJ^2$ below the minimum value of $\tilde{U}$ corresponding to the Lagrange configurations.

fig:HillRegion2

In the equal mass case, the Euler and binary points are equally spaced on the equator while the Lagrange points are placed at the geometric North and South pole, as far as can be from the collinear equator.

4.2. Morse theory and dynamics. Figure 6 depicts the basic picture of Morse theory for $\tilde{V} = -\tilde{U}$ where $\tilde{U}$ is what we have called the normalized potential. This picture govern many of the global features of planar three-body dynamics.

A function is called "Morse" if all its critical points are non-degenerate, meaning that the Hessian of the function at these points is a non-degenerate quadratic form. Morse theory [**98**] reconstructs the topology of a manifold from that of the sublevel sets $f \leq c$ of a Morse function f on the manifold, doing so by paying careful attention to how these sets change as one passes through critical values of the function. $\tilde{U}$ is a Morse function for the three-body problem, which means so is $\tilde{V} = -\tilde{U}$. The sublevel sets $\tilde{V} \leq -c$ are the super-level sets $\tilde{U} \geq c$ and are the shaded domains in figure 6 for three values of c, the value decreasing as the panels descend. To understand what these pictures tell us about dynamics, fix the values of the energy E and angular momentum J, thus defining a subvariety

$$\mathcal{P}_{E,J} \subset \mathcal{P}_0$$

invariant under the planar three-body flow.

PROPOSITION 0.3. *Consider the projection $\mathcal{P}_0 = \mathcal{P}_0(2,3) \to \mathbb{S}^2$ of the centered planar three-body phase space to its shape sphere, namely the map which sends a phase point to the shape of its configuration. Under this projection the energy-momentum level set $\mathcal{P}_{E,J}$, $E < 0$, projects onto the super-level set $\tilde{U} \geq c \subset \mathbb{S}^2$ where $c = |E|J^2$ is the Dziobek value above. If $E \geq 0$ then the level set projects onto all of $\mathbb{S}^2$.*

The proposition provides dynamical constraints on the possible shapes that might arise during evolution. For example, if $-EJ^2 := c >> 1$ corresponding to the top panel of figure 6 then the shape region of the proposition consists of three small

include a proof, or a reference to a proof. -RM disjoint discs, one centered at each of the three collisions. The disc about collision 1, that is, the collision of masses 2 and 3, is characterized by $r_{23} << r_{13}, r_{12}$. It follows that for c in this range the dynamics is always that of a "tight binary": two close masses, spinning fairly quickly about each other, with the third mass remaining relatively far away throughout the evolution. Which pair is tight cannot change with time since the discs are disconnected.

DEFINITION 0.17. *The Hill region for energy E and angular momentum J is the projection of $\mathcal{P}_{E,J}$ to configuration space $\mathbb{E}_0$. The Hill shape region is its projection to shape space. The normalized Hill shape region is its further projection to the shape sphere.*

Thus, the proposition describes the normalized Hill shape region in terms of $\tilde{U}$.

The middle panel of figure 6 depicts a bifurcation occuring at the value of c for which $\tilde{U} = c$ passes through the three Euler points. The critical values of $\tilde{U}$ are the Euler points and the Lagrange points: the central configurations of the Tour. These critical values are precisely those for which the topology of $\mathcal{P}_{E,J}$ changes, and are simultaneously the values of $c = -EJ^2$ for which $\mathcal{P}_{E,J}$ is not a smooth manifold. These critical values, generalized to higher N, are the concern of question 1, chapter 1 of this book.

some words on sundman's zero ang mom theorem- forbids regions from connecting through the cone point... ? -RM

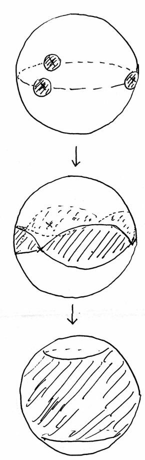


FIGURE 6. Bifurcations of the normalized Hill Shape Region as a function of the Dziobek parameter $|E|J^2$. The top panel is for $Dz >> 1$. Bifurcations occur in the shape as the value of $\tilde{U}$ at the Euler configurations are crossed, and again, as the Lagrange value is passed. The bottom panel is for Dz small.

fig: shape sphere morse

FROM RICK W CREDIT : Hill regions over shape sphere

as functions of Dz -RM

5. Reduction and a shape sphere for more bodies

sec: reduction0

The Morse-theoretic picture just described for the planar three-body problem generalizes to the N-body problem. The perspective to take is that the N-body problem is not one problem, but rather a family of problems parameterized by energy and angular momentum. We want to get a sense of changes in dynamics due to topological changes in the corresponding phase space occuring as we vary energy and angular momentum.

We focus on the case of the planar N body problem, where this picture is most complete. Our first task is to find the analogue of the shape sphere as the arena for drawing our pictures. When we think of points of the shape sphere as oriented similarity classes on triangles, the correct analogue becomes clear: the space of oriented similarity classes of planar N-gons. Group-theoretically this space is formed by letting the group of rigid motions act on configuration space, deleting the total collisions, and forming the quotient. We can form the quotient in stages, first by translation, then by rotations, then by scaling. The quotient by translations has already been achieved by going to center-of-mass coordinates, and brings us to $\mathbb{E}_0(2,N) = \mathbb{C}^{N-1}$ with origin representing total collision. When combined, the group of rotations and scalings forms the group $\mathbb{C}^*$ of nonzero complex scalars acting on on $\mathbb{C}^{N-1}$ by scalar multiplication. Upon deleting the origin and forming the quotient we get the standard representation of projective space. In this way we are led to the higher N-version of the shape sphere as $\mathbb{CP}^{N-2} = (\mathbb{C}^{N-1} \setminus \{0\})/\mathbb{C}^*$. Note that $\mathbb{CP}^1 = \mathbb{S}^2$, thus recovering the shape sphere for $N = 3$. We record the preceding discussion as a definition.

Definition 0.18 (Normalized Shape Space). *Normalized shape space for the planar N-body problem is the space of oriented similarity classes of planar N-gons, identified with $\mathbb{CP}^{N-2}$, as just described.*

As we have seen, upon going to center-of-mass coordinates the phase space $\mathcal{P}_0$ of the planar N-body problem becomes $T^*(\mathbb{C}^{N-1} \setminus \Delta)$. We then have the shape projection

$$\pi : \mathcal{P}_0 = T^*(\mathbb{C}^{N-1} \setminus \Delta) \to \mathbb{C}^{N-1} \setminus \Delta) \to \mathbb{CP}^{N-2} \setminus \Delta.$$

(In the last line Δ denotes the projectivized collision locus.) Now the N-body flow leaves invariant the functions of total energy E, and angular momentum J, so that $\mathcal{P}_0$ is foliated by invariant codimension two-subvarieties $\mathcal{P}_{E_0,J_0} = \{E = E_0, J = J_0\} \subset \mathcal{P}_0$. Consider their images under shape projection :

$$\Omega_{E,J} := \pi(\mathcal{P}_{E,J}) \subset \mathbb{CP}^{N-2}$$

The Morse-theory inspired question we want to ask is: How do the domains $\Omega_{E,J}$ depends on E and J? And how does this dependency affect dynamics on the $\mathcal{P}_{E,J}$?

Inspired by the usual notion of "Hill region" in classical mechanics, we will call $\Omega_{E,J}$ the 'normalized Hill shape region" for the given choice of E and J. We will describe these regions in terms of the normalized potential $\tilde{U}$ defined earlier (definition 48). We recall that $U = \frac{1}{r}\tilde{U}$ where $r^2 = \|q\|^2$. This function $\tilde{U}$ is invariant under rotations and scalings so can be viewed as a continuous function on normalized shape space:

$$\tilde{U} : \mathbb{CP}^{N-2} \to \mathbb{R} \cup \infty.$$

Here we added in ∞ to the range of the potential so that $\tilde{U}(\Delta) = \infty$. ($\tilde{U}$ is analytic away from Δ.) Recall (equation (45) the scale-invariant Dziobek constant.

THEOREM 0.3. *The normalized Hill shape regions are superlevel sets of the normalized potential $\tilde{U}$ as follows. If $EJ^2 < 0$ then*

$$\Omega_{E,J} = \{s \in \mathbb{CP}^{N-2} : +\infty > \tilde{U}(s) \geq \sqrt{2|E|J^2}\}.$$

If $EJ^2 \geq 0$ then $\Omega_{E,J} = \mathbb{CP}^{N-2} \setminus \Delta$

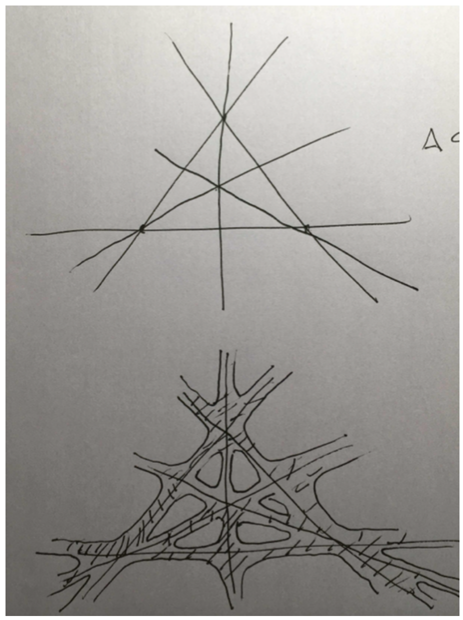


FIGURE 7. A slice through the 4-body Hill region for c large. The slice represents $\mathbb{RP}^2$, the collinear 4-body problem. The 6 lines are the intersections of Δ with the collinear slice. The region is a deleted neighborhood of Δ.

fig: 4bdyHill

REMARK. It can be useful to include Δ in to the Hill region at times so that $\Omega_{E,J} \cup \Delta = \{s \in \mathbb{CP}^{N-2} : \tilde{U}(s) \geq \sqrt{2|E|J^2}\}$ for $E < 0$.

PROOF. The decomposition (50) remains valid, where K_{sh} is now a positive definite quadratic form in shape momenta and energy is $E = K - U$ and $U = \frac{1}{r}\tilde{U}(s)$. The shape variables themselves are coordinatized by $(r,s)(0,\infty) \times \mathbb{CP}^{N-2}$ with rs playing the role that $w \in \mathbb{R}^3$ played in describing three body shape space. Shape space itself is the quotient of configuration space by rigid motions. (In an appendix we explain these coordinates in more detail and verify that shape space is the cone over $\mathbb{CP}^{N-2}$. See appendix XX for details.) Freeze $J = J_0$ and set $V_{eff} = \frac{1}{2}\frac{J_0^2}{r^2} - U$ so that, with J frozen we have that $E = K_{sh} + V_{eff}(r,s;J_0)$. If we now fix $E = E_0 =$ *const* then we must have that $V_{eff}(r,s;J) \leq E_0$ since $K_{sh} \geq 0$. Conversely, since the shape momenta vary over a vector space, we can always arrange that $K_{sh} = E_0 - V_{eff}$ provided $E_0 - V_{eff} \geq 0$. This proves that $\Omega_{E_0,J_0} = \{s \in \mathbb{CP}^{N-2} : \exists r \quad V_{eff}(r,s,J_0) \leq E_0\}$. Now freeze s as well as E_0, J_0 and look at the condition $V_{eff}(r,s,J_0) \leq E_0$ as a constraint on the positive radial variable r. Multiplying through by r^2 and rearranging yields $E_0 r^2 + r\tilde{U}(s) - \frac{1}{2}J_0^2 \geq 0$. The left hand side of this inequality is

fill in ref; or fill in early Saari decomp ref and refer to that ... -RM a quadratic polynomial in r whose discriminant is $\Delta = B^2 - 4AC = \tilde{U}^2 + 2E_0 J_0^2$. If $E_0 \geq 0$ the polynomial is eventually positive provided we take r sufficiently large and hence every $s \in \mathbb{CP}^{N-2}$ lies in Ω_{E_0,J_0}. On the other hand, if $E_0 < 0$ then this polynomial tends to negative infinity with increasing r and there may or may not be solutions to the inequality. By plotting the parabola which is the graph of $r \mapsto E_0 r^2 + r\tilde{U}(s) - \frac{1}{2}J_0^2$ we see that there are positive solutions to the inequality if and only if the roots to this quadratic polynomial are real, which is to say if and only if $\Delta \geq 0$. The latter holds if and only if $\tilde{U}(s) \geq \sqrt{2E_0 J_0^2}$, establishing the claimed characterization of Ω_{E_0,J_0}. QED

Corollary 0.1. *The normalized shape Hill regions change topology only at values $c = -EJ^2$ which are critical values for $\tilde{U}$. The corresponding critical points are the only possible points at which $\{\tilde{U} = c\}$ might fail to be a smooth hypersurface.*

Proof. The proof is application of the basic ideas of Morse theory. See [**98**]. Set $\Sigma_c := \{\tilde{U} \geq c\} \subset \mathbb{CP}^{N-2}$ and look at how these sets change as a function of c. We have $c_1 < c_2 \implies \Sigma_{c_1} \supset \Sigma_{c_2}$, $\Sigma_0 = \mathbb{CP}^{N-2}$ and $\Sigma_\infty = \Delta$. As we drop down from $c = +\infty$ to $c = 0$ the subsets are diffeomorphic as long as we do not encounter any critical values of $\tilde{U}$. In other words, if the interval $[a,b]$ is free of critical values, then all the sets $\Sigma_c, c \in [a,b]$ are diffeomorphic. So, every time the Dziobek parameter $c = -EJ^2$ crosses a critical value of $\tilde{U}$, we expect a topological change in $\Omega_{E,J}$ and hence in $\mathcal{P}_{E,J}$. That the hypersurfaces are if c is not critical is just the implicit function theorem. QED

In the next section we will see that the seeds of these topological changes, namely the critical points of $\tilde{U}$, have their own dynamical importance. They are relative equilibria, the higher N analogue of the solutions of Euler and Lagrange described in the Tour. Counting them is the subject of our first open problem.

END ...

iodic implies q periodic

add some more on Saari decomp; bifurcations, perhaps Smale results -RM

5.0.1. *Quasi-periodicity.* Proof that 'relative periodic implies quasiperiodic".

include here or in appendix? -RM

We give the proof in the planar case.

Step 1. Getting the monodromy to be orientation preserving. Suppose that $q(t+T) = g_0 q(t)$. We may assume that $g_0 \in SO(2)$, for if $g_0 \in O(2) \notin SO(2)$ then $g_0^2 \in SO(2)$ and we have $q(t+2T) = g_0^2 q(t)$. So we just go 'twice around" q.

Step 2. $\mathbb{E}_0 \cong \mathbb{C}^{N-1}$ in the planar case and the projection $\pi : \mathbb{E}_0 \to \mathbb{E}_0/\mathbb{S}^1 = Cone(\mathbb{CP}^{N-2})$ gives $\mathbb{E}_0$ the structure of a principal circle bundle away from 0. Our solution curve $q(t)$ avoids total collision by assumption. If our curve is relatively periodic then its projection to shape space is periodic so yields an an embedded $\mathbb{S}^1 \subset Cone(\mathbb{CP}^{N-2})$. Restricted to this shape curve, our projection π yields a principal circle bundle over a circle. Any such bundle is trivial, so we can identify π^{-1}(shape curve) with $\mathbb{S}^1 \times \mathbb{S}^1$ In these coordinates $q(t)$ is coordinatized as $(\theta_1(t), \theta_2(t))$ with $\theta_1(T) = \theta_1(0)$ while $\theta_2(T) = \omega + \theta_2(0)$. A bit of work now yields that we can find a diffeomorphism of the two torus $\mathbb{S}^1 \times \mathbb{S}^1$ turning this curve into one for which $\theta_2(t) = t + \omega\frac{t}{T}$ and so $q(t)$ is quasi-periodic with at most two frequencies.

I removed the section "perturbation theory"; it was just 6 lines , one paragraph. I've relegated it to the Appendix and probably will leave it out of the book -RM

6. Limitations

sec: finite size

moved section on Lagrangian and Hamiltonian formalism to the appendix. -RM

6.1. Finite size, harmonic functions and $1/r$. Planets and stars are balls of matter, not points. It is a remarkable and beautiful theorem of Newton that the combined gravitational forces exerted by the masses of a spherical mass distribution on a point exterior to the distribution is exactly what we would get if we simply placed all of the mass at a point at the sphere's center. (See Chandrasekhar, chapter 15. Newton, Prop. LXX1, Theorem XXX1.) This theorem is basic to the power of Newton's law of gravitation and is essentially equivalent to the fact that the fundamental solution to the Laplacian in $\mathbb{R}^3$ is a constant times $1/r$.

The fundamental solution for the Laplacian in $\mathbb{R}^d$ is c/r^{d-2} and so one arrives at the same theorem regarding spherically symmetric mass distributions in $\mathbb{R}^d$, provided the gravitational force is taken to be $F(x,y) = -m_a m_b \nabla_a (1/|x-y|^{d-2})$, which is to say, a $1/r^{d-1}$-force. For this reason, many people object to power law potentials with arbitrary exponents α in arbitrary dimensions d. They exclaim to me that if you insist on working in d dimensions then you have to take

$$\alpha = d - 2$$

EXERCISE 0.13. *Suppose we take the pair potential between two bodies in $\mathbb{R}^d$ to be of the form $m_a m_b f(r_{ab})$ so that the force on body a is $-m_a \nabla_a \Sigma m_b f(r_{ab})$. Replace the finite collection $\{m_b, b \neq a\}$ of masses by a a mass distribution $\rho(b)db$ so that the force of the mass distribution on a body x exterior to the mass distribution is, up to a constant, minus the gradient of $W(x) = \int_B f(|x-b|)\rho(b)db$. Suppose that the distribution is spherically symmetric so that $\rho(b)db = \rho(|b|)db$ with db being Lebesgue measure in d-dimensions. Prove that $\nabla W(x) = c\nabla f(r)$ for some constant c if and only if $f = c_2 r^{d-2}$.*

With this perspective taken, we could formulate the N-body problem on any non-compact symmetric space, taking the pair potential to come from the fundamental solution to the Laplacian. Lobachevsky himself stated the Kepler problem on hyperbolic space. See [**83**].

EXERCISE 0.14 (Open problem?). *Investigate the Kepler problem in complex hyperbolic space*

6.2. Relativitistic limitations. Special relativity tells us the speed of a massive object must be less than the speed c of light. But as our point masses approach each other the forces between them increase so that as they tend to collision, their velocities tend to infinity, in clear violation of the dicates of special relativity. To be concrete, consider N equal masses with mass m engaged in a zero energy solution. From $H = K - U$ we get $K = U$. Now $K = \frac{m}{2}\Sigma|v_a|^2$ while $U = \Sigma G m^2/r_{ab}^2$. If v_{max} is the maximum velocity of the bodies and r_{min} is the minimum of the interbody distances r_{ab} then $N m v_{max}^2 \geq 2K$ while $U \geq Gm^2/r_{min}^2$ so that $2Nmv_{max}^2 \geq Gm/r_{min}^2$, or $v_{max}^2 \geq Gm/2Nr_{min}^2$. The relativistic constraint $c^2 \geq v_{max}^2$ yields $r_{min} \geq \sqrt{G/2N}(m/c)$. Imposing special relativity limits approach to collision in gravity. This discussion ignores the relativistic coupling of mass with velocity as measured in a fixed inertial frame. As the velocity of a massive particle increases so does mass, increasing without bound as the speed of light is approached. We will not even bother trying to figure out what to do with that massive can of worms in relation to any N-body type equations.

A more serious philosophical objection to Newtonian gravity raised by special relativity is that forces cannot be transmitted instantaneously. This is the property of "action at a distance". In equation 5, 2 a tiny change in the position of one body instantaneously affects the forces on all bodies, even those light-years away. It seems to be impossible to write down N-body type ODEs which are invariant under the Lorentz transformations of special relativity. One can imagine trying to pose N-body type problems using differential delay equations, staggering the time at which forces are applied according to the speed of light. This has never been done in a useful consistent way. (The interested reader may wish to consult Feynman and Wheeler [**41**] who argue that one must apply a "Jacob's ladder" of retarded and advanced potentials -forward and backwards delays as it were– in order to formulate Maxwell's equations as an N-body variational principle.)

Finally, Newton's N-body equations are a limiting case of general relativity to which we refer the interested reader to the beauiful short book by Einstein [**33**]. The equations describing general relativity are PDEs known as Einstein's equations. The general relativistic two-body problem is that of two black holes moving under each other's gravitationally fields and cannot be solved in closed form. Numerical solutions for this two-body problem proved crucial in interpreting the LIGO observations of gravitational waves from a pair of black holes merging.

7. Notes.

Pollard. Singer. Landau-Lifshitz.

The collision locus **??** is often called the fat diagonal.

In my N-body career I've avoided perturbation theory like the plague. I believed it hopeless to try to compete with its many illustrious practitioners: Lagrange, Laplace, Poincaré, Arnol'd, Meyer, etc. and have instead settled myself into some unexplored niches of celestial mechanics, notable variationally motivated crevices of the subject.

incompleteness:

The flow is incomplete: some solutions fail to exist for time t due to collisions. The set of inital conditions yielding incomplete solutions is non-empty but measure zero. CITE SAARI.

Shape space and Iwai. Clearly stated in [**?**] : Topology and Mechanics II.

Quotients: Bierstone.

Center-of-mass frame. Albouy advocates for a different way to quotient out by translations which he has christened the method of "dispositions". See [**?**], [**?**], [**7**]. His method is based on intrinsic linear algebra. It has proved especially well-suited for attacking open question 1 here, that of counting central configurations.

Formalisms of classical mechanics. The Lagrangian and Hamiltonian formalisms are covered in most books on classical mechanics. For treatments with significant mathematical depth I recommend [**?**] and Arnol'd.

Part 2

Questions

CHAPTER 1

Is the Number of Central Configurations Finite?

Open Question 1.1. *Is the number of central configurations (mod symmetries) finite?*

1. Why care?

When teaching dynamics I will hand my students a polynomial vector field to analyze for their homework. I teach them to start by locating the zeros of the vector field - its equilibrium points. Once found, linear analysis comes to the fore and a whole game unfolds.

Newton gave us homework in 1687 in the form of equation (5). But the N-body system has no equilibria. N stars cannot just sit in space, unmoving. They attract each other! It seems that we cannot start the game.

The next best thing to an equilibrium is a relative equilibrium : an equilibrium modulo symmetries. Upon forming the quotient of phase space by the symmetry group of Galilean motions (section **??**) the relative equilibria project to equilibria for the quotient dynamics. We can use these quotient equilibria as replacements for equilibria, and can begin Newton's homework assignment.

def:relative equilibrium

Definition 1.1. *A relative equilibrium for the N-body problem is a solution to the problem having the form*

$$q(t) = g(t)q_*$$

where $q_* \in \hat{\mathbb{E}}$ is a fixed collision-free configuration and where $g(t) \in SO(d)$ is a one-parameter group of rotations of $\mathbb{R}^d$. A relative equilibrium configuration q_* is a configuration through which passes a relative equilibrium.

We have already the seen relative equilibria for the 3-body problem. They are contained in the solutions found by Euler and Lagrange as described in the Tour in subsection 3.1.

On the other hand

def:cc

Definition 1.2. *A central configuration for the N-body problem is a configuration which, dropped from rest, evolves by shrinking to total collision at its center of mass point, all the time remaining in its same shape and orientation. In other words, if the initial configuration q_* has center of mass is zero, then q_* is a central configuration if and only if the solution $q(t)$ to the N-body equations having initial conditions $(q, v) = (q_*, 0)$ has the form*

{ansatz_cc}

$$q(t) = r(t)q_*, r(t) \in \mathbb{R} \tag{52}$$

We sometimes refer to such solutions as "dropped central solutions".

In the planar N-body problem central configurations are the same as relative equilibrium configurations.

prop: CCS=REs

Proposition 1.1. *For the planar N-body problem the set of central configurations coincides with the set of relative equilibrium configurations.*

For the proof of Proposition 1.1 see subsection 3.1 below.

Due to the rotational and scaling symmetries of Newton's equations, if q_* is a central configuration then so is any rotate or scaling of q_*. Central configurations come in whole orbits with respect to the action of the Lie group of rotations and scalings and so cannot form a finite set. When counting central configurations we must count each orbit as a single point in order to have a chance of getting a finite number. Our open question asks:

Completed Open question. Is the set of **similarity classes** of central configurations a finite subset of its normalized shape space?

1.1. Relations between CCs and REs for higher d. For dimensions $d \neq 2$, the central configurations and relative equilibrium configurations are no longer equal. The relations between these two sets of configurations for the N-body problem in d dimensions are as follows.

prop: CCS=REs2

Proposition 1.2. *Write RE for the set of relative equilibrium configurations and CC for the set of central configurations for the N body problem in d-space. Then*

a) If d is even then $CC \subset RE$.
b) If d is odd then $CC \setminus RE \neq \emptyset$.
d) If $d \geq 4$ then $RE \setminus CC \neq \emptyset$.

The proof of Proposition 1.2 is scattered through the chapter and help serve as a kind of guide to the chapter. For the proof of (a) go to the subsection 3.3. For the proof of (b) see Proposition 1.5 below where we show that the regular simplex in any dimension is a central configuration. In odd dimensions the simplex cannot be rotated so as to yield a relative equilibrium. For example, the regular tetrahedron cannot be a relative equilibrium for the 4 body problem in 3 space since any one-parameter subgroup of rotations has an axis of rotation and hence is the rotation of some plane. The entire simplex does not sit in the plane so one of the bodies is left doing circles above or below the plane which contradicts the laws of gravity: the result cannot solve the 4 body problem. For (c) see section 1.6.

Remarks on terminology. 1. In chapter 0 we gave another definition of "relative equilibrium" : a solution under which the mutual distances remain constant. See 0.11. It is a simple exercise to show that these two definitions are equivalent.

2. Albouy and Chenciner [**7**] wrote a seminal work on expressing the N-body problem solely in terms of mutual distances. There they introduced the phrase "balanced configuration" as a synonym for "relative equilibrium configuration", part of the idea behind the word "balance" being that gravitation and centrifugal forces must "balance" in order to be able to spin such a configuration and get it to be a solution.

subsec: motivations
reasons

1.2. More reasons to care.

- (1) Central configurations correspond to the *actual equilibria* of a partial compactification of the N-body problem known as the McGehee blow-up. See theorem 1.2.
- (2) Central configurations are the only roads in and out of total collision. See proposition **??**
- (3) Relative equilibria are the critical points of the energy-momentum map. Consequently, the behaviour of these maps at relative equilibria provide the dominant input as to the topology of these manifolds and the corresponding symplectic and contact reduced spaces.
- (4) Central configurations and relative equilibria generate the only solutions to the N-body problem for which we have explicit formulae.

Reason (3) is explained in proposition **??**. Reason (4) is explained below in lemma 1.1. For explanations of reasons (2)-(4) go to section 1.4 for details.

start off refering to CCs as being critical pts.. refer chapter 0 re U -RM

1.3. CCs as critical points. We have seen (equation (35) that Newton's equations are succinctly encoded by the function U, the negative of the potential. U has no critical points, however, the normalized potential of definition (48

{eq: scaled U}

$$\tilde{U} = \sqrt{I}U : \hat{\mathbb{E}}_0 \to \mathbb{R} \tag{53}$$

has critical points and these critical points are *precisely* the central configurations whose center of mass is at zero. Here $I(q) = \|q\|^2$ is the moment of inertia. See exercise 1.1 below.

$\tilde{U}$ is invariant under scaling and rotation and hence descends to yield a function on the normalized shape space, the quotient of $\mathbb{E}_0$ by scalings and rotations. (See figures 4 and 6 for depictions of its contours when $N = 3$). The open question then asks: *Does $\tilde{U}$, viewed as a function on the normalized shape space, have a finite number of critical points?*

so: have to define normalized shape space in intro ch -RM

2. What's known?

We saw all of the central configurations for the planar three-body problem in the tour of chapter -1. Modulo oriented similarity there are five: the three collinear solutions of Euler and the two equilateral solutions of Lagrange.

Moczurad and Zgliczyński [**99**] compiled the table below for the equal mass case up to $N = 7$. The table lists the number "#' of oriented similarity classes of central

N	unlabelled #	labelled #
2	1	1
3	2	5
4	4	50
5	5	354
6	9	3624
7	14	53640

TABLE 1. Counting central configurations in the equal mass case. See [**99**]. The middle column is the number of unlabelled distinct configurations. The last column counts the number after labelling.

tab:equalmassCCs

configurations in the equal mass planar N-body problem as a function of the number N of bodies. When the masses are all equal we can permute the masses of a central configuration to get another central configuration. The middle column "unlabelled # " of table 1, counts oriented similarity classes of central configurations modulo this action of the permutation group: it counts the number of distinct shapes of N-gons in the plane, ignoring the mass labels at the vertices. The second column "labelled # " accounts for labelling. Euler and Lagrange found the 2 shapes for $N = 3$. Albouy [**2**] established the four unlabelled shapes as depicted in figure 18 and the 50 labelled shapes for $N = 4$.

The next table summarizes known finiteness results for the planar N-body problem.

N	#	authors	year
2	1	Newton; Kepler	1667
3	5	Euler, Lagrange	1767, 1772
4	$\leq 50^*$	Hampton-Moeckel; Simo	2006 ;1978
5	generically finite	Albouy-Kaloshin	2012

tab:label

TABLE 2. Established bounds for the number of planar central configurations.

N= 4. The asterisk by the result "50" for $N = 4$ in the table is there to indicate that the rigorously established upper bound ([**52**]) is not 50. Simo, [**99**], Moeckel, and others have done careful numerics which make it nearly certain that 50 is the actual least upper bound. But the rigorously established upper bound is 8472.

N = 5. In the table's last row "generically finite" means that the number of central configurations mod symmetry is finite provided the vector of masses $\vec{m} = (m_1, \ldots, m_5)$ comes from a generic set, specifically from a Zariski-dense subset of the positive part of $\mathbb{R}^5$. masses.

The following counterexample discovered by Gareth Roberts [**?**] underlines the difficulty of the problem.

THEOREM 1.1. *In the five body problem if we allow one of the five masses to be negative then the answer to the open question is no. Specifically if the mass distribution is* $(4, 4, 4, 4, -1)$ *then there exist a curve's worth of inequivalent central configurations within the planar* 5*-body normalized shape space.*

Robert's example is not a counterexample to the conjectured finiteness since for gravity we require masses to be positive. Understanding Robert's example in a deep way proved essential to Albouy and Kaloshin [**6**] for their $N = 5$ result.

2.1. A few general results.

DEFINITION 1.3. *Write* $\#(N, \mathbf{m}, d)$ *for the number of central configurations for the* N *body problem with mass distribution* $\mathbf{m}$*, and the bodies moving in* d *space.*

subsubsec: Moulton

2.1.1. *Moulton's Collinear result.*

$$\#(N, \mathbf{m}, 1) = \frac{1}{2} N!$$

Moulton [**120**] established this fact by showing that there is exactly one linear central configuration for each ordering of the N masses on the line. See proposition 1.4 below for a restatement and a proof. There are $N!$ such orderings, hence the count. If we view these collinear central configurations within the plane instead of the line then an extra symmetry acts: rotation of the line by 180 degrees within the plane which acts by $O(1) = \mathbb{Z}_2$ on the line, flipping the ordering of the masses on the line. It is standard to count a linear central configuration and its flip as the same. We thus get $N!/2$ linear Moulton central configurations for any given mass distribution. When $N = 3$ these are the $3!/2 = 3$ central configurations found by Euler.

subsubsec: Xia

2.1.2. *Xia's open finiteness results.* Xia [**?**] has established, for any N, the existence of an open set of mass distributions $\mathbf{m} = (m_1, \ldots, m_N)$ such that

$$\exists m_1, \ldots, m_N : \#(N, m, 2) = (N-2)!((N-2)2^{N-1} + 1),$$

thus establishing finiteness for an open set of mass distributions. Xia's construction is iterative, assuming m_1 and m_2 are close to equal, then imposing conditions of the form $m_i >> m_{i+1}$ on successively introduced masses.

2.1.3. *Palmore's Morse estimates.* As described above (see also exercise 1.1) central configurations can be characterized as the critical points of the function $\tilde{U}$ on shape space. (See equation (53). This function depends parametrically on the mass distribution $\mathbf{m} = (m_1, \ldots, m_N)$.

Proposition 1.3 (Palmore; see Moeckel, Prop. 25). *Suppose that for some mass distribution* $\mathbf{m}$ *the scaled potential* $\tilde{U}$ *is a Morse function on the planar N-body shape space* $\mathbb{CP}^{N-2} \setminus \Delta$. *Then, for that mass distribution, the answer to the Open Question is 'yes' and, moreover, we have the Morse lower bound*

$$\frac{3}{4}N! - \frac{1}{2}(N-1)! \leq \#(N, \mathbf{m}, 2)$$

for the number of central configurations.

For example, for $N = 10$ the Moulton number is $1,814,400$, while the Morse lower bound of Palmore is $2,540,160$.

3. Central Configurations as Critical Points

sec:CC eqns

We derive the fact that central configurations are critical points of functions built from U and I (equations 24, 40).

View equation (52) from Definition 1.2 as an ansätz for a solution to Newton's equations and plug it into the condensed form (35) of the N-body equations to get

{cc0}
$$(\ddot{r})q_* = \frac{1}{r^2}\nabla U(q_*) \tag{54}$$

where we used $r^2 = \|q\|^2$ and that ∇U is homogeneous of degree -2. Take the *mass* inner product of both sides of this equation with q_* to arrive at

{cc1}
$$\ddot{r} = -\frac{\mu}{r^2} \tag{55}$$

where

{KeplerConstant}
$$\mu = \frac{U(q_*)}{I(q_*)} \tag{56}$$

In arriving at (56) we've used $r^2 = I(q) = \langle q, q\rangle$ and $-U(q) = \langle q, \nabla U(q)\rangle$ the latter following from U being homogeneous of degree -1 and Euler's identity. Equation (55) is the 1-dimensional Kepler problem. (See section 0.2.) If $r(t)$ solves equation (55) then $\ddot{r}r^2 = -\mu = const$ so equation (54) yields

{cc2}
$$\nabla U(q) = -\mu q; \text{ for } \qquad q = q_*. \tag{57}$$

We will call this equation the *central configuration equation* .

These steps are reversible. Suppose that $q = q_*$ solves the central configuration equation (57). Form the one-dimensional Kepler problem (55) with Kepler constant μ given by equation (56) and take the solution $r(t)$ having inital conditions $r(0) = 1, \dot{r}(0) = 0$ to arrive at a dropped central solution as per definition 1.2. Thus central configurations are the solutions to the central configuration equation (57).

We can rewrite the central configuration equation (57) as the Lagrange multiplier equation

{cc3}

$$\nabla U(q) = -\lambda \nabla I(q) \qquad \lambda = \mu/2 \tag{58}$$

since $\nabla I = 2q$. Equation (58) says that q is a critical point of the function U restricted to the sphere $I = I(q_*)$.

We have just shown that central configurations are critical points of the function U upon restricting it to a sphere $I = const.$. Here are two additional characterizations of central configurations as critical points.

ex: CC1

EXERCISE 1.1. *A. Show that the central configurations are the critical points of the normalized potential $\tilde{U} = \sqrt{I}U$. (See definition 48.)*

B. Show that critical points of the function $U + I$ are central configurations and that any central configuration can be scaled so as to be a critical point of this function.

ex: CCs

subsec: planar case

3.1. The planar case. Proof of Proposition 1.1. By making slight variations in the derivation just given of equations 58 and 57 we will prove Proposition 1.1 which says that central configurations and relative equilibrium configurations are the same thing in the planar N-body problem.

Identify $\mathbb{R}^2$ with $\mathbb{C}$ in the usual way, so that (x, y) is sent to $(x+iy)$ and rotation by θ becomes multiplication by $\lambda = e^{i\theta}$. Then $\mathbb{E} \cong \mathbb{C}^N$. Replace the real scalar $\lambda(t)$ in our ansätz with the complex scalar $\lambda(t) \in \mathbb{C}$. Observe that $\nabla U(\lambda q) = \frac{\lambda}{|\lambda|^3} \nabla U(q)$ for $\lambda \in \mathbb{C}$. Now go through precisely the same manipulations as we went through above, to see that our modified ansätz with $\lambda(t) \in \mathbb{C}$ satisfies Newton's equations if and only if equation 57 is satisfied together with the 2-dimensional Kepler problem $\ddot{\lambda} = -\mu \frac{\lambda}{|\lambda|^3}$. The Kepler constant μ is still given by equation (56). Since our planar Kepler problem admits the circular solution $\lambda(t) = e^{i\omega t}$ with frequency ω satisfying $\omega^2 = \mu$ we have proved the proposition. QED

A moment's reflection shows we have proven more:

lem:CCtoRE

LEMMA 1.1. *In the planar N-body problem a complex one-dimensional family of solutions of the form $q(t) = \lambda(t) q_*$ passes through each central configuration $q_* \in \mathbb{C}^N$ where the $\lambda(t)$ solve the corresponding planar Kepler problem $\ddot{\lambda} = -\mu \frac{\lambda}{|\lambda|^3}$ described above. The complex parameter can be taken to be the initial Keplerian velocity $\dot{\lambda}(0)$ at time 0 of the initial conditons $(\lambda(0), \dot{\lambda}(0))$ to our Kepler problem, where we take $\lambda(0) = 1$ to insure the family passes through q_* at time $t = 0$.*

These Kepler families are the exact solutions which central configurations give rise to, referred to as the 4th ‘reason to care] in subsection 1.2.

If, instead, we label the central configurations by their shapes $[s] \in \mathbb{CP}^{N-2}$ then we get a 3-parameter family of solution curves (modulo time translation) for each central configuration shape $[s]$. As parameters we can use the energy, angular momentum, and orientation of the solution.

3.2. Consequences of Criticality. The characterization of central configurations as critical points has a number of consequences. We describe three now.

prop: Moulton

PROPOSITION 1.4 (Moulton). *Considered modulo scaling and translation, the collinear N-body problem has exactly $N!$ central configurations one for each choice of orderings of placements of the masses on the line, and hence $N!/2$ when viewed as planar CCs modulo rotations, translations and scalings.*

In order to describe the next consequence we introduce the *rank* of a configuration. This rank of a configuration $q = (q_1, \ldots, q_N) \in \mathbb{E}(d, N)$ is the dimension of the affine span $\mathbb{A} \subset \mathbb{R}^d$ of its vertices $q_1, \ldots, q_N \in \mathbb{R}^d$. Recall that $\mathbb{A}$ is the smallest affine subspace of $\mathbb{R}^d$ containing the N points q_a and can be characterized as the collection of all points P expressible in the form $P = \Sigma t_a q_a$ as $(t_1, \ldots, t_N)$ varies over $\mathbb{R}^N$. We can find an isometry of $\mathbb{R}^N$ whose fixed point set is $\mathbb{A}$. It follows that any solution to the N-body problem with initial condition $(q, 0)$ lies in $\mathbb{A}$. The rank of q is an integer which is less than or equal to both d and $N-1$.

prop: simplex

PROPOSITION 1.5 (simplex). *The only central configuration for the N-body problem which has rank $N-1$ is the regular $N-1$-simplex, the unique-up-to-similarities configuration for which all side lengths r_{ab} are equal. The regular simplex is a central configuration regardless of the mass distribution $m_a, a = 1, \ldots, N$, and the exponent α of the power law, for any power law N-body problem, including the gravitational one.*

In order to state the third consequence we recall that a critical point of a smooth function is called non-degenerate if its Hessian - that matrix of second partial derivatives - is non-degenerate. A Morse function is a smooth function all of whose critical points are non-degenerate. Morse theory [**98**], one of the most important developments of 20th century mathematics, relates the topology of a manifold to the nature and number of critical points of a Morse function on the manifold.

prop: Palmore

PROPOSITION 1.6. *If, for a given choice $\mathbf{m} = (m_1, \ldots, m_N)$ of masses, $\tilde{U}$ is a Morse function on the normalized shape space for N-bodies in d-space, then the answer to the open question of this chapter is 'yes' for these masses: $\#(N, \mathbf{m}, d) < \infty$. Moreover, in the case $d = 2$ of the planar N-body problem, we have Palmore's lower bound*

$$\#(N, \mathbf{m}, 2) \geq \frac{3}{2} N! - 2(N-1)!.$$

We now prove these consequences of criticality.

PROOF FOR PROPOSITION 1.4 [ON THE MOULTON CONFIGURATIONS]. We make use of the function $U + I$ of part (B) of exercise 1.1.

The configuration space of the collinear N-body problem is $\mathbb{R}^N$. $\mathbb{R}^N \setminus \Delta$ consists of $N!$ components, corresponding to the $N!$ possible orderings of N points on the line. Each component is itself convex. The restriction of $U+I$ to a single component is easily checked to be convex and since it blows up on Δ it blows up on the boundary of any component. (See the proof of Moulton's theorem on p. 33-35 of [**101**] for the computation showing that this function is strictly convex.) A convex function on a convex domain which blows up on the domain's boundary has a unique minimum in the interior of the domain. QED

PROOF OF PROPOSITION 1.5 [ON THE REGULAR SIMPLEX]. We do the proof in the power law case of $U = G\Sigma m_a m_b / r_{ab}^{\alpha}$

We may suppose that q is centered and that $d \leq N-1$ since the space in which the dropped solution $q(t)$ with initial condition $(q,0)$ evolves is the affine space $\mathbb{A}$ whose dimension is $rank(q) \leq min(d, N-1)$. That is, we may assume that $\mathbb{R}^d = \mathbb{A}$. If the configuration q has rank $N-1$ we have that $d = N-1$ and no element of the linear isometry group $O(d)$ leaves q fixed. It follows that the rotational orbit $SO(d)q$ of q admits a neighborhood U on which $O(d)$ and consequently $SO(d)$ acts freely. We may take $U = SO(d)U$ to be an $SO(d)$-invariant neighborhood. Then $U/SO(d) \subset \mathbb{E}/SO(d)$ has dimension $\binom{N}{2}$ and the $\binom{N}{2}$ side lengths r_{ab} form a set of coordinates for this neighborhood $U/SO(d)$ of the quotient space. We have $U = G\Sigma m_a m_b / r_{ab}^{\alpha}$ while $I = \frac{1}{M}\Sigma m_a m_b (r_{ab})^2$. Writing out the Lagrange multiplier equation $dU = -\frac{\mu}{2} dI$ in these variables we get

$$-\alpha \Sigma m_a m_b r_{ab}^{-\alpha-1} dr_{ab} = \frac{-\mu}{2} \frac{2}{M} \Sigma m_a m_b r_{ab} dr_{ab},$$

or $\Sigma m_a m_b (\alpha r_{ab}^{-(\alpha+1)} - \frac{\mu}{M} r_{ab}) dr_{ab} = 0$ Since the dr_{ab} are linearly independent of $U/SO(d)$ we can cancel the common factor of $m_a m_b$ and are left with the $\binom{N}{2}$ equations $\alpha r_{ab}^{-(\alpha+1)} - \frac{\mu}{M} r_{ab} = 0$. This leaves us with $\frac{M\alpha}{\mu} = r_{ab}^{\alpha}$ so that all the r_{ab} must have the same length $(\frac{M\alpha}{\mu})^{1/\alpha}$.

QED

SKETCH OF PROOF OF PROPOSITION 1.6 [ON MORSE AND PALMORE'S ESTIMATE].

By the Morse lemma ([**98**]), the critical points of a Morse function are discrete. Hence to show that the critical points of a Morse $\tilde{U}$ are finite it is enough to show that the critical lie in a compact subset of the normalized shape space minus collisions. This fact was proved by Shub.

$\boxed{\texttt{Shub}}$ LEMMA 1.2 (Shub). *For fixed positive masses $m = (m_1, \ldots, m_N)$ there is a conical neighborhood of the collision locus in configuration space which is free of collisions.*

For a proof Shub's lemma see Proposition 15 on p. 29 of [**101**] or the original work [**143**].

To prove Palmore's estimate one combines the Morse inequalites with Moulton's theorem. The Morse inequalities are inequalities between weighted sums of critical points of a Morse function and sums of Betti numbers of the underlying manifold. Moulton asserts that $\tilde{U}$ has exactly $N!/2$ critical points within the collinear configurations $\mathbb{RP}^{N-2} \setminus \Delta \subset \mathbb{CP}^{N-2} \setminus \Delta$. They are all absolute minima for the restrictiion of $\tilde{U}$ to the collinear space. In the orthogonal directions $iT_{[s]}\mathbb{RP}^{n-2} = (T_{[s]}\mathbb{RP}^{N-2})^{\perp}$ to the collinear space the critical points are all local maxima. This proves that the number of minus signs - the index of the critical points - are $N-2$. One puts this data into the Morse inequalities to achieve the result. See Moeckel [**101**], p. 45, the proof of Proposition 25, or the original reference [**124**] for details.

add on theorem 2: Palmore's index $\leq N-2$ result ? -RM

REMARK. A generic smooth function is Morse. This implies, in particular, that arbitrarily close to any function is a Morse one. Thus we expect $\tilde{U}$ to be Morse for most masses. Being Morse is an open condition, and Xia established that the set of masses for which $\tilde{U}$ is Morse is non-empty. We expect that the subset of the mass distribution space $\mathbb{P}R_+^{N-1}$ on which $\tilde{U}$ fails to be Morse is an algebraic subvariety - in the usual spirit of singularity theory. If true, this would establish generic finiteness. Apparently, this expectation is hard to prove and is probably the wrong approach to the problem.

subsec: even ds

3.3. Even-dimensions and almost complex structures. The planar results extend to the even-dimensional N-body problem by viewing multiplication by i acting on $\mathbb{R}^2 \cong \mathbb{C}$ ias an "almost complex structure" and generalizing appropriately. Recall that an almost complex structure on $\mathbb{R}^d$ any skew-symmetric transformation $\mathbb{J} : \mathbb{R}^d \to \mathbb{R}^d$ such that $\mathbb{J}^2 = -I$. We will insist that $\mathbb{J}$ be skew-symmetric as well so that $exp(t\mathbb{J})$ is a one-parameter subgroup of $SO(d)$. If $\mathbb{R}^d$ admits an almost complex structure then $d = 2k$ is even, and the structure yields an isomorphism $\mathbb{R}^d \cong \mathbb{C}^k$. Return to the proof of (a) of Proposition 1.2 given above. Replace i everywhere by this $\mathbb{J}$. The solution ansätz is then $q(t) = exp(\omega \mathbb{J} t) q_*$ where ω is a scalar. We have proven (a) of Proposition 1.2: central configurations in even dimensions yield relative equilibria.

this subsection seems misplaced. where to put it? -RM

Example 1.1. *The regular tetrahedron is a central configuration for the 4-body problem in $\mathbb{R}^3$. See proposition 1.5. The regular tetrahedron cannot be made into a relative equilibrium for the problem: it cannot be spun about an axis in space and turned into a solution to Newton's equations. However $\mathbb{R}^4 = \mathbb{C}^2$ admits an almost complex structure. By taking the standard embedding $\mathbb{R}^3 \subset \mathbb{R}^4$ we promote the regular tetrahedron from a central configuration to a relative equilibrium configuration for the 4-body problem in $\mathbb{R}^4$. A similar trick works for the regular N simplex for any even N.*

motivation

4. Exploring the Motivations

In the beginning of this chapter (subsection 1.2) we listed four reasons, labelled (1)-(4) there, for studying central configurations. We gave a detailed exploration of reason (4) in lemma 1.1 Here we go into some detail regarding the other reasons.

4.1. Equilibria out of blow-up. This was the first reason we listed. Write

$$r = \sqrt{\langle q, q \rangle}$$

so that $r = 0$ corresponds to total collision and set

$$q = rs; \langle s, s \rangle = 1.$$

We call s the normalization of q and (r, s) 'spherical coordinates' on configuration space. The McGehee blow-up is the change of variables:

McGehee

$$q = rs \tag{59}$$
$$v = r^{-1/2} y \tag{60}$$
$$dt = r^{3/2} d\tau \tag{61}$$

These relations define a *transformation* $(q, v; t) \mapsto (r, s, y; \tau)$ which is invertible and well-defined as long as $r > 0$. McGehee's blow-up changes both the dependent $((q, v))$ and independent (t) variables of Newton's equations. The McGehee process creates, as if by magic, equilibria for Newton's equations on the previously invisible collision manifold $r = 0$.

thm: McGehee

THEOREM 1.2. *When rewritten in McGehee variables the vector field defining the N-body equations extends to an analytic vector field (62) defined on the collision manifold $r = 0$. Equilibria arise on the collision manifold and are in 2-to-1 correspondence with normalized central configurations. Each dropped central solution (definition 1.2) is associated to a pair of equilibria, being heteroclinic between the pair.*

This theorem enables us to return to the start of Newton's homework for us. At last we have equilibria. We can linearize about them. The linearized flow governs near-collision honest ($r > 0$) solutions to Newton's equations. Linearization and stable-unstable manifold results so obtained have been among the deepest results achieved in the last 40 years on the N-body problem.

The idea behind the McGehee transformation is to make Newton's equations scale invariant. The idea cannot literally succeed since the equations are not invariant under scaling, but we can isolate all scale information into the single radial variable $r = \|q\|$. Scaling by $\lambda > 0$ acts by $q \mapsto \lambda q, \lambda \in \mathbb{R}$ so that the basic variables r and $q \in \mathbb{E}$ are homogeneous of degree 1. The variable s is scale invariant. Under scaling $U \to \lambda^{-1}U$. The variable guiding principle for the remainder of the transformation is that the total energy $E = K(v) - U(q)$ must scale homogeneously. Since K is homogeneous of degree +2 in velocities v the guiding principle requires that $v \mapsto \lambda^{-1/2}v$ so that K is also homogeneous of degree -1. (For α-force laws we would get $v \mapsto \lambda^{-\alpha/2}v$.) We can effect this scaling of v by insisting that time scale according to $t \mapsto \lambda^{3/2}t$, for then $v = \frac{dq}{dt} \mapsto \frac{\lambda dq}{\lambda^{3/2}dt} = \lambda^{-1/2}v$. This leads directly to the transformation equations (61) of McGehee.

The variable s lies on the unit sphere $\mathbb{S} = \{r = 1\} \cong S^{d(N-1)-1} \subset \mathbb{E}_0$ of configuration space.

$$s \in \mathbb{S} = S^{dN-1} = \{r = 1\} \subset \mathbb{E}$$

EXERCISE 1.2. *Write $'$ for $\frac{d}{d\tau} = r^{3/2}\frac{d}{dt}$. Show that McGehee's transformation transforms Newton's equations (35) to the equations*

{blowup}

$$\begin{aligned} r' &= r\nu \\ s' &= y - \nu s \\ y' &= \nabla U(s) + \frac{1}{2}\nu y \end{aligned} \tag{62}$$

where $\nu = \langle s, y\rangle$. These equations are the McGehee blown-up equations. In the blown up variables the energy is given by $H = r^{-1}(K(y) - U(s))$, and the angular momentum by $r^{1/2}J(s,y)$.

Proof of theorem 1.2. A glance at the equations shows that they make sense when $r = 0$ and that $r = 0$ is an invariant submanifold. To find their equilibria set the left hand sides of the three equations of (62) to zero. Get $r\nu = 0, y = \nu s$ and $\nabla U(s) = -\frac{1}{2}\nu y$. These last two equations combine to give $\nabla U(s) = -(\frac{1}{2}\nu^2)s$ which is the equation for s to be a central configuration with eigenvalue parameter $\mu = \frac{1}{2}\nu^2$.

The energy equation yields $rE = K(y) - U(s)$ which, upon setting $r = 0$ yields $\frac{1}{2}\nu^2 - U(s) = 0$. It follows that $\nu = \pm\sqrt{2U(s)}$ at the equilibria which established the claimed 2:1-ness. (Recall $U > 0$ on the sphere $\mathbb{S}$.) The $-$ root represents collisions, since $r' < 0$ nearby according to the first equation of equations (62). The $+$ roots are time-reversed collsions: explosions.

Re-iterating, the equilibria are at $(r, s, y) = (0, s, \nu s)$ where s is a normalized central configuration and $\nu = \pm\sqrt{2U(s)}$. The dropped central configuration solutions having normalized shape s and energy $E < 0$ are heteroclinic connections between these two equilibria. This can be seen directly from the definition of dropped central configuration solution and the one-dimensional Kepler equation. The energy $E = -\frac{1}{r_M}U(s)$ where r_M is the maximum of r achieved along the dropped orbit. QED

4.2. Saari decomposition and the homographic family. The dropped central configurations and relative equilibria associated to a normalized central configuration s_0 form two extremes of the three-parameter family of homographic planar N-body solutions $\lambda(t)s_0$ described by the ansätz of lemma 1.1. In this section we will visualize this family as it sits within the McGehee phase space. See figure 4.2. This visualization will be very useful later on

put ref in:chapter braids, section..? -RM

The three parameters of the homographic solutions are the energy, angular momentum and inertial orientation of the Keplerian ellipses defining the solution. If we fix the energy $E < 0$ and mod out by rotations we are left with one-parameter family parameterized by angular momentum J. A one-parameter family of curves forms a surface. The closure of our surface is the disc depicted in figure 4.2 with the curves foliating it. This disc contains one new orbit which is not part of the ansätz: the heteroclinic connection at the bottom of the disc, lying on the collision manifold and joining the two equilibrium points associated to s_0.

fig: rnuplots

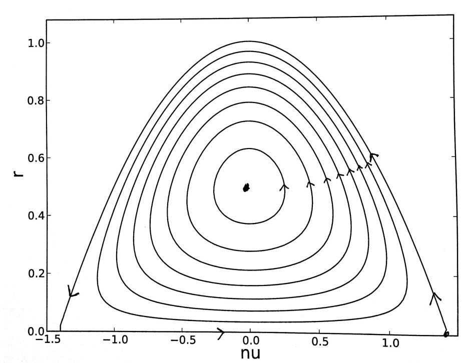


FIGURE 1. A central configuration family with fixed energy projected onto the ν, r plane. The arch and 'floor' $r = 0$ comprise the rest cycle

The energy equation

{eq: ccEnergy}
$$rE = \frac{\nu^2}{2} + \frac{J^2}{2r} - U(s_0), \tag{63}$$

valid for the curves in the ansätz of lemma 1.1 for s_0, allows us to draw the figure. We will establish this equation as a special case of the "Saari decomposition" below. To understand how the equation 63 leads to the figure, fix E and s_0 in the equation. Now the equation defines a one-parameter family of curves in the (τ, r) plane, J being the parameter. These are the curves plotted in figure 4.2. The bounding top arch corresponds to $J = 0$ which is the dropped solution. The fixed center, looking like a stable equilibrium, corresponds to the relative equilibrium.

4.2.1. *Velocity (Saari) decomposition.* For the planar N-body problem, recall that we identify$\mathbb{E}$ with $\mathbb{C}^N$ and thence $\mathbb{E}_0$ with the codimenson one complex subspace $\Sigma m_a q_a = 0$ of $\mathbb{C}^N$. We can multiply any $q \in \mathbb{E}_0$ by a complex number λ, yielding the subspace $\mathbb{C}q \subset \mathbb{E}_0$.

Definition 1.4. *The horizontal space at q, also called the space of shape directions, is the orthogonal complement (rel. the mass metric) of the $\mathbb{C}$-span of q.*

$\mathbb{C}q$ further splits into the orthogonal direct sum of two real lines, namely $\mathbb{R}q$ and $\mathbb{R}iq$. The line $\mathbb{R}q$ is tangent to scalings of q. The line $\mathbb{R}(iq)$ is tangent to the circle $e^{i\theta}q, \theta \in \mathbb{R}$ of rigid rotations of q. We call these lines (scale) and (rotation). In this way get the orthogonal decomposition:

eq: ccEnergy
$$\begin{aligned} T_q\mathbb{E}_{cm} &= \text{(scale)} + \text{(rotation)} + \text{(horizontal)} \qquad (64)\\ &= \mathbb{R}q \oplus i\mathbb{R}q \oplus \{v : J(q,v) = 0, \nu(q,v) = 0\} \qquad (65)\end{aligned}$$

where in the last term we used that $\langle q, v\rangle = r^{1/2}\nu(s,y)$ while $\langle iq, v\rangle = J(q,v)$.

Apply this decomposition to a vector $v = \dot{q} \in \mathbb{E}_0$, viewed as a tangent vector at $q = rs$, we get the orthogonal decomposition:

$$v = \lambda s + i\omega s + v_{hor}$$

with $\lambda = \langle v, s\rangle$ and $\omega = \langle v, is\rangle$. Use the McGehee relations $v = r^{-1/2}y$ and $q = rs$ to see that $\lambda == r^{-1/2}\nu(y,s)$ and $\omega = \frac{1}{r}\langle iq, v\rangle = \frac{1}{r}J(q,v)$. (Recall that $\langle s, s\rangle = 1$.) It follows that the kinetic energy $K = \frac{1}{2}\langle v, v\rangle$ decomposes as

{KEdecomp}
$$K(q,v) = \frac{1}{2}\frac{\nu^2}{r} + \frac{1}{2}\frac{J^2}{r^2} + K_{shape}(v_{hor}) \tag{66}$$

The final term K_{shape} is formed by computing the squared length of v_{hor}. It turns out be equal to the kinetic energy of the standard (Fubini-Study) metric on the shape space $\mathbb{CP}^{N-2}$ times the factor r^2.

Establishing the energy relation, equation (63). For any one of the solutions $q(t) = \lambda(t)s_0$ coming from our central configuration, the shape $[q(t)] = [s_0] \in \mathbb{CP}^{N-2}$ does not change, hence the shape velocity v_{hor} and K_{shape} are both-zero. This yields $K = \frac{1}{2}\frac{\nu^2}{r} + \frac{1}{2}\frac{J^2}{r^2}$ and thence, using $E = K - U$ to our energy relation.

Remark on notation We call this composition the "Saari decomposition" after work of Don Saari [**135**] pointing out the importance of this group-induced splitting to celestial mechanics. The last factor is called the horizontal space because if we look at the projection $\mathbb{E}_0 \to \mathcal{P}(\mathbb{E}_0)$ as a Riemannian submersion, then

ref?? -RM

the family of horizontal spaces is precisely the family corresponding to the standard connection for the Hopf fibration $\mathbb{S} \to \mathbb{CP}^{N-2}$. Under the differential of the projection the horizontal space maps linearly isomorphically onto the tangent space to the normalized shape space $\mathbb{CP}^{N-2}$.

4.3. Roads to total collision. We explore reason (2) of the four motivations listed in subsection 1.2: the fact that as we tend to total collision we do so along central configurations.

A centered solution $q(t)$ suffers a total collision at time $t = 0$ if $\lim_{t\to 0} q(t) = 0$. What is its limiting shape? The question is best answered by using McGehee blow-up. We note that it requires infinite McGehee time τ to reach collision.

thm: Sundman

THEOREM 1.3. *Suppose $q(t)$ is a total collision solution and let $(r(\tau), s(\tau), y(\tau))$ be the solution written in McGehee variables so that $r(\tau) \to 0$ as $\tau \to \infty$. Then the accumulation points of the normalized configuration $s(\tau)$ as $\tau \to \infty$ form a closed connected subset of the set of normalized central configuratons. If all central configurations for the given mass distribution are non-degenerate then this subset is a singleton: there is a normalized central configuration s_0 such that $(s(\tau), y(\tau)) \to (s_0, \nu_0 s_0)$ as $\tau \to \infty$, with $\nu_0 = -\sqrt{2U(s_0)}$.*

It is unknown whether there exist total collision solutions for which $s(\tau)$ does *not* have a unique limit, i.e. has more than one accumulation point as $\tau \to \infty$. This question is intimately linked with the question of infinite spin to be described later. The last part of the theorem asserts that as long as $\tilde{U}$ is Morse, $s(\tau)$ does have a unique limit and it is one of the finite number of central configurations.

PICTURE IN the N = 3 case here!

Sundman was the first to prove a version of theorem 1.3. In the process of obtainining his theorem he proved an important auxiliary result.

THEOREM 1.4. *Any solution suffering total collision has zero angular momentum.*

SKETCH OF PROOF OF THEOREM 1.3. As $q(t)$ comes in to total collision, where can $s(\tau), y(\tau)$ go? Common sense suggests that if they settle down, then they must do so onto a particular equilibrium. But do they settle down? The first part of the theorem asserts that the ω-limit set of $(r(\tau), s(\tau), y(\tau))$ lies within the equilibria. Recall that the equilibria consist of two branches according to whether $\nu < 0$ or $\nu > 0$. Ours must lie in the negative branch since as we approach collision r decreases and we have $r' = r\nu$. To verify that the ω-limit set in fact does lie in the equilibrium manifold one uses the fact, easily verified, that $-\nu$ acts like a Liapanov function on the collision manifold. In other words, upon restricting the dynamics to the collision manifold we have $-\nu' \geq 0$ with $\nu'(0, s, y) = 0 \iff (0, s, y)$ is an equilibrium. See MY BLOW UP TEXT ... or Rick or Alain ...

If all the central configurations are non-degenerate, then, viewed on normalized shape space, they form isolated points. It follows that the projection $[s(\tau)]$ of $s(\tau)$ to normalized shape space has a unique limit. (The set of accumulation points forms a connected set.) Then the set of accumulation points of $s(\tau)$ must be a connected set in the rotation-group fiber over $[s_0]$ relative to the fibration $\mathbb{S} \to \pi(\mathbb{S}) \subset Sh$. In order to proceed to existence and uniqueness of the limit we will use the nondegeneracy assumption. There are two ways to proceed. find a theorem in a dyn sys book asserting this connectedness assertion - RM

Way 1. Invoke Hirsch-Pugh-Shub theorem 4.1, (e) 'Lamination', p 39. The non-degeneracy assumption implies that the fiber is a normally hyperbolic set of rest points. This theorem then asserts that its unstable manifold, $W^s(\text{fiber})$ fibers as $W^s(p), p \in$fiber. Now our solution lies in $W^s(\text{fiber})$, hence in some particular $W^s(p)$.

Way 2. Invoke the theorem of Sundman on zero angular momentum and use that imposing zero angular momentum defines a connection for $\mathbb{S} \to \pi(\mathbb{S})$. We can form the reduced McGehee dynamics by dividing out by rotations. The reduced collision manifold $r = 0 = \tilde{H}$ has dimension $2(N-2)$ the topology of $T(\mathbb{CP}^{N-2} \setminus \Delta)$. The nondegeneracy assumption implies that the equilibrium points, viewed in the reduced dynamics, are hyperbolic. Indeed, following a trick by Devaney, Moeckel shows that the eigenvalues of the linearized matrix are $k = \frac{1}{4}(-\nu_0 \pm \sqrt{\nu_0^2 + 16\lambda})$ where the λ are the eigenvalues of the Hessian of $\tilde{U}$ relative to the Fubini-Study (shape) metric on $\mathbb{CP}^{N-2}$. (See p. 229 eq 3.5 in [?] and correct for a typo in applying the quadratic formula to a quadratic expression for the eigenvalues k two lines above.) Since $\nu_0 < 0$, and nondegeneracy is equivalent to $\lambda \neq 0$, the real part of μ is never zero. A curve tending to total collision along $[s_0]$ must be in the stable manifold of the fiber over $[s_0]$ which means that $Re(\mu) < 0$ and, since $\nu_0 < 0$ that consequently $\lambda > 0$. Thus, the eigenvalues of interest for us are of the form $-a + ib$ with $a > 0$.

double check eigenvalue equations -RM

saved reference on correceted evalues from Rick in file BookN-body/Correspondences/RickEvalcomps.pdf sent to me 3/29/22. Put in an appendix? or ? -RM

It follows that projected shape curve of our collision solution has the complex coordinate form $c(\tau) = (Z_1(\tau), Z_2(\tau), \ldots, Z_{N-2}(\tau)) + O(|Z|^2)$ with the $Z_i(\tau) \in \mathbb{C}$ being real linear combinations of curves $W(\tau) = Aexp(k\tau)$ with $k = -a + ib$, $a > 0$. The Euclidean length of any positive infinite end ($\tau > \tau_0$) of a spiral of the form $W(\tau) = exp((-a+ib)\tau)$ with $a > 0$ is finite as is a linear combination of such spirals. The shape space metric is a metric, so locally Lipshitz equivalent to Euclidean space and this same finiteness thus holds for our projected shape curve. The horizontal lift of a smooth curve with finite length has a well defined limit as $\tau \to \infty$. This endpoint is our desired point $\lim_{\tau\to\infty} s(\tau)$.

QED

In case the Hessian is degenerate, for example at one of Palmore's 4 body central configurations, the argument falls apart. There could be collision curves whose projection to normalized shape space have infinite length -something like a spiral of Archimedes. The horizontal lift of such a curve could have ω-limit set the entire circle fiber over $[s_0]$. Whether such curves exist is one aspect of the problem of infinite spin. The other aspect is present only if the answer to the Open Question here is 'no' and there exist continua of central configurations in normalized shape space.

4.4. Bifurcations. Here we explore reason (3) of the four motivations listed in subsection 1.2: bifurcations of the energy-momentum map. We explored this reason earlier on in the book around figure 6 for the case $N = 3, d = 2$. The Morse-theoretic perspective expressed in that picture persists for $N > 3$ and, with numerous complications, for $d > 2$.

We will limit ourselves to the planar case, leaving the general case to some remarks at the end. We work with the centered problem. The phase space $\mathcal{P} = \mathcal{P}(2, N)$ is the open set $(\mathbb{C}^N \setminus \Delta) \times \mathbb{C}^N$ of $\mathbb{R}^{4N} = \mathbb{C}^{2N}$.The energy E and angular momentum J are conserved, so that for each choice of E and J, we get an invariant subsystem $\mathcal{P}_{E,J} \subset \mathcal{P}$ of codimension 2. Put the energy and angular momentum together to form a map $\mathcal{P} \to \mathbb{R}^2$, the energy-momentum map. Then $\mathcal{P}_{E_0,J_0}$ is

ok , then you best make those remarks at the end! -RM the inverse image of a poin $(E_0, J_0) \in \mathbb{R}^2$ under this map. As long as E_0, J_0 are a regular value of the energy-momentum map then $\mathcal{P}_{E_0,J_0}$ is a smooth manifold. And if E, J vary over a connected open set of regular values then all the $\mathcal{P}_{E,J}$'s are diffeomorphic. The bifurcation set is the set of singular values of the energy-momentum map. A point (E_0, J_0) is a singular value if and only if it is the image of critical point z_0 of the energy momentum map, i.e. if and only if there is some $z_0 \in \mathcal{P}_{E_0,J_0}$ with $dE(z) = \lambda dJ(z)$ and $E(z_0) = E_0, J(z_0) = J_0$.

LEMMA 1.3. *A point is a critical point for the energy momentum map if and only it is a relative equilibrium.*

PROOF. Let ω be the symplectic form so that the Hamiltonian vector field is $X_H = \omega^{-1} dH$. By definition of momentum map, $\frac{\partial}{\partial\theta} = \omega^{-1} dJ$ where $\frac{\partial}{\partial\theta}$ is the infinitesimal generator of rotation. For the sake of clarity we have used H for E, the energy viewed on the cotangent side. We then have, by linearity $dE(z_0) = \lambda dJ(z_0) \iff X_H(z_0) = \lambda \frac{\partial}{\partial\theta}|_{z_0}$ which in turn says that the time t dynamical evolution of the state z_0 is the same as rotation of the configuration by $exp(i\lambda t)$. QED

Since in two-dimensions relative equilibria and central configurations are the same, we have shown that

Consequently, to understand the topology of $\mathcal{P}_{E,J}$ subvariety of phase space where energy and angular momentum have these given values. We can quotient $\mathcal{P}_{E,J}$ by the action of the rotation group and we get a variety on a contact manifold of dimension $4N - 7$ on which the dynamics lives and where all indications suggest that we cannot further reduced the dimension. Scaling takes E, J to $\frac{1}{\lambda}E, \lambda^{1/2}J$, leaving $J^2E = Dz$ invariant.

we cannot subset of and s, which, as we have seen , is the same as forming the quotient by translations and Galilean boosts. As we Smale XXX pushed us to think of the N-body problem as a family of dynamical systems depending parametrically on the energy E and angular momentum J. Write $M_{E,J}$ for the subspace of phase space at which the energy has the value E and the angular momentum has the value J. The map $(E, J) : \mathcal{P} \to \mathbb{R}^2$ is called the energy momentum map. As Sard guarantees, almost every value of the plane $\mathbb{R}^2$ is a regular value of the Energy momentum map. The scaling symmetry $(q, p) \mapsto (\lambda q, \lambda^{-1/2}p)$ maps $M_{E,J}$ diffeomorphically onto $M_{\lambda^{-1}E, \lambda^{1/2}J}$.

(2). The topology of a manifold strongly influences the type of dynamics it can support. Because energy and momentum are conserved, we have, with N-body dynamics, a family of dynamical systems for each choice of energy and momentum. It is of interest then, to understand how the topology changes as we change energy E and angular momentum J. Values of E and J where such a change occurs are bifurcation values. As in Morse theory, we expect bifurcations to happen in concert with critical values of the energy momentum map $(E, J) : \hat{P} \to \mathbb{R} \times \Lambda^2\mathbb{R}^d$. As we will see momentarily, the critical points of the energy momentum map are relative equilibria, so, at least in the planar case, they are associated to central configurations. (Because $\hat{P}$ is non-compact we must also worry about "critical points at infinity" in addition to finite critical points. See Albouy REF)

(3). We already saw this for $N = 3$ with the Euler and Lagrange solutions. The explicit 1-parameter family of solutions associated to any planar central configuration is described below in lemma 1.1.

5. Some words on the Hampton-Moeckel and Albouy-Kaloshin proofs

Both papers begin by rewriting the conditions to be a central configuration as a system of real polynomial equations. They then complexify their systems by allowing the variables to be complex, and in this way arrive to a complex affine algebraic variety. Their goal is to show that this variety is a finite point set. The method of achieving that goal boils down to proof by contradiction.

5.0.1. *Hampton-Moeckel.* Albouy and Chenciner wrote down the following remarkable set of $\binom{N}{2}$ equations in the $\binom{N}{2}$ mutual distance variables r_{ij} which are equivalent to the condition of being a central configuration [**7**].

{eq: Albouy-Chenciner}

$$[\Sigma_k m_k S_{ik}(r_{jk}^2 - r_{ik}^2 - r_{ij}^2)] + (i \leftrightarrow j) = 0 \tag{67}$$

where '$+(i \leftrightarrow j)$ means add the same expression but with the roles of i and j switched. Here S is the matrix with entries

$$S_{ik} = \frac{1}{r_{ik}^3} - \frac{\lambda}{M}, i \neq k, \qquad S_{ii} = 0$$

while $S_{ii} = 0$, and where $M = \Sigma m_i$ is the total mass. The parameter λ is the same eigenvalue parameter of equation (58). Since $\lambda = U/I$ it is of degree -3 with respect to scaling so that we can fix the size of the configuration by insisting that $\lambda/M = +1$. (The λ used by [**52**] in their formulation is the negative of ours, so they write $\lambda/M = -1$.) One remarkable thing about the Albouy-Chenciner equations (67) is that the dimension d in which the bodies move is nowhere to be seen. Only the number N of bodies occurs as a finite parameter.

Hampton and Moeckel [**52**] take the Albouy-Chenciner equations (67) as the starting point to describe their $N = 4$ algebraic variety, and add to them equations specifying that the dimension d of the affine space within which the bodies are collapsing is 2. The logic goes like this. Recall (proposition 1.5) that for $N = 4$ bodies all central configurations have rank ≤ 3 and there is only one central configuration of rank 3: the regular tetrahedron. On the other hand, by Moulton, there are precisely $4!/2 = 12$ central configurations of rank 1. It remains then to deal with the rank 2 configurations. Dziobek came up with Plucker-type relations

$$S_{12}S_{34} = S_{13}S_{24} = S_{14}S_{23}$$

which characterize rank two-ness for four bodies. By adding in these equations to the 6 Albouy-Chenciner equations they get a system of 9 polynomial equations in the 6 unknowns r_{ij} and this defines their algebraic variety.

From here they proceed by contradiction. They allow the r_{ij} to be complex. If their variety is not a finite point set then it forms an affine variety of positive dimension in $\mathbb{C}^6$. A general fact about an affine algebraic variety $A \subset \mathbb{C}^M$ of positive dimension is that its projection onto almost any line $\mathbb{C} \subset \mathbb{C}^M$ is 'dominant' - meaning that the image of A on the line is the entire line minus a finite set of points. One can then use the parameter t of this line to 'parameterize' the variety using Puiseux expansions. Think of the variables as $(t, z_2, \ldots, z_M) \in \mathbb{C}^M$. The space of rational functions in $z_2, \ldots, z_M$ is an algebraically closed field $\mathbb{K}$. We can then view the defining polynomials for A as polynomials in t, i.e. $p(t) \in \mathbb{K}[t]$. Following another one of Newton's great ideas - what is today called the Newton-Puiseux expansion, one can iteratively solve the equations $p(t) = 0$ getting an expansion describing parts of A. A beautiful involved set of ideas ensues from here, ideas based on the Newton polytopes built from the 9 defining polynomials (the $p(t)$'s) which define

check if this fact is true. ?? and if this is the fact that they use -RM their A. The methods include what is known as BKK theory, the Minkowski mixed volumes of convex sets and, at the end of the day, extensive computer algebra as their polytopes have 1,000s of vertices. We leave the intrigued reader to further pursue the topic by refering to the original paper.

5.0.2. *Albouy-Kaloshin.* If a complex affine algebraic variety is not finite, then it must be non-compact. One can think of this assertion as a version of the maximum principle from complex analysis. This non-compactness is the basic principle used by Albouy and Kaloshin. In both this paper, and the Hampton-Moeckel paper one must, by some method, work on the quotient space by rotations, translations and dilations. Hampton-Moeckel, achieve the quotient by rotations and translations by working with the mutual distance r_{ij} variables. Albouy-Kaloshin instead work with the original configuration variables $q_1, \ldots, q_N$ and impose the slice condition that $q_1 - q_2$ lie along the x-axis to get rid of rotations. They get rid of translation by imposing the center of mass condition $\Sigma m_a q_a = 0$. In addition to these now $2(N-1)-1$ variables (2 being the 'd' of the planar N-body problem) they throw in $\binom{N}{2}$ variables corresponding to $\frac{1}{r_{ij}}$. They rewrite the condition of being a central configuration, equation (58) in terms of these variables. To proceed further they suppose that $N \leq 5$. They show, by a kind of combinatorial analysis of cases involving plane diagrams, that it is impossible for any one of their variables to tend to infinity while remaining on their variety. According to the basic principle then, this completes the proof of finiteness. In their combinatorial analysis they run into a few impossible special cases corresponding to Gareth Robert's counterexample, or generalizations of it. These are cases where some variables might go to infinity. They then exclude these special cases by imposing polynomial conditions on the masses, hence their "generic finiteness". Again, I refer the intrigued reader to the original paper.

5.1. Stable and Unstable manifold controlling the flow. What is more important about all this work is what flows out of total collision. Recall that the stable manifold ... XXX....

ete etc .. . hint of Moeckel thesis At fixed energy ....

A bit more work, using the property that XX is a Liapanov function on the total collisionmanifold. With a little bit of work then one can show that if $q(t)$ tends to total collision then it must approach the manifold of equilbria i.e. the locus of central configurations. If this locus is finite, mod G, then it limits onto a particular central configuration, thus proving the theorem of Sundman.

problem of infinite spin...

6. Relative equilibria in four dimensions

sec: higher ds

In dimension $d = 4$ and higher there exist relative equilibria that are not central configurations. This possibility was investigated by Albouy and Chenciner [REALLY?] and in dimension 4 is based on the normal form:

$$\Omega = \begin{pmatrix} 0 & -\omega_1 & 0 & 0 \\ \omega_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\omega_2 \\ 0 & 0 & \omega_2 & 0 \end{pmatrix}$$

for skew-symmetric 4 by 4 matrices. If we identify $\mathbb{R}^4$ with $\mathbb{C}^2$ then this is the matrix for the real linear transformation $(z_1, z_2) \mapsto (i\omega_1 z_1, i\omega_2 z_2)$.

Perhaps the simplest family of examples of relative equilibria that are not central configurations occurs for the equal mass three-body problem (1, 2 in 4 dimensions.

PROPOSITION 1.7. *Every isosceles triangle is a relative equilibrium configuration for the equal mass 3-body problem in* $\mathbb{R}^4$*. In the corresponding relative equilibrium solution the two Jacobi vectors travel in orthogonal two-planes at different frequencies, the ratio of these frequencies being a function of the ratio of the side lengths of the isosceles triangle as described by equation (75) below. (See figure.)*

PROOF OF PROPOSITION. The proof can be achieved by a direct computation. Take $m_1 = m_2 = m_3 = 1$. Then we can write the equations defining this equal mass three-body problem as :

{N3R4}

$$\begin{aligned} \ddot{q}_1 &= \frac{q_2 - q_1}{r_{21}^3} + \frac{q_3 - q_1}{r_{31}^3}, \\ \ddot{q}_2 &= \frac{q_1 - q_2}{r_{21}^3} + \frac{q_3 - q_2}{r_{32}^3},, \\ \ddot{q}_3 &= \frac{q_1 - q_3}{r_{13}^3} + \frac{q_2 - q_3}{r_{23}^3},. \end{aligned} \tag{68}$$

Suppose the triangle to be isosceles with vertex 3 so that $r_{23} = r_{13}$. Set

$$a = r_{23} = r_{13} \text{ and } c = r_{12}.$$

The two Jacobi vectors corresponding to the choice of vertex 3 are $\xi_1 = q_1 - q_2$ and $\xi_2 = q_3 - \frac{1}{2}(q_1 + q_2)$. If we subtract the second equation of (68) from the first and use the isosceles condtion we get

$$\ddot{\xi}_1 = -(\frac{2}{c^3} + \frac{1}{a^3})\xi_1 \tag{69}$$

If we add the first equation to the second and subtract that half that result from the third equation of (68) we get

$$\ddot{\xi}_2 = -\frac{3}{a^3}\xi_2. \tag{70}$$

These pair of equations suggests the ansätz

{ansatz_RE4}

$$\begin{aligned} \xi_1(t) &= e^{i\omega_1 t}\xi_1(0) \\ \xi_2(t) &= e^{i\omega_2 t}\xi_2(0) \end{aligned} \tag{71}$$

which solves Newton's equations, rewritten in Jacobi vectors, provided that

{freqs}

$$\begin{aligned} \omega_1^2 &= (\frac{2}{c^3} + \frac{1}{a^3}) \\ \omega_2^2 &= \frac{3}{a^3}. \end{aligned} \tag{72}$$

and that a, c remain constant. To guarantee this constancy recall that $\xi_1(0), \xi_2(0) \in \mathbb{C}^2 = \mathbb{R}^4$. Take the two vectors to lie in Hermitian orthogonal complex subspaces:

{ansatz_RE4b}

$$\begin{aligned} \xi_1(0) &= (z_1, 0), \\ \xi_2(0) &= (0, z_2). \end{aligned} \tag{73}$$

so that at every instant $\xi_1(t)$ and $\xi_2(t)$ are perpindicular. Then

$$c = r_{12} = |\xi_1| = |z_1|$$

while

$$a = \sqrt{|z_1|^2 + \frac{1}{4}|z_2|^2}$$

which are indeed constant. The equation for a is seen by noting that

$$\begin{aligned} q_3 - q_1 &= q_3 - \frac{1}{2}(q_1 + q_2) - \frac{1}{2}(q_1 - q_2) \\ &= \xi_2 - \frac{1}{2}\xi_1 \end{aligned} \tag{74}$$

and similarly

$$q_3 - q_2 = \xi_2 + \frac{1}{2}\xi_1$$

and then using the orthogonality of ξ_1, ξ_2 and the fact that $|\xi_i| = |z_i|$ to compute

$$a^2 = \frac{1}{4}|z_1|^2 + |z_2|^2.$$

Summarizing, the solution ansätz defined equations (71) and (73) yields a relative equilibrium solution to Newton's equations provided the frequencies satisfy relations (72) with $c = |z_1|, a = \sqrt{\frac{1}{4}|z_1|^2 + |z_2|^2}$.

If we divide the two frequency relations we get

{freq_ratio}

$$\frac{\omega_1^2}{\omega_2^2} = \frac{1}{3}(\frac{2a^3}{c^3} + 1) \tag{75}$$

Note, in particular, that $|\omega_1| = |\omega_2|$ if and only if $a = c$, which is to say, if and only if the triangle is equilateral (and thus central).

QED

******

More generally, suppose that $g(t) = exp(t\Omega)$ for the old ansätz in the definition of a central configuration: $q(t) = g(t)q_*$, with $\Omega \in so(d)$ a skew symmetric matrix. Plugging this in to Newton's equations and usign $F(gq) = gF(q)$ we get

$$-Sq_* = F(q_*); S = -\Omega^2 = \Omega\Omega^T$$

These are the general equations for a relative equilibrium. Note that S is symmetric positive semi-definite, but is not an arbitrary matrix of this type. Due to the normal form for a skew-symmetric matrix, the positive eigenvalues of S all have multiplicity two, corresponding to the two-by-two rotational blocks of Ω.

7. Other questions and projects

1. This conjecture was posed by Rick Moeckel.

Moeckel's conjecture. For all N there is a neighborhood of equal masses in the space of mass distributions for the planar N-body problem with the property that no relative equilibria is linearly stable for the masses in this neighborhood.

In the three-body problem, the conjecture is established. The Euler relative equilibria are never linearly stable. The Lagrange relative equilibria are only stable when one mass is dominant. See FIGURE where the shaded region indicates the set of masses for which the Lagrange relative equilibrium is linearly stable.

2. Do the Albouy-Chenciner isosceles relative equilibria generalize to four bodies forming a tetrahedron? To N bodies? The regular tetrahedron with all side lengths equal is a central configuration for the 4-body problem in 3-space. (See 1.5.) We saw that it can be made into a relative equilibrium in $\mathbb{R}^4 = \mathbb{C}^2$. Can we,

instead, 'spin it' within a 6-dimensional space $\mathbb{R}^6 = \mathbb{C}^3$ (a two-plane $\mathbb{C}$ for each of its three Jacobi vectors), and then deform it so as to arrive at tetrahedral relative equilibria whose side lengths are not all equal? Do we really need to go to 6 dimensions, or does 4 dimensions provide enough space to spin and deform the regular tetrahedron?

8. Historical notes.

mostly refer to Albouy and to Moeckel

I learned of this theorem from the beautiful paper 'A l'infini en temps fini' by Chenciner REF. The idea is based on McGehee blow-up.

CHAPTER 2

Are there any stable periodic orbits?

Open Question 2.1. *Are there any stable periodic orbits?*

1. Really?!

How could such a basic question remain un-answered? But if "stable" means "Lyapunov stable" then this question is wide open *for the planar 3-body problem.*

Definition 2.1. *An orbit for a dynamical system is Lyapunov stable if every neighborhood U_1 of the orbit contains a neighborhood U_2 such that all solutions starting in U_2 remain in U_1 for all future times.*

We must interpret the open question in the center-of-mass frame or upon reduction by the Galilean group. Otherwise, the answer becomes an immediate 'no'. Take any periodic solution $q_a(t)$ and boost it by adding the same $v \in \mathbb{R}^d$ with $|v| \leq \epsilon$ to all its initial velocites. We get a new solution $q_a(t) + tv$ which linearly diverges from our starting solution with time, despite having arbitrarily close initial conditions. Also, let us recall that a relative periodic solution is one for which the relative distances $r_{ab}(t)$ are periodic functions of t. Such an orbit is periodic or quasi-periodic in the center-of-mass frame. Reformulating the question then: figure here... -RM

Are there any Lyapunov stable relative periodic orbits for the planar three-body problem when viewed in the center of mass frame?

$\diamond$

Quite close to the open question above is the

Oldest question in dynamical systems: *Arbitrarily close to any relative periodic solution for the planar three-body problem are there initial conditions whose corresponding solutions are unbounded?*

An orbit is unbounded if at least one of its mutual distances $r_{ab}(t)$ is an unbounded functions of t. (See definition 0.11.) In terms of the Hill region picture at negative energy for the planar three body problem as depicted in figure 5, these are solutions whose projections to shape space wind down one of the three arms of our plumbing fixture, heading off to infinity down that arm. A 'yes' answer to the oldest question would give a dramatic 'no' to our open question.

2. What's known?

2.1. Weaker types of stability. If solutions cannot be shown to enjoy Lyapunov stability, what types of stability can they have? The main types of stability achievable for solutions arising in celestial mechanics are linear, KAM and Nekhoroshev stability. We will sketch these notions of stability below. The implications amongst them are

$$\text{Nekhoroshev} \implies \text{KAM} \implies \text{linear} \tag{76}$$

Usually, when a celestial mechanician says that an orbit is stable they mean that it is linearly stable. For example, the figure eight orbit is Nekhorshev stable, hence KAM stable and linearly stable.

2.2. Instabilities.

2.2.1. *Arnold Diffusion.* "Arnol'd diffusion" is the catch-all phrase for mechanisms for escaping arbitrarily small neighborhoods of KAM stable orbits. . A "no" answer to the open question would imply that every KAM stable periodic orbit for the N-body problem admits Arnol'd diffusion. A 'yes' answer to the Oldest Question would go further and imply that each of these Arnol'd diffusions actually diffuses all the way out to infinity. Arnol'd diffusion is believed to occur generically in Hamiltonian systems. But the vector field defining the N-body problem is just itself, a particular vector field, not a generic one. Generic forays into the empire of Arnol'd diffusion do not seem to offer any help in answering the open question.

2.2.2. *Varying the power law: strong force instability.* If $\alpha = 2$ in the power law potential , then the Lagrange-Jacobi identity for that case (see equation (49)) asserts that $\ddot{I} = 4H$ where H is the energy. It follows that periodic orbits only exist when $H = 0$ and $\dot{I}(0) = 0$. If either condition is violated then I either increases without bound or goes to zero indicating total collision. Since small perturbations change the energy and $\dot{I}(0)$ we can say that we've answered the oldest question in this case: a small perturbation can make a zero energy initial condition positive and hence arbitrarily nearby to any periodic solution there is an unbounded solution.

We can keep the $\alpha = 2$ problem interesting by restricting our perturbations to those which respect $H = 0$ and $\dot{I} = 0$. There are many relative periodic orbits having $H = 0$ and $\dot{I} = 0$. They play important role in the next chapter. If we further fix the angular momentum J to be zero, and if all of the masses are equal, then we have shown that all of these solutions are linearly unstable. In a sense then, even here all orbits lead to unbounded solutions , but now "unbounded" means, somewhat paradoxically, some pair tends to collision. In the instability, the REF

Following Moore [**119**] we can wonder at the panoply of behaviour as we vary the power law exponent through the interval $[-2, 2]$, sweeping α from the integral case $f(r) = r^2$ of the harmonic oscillator where $\alpha = -2$ and all periodic orbits are Lyapunov stable, to the strong force case $\alpha = -2$ where $f(r) = 1/r^2$ and all orbits are unstable. In between we hit the Newtonian case. As we do so, periodic orbits come and go, disappearing and colliding. Is any of this any help in understanding stability in the Newtonian case? So far, not so much.

2.3. And our solar system? Is our solar system stable? Read [**77**] for a short engaging and personal history of this problem. Quoting : " The question of the stability of the Solar System was thus a question of limiting or not the divine power, and was one of the main scientific questions of the 18th century." Laskar's work itself [**77**] demonstrated that the dynamics in our solar system is actually chaotic, with a Lyapanov time scale of $.5 \times 10^{-6}$ in units of 1/year. One has mixing and indeterminacy of of some of the basic coordinates describing the orbital motions on that time scale. Indeed, there is no one single periodic orbit to talk about for the system to to be "stable about"! One can go on with the question though and ask, for instance, will the earth be at roughly the same distance from the sun, in, say 3 billion years? Lagrange argued, based on using averaging to write down an approximating system, that the semi-major axes of the planets vary much less and much slower than their orbits' eccentricites. To get significant drifts in semi-major axes takes 100s of millions of years, and then it only happens in the inner solar sytem. To get planetary orbits to cross and hence completely disrupt the inner solar system takes on the order of the age of the solar system on numerical runs, and due to the intrinsically chaotic nature of the dynamics, that question has to be approached statistically.

2.4. 4 dimensions stablize. Allow our bodies the freedom to move in 4 dimensions and we get a surprise.

Albouy-Dullin

THEOREM 2.1. *Some of the relative equilibria for the three-body problem in 4-space described at the end of the previous chapter are Lyapunov stable.*

The proof of this theorem was published in 2020 by Albouy and Dullin [**3**] and Dullin and Scheurle [**32**] REF.

2020? 2015? -RM

A philosophical rationale underlies the appearance of stable orbits for the three-body problem in $d = 4$ dimensions and their apparent lack when $d = 2$ or 3. A general physical principle asserts that solutions seek out states of lowest energy. In celestial mechanics both angular momentum and energy are conserved but angular momentum beats out energy.

PROPOSITION 2.1. *In the 2-body (Kepler) problem, at fixed non-zero angular momentum, the energy is bounded below and the solutions which minimize energy are the circular orbits.*

This proposition is held responsible for the near-circular nature of the orbits of the planets. The idea is that during planetary formation, dissipative forces were involved, but ones which conserved angular momentum, so that over millions of years the orbits of the planets around the sun settled down to near circular orbits. REFER TO GRAPH OF EFFECTIVE POTENTIAL

put graph here or earlier -RM

PROPOSITION 2.2. *Fix the value of the angular momentum $J_0 \in \Lambda^2\mathbb{R}^d$ for the three-body problem in d-space. Write $\mathcal{P}_{J_0} \subset \mathcal{P}$ for the subvariety of centered phase space $\mathcal{P}$ having angular momentum J_0. If $d = 2, 3$ then the energy is unbounded below on $\mathcal{P}_{J_0}$: its infimum is negative infinity. However, if $d = 4$ and if the rank of J_0 is 4 then the energy is bounded below on $\mathcal{P}_{J_0}$.*

The Lyapunov stable relative equilibria of theorem 2.1 correspond to absolute minima of the restricted energy of this proposition. This is a beautiful, simple and surprising result. It does not seem to help us at all with the real question, Lyapunov stability for orbits in the spatial or planar three-body problem.

⋄

2.5. Chapter plan. I will define and explain the other types of stability in the rest of this chapter, focusing on the fact, not evident in the literature, that the stability criteria are finite in nature and so can be verified by good numerics. After our discovery of the eight it took me months to properly digest the stability type of the eight and what others were telling me about it, particularly Carles Simó.

Simó's pictures of the KAM torii near the figure eight, and staring at pictures of the standard map were especially helpful. For the reader interested in a broad yet succinct outlook on stability in celestial mechanics and KAM I cannot hope to equal [**87**] and highly recommend it.

sec: linear stability

3. Linear Stability

This section is inspired in good measure by chapter 6 of Meyer [**94**].

Suppose $\gamma(t)$ is a periodic orbit for a vector field X on a manifold $\mathcal{P}$. Let Φ_t be the flow of the vector field so that $\gamma(t) = \Phi_t(z_0)$ where $z_0 = \gamma(0)$ and $\Phi_T(z_0) = z_0$ where T is the period of γ. The linearized flow map

$$A = d\Phi_T(z_0) : T_{z_0}\mathcal{P} \to T_{z_0}\mathcal{P}$$

contains crucial stability regarding the orbit.

Definition 2.2. *The linear map A just defined is called the monodromy matrix of the periodic orbit.*

We can compute A by linearizing our nonlinear vector field $z \mapsto X(z)$ along the orbit. Write our original ODE as $\dot{z} = X(z)$ where, for simplicity, $\mathcal{P} = \mathbb{R}^m$ so that $X : \mathbb{R}^m \to \mathbb{R}^m$. Consider a family of solutions $z_\epsilon(t)$ to this ODE depending smoothly on some parameter ϵ. For example, ϵ might parameterize a curve of initial conditions. Setting $\delta z(t) := \frac{d}{d\epsilon}|_{\epsilon=0} z_\epsilon(t)$ and differentiating the ODE we arrive at the linear time dependent ODE

$$\delta\dot{z} = DX(z(t))\delta z(t)$$

which governs the variation of solutions and supplements our original ODE. If $\Phi_t : \mathbb{R}^m \to \mathbb{R}^m$ is the time t flow of our ODE then $z_\epsilon(t) = \Phi_t(z(0) + \epsilon\delta z_0)$ is such a family of solutions. Differentiating this last relation with respect ot ϵ shows that the linearization of the flow is given by

$$D\Phi_t(z(0))(\delta z_0) = \delta z(t)$$

where $\delta z(t)$ solves the linearized ODE and has initial condition δz_0. Thus $A(\delta z_0) = \delta z(T)$. This relation between the linearized ODE and linearized flow provides a useful numerical trick for computing linearizations of orbits on the fly and is essential for estimating linear stability.

Since A is linear, the zero vector is a fixed point for A, viewed as a map. The Lyapunov stability of the zero vector for A relates to the Lyapunov stability of our non-linear orbit γ. The following fact plays an organizing role:

Proposition 2.3. *If A has an eigenvalue, complex or real, with modulus greater than* 1, *then the orbit is not Lyapunov stable.*

Observe that such an eigenvalue guarantees the Lyapunov *instability* of 0 for A as a map. Indeed, supposing this eigenvalue $\lambda > 1$ positive real for simplicity and let v be its eigenvector. We get $A^n v = \lambda^n v \to \infty$ as $n \to \infty$. Since the linearized flow is a first order approximation to the flow, it is not a surprise that the flow itself moves away from the orbit in the direction of v. This intuition is turned into a proof through the magic of the "unstable manifold". The eigenspace for such an unstable eigenvalue corresponds to a growing mode for the dynamics and and lies in the tangent space to the unstable manifold to this orbit. For a full proof of the proposition and the notion of unstable manifold see any graduate book in dynamical systems.

In our situation Φ_T is symplectic so A is a linear symplectic map. The spectrum $\sigma(A) \subset \mathbb{C}$ of such a map enjoys the symmetries: $\lambda \in \sigma(A) \iff \frac{1}{\lambda}, \bar{\lambda}, \frac{1}{\bar{\lambda}} \in \sigma(A)$. See for example section 42, p. 226 of [**10**]. It follows that if our vector field X is autonomous Hamiltonian then the spectrum of the orbit's monodromy $\sigma(A)$ must satisfy $\sigma(A) \subset \mathbb{S}^1$ if the orbit has a chance of being stable. Hence the following definition is reasonable.

DEFINITION 2.3. *The periodic Hamiltonian orbit is called spectrally stable or elliptic if all the eigenvalues of its monodromy lie on the unit circle.*

If our orbit is elliptic then we can write the eigenvalues of its monodromy as

$$\lambda_j, \bar{\lambda}_j = exp(\pm i 2\pi \omega_j).$$

We call ω_j the *frequencies* of the orbit.

3.1. Symmetry-induced 1's in the spectrum. Having 1's in the spectrum of a linear map are an indication of instability, even though 1 lies on the unit circle. The hypothesis of the standard version of the KAM stability theorem require no 1's in the spectrum of a linear operator closely related to the monodromy operator A described above, namely the differential of the return map described in the next section. But several 1's always occur in the spectrum of the monodromy map A for any periodic orbit of the N-body problem. We will need to understand *how* these potentially unstabilizing 1's arise from symmetries in order to dispense with them and proceed to a reasonably useful version of KAM stability.

To see that we have at least one 1 as an eigenvalue, use the definition of the period T of our orbit $\gamma(t+T) = \gamma(t)$, which is to say $\Phi_T(\gamma(t)) = \gamma(t)$. Differentiate this last identity with respect to t at $t = 0$ to obtain $Av = v$ where $v = \dot{\gamma}(0) = X(z_0) \neq 0$. Thus any monodromy A for the periodic orbit of any autonomous flow, symplectic or not, has at least one occurrence of 1 as an eigenvalue.

We will now show that A has a Jordan block of the form

$$\begin{pmatrix} 1 & 0 \\ -3/2 & 1 \end{pmatrix}$$

associated to energy-time. Such a block guarantees that 0 is not Lyapunov stable for the linear map A, indicating subtleties in the relation between the stability of a flow and that of its linearization, subtleties we will need to understand.

$AX(z_0) = X(z_0)$ (see two paragraphs up) accounts for the last column of our claimed Jordan block. The vector $(0,1)$ relative to that block structure represents $X(z_0)$. To account for the first column of the block, we will show that the scaling symmetry of Newton's equations leads to a vector field V having $AV(z_0) = V(z_0) - \frac{3}{2}X(z_0)$ which is the information encoded in the first column of the block structure. The space-time dilational symmetry as described in subsection 3.4 is $q(t) \mapsto \lambda q(\lambda^{-3/2}t)$. At the phase space level this symmetry is $S_\lambda((q,p)) := (\lambda q, \lambda^{-1/2}p)$. If $\Phi_T(z_0) = z_0$ then the flow through $S_\lambda z_0$ is periodic with period $\lambda^{3/2}T$, which is to say that $\Phi_{\lambda^{3/2}T}(S_\lambda z_0) = S_\lambda z_0$. Now set $\lambda = exp(\epsilon)$ and differentiate this identity at $\epsilon = 0$ to find that $AV(z_0) + \frac{3}{2}X(z_0) = V(z_0)$ where $V(z)$ is the infinitesimal generator of the dilation, ie. $V(z) = \frac{d}{d\epsilon}|_{\epsilon=0} S_{exp(\epsilon)} z$ or $V((q,p)) = (q, -\frac{1}{2}p)$.

An analogue of the generalized eigenvector V which is 'dual' to the time direction X exists for periodic orbits in any autonomous Hamiltonian system. This additional vector corresponds to the fact that we can usually continue the periodic orbit on to nearby energy levels. As long as the period of the continued orbit changes to first order with the change in energy level then the corresponding two-by-two matrix has such a non-trivial Jordan block.

We have just instantiated an instance of a theorem in symplectic linear algebra: if 1 is an eigenvalue of a linear symplectic map then it comes with even multiplicity. The proof of this simple theorem proceeds by verifying that the generalized eigenspace for 1 is a symplectic subspace. For the N-body problem these pairs of 1's correspond to conjugate pairs of the symmetry-related variables. We just constructed the time-energy pair above, with its Jordan block displayed. The remaining pairs for the centered N-body problem are rotation-angular momentum pairs.

These pairs of 1's for Hamiltonian systems with symmetry like the N-body problem are a kind of linear shadow of Noether's theorem. Recall that associated to each infinitesimal symmetry of such a system there is a conservation law. Write an infinitesimal symmetry as V and the corresponding conserved quantity as f. They commute: $df(V) = 0$. Assume that V is independent of X a at the base point z_0 of our periodic orbit. Since the Hamiltonian flow commutes with the flow of V and preserves f, we have that $A(V(z_0) = V(z_0)$ and $A^* df(z_0) = df(z_0)$. Our linear algebraic situation is that A preserves the nonzero vector $V(z_0)$ and the nonzero covector $df(z_0)$ whose kernel contains v. We can choose a basis $e_1, \dots, e_n$ for our vector space such that $e_1 = v$ and $df(z_0) = \theta_2$ where $\theta_1, \theta_2, \dots, \theta_n$ is the dual basis to our chosen basis. In this basis A has the form

$$A = \begin{pmatrix} 1 & * & \dots & * & * \\ 0 & 1 & * & \dots & * \\ \vdots & * & \cdots & \vdots & * \\ 0 & * & & \cdots & * \end{pmatrix}$$

exhibiting 1 as a double root for A's characteristic polynomial, and the possibility of a corresponding two-by-two Jordan block. This Jordan block is distinct from the one exhibited for time-energy described above. So it yields another pair of 1's.

Concretely, consider the case of the planar N-body problem. We can rotate our periodic solution $q(t)$ to form $R_\theta q(t)$ without changing its period T. Here R_θ denotes rotation about θ radians. Differentiating the consequent relation $\Phi_T(R_\theta(z_0)) = R_\theta(z_0)$ with respect to θ leads to $A(Jz_0) = J(z_0)$ where $z \mapsto J(z)$ is the infinitesimal generator of rotation. As we have seen earlier, we can identify the planar N-body phase space with center-of-mass fixed with $\mathbb{C}^{2(N-1)}$, writing its points as $z = (q,p) \in \mathbb{C}^{N-1} \times \mathbb{C}^{N-1} = \mathbb{C}^{2(N-1)}$ and $\mathbb{C}^{N-1} = \{q = (q_1, \dots, q_N) \in \mathbb{C}^N : \Sigma m_a q_a = 0\}$. Then $R_\theta(z) = e^{i\theta} z$ so that $J(z) = iz$ is the infinitesimal generator of rotation in the plane. The generalized eigenvector 'dual' to iz_0 corresponds to conservation of angular momentum.

For the *centered* planar N-body problem there are only these 4 mandatory symmetry induced 1's: the time-energy pair and the rotation-angular momentum pair. For the centered N-body problem in d-dimensions, $d > 2$ there are more mandatory 1's corresponding to the rest of $SO(d)$. This total number of mandatory 1s can be computed from the dimension of the symplectic reduced space and is given in section 6.7 of [**94**].

4. The Return Map and Reduction

The return map, appropriately adapted to the notion of symplectic reduction, allows us get rid of all of the mandatory symmetry-induced 1's in the spectrum of the monodromy. Consequently a deeper understanding of the return map and its symplectic properties is mandatory for stating the hypothesis of the KAM stability theorem. We proceed to its description.

Let γ be a periodic orbit of the smooth vector field X as before. Choose a small disc $\mathbb{D}$ transverse to the orbit at the point $z_0 = \gamma(0)$. View the points of $\mathbb{D}$ as initial conditions for the flow and flow forward around until the orbits first pierce the disc again, thus defining the return map

$$R : (\mathbb{D}, z_0) \dashrightarrow (\mathbb{D}, z_0).$$

which has z_0 itself as a fixed point. We use the broken arrow notation for the map R to indicate that its domain U and range V need not be all of $\mathbb{D}$ but rather open neighborhoods of $\mathbb{D}$ containing z_0. Other names for the *return map* are Poincaré return map, Poincaré section or section map.

Since X is nonzero near z_0, the flow-box theorem tells us that $\mathbb{D} \times I$ provides a model for the flow near z_0, with X tangent to the direction of the interval I. Recall that $X(z_0)$ is an eigenvector for the monodromy map A. It follows that A induces a linear map $T_{z_0}\mathcal{P}/\mathbb{R}X(z_0) \to T_{z_0}\mathcal{P}/\mathbb{R}X(z_0)$. The differential dR_0 realizes this quotient map, with $T_0\mathbb{D}$ realizing the quotient vector space.

There were two 1's in A corresponding to energy-time. We just got rid of one of them. To get rid of the other 1 in our Hamiltonian setting, we choose a $\mathbb{D}$ of one less dimension, one which lies in the energy level h_0 of the orbit, and remains transverse to the orbit. (We could, for example, intersect our previous disc with the energy level set.) Thus $\mathbb{D} \subset \{E = h_0\} \subset \mathcal{P}$. We will call such a $\mathbb{D}$ a *symplectic slice*.

In symplectic reduction, as explained in appendix A, we fix the value of the momentum map and quotient out the result by the group action. We can view E as the momentum map for the $\mathbb{R}$-action which is the Hamiltonian flow, so that the corresponding reduced space is, formally $\{E = h_0\}/\mathbb{R}$. This space is rarely a manifold, but , by the flow-box theorem it is locally a manifold near points such as z_0 where $X(z_0) \neq 0$. Since the disc is transverse to the flow and lies in the energy level set, it forms a realization of this local quotient. In particular, the symplectic form, restriced to the disc, is symplectic and the return map R is a symplectic map of a disc whose codimension is 2 within the full phase space.

We saw that the other mandatory symmetry-induced 1's for the N-body problem were associated with rotation and angular momentum. We can get rid of these by working in a local realization of the symplectic reduced space. Again, an appropriate disc passing through z_0 will do the trick. Generally speaking, suppose we have conservation laws $f_1, \dots, f_k$ which fit together so as to form the momentum map $f = (f_1, \dots, f_k)$ for some G action on our phase space. Suppose also that f is equivariant, as is angular momentum. Then elements of G move f around, typically. Write $\mu = f(z_0) \in \mathfrak{g}^*$. Let G_μ be the Lie algebra of the stabilizer of μ. The standard Marsden-Meyer-Weinstein construction (section A.7 of Appendix A) of the symplectic reduced space tells us that this reduced space is $f^{-1}(\mu)/G_\mu$. As a local model for this reduced space we take a disc $\mathbb{D} \subset f^{-1}(\mu)$ which is transverse to the G_μ-orbits. Intersect this disc with our energy disc from the previous paragraph and we get a symplectic disc and a symplectic return map R as above, which locally respects reduction, energy and the conservation laws. (Compare with subsection 7.2 of Appendix A.) This is our final disc, and on it dR_0 represents the linear map A modulo the generalized eigenspace of all of our "mandatory 1's.

Instead of proceeding as above, in some instances it is preferable to work directly on the reduced space, with its flow. Suppose that the orbit through z_0 upstairs in the full phase space is not periodic, but rather is relatively periodic. Then downstairs on the reduced space it is represented by a bona-fide periodic orbit. If z_0 is not a relative equilibrium then the downstairs orbit is not a zero of the vector field. We then take our disc to be in the the reduced space, *within the reduced energy level* there, and transverse to the reduced vector field near the image of z_0. We thus get a symplectic return map R as above and proceed as before.

5. KAM stability

The KAM stability theorem as we describe it requires some knowledge of the Poincaré return map associated to a periodic orbit. Unfamiliar terminology used in the theorem will be described immediately following.

KAM

THEOREM 2.2. *[KAM] Suppose that a periodic orbit for a smooth autonomous Hamiltonian system has return map which*

- *(L) is linearly elliptic with eigenvalues which satisfy no resonance conditions of order 4 or less (see equation (77 below) and*
- *(T) is such that its third order Taylor expasion at the origin satisfies the twist condition (see equation (83 below)*

Then that orbit lies in a tubular neighborhood inside its energy level which is partially filled by irrational Lagrangian torii invariant under the flow. These are the "KAM torii". The Lebesgue measure of the union of these torii is positive and their density tends to 1 *as we approach the orbit.*

thm:KAM

FIGURES HERE

The hypothesis of the KAM stability theorem, conditions (L) and (T) are detailed in the next section. They are conditions on the Taylor expansion of the return map R at the origin, the origin representing z_0, the point where the orbit crosses the section $\mathbb{D}$. Condition (L) (for linear) is a condition on R's first derivative. Condition (T) (for twist) is a condition on R's third jet. Both are open conditions.

def: KAM stable

DEFINITION 2.4. *An orbit satisfying the hypothesis of the theorem is called a KAM stable orbit.*

We pause to describe some of the terminology used in the conclusion of the KAM stability theorem above. Phase space $\mathcal{P}$ comes with a symplectic form ω and the N-body vector field is a Hamiltonian vector field relative to this form. See section A.1. Let $2m$ be the dimension of P. Then m is called the number of degrees of freedom of the system. We call a submanifold of a phase space "Lagrangian" if its dimension is m and if the restriction of ω to the submanifold is zero. (A submanifold on which $\omega = 0$ must have dimension m or less.) This gives meaning to the term "Lagrangian torii". We say that a flow on a torus is "linear" if there exist angular coordinates θ_i for the torus such that the defining vector field is $\dot{\theta}^i = \nu^i = const.$ If the list of frequencies $(\nu_1, \dots, \nu_m)$ is independent over $\mathbb{Q}$ we say that the linear flow, or by abuse of notation, the torus, is "irrational". Here, that flow is the restriction of the given vector field to the invariant torus. If U is a measurable set and $\gamma \subset U$ is a smooth curve then the "density of U as we approach γ is $\lim_{\epsilon\to 0} v(N_\epsilon(\gamma) \cap D)/v(N_\epsilon(\gamma)$ where v denotes volume, i.e. Lebesgue measure, and $N_\epsilon(\gamma)$ means the ϵ neighborhood of the set γ, with respect to any distance which gives the same topology as the manifold topology.

KAM stability is a kind of probabalistic form of Lyapunov stability. Suppose that you start within ϵ of a KAM stable orbit. With high probability you will land on a a KAM torus and then you stay on it for all time. Thus the *probability* that you wander away from your ϵ-neighborhood tends to 0 with ϵ.

We cobbled together the statement of theorem 2.2 primarily from section 6, p. 411 of Appendix 8 of Arnol'd's mechanics book, [**10**] with some help from section 5 as well as conversations and other sources. See the notes section near the end of this chapter.

6. Resonance, twist and an oscillator model.

We now state the two KAM hypothesis (L) (for 'Linear') and (T) (for Twist), and provide a local picture of the KAM torii. Identify the KAM orbit in question with the circle $\mathbb{S}^1$ and the transverse symplectic disc with a $2n$-dimensional disc $\mathbb{D} \subset \mathbb{C}^n = \mathbb{R}^{2n}$ where $2 + 2n$ is the dimension of the phase space and the disc lies in an energy level. A tubular neighborhood of the orbit within the energy level set is diffeomorphic to $\mathbb{S}^1 \times \mathbb{D}^{2n}$ with $\mathbb{S}^1 \times 0$ corresponding to the orbit, and with $\mathbb{D}^{2n}$ symplectically diffeomorphic to the slice. The symplectic structure on the disc $\mathbb{D}^{2n}$ is the standard one induced from by $\mathbb{C}^n$, namely $\omega = \frac{1}{2i}\Sigma d\bar{z}_j \wedge dz_j$. The 'linearly elliptic condition" implies that we can coordinatize the $\mathbb{C}^n$ factor so that the linearized return map at the origin takes the form

$$dR_0(z_1, \dots, z_k) = (\lambda_1 z_1, \dots, \lambda_n z_n), \text{ where } \lambda_j = exp(2\pi i \omega_j) \in S^1$$

(Recall that dR_0 can be viewed as arising from the monodromy map A by a quotient.)

Condition (L). Condition (L) in theorem 2.2 is the condition that there are **no** nonzero integer vector solutions $(k_0, k_1, \dots, k_n)$ to the linear equations:

{resonance} (77) $$k_1\omega_1 + k_2\omega_2 + \dots k_n\omega_n = k_0 \text{ having } 0 < \Sigma_{j=1}^n |k_j| \le 4,$$

A linear equality among frequencies ω_j involving integer coefficients k_j is called a "resonance". The size of the integers involved is the 'order' of the resonance. So condition (L) asserts that there are no resonances of order 4 or less. In particular condition (L) excludes any of the λ_j from being 1, hence all our discussion around "mandatory 1's.

For any choice of positive real constants c_j, the linearized return map dR_0 leaves the polydiscs

$$z : |z_j| \le c_j$$

invariant, together with their "boundaries" the n dimensional torii given by $|z_j| = c_j$. Then these torii are all Lagrangian. A subcollection of these torii will correspond to the KAM torii of the KAM theorem.

Oscillators and Normal Forms. The Harmonic oscillator with frequencies ω_j is encoded by the quadratic Hamiltonian $H_{osc}(z_1, \dots, z_n) = \frac{1}{2}\Sigma\omega_j|z_j|^2$ where the z_j are as above and $z_j = q_j + ip_j$ with q_j, p_j Darboux coordinates so that $\omega = \Sigma dq_j \wedge dp_j$. The Hamiltonian vector field for the harmonic oscillator is linear with eigenvalues

$\pm i\omega_j$ and its time 2π flow is dR_0 above. Define new coordinates I_j, θ_j on $\mathbb{C}^n$ called "action-angle coordinates" so that

$$H_{osc} = \Sigma \omega_j I_j. \tag{78}$$

The explicit change of coordinates is

$$I_j = \frac{1}{2}|z_j|^2, \theta_j = Arg(z_j)$$

These are Darboux coordinates: $\omega = \Sigma dI_j \wedge d\theta_j$. Their Darboux nature is seen by recalling that $dx \wedge dy = d(\frac{1}{2}r^2) \wedge d\theta$ expresses the planar area form in polar coordinates. Then q_j, p_j play the role of the Cartesian coordinates (x, y) while (I_j, θ_j) of polar coordinates r, θ.

If a Hamiltonian $H = h(I_1, \ldots, I_n)$ on $\mathbb{C}^n$ is a function of actions alone then its time t flow leaves the I_j invariant and acts on the θ_j by $\theta_j \mapsto \theta_j + t\nu_j(I)$ where $\nu_j = \frac{\partial h}{\partial I_j}$. We still call the ν_j frequencies. We call the map $I = (I_1, \ldots, I_n) \mapsto \nu(I) = (\nu_1(I), \ldots \nu_n(I))$ the frequency map. We will call a symplectic map obtained by setting $t = 1$ in such a flow "an integrable oscillator map". The integrable oscillator maps relevant for (T) are those whose Hamiltonian $H = H_{model}$ is quadratic in actions: $H_{model} = h(I_1, \ldots, I_n)$ where

{eq: modelOsc}

$$H_{model} = H_{osc} + h_2 \tag{79}$$

where

$$h_2 = \frac{1}{2}\Sigma a_{jk} I_j I_k \tag{80}$$

and the a_{jk} are the entries of a symmetric matrix. Its corresponding frequency map is $\nu_j(I) = \omega_j + \Sigma a_{jk} I_k$. Let us write $R_{model} : \mathbb{D} \to \mathbb{D}$ for the time 1 map associated to H_{model}. Then we have:

{eq: BirkN1}

$$R_{model} : (I_j, \theta_j) \mapsto (I_j, \theta_j + \nu_j(I)) \text{ where } \nu_j(I) = 2\pi\omega_j + \Sigma a_{jm} I_m. \tag{81}$$

The Birkhoff normal form theorem is, at its core, a constructive iterative algorithm for getting an elliptic Hamiltonian maps into the above model integrable form plus a higher order error. Under our non-resonance hypothesis (L) the algorithm yields a symplectic polynomial diffeomorphism fixing the origin and converting the *3rd jet* of our given return map R to that of an integrable oscillator map of the above quadratic type of H_{model}. See Arnol'd [**10**], Appendix 7 B, particularly p. 388 and Theorem 3 there regarding the Birkhoff normal form. So what this theorem says is that by a polynomial change of variables we can put R into the form:

{eq:BirkNormForm}

$$R = R_{model} + O(r^3); \tag{82}$$

where R_{model} is as above. The term h_2 of H_{model} is called the 2nd order Birkhoff invariant. [REF?? Sternberg Cele Mech? Guill-Stern?] The 'r^3' of the error term refers to the r of $r^2 = 2\Sigma_j I_j = \Sigma |z_j|^2$. Under assumption (L) the spectral invariants of the symmetric matrix a_{jk} become symplectic invariants – unchanged by symplectic conjugacies, which is to say, by choice of Darboux coordinates.

Condition (T).

The 'twist' condition (T) is the condition that the matrix arising in the 2nd order Birkhoff normal form h_2 be non-degenerate

{twist}

$$det(a_{jm}) \neq 0. \tag{83}$$

Physically, the twist condition asserts that the frequencies vary with amplitudes and that this variation is non-degenerate: the frequency map is a local diffeomorphism near $I = 0$.

EXERCISE 2.1. *Show that relative to Cartesian coordinates the twist condition is a condition on the 3rd order Taylor expansion of R at the origin. Hint: trace through the changes of variables* $q_j = \sqrt{2I_j}\cos(\theta_j), p_j = \sqrt{2I_j}\sin(\theta_j)$ *and lookk at the normal form in Cartesian coordinates.*

The KAM torii guaranteed to exist correspond to those for which the frequencies $\nu(I)$ are "sufficiently irrational". (See [**10**], [**37**]). This condition of being "sufficiently irrational" selects out a frequency set $F \subset \mathbb{R}^n$ of positive measure. On these torii the Hamiltonian flow becomes, after a time reparameterization, the flow governed by $\dot{\theta}_0 = 1$, $\dot{\theta}_j = \nu_j(I)$, where θ_0 is the coordinate around the orbit circle S^1. The corresponding flow, restricted to such a torus is the standard irrational linear flow on a torus, an ergodic flow all of whose orbits are dense in the torus. This explains the meaning of "irrational torii" in the theorem statement. The non-degeneracy condition on the frequency map guarantees that the inverse image $K := \nu^{-1}(F)$ of this frequency set in action space continues to have positive measure. K has the structure of a fat Cantor set in action space, that is, topologically a Cantor set, but one whose Lebesgue measure relative to $dI_1 dI_2 \dots dI_n$ is positive. The set of invariant torii whose existence the KAM theorem guarantees is diffeomorphic to $S^1 \times K \times \mathbb{T}^n = K \times \mathbb{T}^{n+1}$ where the torus $\mathbb{T}^n$ is coordinatized by the θ_j, $j > 0$ now. We call these the "surviving torii". Lebesgue measure on the energy surface is $dI_1 \dots dI_n d\theta_1 \dots d\theta_n d\theta_0$, showing that the measure of the set of surviving torii is positive. Divide the measure of set of surviving torii having $|I| \le \epsilon$ by the measure of the subset of phase space having $|I| \le \epsilon$. The assertion of "density 1" in theorem 2.2 is the assertion that this relative measure tends to 1 as $\epsilon \to 0$.

Chenc suggests emphasizing the philosophy underlying KAM. KAM torii being CLOSURES of orbits, are hard to destroy. Rational torii, not being closures of any one of their orbits are easily destroyed by perturbation - RM

7. Arnol'd diffusion: dodging sticky torii.

Lagrangian manifolds have half the dimension of the phase space $\mathcal{P}$ within which they sit. Let us write $2(n+1)$ for the dimension of this phase space, so that $n+1$ is the number of degrees of freedom, or half the dimension of phase space. Now suppose we have a KAM stable periodic orbit for an autonomous Hamiltonian system on $\mathcal{P}$. Since the KAM torii are Lagrangian and live inside the same energy level as their mother orbit, they form a family of $n+1$ dimensional submanifolds embedded within that $2n+1$-dimensional energy surface containing the orbit. If $n = 1$ each torus is 2-dimensional within its 3-dimensional energy level set and separates the energy level into two components, one containing the orbit , thereby preventing escape from the orbit and the concentric family of torii force Lyapunov stability. The circular restricted three-body problem has 2 degrees of freedom and Lyapunov stability has been established by this means for some of its orbits. See for example [**97**] and references therein. As soon as the number of degrees of freedom is greater than 2 this topological separation and blocking of motion by torii fails. For example, if $n = 2$ then the torii are 3-dimensional within a 5-dimensional energy level set and the complement of any torus will be a connected subset of the energy level set. A clever trajectory can wind around and dodge all the many KAM torii, eventually escaping any given neighborhood of the orbit. This escape process, this evading the Cantor net of KAM torii is known as "Arnol'd diffusion" and is currently the only approach towards a "No" answer to our Open Question, or its unbounded variant.

8. Nekhoroshev Stability

Although escape from the KAM torii surrounding an orbit can happen, when it does it takes a long long time. An imprisoned orbit may eventually escape the jail of its stability but a Great Escape may take many lifespans.

THEOREM 2.3 (Nekhoroshev). *Assume that a periodic orbit is KAM stable and that moreover the twist matrix of hypothesis (T) for the orbit's return map (equation (83) is positive definite. Let d be a Riemannian distance on phase space and $2n$ the dimension of phase space. Then there is an $\epsilon_0 > 0$ and a positive number C with the following signficance. Any solution starting within a distance ϵ of the orbit, with $\epsilon \leq \epsilon_0$, stays within $C\epsilon^{1/2n}$ of the orbit over a time interval $|t| \leq Cexp(\epsilon^{-1/2n})$.*

Oft-stated versions of the theorem says that the initial actions, as per our oscillator model above, assert that the individual action I_j as per our oscillator model do not vary by more than $C\epsilon^a$ over times of order ϵ^{-b}. Pöschel and others got these bounds so that $a = b = 1/2n$. The distance of a point from the orbit is of order $\sqrt{\Sigma I_j}$, and in the theorem is assumed to initially be ϵ. Since $\epsilon << \epsilon^{1/2n}$, these oft-stated versions of Nekhoroshev's theorem imply that the distance from the orbit remains of order $\epsilon^{1/2n}$ for t of order $\epsilon^{-1/2n}$, as we have stated.

Proofs of the theorem are based on estimating travel time near "resonance webs" -collections of hypersurfaces where the resonance conditions $\Sigma k_a \nu_a(I) = k_0$ hold with k_a, k_0 integer. All escaping orbits must escape by travelling near resonance webs. The analysis is deep and complicated and I refer interested readers to the literature, particularly [**127**].

8.1. KAM Heuristics. Irrational tori are hard to destroy. This principle drives KAM. The Birkhoff normal form R_{model} (81) leaves each of the tori $I_j = const$ invariant. Thanks to the Birkhoff normal form theorem, (81), (**??**) our return map R is a perturbation of this normal form. As we get closer to the orbit ($r \to 0$) the perturbation $R - R_{model}$ tends to zero. The principle tells us to expect that more and more of the normal form's irrational torii to survive as $r \to 0$.

What makes irrational torii hard to destroy while rational torii are relatively ephemeral? Each irrational torus comes to us as a solid block, being the closure of any one of its orbits. In contrast, each orbit on a rational torus is already closed, making the whole $n-1$-parameter ensemble of these orbits which make up a rational n-torus rather floppy and weakly held together.

The principle that irrational torii are sturdier than rational ones is best illustrated by the theory of circle diffeomorphisms and how it relates to area preserving maps R of the disc which take the origin to the origin. When $n = 2$ we just have one action $I = I_1$ and one angle $\theta = \theta_1$. We can view the model map R_{model} as a one-parameter family of circle maps $\theta \mapsto R_I(\theta) = \theta + \nu(I)$ parameterized by I. The circle labelled by I is irrational or rational depending on whether or not $\nu(I) = \omega + aI$ is. In the theory of circle diffeomorphisms every circle map f has attached to it a rotation number ν. If ν is irrational and f is sufficiently smooth then the theory supplies us with a a change of variables turning f into the map $f_\nu : \theta \mapsto \theta + \nu$ of irrational rotation by ν. On the other hand, if the rotation number is rational so that $\nu = p/q$ with $p, q \in \mathbb{Z}$ then typically the map f has a finite number of of period-q

orbits, alternating around the circle between being stable and unstable. The points between these periodic orbits, upon being iterated under the rational circle map, accumlate onto one or another of the finite number of stable periodic orbits.

Now we turn our perturbation on, replacing the normal form R_{model} of the 2-disc by the actual map R. We try to think of R as a one-parameter of circle maps, knowing that this is putting the horse before the cart -assuming existence of invariant circles – but let us see where it takes us. The circle theory above would tell us that each irrational circle gets perturbed to a nearby circle which is conjugate to it, and hence on which the map yields one conjugate to irrational rotation by $\nu = \nu(I)$. Thus the irrational circles near the origin wiggle a bit, but mostly remain. At this point, the circle theory suggest a key idea occuring in the proofs of KAM: focus on the rotation number and not the action. As we turn the perturbation on, try to follow the invariant circles according to their rotation number labels, not the values of their actions.

What happens to the rational circles upon perturbation? They do not remain whole, but instead, their fictitious 'perturbed circle maps" have split up into a finite number of orbits some of which are stable and some unstable. Being in the symplectic world, we have restrictions on the eigenvalues at these finite number of orbits. If an orbit has an unstable direction (eigenvalue outside the unit circle) it also has a stable one (eigenvalue inside the unit circle). The rational circle has bifurcated, splitting up into a finite number of elliptically stable periodic orbits and an equal number of hyperbolic ones. The dynamics near the hyperbolic ones, by necessity, must wander off into two dimensions and cannot stick to any one curve. The rational circles have been destroyed.

More happens with the hyperbolic ones. Typically they have homoclinic or heteroclinic tangles associated to them – intersections of stable and unstable manifolds, leading to Hamiltonian chaos in the interstices away from the allegedly surviving irrational circles. The elliptic periodic orbits on the other hand provide us with opportunities to try out KAM again and are typically surrounded by what are called "elliptic islands". We can try applying KAM to these, and the pattern repeats around each of them, leading to an overall fractal type of structure.

The cartoon model for all of this is the standard map pioneered by Chirnikov.

Pictures of standard map.

REcommended REFS: Moser. Arnold. De la Llave .... and unstable ones, with most orbits of neither type. Meiss ... Fejoz

9. Application to N bodies. Resonances and necessity of reduction.

Due to the 'mandatory 1's' described above, forced on us by its symmetries, the periodic orbits of N-body systems *never* satisfy hypothesis (L) of the KAM stability theorem. As we described in that section, can get rid of these 1's by the projecting the orbit to the relevant symplectic reduced space , or, what is the same, adapting the return-map discs $\mathbb{D}$ to angular momentum as well as energy, and making sure that they are transvserse to relevant subgroup (G_μ) of G.

Let γ be a solution to a Hamiltonian system with symmetry. Let J be the momentum map for the Hamiltonian action –so angular momentum for the N-body problem in center-of-mass frame, and $J_0 = J(\gamma(0)) = J(\gamma(t))$ the value of the momentum map on the orbit. When we say "γ's reduced space" we mean the corresponding symplectic reduced space $\{J = J_0\}/G_{J_0}$ together with the projected orbit $\pi(\gamma(t))$ where $\pi : \{J = J_0\} \to \{J = J_0\}/G_{J_0}$. (Here as usual G_{J_0} is the co-adjoint stabilizer of that orbit.)

DEFINITION 2.5. *We will say that the relative periodic orbit γ is reduced KAM stable if its projection to its reduced space satisfies the hypothesis (L) and (T) of 2.2 for the Hamiltonian flow induced on the reduced space.*

thm: KAM stable

THEOREM 2.4. *[Reduced KAM stability; planar case] Let $\gamma(t)$ be a relative periodic orbit for the* **planar** *N-body system which is reduced KAM stable. Then a rotationally invariant neighborhood of γ in the full phase space is partially filled with irrational invariant torii whose density tends to 1 as we approach the orbit. On these torii the relative distances r_{ab} are quasi-periodic and hence bounded functions of time.*

PROOF.
Apply the KAM stability theorem to the reduced orbit. One gets KAM stability at the reduced level, for that value of energy and momentum, with consequent slice map defined on a disc of dimension $2(N-2)$. Now vary initial conditions, allowing energy and momentum to vary. We can follow the relative periodic orbit by the IFT into these new values, getting a new orbit in a new reduced space. Because of the planarity assumption, these reduced spaces do not jump in dimension and vary smoothly with initial condition. Since the KAM conditions (L) and (T) are open conditions they persist into nearby values of angular momentum and energy, yielding KAM stable orbits within each reduced space, and in particular being of positive measure and density 1 nearby. Now pull back up to the total phase space to see that the rotationally invariant family of nearby relative periodic orbits are surrounded by torii (no longer Lagrangian) whose total measure is positive and whose density tends to 1 as we approach any one of the family.

10. Case Study: the Eight

sec: case study

describe Simo's analysis - RM

This section is inspired by Simós' article [**144**].

The figure eight orbit (**??**) is KAM stable. At a practical numerical level, take the initial conditions that Simó culled and integrate forward with a good numerical integrator for a 100,000 periods. Observe that the orbit continues to look the same.

To obtain a rigorous proof of KAM stability, one can verify the KAM hypothesis (L) and (T) as follows. The reduced space for the planar three body problem at angular momentum zero (or any angular momentum) has dimension 6. Subtract 2 for the energy-time 'mandatory 1's' to see that the relevant return map for investigating near-eight behaviour is a symplectic map $R : (\mathbb{D}^4, 0) \to (\mathbb{D}^4, 0)$ of a 4-disc. Its differential dR_0 has four eigenvalues. It follows that if we find eigenvalues $e^{i\omega_1}, e^{-i\omega_1}, e^{i\omega_2}, e^{-i\omega_2}$ with ω_1, ω_2 clearly distinct, then we have found all the eigenvalues for dR_0 and that the orbit is elliptically stable.

check if Simo's are $e^{i\omega}$ or $e^{2\pi i\omega}$; write those other numerical guys. -RM

Simó [**144**] investigated the stability of the eight and the dynamics of its surroundings with great care numerically. He computed the eigenvalues of the differential of the return map to be $e^{i\omega_1}, e^{-i\omega_1}, e^{i\omega_2}, e^{-i\omega_2}$ where

$$\omega_1 = .008422724708131, \omega_2 = .298092529004750$$

(The eigenvalues are independent of the period or size chosen for the eight since scaled versions of the eight yield conjugate return maps, hence identical eigenvalues.) To verify KAM condition (L) one checks that the 64 numbers $k_1\omega_1 + k_2\omega_2$

with $k_i = -4, -3, -2, -1, 1, 2, 3, 4$ are not integers. (We left $k_i = 0$ out of the sum set since it is obvious that $n\omega_i \neq \mathbb{Z}$ for $n = 1, 2, 3, 4$.)

The reader might object at this point. There is no closed form expression for the eight and so approximate representations of its monodromy matrix A and linearized return map dR_0 must be found with the aid of numerical integration. We do not (and probably cannot) have exact closed form expressions for A or its eigenvalues. When the machine is telling us that one of its eigenvalues is $\lambda = e^{i\omega_1}$ could it actually be that $\lambda = (1/.9999...9)e^{i\omega_1}$ and hence the orbit would fail to be spectrally stable? No. Due to the spectral symmetry REF , if one eigenvalue $\lambda = re^{i\omega}$ were to lie outside but very close to the unit circle then it must arise in a full quartet: $\lambda, 1/\bar{\lambda} = \frac{1}{r}e^{i\omega}, \bar{\lambda}, 1/\lambda$ are all in the spectrum. If the closeness of r to 1 were below machine precision we could not tell λ from $1/\bar{\lambda}$ or $\bar{\lambda}$ from $1/\lambda$, but, regardless, taken together $\lambda, 1/\bar{\lambda}, \bar{\lambda}, 1/\lambda$ would have to account for a 4-dimensional eigenspace, exhausting our spectrum, and making the occurence of the other found pair of eigenvalues $e^{\pm i\omega_2}$ impossible. So it is enough to find 4 distinct eigenvalues lying within machine precision of the unit circle and arising in two complex conjugate pairs to **guarantee** that our orbit is linearly stable.

Simó numerically approximated the Taylor expansion of the return map of the eight, thereby establishing the twist condition (T) and convincingly establishing that it is KAM stable. In that paper you can find pictures of some of the KAM torii around the eight. Later Simo and Kapela [**66**] used interval arithmetic to rigorously establish the two KAM conditions (L) and (T), thereby guaranteeing the eight's KAM stability.

Gareth Roberts [**129**] gave a separate proof of linear stability of the eight based on its D_6-symmetry of the eight. Recall OOP – do in TOUR? – that the eight is determined by one-twelfth of its orbit. Using these symmetries Roberts could reduce the spectral stability computation to the analysis of a single two-by-two matrix culled from this one-twelfth and the symplectic realizations of the symmetry operations.

◇

Domain of stability.

The eight is KAM stable, but how far can we push it and retain stability? There are many directions we can push in. We can change initial conditions, keeping the masses the same. Or we can vary the masses, using the implicit function theorem to follow the orbit as a relative periodic orbit and ask for the new orbit's stability. Simó investigated the results of both types of pushing. For the first, he took a certain two-dimensional slice of the Poincaré disc and integrated forward. As he integrated he kept track of the minimum and maximum of the interbody distances and threw orbits out as 'unstable' when these distances went below or above some tolerance. He kept track of up to 30,000 crossings and then plotted the resulting "Julia set" of 'stable' orbits. The result in indicated in figure 1.

The specific slice Simó took was as follows. He insisted the initial positions were collinear. After reduction, the set of phase points with angular momentum zero and energy 1/2 is 4 dimensional and can be parameterized near the eight by the initial velocites $\dot{x}_2, \dot{y}_2, \dot{x}_2, \dot{y}_3$ of bodies 2 and 3. His slice was determined by the eight relation $\dot{x}_2 = -\frac{1}{2}\dot{x}_3, \dot{y}_2 = -\frac{1}{2}\dot{x}_3$. In the figure the axes are then the inital values of $\dot{x}_3, \dot{y}_3$. The black regions are stable according to the numerical tests. The initial

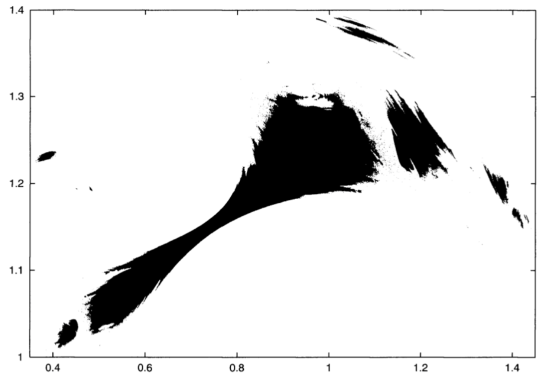


FIGURE 1. Stability Zone around the eight. Courtesy of C. Simó. [**144**] fig: Julia

condition for the eight is close to the center of the plot, in the narrow pinched section of the main black region.

In order to follow the orbit as we vary masses, let's recall how to use the implicit function theorem to follow a periodic orbit. Let R denote the return map, now viewed as a function of the masses $m \in U \subset \Delta_2$ as well as the initial positions $x \in \mathbb{D}^4$, with the masses entering through Newton's defining equations. Here $\Delta_2 = \{(m_1, m_2, m_3) : m_i > 0, m_1 + m_2 + m_3 = 3\}$ is the 'mass simplex" and U is some small open subset near the point $m_* = (1,1,1)$ of equal masses. Thus $R : \mathbb{D}^4 \times U \to \mathbb{D}^4$. A periodic orbit is a fixed point of R so a pair (x, m) with $R(x, m) = x$. We've set our coordinates up so that $x = 0, m = m_*$ corresponds to the figure eight. Rewrite the fxied point equation as $F(x, m) = 0$ where $F(x, m) = R(x, m) - x$. The inverse function theorem asserts that we can 'follow" the zero to a family of orbits $x(m)$ with $F(x(m)) = 0$ provided $\frac{\partial F}{\partial x}|_{0,m_*} \neq 0$. Here, the partial means the derivative with respect to all directions $x \in \mathbb{D}^4$ and saying that the partial derivative is not zero means that it is invertible. But $\frac{\partial F}{\partial x} = DR_0 - Id$ with $m = m_*$. This matrix is invertible **exactly** when 1 is not an eigenvalue of dR_0. The condition (L) on the reduced eight, substantiated by Simó's eigenvalue computations, guarantees this invertibility of $\frac{\partial F}{\partial x}|_{0,m_*}$. So, we can follow the orbit uniquely to get a nearby family of near-eights. Simo followed these mass-dependent eights, keeping track of their stability. He found these periodic orbits became linearly unstable when he varied the masses by 10^{-5} away from the equal mass case.

◇

Are there figure eight triple star systems in our universe? Heggie [**53**] investigated this question numerically. He then threw pairs of equal mass binary systems at each other numerically to get a sense of how often figure eights might form in a galaxy. He writes that "a small fraction of one percent" of the experiments led to eights that lasted a few periods. If the masses are not equal, the creation of eights becomes even less likely, since as Simó discovered, moving a fraction of a percent away from equal masses leads to instability. Heggie summarized his investigations by asserting in a conversation that "there are somewhere between one figure eight orbit per galaxy and one per universe"

⋄

The eight has zero angular momentum. Turn on a bit of angular momentum. Rest of this section in flux CITE 'rotating eights' The eight persists but must rotate. In the rotating frame the new eight looks stable. FIGURE . Back in the inertial frame this perturbed eight rotates – so cannot stay close to the mother eight that birthed them. So it does not make sense to talk about the eight being "stable" in inertial space under perturbations which change its angular momentum. This discussion shows us again that the correct way to speak about N-body stability is in the Poisson reduced phase space $\mathcal{P}/G$.

include discussion of rotating eights?? -RM

If we view the eight as a solution to the planar 3-body problem then it is reduced KAM stable: these rotating eights are KAM stable. They have their own KAM torii , torii which are close in shape space and also in reduced phase space to the mother orbit than birthed them. Viewed from the perspective of the spatial 3-body problem they are also stable. The eight, viewed in space, has three axes of symmetry, one of which represents rotations in its plane. Along all three axes we can follow eights, the different eights unfolding in different ways. For very small perturbations all these new eights will be KAM stable.

okay! new tools - Poisson reduction -RM

11. More KAM stable celestial situations

One motivation behind the birth of the KAM theorem was to prove or at least better understand the stability of the solar system. I urge the reader to consult Appendix 8 of Arnol'd's [**10**] for a beautiful treatment which I cannot hope to match. Here is a very abbreviated catalogue of situations within the N-body problem to which KAM stability theorems have been applied.

For the solar system itself, or what are known as "planetary systems" there is a vast literature. The underlying KAM theorem used here is closer to the original version which is explicitly a perturbation theory. The ratios of the masses of the planets to that of the sun are proportional to the perturbation parameter. So select out 1 mass as large, call it "m_0 for the sun, and scale the others by ϵ to arrive at a family of $N+1$-body problems with mass ratios $1 : \epsilon m_1 : \epsilon m_3 : \ldots \epsilon m_N$. If we formally let $\epsilon \to 0$ we get N uncoupled Kepler problems. Look at the situation where the Keplerian orbits are arranged in concentic circles, well separated. Now turn ϵ back on. We are in the realm of KAM theory. A new problem arises not discussed above. For each planet a new cluster of 1's arise in the eigenvalue of the return map not accounted for amongst the mandatory 1's we described above. These arise due to the fact that **every** orbit of the negative energy Kepler orbit is periodic and moreover if we fix their energy then their periods are all the same. This is refered to as the $1 : 1$ resonance of Kepler: perturbed Kepler orbits come back to where they started. Many acre-feet of prose have gone into getting around or dealing with these 1's, thus insuring that some version of the conclusion of the KAM theorem holds in the planetary situation. Due to this problem, an entire subfield of KAM theory has developed, known as "degenerate Hamiltonian KAM theory". For a few post-Arnol'd results and references on the planetary problem I urge the interested reader to consult Féjoz [**37**] and Chiercia [**26**].

The Lagrange orbit in the three-body system is an interesting fairly well-studied case. A few references in historical order which arrive at various KAM stability results around the Lagrange orbit are [**29**], [**97**], [**11**], and [**159**]. The first two concern the planar restricted problem and the third concerns the spatial restricted problem.

12. Global Instability with Strong Force

If we go to the strong force ($\alpha = 2$ power law) case of the N-body problem then the answer to the Open Question becomes "no". Indeed the Lagrange-Jacobi identity in this case (49) asserts that $\ddot{I} = 4E$ in this case, where E is the energy. It follows that if $E < 0$ then every orbit begins and ends in total collapse: $I = 0$, while if $E > 0$ every orbit is unbounded:$I \to \infty$ as $t \to \pm\infty$ (FIGURES XX) So for an orbit to be periodic it must have $E = 0$. In that case $\ddot{I} = 0$ so $\dot{I} = const.$ along any orbit. But if the orbit is periodic we must have $\dot{I} = 0$. In summary, strong force periodicity requires both $\dot{I}(0) = 0$ and $E = 0$ of the initial conditions. In this case the value of I remains constant along the orbit. By a slight perturbation of initial conditions we change E or $\dot{I}(0)$, so there are no Lyapunov stable orbits.

The same arguments tell us that the only relative periodic orbits arise when $E = 0 = \dot{I}$. We can search for stable relative periodic orbits *within* the invariant manifolds $H = 0$ and $\dot{I} = 0$. They appear to be unlikely.

thm: unstable alpha =2

THEOREM 2.5. *For the case of the planar strong force ($\alpha = 2$) three body problem with equal masses, and within the invariant submanifold $H = \dot{I}(0) = J = 0$ every relative periodic orbit is linearly unstable and hence cannot be Lyapunov stable within this invariant submanifold.*

The theorem is proven in REF where we show that the flow modulo rotations, after reparameterization, becomes smoothly conjugate to geodesic flow on a complete surface of negative curvature. For such geodesic flows instability is well known.

See problem 6 in the next section for more in this direction.

◇

13. More approachable Questions.

The Newtonian planar 3 body problem can be viewed as a parameterized family of problems, the parameters being the masses, the energy E, and the angular momentum J. A relative periodic orbit for the problem must have negative energy E, by the Lagrange-Jacobi identity REF. The figure eight provides a KAM stable periodic orbit for $J = 0$.

1. Are there KAM stable relative periodic orbits having angular momentum zero for any mass distributions?

Scaling symmetry preserves the combination EJ^2, known as the Dziobek constant.

2. Are there KAM stable orbits for all negative value of the Dziobek parameter EJ^2 for the equal mass planar 3-body problem? For any masses?

To begin on this problem the reader might want list out what is known by undertaking a literature search. In Moeckel [**100**] you can find some stated results around periodic orbits.

3. Prove Arnol'd diffusion occurs for the figure eight orbit, thus establishing that it is not Lyapunov stable. A believable scenario to investigate would be one of "diffusion" from arbitrarily close to the eight out to near to one of the three companion hyperbolic eights that Simó found [**144**]. Once near the companion hyperbolic eight one could show follow that eight's unstable manifold very far away.

4. When the figure eight is viewed as a solution to the **spatial** three body problem is it still KAM stable?

The proof we gave for theorem 2.4 fails for the spatial problem at zero angular momentum since the reduced spaces jump from 6 to 8 dimensions. It seems likely that some additional condition will need to be verified on some "torsion" arising from a now-6 dimensional return map in order to invoke the usual KAM technology and proceed to a proof.

carefully define what you mean by "KAM stable" for a rel period orbit, either HERE or , probably better, earlier! -RM

5. Generalize the Albouy-Dullin-Scheule stable example to any number N of bodies as follows. The equilateral triangle form the vertices of a regular 2-simplex and is a central configuration for $N = 3$. See proposition 1.5. Take instead the regular simplex central configuration solution for any number $N > 3$ ofbodies, placing them in a Euclidean space $\mathbb{E}$ of dimension $d = 2(N-1)$. We saw in the previous chapter that the configuration can be 'spun' using an almost complex structure $\mathbb{J}$ on $\mathbb{E}$ so as be made into a relative equilibrium. Any element $\Omega \in so(d)$ can be put in a two-by-two block diagonal form corresponding to splitting $\mathbb{E}$ into $N-1$ orthogonal 2-planes and 'spinning' each two plane independently at a different frequence. $\mathbb{J} \in so(d)$ corresponds to all frequencies equal, say $1 : 1 \ldots : 1$. Show that by deforming the side lengths with the frequencies, the simplex can be deformed so as to form a family of quasi-periodic "isosceles" relative equilibrium solutions, each of its $N-1$ Jacobi vectors spinning in one of the 2-planes. Show that if "spun fast enough" some of these relative equilibria become Lyapunov stable.

For the strong force $\alpha = 2$ power law N-body problem, we saw in the previous section that relative periodic orbits only exist when $E = 0$ and $\dot{I} = 0$. If we fix $E = 0, \dot{I} = 0$ there are many relative periodic orbits. If, in addition, $J = 0$ and the masses are all equal, then all of these orbits are linearly unstable.

6. Consider the $\alpha = 2$ problem within the invariant manifold $E = 0, \dot{I} = 0$. Are all relatively periodic orbits linearly unstable, regardless of the value of the angular momentum and the mass ratios?

14. Notes.

Section 0. When Chenciner and I stumbled over Moore's temporarily forgotten figure 8 orbit we were surprised to find that it was stable. Or at least that's what numerical integrations on the computer and our friends who were experts in numerical computations told us. Back then, I had worked in control theory and in Hamiltonian systems so I knew a bit about "asymptotic stability" and "stability by the energy-Casimir method" but I was quite ignorant of the notions of stability available in "real" Hamiltonian dynamics, which is to say, in celestial mechanics.

Section 1. Work of Laskar and colleagues suggests that the solar system might not be stable. REF. In a few billion years there is a few percent chance that the eccentricity of Mercury's orbit increases to the point that it intersects the orbit of Venus, to cataclysmic effect.

Hermann in turn attributes this ‘Oldest Question’ to Newton.

Section 2. See p. 74 of chapter 6 of [**94**] for a nice summary of various types of stability.

Linear stability. The eigenvalues of the monodromy are often refered to as the “multipliers”. refer to Herman,

for KAM Arnol’d , that KAM book, Stable and Random Motion ,and some basic dyn sys texts note (T) can be relaxed and allow some degeneracy

note usual statements of the KAM theorem require irrationality and linear dependence over the reals of the frequencies. Such assertions are numerically unverifiable. In contrast, the hypothesis (L) and (T) are finite in nature, and can be verified rigorously on a machine using, for example, interval arithmetic.

The diffeomorphism of the first step of the Birkhoff normal form takes the form $Id + Q(z, \bar{z})$ with Q homogeneous quadratic

The version of the Nekhoroshev theorem stated is an improvement of the original due to Pöschel REF.

On ‘case study”. 2nd paragraph. Another way to look at this implication that imprecise numerics can give exact spectral stability is as follows. Distinct, but numerically indistinguishable eigenvalues very close to, but not on the unit circle might yield a single eigenvalue in the numerics but it would have to have multiplicity two, while the multiplicities can be computed and are found to be one (two for each complex conjugate pair).

CHAPTER 3

Is every braid realized?

OPEN QUESTION 3.1. *Is every braid type on N strands realized by a periodic solution of the planar Newtonian N-body problem?* `braid_question`

The following special case inspired much of my work and remains open.

OPEN QUESTION 3.2. *Is every relative braid type on 3 strands realized by a relative periodic solution of the equal mass zero angular momentum Newtonian 3-body problem?* `braid_question`

1. Braiding bodies

When three points move in the plane without colliding they generate a braid on three strands. Take time as a third axis orthogonal to the plane. Then the three moving points $q_a \in \mathbb{R}^2 = \mathbb{C}, a = 1, 2, 3$ sweep out three non-intersecting curves $t \mapsto (q_a(t), t)$ in Galilean space-time $\mathbb{R}^2 \times \mathbb{R}$ whose union forms a traditional braid on three strands. Figure 1 shows the braid generated by the figure eight solution to the three-body problem while figure 2 shows the braid made when 1 and 2 form a tight binary system circling each other while 3 is far away. These space-time braids provide a means of visualizing free homotopy classes in $\mathbb{C}^3 \setminus \Delta$, the configuration space of the three-body problem.

By a *braid type* on N strands we mean a free homotopy class of loops for the planar collision-free N-body configuration space, $\mathbb{C}^N \setminus \Delta$. Braid types are in bijective correspondence with conjugacy classes of the fundamental group of $\mathbb{C}^N \setminus \Delta$, a group which is known as the *pure braid group* or *colored braid group* on N-strands and which forms a well-known finite index subgroup of the more famous Artin braid group. We cover some basics regarding braid groups below in section C. See also [**14**] Birman, 1974, chapters 1 and 2.

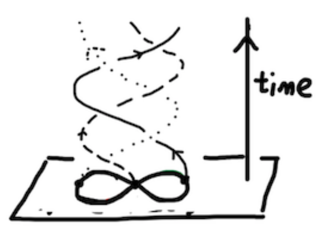


FIGURE 1. The braid generated by the figure 8 orbit. `fig8braid`

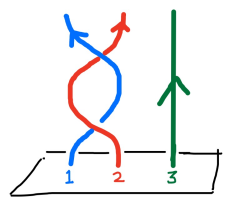


FIGURE 2. The braid A_{12} generated by a tight binary forms one of the generators of the pure braid group.

fig:braidtightbinary

We also stated a relative version of the open question. Recall that a planar N-body solution is called "relative periodic" of period T if $q(t+T) = Rq(t)$ for some rotation matrix $R \in SO(2) = \mathbb{S}^1$. The rotation group generates the center Z of the pure braid group by taking a fixed base configuration $q_0 \in \mathbb{C}^N \setminus \Delta$ and rotating it one full revolution. Since the N-body configuration space modulo rotations is homotopic to shape space, P_N/Z is the fundamental group of N-body shape space.

def: rel braid

DEFINITION 3.1. *A relative braid type is a free homotopy class in the collision-free planar N-body shape space, or, what is the same, a conjugacy class in the projective pure braid group P_N/Z.*

q: relative braids

QUESTION 3.1. *Is every relative braid type realized by a relative periodic orbit?*

1.0.1. *Links and the Borromean Rings.* (Just for fun.) If the period of $q(t)$ is T then the space-time braid connects the plane $t = 0$ to the plane $t = T$. The three starting points at $t = 0$ lie directly below the three ending points at $t = T$. Reconnect ending points to corresponding starting points by arcs, thereby forming a link. Applied to the braid for the figure eight braid, the link we get is the Borromean rings. See figure **??**

1.1. Eclipse sequences. As our three points move about in the plane, every once in a while they become collinear. An *eclipse* has occurred. Label that eclipse

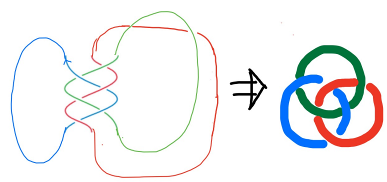


FIGURE 3. The braid for the figure 8. fig: BorromeanRings3

by the occluding mass. Our collision-free triple of curves then ticks off an *eclipse sequence* - a 'word' in the three 'letters' 1, 2, and 3, as time progresses. (In many references the eclipse sequence is called the 'syzygy sequence'.) For example, the figure eight solution makes the eclipse sequence 123123 in one period. This eclipse sequence encodes the curves' relative braid type. Correspondingly we have the following modified version of this chapter's open question.

eclipse_question QUESTION 3.2. *Is every periodic eclipse sequence realized by a relative periodic solution of the planar 3-body problem?*

The Lagrange relative equilibrium solution (3.1) consists of a rotating equilateral triangle. Suffering no eclipses, its eclipse sequence is empty. Nevertheless, the pure braid it generates is non-trivial, and indeed is the generator of the center of pure braid group. The following lemma formalizes the relations between the original Open Question of this chapter and the follow-up question 3.2.

lemma: eclipse rep LEMMA 3.1. *Let Z be the center of the pure braid group on 3 strands. Then Z is generated by rotating the base configuration one full revolution. With this in mind, the following sets are naturally isomorphic:*

- *(1)conjugacy classes in P_3/Z.*
- *(2) free homotopy classes in the shape sphere minus collisions*
- *(3)signed reduced periodic eclipse sequence*

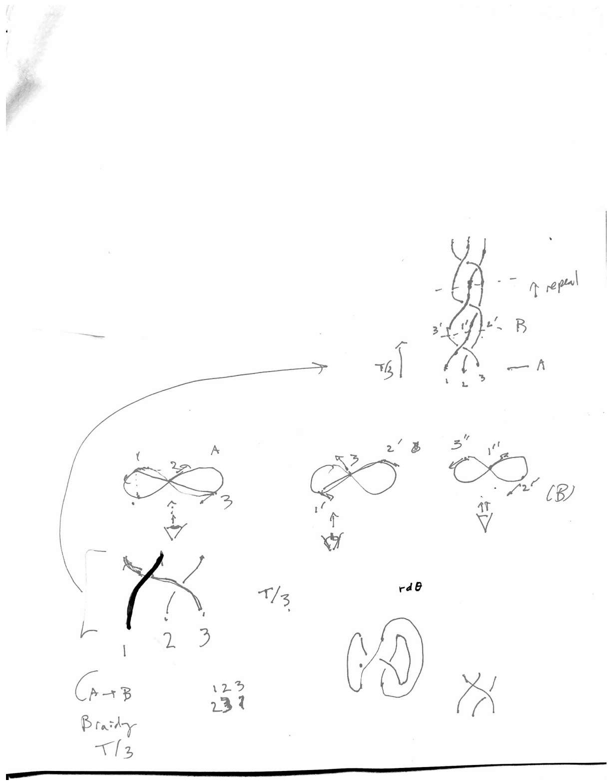


FIGURE 4. From the 8 to the Borromean braid. fig: BorromeanRings1

We will call an element of any one of these sets a **relative free homotopy class***.*

In the process of proving the lemma we will define what we mean by an eclipse sequence being "signed"' , "reduced" and "periodic".

PROOF AND EXPLANATION OF LEMMA. The Lagrange solution, projected to the shape sphere is trivial homotopically: it is a constant loop. A loop in the shape sphere $\mathbb{S}^2$ represents a *relative* periodic curve: a curve which closes up modulo a rotation. Two different lifts of the same shape curve differ by a rotation, hence, if each lift is closed then the difference of their homotopy classes is represented by a loop generated by the action of the rotation group and such rotational loops lie in the center of the pure braid group. Thus P_3/Z is the fundamental group of $\mathbb{S}^2$ minus three points. This realization provides the equivalence (1) $\iff$ (2).

To get (2) $\iff$ (3) observe that the three binary collision points $(12), (23), (31)$ cut the collinear equator of shape sphere into three arcs, labelled 1, 2 and 3, arc a representing the locus of collinear configurations of eclipse type a. FIGURE. When a curve crosses the arc labelled a an eclipse of type a has occured. Thus we can read off the eclipse sequence of a solution by projecting the solution to the shape sphere and counting off its intersections with the three collinear arcs.

subsec:signed

1.1.1. *Signing the Sequences.* We write 1^+ if the curve crosses arc 1 from the upper hemisphere to the lower hemisphere - or equivalently if the triangle goes from positively oriented to negatively oriented. We write 1^- if it crosses arc 1 from the

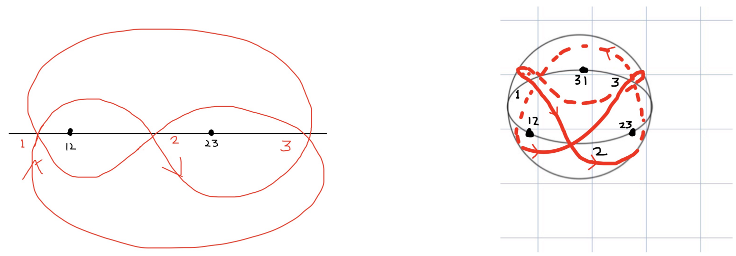


FIGURE 5. Eclipse sequence for the eight. Count crossings of the labelled equatorial arcs to get sequence 123123. The left figure is the stereographic projection of the right one, with the 31 collision placed at infinity fig:eclipses

lower hemisphere heading into the upper hemisphere. In this way the solution ticks off a sequence such as $1^+2^-3^+2^-\dots$. The superscripts + and − must alternate so that a choice of one sign on any one symbol determines the rest of the signs, **We have described the meaning of a "signed" eclipse sequence.** There are exactly two signed eclipse sequences for a given unsigned eclipse sequences. If we choose a relative periodic curve $q : \mathbb{R} \to \mathbb{C}^3 \smallsetminus \Delta$ representing one of these sequences then the other sequence is represented by the curve obtained by applying a planar reflection (for example $q_a \mapsto \bar{q}_a$) to this curve. On shape sphere, such a reflection acts by reflecting the shape image about the equator. Due to this simple 2:1 nature, subsec:stutters in what follows we do not bother with signs decorating eclipse sequences.

1.1.2. *Reducing Stutters.* Crossing the same arc twice in a row is a homotopically removable "mistake". We call such multiple crossings of the same arc "stutters" . Stutters can be removed in pairs. FIGURE. For example, homotopically speaking 122213 = 123 and 12213 = 113 = 3. We call an eclipse sequence "reduced" if it has no stutters.

1.1.3. *Cycling sequences.* Since our shape curve is periodic so is its eclipse sequence. So we can think of its symbols $1, 2, 3$ as written cyclically on a circle. Written cyclically it must continue to have no stutters if it is to be reduced. For example, 1231, when viewed as a periodic eclipse sequence, is not reduced, since when wrapped around on the circle it is the same as 1123.

QED .

It is worth noting that the fundamental group of $\mathbb{S}^2 \smallsetminus \Delta$ is the free group on two letters, with these letters being represented by simple loops enclosing any two of the three binary collision points.

EXERCISE 3.1. *Show that modulo particle relabelling (permutations of the three masses) and reflection there are less than 2^{M-2} distinct reduced eclipse sequences of length M.*

2. Motivation: Variational methods.

The following success story motivated the question of this chapter.

THEOREM 3.1. *If X is a compact manifold endowed with a Riemannian metric, then every free homotopy class of loops in X is realized by a periodic geodesic whose length minimizes the length of all closed curves in this class.*

SKETCH PROOF. The Riemannian metric allows us to measure the length of sufficiently smooth paths in X. Fix a free homotopy class. Minimize the length of loops representing this class. Geodesics are extremals of the length functional so in particular the minimum length loop will be the desired representative geodesic.

But does this minimizer actually exist? Compactness of X combined with the Arzela-Ascoli theorem insures existence of such a length minimizer. A bit of functional analysis shows that the minimizer is smooth enough that the derivation of the Euler-Lagrange equations goes through: the minimizer must be a geodesic representing the class. QED

The proof just sketched is an example of *the direct method of the calculus of variations*. We will try to apply the method to the N-body problem. To begin we require a variational principle: a functional on paths, like length, whose critical points are the sought-after solutions. The N-body problem at fixed energy E does admit a Riemannian metric (most of) whose geodesics are solutions to Newtons' problem at that energy E. See the Appendix on the Jacobi-Maupertuis metric . However the following variational principle has proved more useful than the JM length for the N-body problem.

write, and refer to, appendix -RM

DEFINITION 3.2. *The action A of a path $q : I \to \mathbb{E}$ is the integral $A(q(\cdot)) = \int_I L(q, \dot{q}(t))dt$ where $L(q, v) = K(v) + U(q)$ is called the "Lagrangian".*

Critical points of the action are solutions to Newton's equations, **provided** the critical points are collision-free. This fact is known as the *action principle*

Remark The Lagrangian is $K + U$. The Hamiltonian (conserved energy) is $K - U$. Compare them. Do not confuse them.

In the action-minimizing approach to the Open Question we fix the period T (rather than the energy E as we would have to do in using the JM metric) and the *braid type* of the competing curves and tries to copy the compact Riemannian proof above. One would like to say the action-minimizer exists and is our desired curve. Without additional assymptions made on the competing curves this direct approach fails due to the non-compactness of the N-body configuration space $\hat{\mathbb{E}} = \mathbb{C}^N \setminus \Delta$. There are two sources of non-compactness: $r_{ab} \to 0$ and $r_{ab} \to \infty$. We might call these the "infrared" ($r \to \infty$) and "ultraviolet" ($r \to 0$) divergences, following the terminology of quantum field theory. The more difficult of the two is the ultraviolet: particle collisions.

infrared? Ultraviolet? too corny? -RM

The standard gravitational N-body problem has solutions which end in collision in finite time and these have finite action. (See exercise XXX at the end of this chapter regarding the finiteness of the action.) By passing through collisions we can jump from one braid type to another while the action changes continuously.

then , include exercise! ? -RM

Nothing seems to prevent an action minimizing sequence of paths within a braid type to tend to a collision path. In 1977 W. Gordon [**46**] showed that this 'jumping phenomenon' actually happens for the standard two-body problem.

THEOREM 3.2 (W. Gordon). *Consider the braid type in which the two bodies wind twice around each other in the gravitational two-body problem. (See figure 2.) Fix the period T. Then any action minimizing sequence within this braid type and period tends to a path with collisions.*

On the other hand, if the braid type is that in which the two bodies wind once around each other, then the minimizer is realized and indeed a single period of every Keplerian solution of the two-body problem having the given period T realizes this minimizer.

To underline both the subtlety and the main issue, consider the limiting collision-ejection solution to Kepler's problem, the one with zero angular momentum. FIGURE ? Its action is the same as that of all the other T-periodic Keplerian solutions and so also realizes the minimum of the second part of Gordon's theorem, despite having a collision.

To get the direct method to work for N-body problems one must either leave the Newtonian $1/r$ potential or one must supplement the topological braid type constraint on loops with additional conditions.

Remark. The Jacobi-Maupertuis [JM] length is a direct extension of the length functional in Riemannian geometry to mechanical problems. See appendix D. The JM length is the path length as measured by a degenerate Riemannian metric known as the Jacobi-Maupertuis metric, a metric which depends on the choice of energy E, and is defined on the Hill region: $\{U \geq -E\} \subset \mathbb{E}$. In the interior of the Hill region this metric is an honest Riemannian metric and its geodesics are reparameterizations of the energy E solutions to Newton's equations. It is aesthetically tempting to try to use the JM metric to get interesting periodic solutions for the N-body problems. So far almost all such attempts have failed due largely to the fact that the JM metric degenerates to zero on the Hil boundary $\{U = E\}$. Two notable successes of the JM method in mechanics are Seifert's work [**?**] and Moeckel's work [**?**].

3. What's Known?

3.1. Cheating by changing the force. The direct method works beautifully for the planar N-body problem if we are willing, like Poincaré, to cheat. Poincaré changed Newton's $1/r$ pair potential to a $1/r^\alpha$ power law potential, $\alpha \geq 2$. Following Gordon [**46**] we call these "strong force" potentials.

DEFINITION 3.3 (Gordon). *A strong force potential is a power law potential $U_\alpha = \Sigma \frac{m_a m_b}{r_{ab}^\alpha}$ for which $\alpha \geq 2$*

The following computation underlies Poincaré's success.

LEMMA 3.2. *Let $L = K + U$ be the Lagrangian for a strong force N-body problem U and $A = \int L dt$ the corresponding action. Then any curve in $\mathbb{E} = (\mathbb{R}^d)^N$ which suffers a collision has infinite action.*

lemma:infinite action

PROOF OF LEMMA The key is the observation that for any positive r_1 the integral $\int_0^{r_1} \frac{dr}{r^{\alpha/2}}$ is infinite provided $\alpha \geq 2$. The kinetic energy in center-of-mass frame can be written as $\frac{1}{M}\Sigma_{a<b} m_a m_b |\dot{q}_{ab}|^2$ where $M = \Sigma m_a$ and $q_{ab} = q_a - q_b$. Since

$|\dot{q}_{ab}|^2 \geq \dot{r}_{ab}^2$ it follows that $K \geq \frac{m_a m_b}{M}\dot{r}_{ab}^2$ and so $L \geq m_a m_b(\frac{1}{M}\dot{r}_{ab}^2 + \frac{1}{(r_{ab})^\alpha})$ for any pair a, b. Now suppose that $q(t)$ is a smooth path in $\mathbb{E}$ in which bodies a and b collide at some time $t = t_c$. Write $r = r_{ab}$ so that $L \geq m_a m_b(\frac{1}{M}\dot{r}^2 + \frac{1}{r(t)^\alpha})$. Use $A^2 + B^2 \geq 2|AB|$ to get $L \geq C\frac{\dot{r}}{r^{\alpha/2}}$ or $Ldt \geq C\frac{dr}{r^{\alpha/2}}$ where $C = \frac{2m_a m_b}{\sqrt{M}}$ Use the divergence of the integral $\int_0^{r_1}\frac{dr}{r^{\alpha/2}}$ to conclude that $A(q([t_1, t_c]) \geq C|\int_0^{r_1}\frac{dr}{r^{\alpha/2}}| = +\infty$ where $t_1 < t_c$ is any time at which $r(t_1) \coloneqq r_1 \neq 0$. QED

Recall that $q : \mathbb{R} \to \mathbb{E} = \mathbb{C}^3$ is called 'relative periodic" of period $T > 0$ if $q(t+T) = Rq(t)$ for some rotation $R \in SO(2) = \mathbb{S}^1$. Poincare's theorem, theorem 3.3 concerns relative periodic orbits with $R = exp(i\omega)$. FIGURE FOR POINCARE.

Poinc

Theorem 3.3. *[Poincaré , 1896] Consider the planar 3-body problem with a strong force potential. Fix two integers n_1, n_2, a* **nonzero** *angle $\omega \in \mathbb{R}/2\pi\mathbb{Z}$ and a positive period $T \in \mathbb{R}$. Then there is a collision-free relative periodic solution of period T such that edge 12 rotates by ω radians, edge 13 rotates by $\omega + 2\pi n_1$ radians and edge 23 rotates by $\omega + 2\pi n_2$ radians in one period. The solution minimizes the action over all relative periodic curves satisfying the stated constraints.*

Poincaré's proof proceeds in a manner identical to the proof of the Riemannian geometry theorem. The key point is that lemma 3.2 prevents collisions. Collisions are required if we are to pass from one braid type – or in Poincaré's theorem, one homology class to another. Thus if a minimizer exists it is collision-free and satisfies the given topological constraint. A full proof is given later

Give! place section, reference! -RM

Poincaré's method of proof with almost no changes allows us to answer the open questions of this chapter for strong force potentials.

Call an eclipse sequence "tied" if all three letters $1, 2, 3$ appear in the sequence.

Theorem 3.4 (Montgomery, 1998). *In a strong force planar three-body problem every tied eclipse sequence is realized by a relative periodic action-minimizing solution. This solution has zero angular momentum.*

strong_eclipses

And here is a version for any number of bodies.

Theorem 3.5. *Let β be any relative braid type (definition 3.1) for the planar N-body problem, $R = exp(i\omega) \neq 1$ any non-identity monodromy, $T > 0$ a period, and $U = U_\alpha$ be a strong force potential. Then there is a relative periodic solution to the N-body problem with potential U, which realizes the relative braid type β and has period T and monodromy R. This solution minimizes the action over all relative periodic curves satisfying the constraints specified by R, β.*

thm: rel braids

Poincaré's theorem 3.3 is the homological $N = 3$ version of theorem 3.5 just stated as we now explain. Given a periodic curve in $\hat{\mathbb{E}} = \mathbb{C}^3 \setminus \Delta$ we can count the winding number of each edge $q_a - q_b$. This counting defines a surjective homomorphism $P_3 \to \mathbb{Z}^3$ which is the Hurewicz homomorphism $\pi_1(X) \to H_1(X)$ for the case $X = \hat{\mathbb{E}}$. (See section C below.) The center Z of the pure braid group P_3 on three strands is generated by rotations. The Hurewicz map sends the center to the 'diagonal' winding numbers, $(n, n, n) \in \mathbb{Z}^3$. It follows that upon forming the quotient of $\hat{\mathbb{E}}$ by the rotation group, we arrive at a surjective homomorphism $P_3/Z \to \mathbb{Z}^3/\mathbb{Z} = \mathbb{Z}^2$ realising the Hurewicz homomorphism for the shape sphere (or space) with collisions deleted:

{eq:Hurewicz3}

$$\pi_1(S^2 \setminus \Delta) \to H_1(S^2 \setminus \Delta). \tag{84}$$

The integers (n_1, n_2) of Poincaré's theorem parameterize the elements of $H_1(S^2 \setminus \Delta)$. We could call them "relative homology classes". A conjugacy class in $\pi_1(S^2 \setminus \Delta)$ is a relative braid type on 3 strands and the Hurewicz map is constant on conjugacy classes. So the Hurewicz map takes theorem 3.5 to Poincaré's theorem 3.3.

For a proof of theorem 3.4 and the other theorems discussed in this subsection, see section 3.4.

"Reduced" is perhaps a better adjective? 'Relative' conflicts with the existing use of that adjective in homology -RM

3.2. Imposing symmetries: choreographies and the Eight. We know one braid type that is realized: that of the figure eight orbit (see 3.3 and figures 4, 3), with its eclipse sequence of 123123. The eight was discovered (twice) by applying the direct method of the calculus of variations after imposing symmetry conditions on the closed curves which compete in the action principle. The symmetries require the masses be equal. At about the time that the eight was found, Chenciner and Venturelli [**25**] used the same methods to discover a pleasing solution to the spatial four-body problem which they dubbed the hip-hop (figure **??**). These two works represent the first real success of the direct method of the calculus of variations in the *gravitational* N-body problem. They started an avalanche of subsequent work.

A good way to describe discrete symmetries of periodic solutions to the N-body problem is to view periodic curves $q(t) = (q_1(t), \dots, q_N(t))$ in configuration space as maps

{symmetric rep}

$$q : \mathbb{T} \times [N] \to \mathbb{R}^d \tag{85}$$

by setting $q(t, a) = q_a(t)$. Here $[N] = \{1, 2, \dots, N\}$ is the N element set of mass labels. Points of $\mathbb{T} = \mathbb{R}/T\mathbb{Z}$ represents periodic time, so T is the orbit's period: $q(t + T) = q(T)$ and we call $\mathbb{T}$ the time circle . The group of permutations S_N acts on $[N]$ by permuting the mass labels, preserving the action of q provided the masses are all equal. We henceforth fix the center of mass by assuming the map satisfies $\Sigma_{\{a \in [N]\}} q(t, a) = 0$ at each instant t of time. The one-dimensional isometry group $O(\mathbb{T}) = O(2)$ of the time circle $\mathbb{T} = \mathbb{S}^1$ acts by time translations and reflections. The group $O(d)$ of orthogonal transformations of $\mathbb{R}^d$ acts on inertial space. All three group actions preserve the action A of a loop when the masses are equal, so altogether we have the group $O(\mathbb{T}) \times S_N \times O(d)$ acting on our loops, leaving their action unchanged, and mapping solutions to solutions.

We impose a symmetry condition on a loop by choosing a finite group Γ together with a faithful representation

$$\rho : \Gamma \to O(\mathbb{T}) \times S_N \times O(d)$$

by which we simply mean a homomorphism with empty kernel. Choosing a representation of Γ is the same thing as choosing a way for Γ to act on $\mathbb{T}$, on $[N]$ and on $\mathbb{R}^d$. By slight abuse of notation we write this action as $\rho(g)(t, a, x) = (gt, ga, gx)$ for $(t, a, x) \in \mathbb{T} \times [N] \times \mathbb{R}^d$ and $g \in \Gamma$. Imposing equivariance defines the symmetry condition on loops.

def: symmetric

Definition 3.4. *By a* Γ*-symmetric loop we mean any loop (equation (85)) which is* Γ*-equivariant:*

$$q(gt, ga)) = gq(t, a).$$

The idea is to restrict the action A to Γ-symmetric loops and then minimize the action by the direct method. If we can show that the minimum exists and is collision-free then it is automatically a Γ-symmetric solution to the N-body problem.

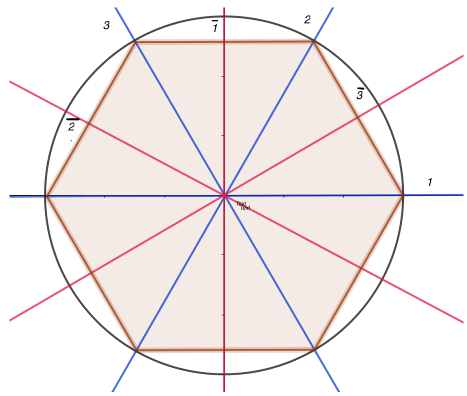


FIGURE 6. The symmetry group of the figure eight is that of a regular hexagon. The reflections about the 6 indicated lines generate this group.

fig: hexagon

It was by this method of combining the direct method with symmetries that the eight and the hip-hop were found.

To obtain the figure eight orbit we impose the symmetry group $\Gamma = D_6$ of a regular hexagon. The relationship of this group to the three-body problem is best seen by looking at the shape sphere with its tesselation into 12 spherical triangles made by the 3 isosceles great circles and the collinear equation. But this does not quite describe its representation on the three-body problem. D_6 has 12 element and is generated by the reflections about the symmetry axes of a regular hexagon. See figure 6.

To describe the representation of D_6 yielding the eight, observe that the symmetry lines of the hexagon come in two alternating flavours: lines which join opposite vertices and lines which join midpoints of opposite edges. See figure 6 where we have colored the two flavours blue and red. Label the blue lines going around the circle counterclockwise as $1, 2, 3$ with 1 corresponding to the x-axis. Label the red lines $\bar{3}, \bar{1}, \bar{2}$, making sure to interleave the labeling to insure that two mass labels (eg $1, \bar{1}$) never are adjacent to each other on the circle, and adding the bars to insure they are not confused with the blue lines. Draw the time circle $\mathbb{T}$ circumscribing the hexagon. Then each reflectional symmetry of the hexagon acts on the time circle by that same reflection. For example, if we coordinatize the circle so line 1 passes through $t = 0$ then the reflection about this line generates the time transformation $t \mapsto -t$. (If T is the period of the time circle so that $\mathbb{T} = \mathbb{R}/T\mathbb{Z}$ then t and $t + T$ are the same point of the time circle, and so the reflection $t \mapsto -t$ is also represented by $t \mapsto T - t$, showing that it also fixes the antipodal point $t = T/2$ to $t = 0$ on the time circle.) The reflection for any unbarred line act on the physical inertial plane $\mathbb{R}^2 = \mathbb{C}$ like a half-twist: $z \mapsto -z$, The reflection about any barred line act on the plane by reflection about the x-axis: $z \mapsto \bar{z}$. Finally, the reflection about a line labelled i or $\bar{i}$ acts on the label space $[3]$ by (jk), which is to say by switching the two remaining elements $j, k \neq i$ of label space. Note, if we let the reflections of figure 6 act on the labelled lines then this action on label space is precisely the action on the labelled lines: reflection about line 1 switches lines 2 and 3 (and $\bar{2}$ and $\bar{3}$) so can also be seen from the figure.

thm: eight

THEOREM 3.6. *The loop obtained by minimizing the action over the space of D_6-symmetric loops is collision-free and consequently solves Newton's equations. This solution is the figure eight solution.*

3.3. Choreographies. A surprising fact about the figure eight solution is that its three masses trace out the same curve, chasing each other around this curve. Such a solution is called a *choreography* . That the eight is a choreography follows rather directly from its D_6 symmetry. Take the product $h \in D_6$ of 4 consecutive reflections from figure 6, for example $1, \bar{3}, 2, \bar{1}$. This h acts by rotation by $1/3$ of a revolution on the time circle, by the identity on the inertial plane and by a cyclic permutation on the labels. Thus h generates the cyclic group $\mathbb{Z}_3 \subset D_6$. If we restrict our representation of D_6 to this $\mathbb{Z}_3$ we obtain a representation of $\mathbb{Z}_3$ which forces the motion to be that of a three-body choreography.

More generally, to impose the choreography condition for N equal masses, take $\Gamma = \mathbb{Z}_N$, the cyclic group of order N. Take $T = N$ so that the time circle is $\mathbb{T} = \mathbb{R}/N\mathbb{Z}$. Define the representation of $\mathbb{Z}_N$ by having the generator of $\mathbb{Z}_N$ act by taking (t, a, x) to $(t+1, a+1, x)$, where addition of the mass labels a is taken mod N. Then $\mathbb{Z}_N$-equivariance is the assertion that $q(t+1, a+1) = q(t, a)$ or $q_{a+1}(t) = q_a(t-1), t = 1, 2, \ldots, N$. In other words, all bodies travel the same curve, staggered in phase from each other by $1/N$th of a period. If we know how any one of the bodies move, we know how all the bodies move.

Soon after the eight was found Simo did a numerical study and found hundreds of other planar three-body choreographies. See figure (**??**) for a few of these.

3.4. Hip-hop. The Hip-hop is a symmetric solution to the equal mass spatial 4-body problem. The symmetry group of the hip-hop is $\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_4$. The generator of $\mathbb{Z}_2$ acts on the time circle by 180 degree rotation: $t \mapsto t + T/2$, trivially on the mass labels, and by $-Id$ on inertial space $\mathbb{R}^3$. The generator h of $\mathbb{Z}_4$ acts trivially on the time circle, by the cyclic permutation $a \mapsto a+1$ on the set $[4]$ of mass labels, and by the orthogonal transformation $B(x, y, z) = (-y, x, -z)$ on inertial space $\mathbb{R}^3$. Note that B projects onto the plane $\mathbb{R}^2$ and this projected map (which agrees with the restriction of B to $z = 0$) is rotation by 90 degrees. Thus the orbit of any point in $\mathbb{R}^3$ under the $\mathbb{Z}_4$ action projects onto a square in the plane.

As a consequence of being $\mathbb{Z}_4$-symmetric the motion of any one body determines that of all the others. If $q_1(t)$ is the motion of body 1 then the full 4-body loop is $q(t) = (q_1(t), B(q_1(t)), B^2(q_1(t)), B^3(q_1(t)))$. The $\mathbb{Z}_2$ symmetry implies that $q_1(t + T/2) = -q_1(t)$.

thm: hiphop

THEOREM 3.7. *The loop obtained by minimizing the action over the space of $\mathbb{Z}_2 \times \mathbb{Z}_4$-symmetric loops is collision-free and consequently solves Newton's equations. This solution is non-planar, with the 4 bodies oscillating symmetrically between two antipodally related tetrahedra, q_* and $-q_*$ forming a planar square ($z = 0$) at the instant half-way between.*

For q_* we take the configuration at which the z-coordinate of $q_1(t)$ is maximized. By a time translation we can assume this happens at $t = 0$. Then $q(T/2) = -q_*$.

3.5. Designer orbits. If we want a solution with a particular shape, or pleasing symmetry, then we can try their luck by choosing a group Γ and a representation of Γ that forces this symmetry. The direct method, applied to the loops having this symmetry type might yield the desired solution. An obvious choice of symmetry is that of the Platonic solids. See [**48**] or [**?**] for orbits having these symmetries, obtained by this method.

Sticking to the plane, one can add more and more bodies. A number of infinite families of such planar orbits have been established in this way by [**39**]. See also [**?**] for a summary of the state of affairs as of 2006.

As an alternative method one can take a slightly different perspective on the eight. In one-sixth of a period it travels from one collinear Euler instant to another, with the order of the masses on the line being permuted. The angle between these two lines is quite particular to the eight and the angle bisector of the angle they make is one of the symmetry axes of the eight. Instead, one could lay out N equal masses on the line in some configuration, then rotate the line and permute the masses. In this way we get a shooting problem from one collinear configuration to another. By varying the lines and the points on the line it may happen that velocities match in such a way that at the next collinearity the same angle and set of configurations repeats. In this way one would have made another type of solution, periodic if the angle between lines is a rational multiple of π. This method has been used by REF to establish a number of surprising orbits.

100s or 1000s have been found by now. We tour some in section XXX

This is the "designer" part of the activity – choosing a good group to act and a good group action. We then restrict the action functional to equivariant loops, namely those loops satisfying

The key to these works is to fix a symmetry class of loops over which to minimize the action A and then to figure out a way to guarantee that collisions do not occur amongst the minimizers.

3.6. Numerical algorithms. Periodic solutions lend themselves to Fourier series. Symmetry conditions put constraints on the coefficients of the Fourier series. One can search for numerical approximations to the minimizing solutions realizing a particular symmetry type by truncating to some order these Fourier series and computing the action of the resulting path. This action becomes a function of a finite number of Fourier coefficients. Apply a gradient search to find the Fourier coefficients that minimize. See [**?**], [**119**], [**?**]. All of this is beautifully done on the fly in the program written by Gregory Minton.

If one wants higher accuracy and more assurances that one has found an actual solution to Newton's equations, one can use the discrete symmetry to slice up phase space into Poincare sections associated to the symmetry type. One can then take the approximate minimizer as a seed for integration, in going from one slice to another, or through a series of slices. A fixed point of the resulting map will be the desired periodic orbit. In this way one gets an appoximate minimizer for the symmetric variational principle. See [**?**], XXX. All of this can be turned into rigorous existence proofs with bounds using the methods of interval arithmetic. REF

3.7. Abandon the method: answer the question! The following result almost solves the Open Question for the Newtonian three-body problem. Its proof uses dynamical systems methods instead of variational methods.

THEOREM 3.8 (Moeckel-Montgomery; 2015). *Every relative braid type on 3 strands is realized by a relative periodic solution to the Newtonian three-body problem* *with equal or nearly equal masses. These guaranteed solutions have small but non-zero angular momentum, come repeatedly very close to triple collision, and their eclipse sequences have many stutters.*

eclipse_main

Recall that a stutter within an eclipse sequence is a repeated occurrence of the same symbol.

We sketch the ideas behind this theorem in section 3.6.

direct method

4. The Direct Method. Proofs and collision obstructions.

In this section we go into some of the technical details around the direct method and use it to prove some of the theorems listed in the previous section. We start by proving Poincaré's theorem (3.3). In this proof we do not have to worry about collisions but the structure of the argument serves as a template for all our other variational proofs.

4.1. Proof of Poincaré's theorem. We will go through the steps carefully, enumerating them in such a way that simple replacements of the ingredients going into the steps will yield the proofs of many of the rest of the theorems of this section.

Step 1. Let Ω denote the space of all absolutely continuous paths satisfying the constraints of the theorem and having finite action. So Ω consists of all absolutely continuous paths $q : \mathbb{R} \to \mathbb{C}^3 = \mathbb{E}$ for which $q(t+T) = exp(i\omega)q(t)$, whose edges $q_1(t)-q_2(t), q_1(t) - q_3(t)$ have winding numbers n_1, n_2 per period T, and whose action $A(q) = \int_0^T L(q(t), \dot{q}(t))dt$ per period is finite. For the path's winding numbers to be defined the edges 12 and 13 must never shrink to zero, and this is guaranteed by ut lemma 3.2. Let A_* be the infimum of this action $A(q)$ as q varies over Ω.

Step 2. Let $q^{(n)} \in \Omega$, $n = 1, 2, 3, \ldots$ be any minimizing sequence : $A(q^{(n)}) \to A_*$ as $n \to \infty$. The heart of the proof consists of establishing that some subsequence of the minimizing sequence converges in the uniform (C^0) topology to a path q_*, that $q_* \in \Omega$ and that q_* is collision-free.

This subsequence is obtained by applying the **Arzela-Ascoli theorem:** any *bounded*, *equicontinuous* sequence of paths in $C^0([0,T], \mathbb{E})$ has a convergent subsequence. We recall the definitions. The sequence $q^{(n)}$ is *bounded* if there exists a positive constant C such that for all n we have $sup_t \|q^{(n)}(t)\| < C$. The sequence is *equicontinuous* if for every $\epsilon > 0$ there exists a δ *independent of* n, such that for all $t_0, s \in [0, T]$ we have $|s - t_0| < \delta$ implies $\|q^{(n)}(s) - q^{(n)}(t_0)\| < \epsilon$. So we must establish the boundedness and equicontiniuity of our minimizing sequence.

Boundedness. The standard way to establish boundedness in the calculus of variations is by verifying that the action with the path constraints is *coercive*.

Definition 3.5. *A pair (A, Ω) consisting of a variational principle A and a domain Ω for A is coercive if $\|q\| \to \infty \implies A(q) \to \infty$.*

Use the C^0 norm $\|q\| = sup_{t\in[0,T]}\|q(t)\|$ on paths in the definition of coercivity. It is then immediate that if (A, Ω) is coercive then any minimizing sequence is bounded.

$\omega \neq 0 \implies$ coercivity. We will only use the kinetic energy part of the action. We have $A(q) \geq \int K(\dot{q})dt := \frac{1}{2}\int \|\dot{q}\|^2 dt$. Cauchy-Schwartz's inequality $\int fg \leq \sqrt{\int f^2}\sqrt{\int g^2}$ applied to $f = \|\dot{q}\|, g = 1$ yields $\sqrt{\int \|\dot{q}\|^2 dt}\sqrt{\int dt} \geq \int \|\dot{q}\|dt := \ell(q)$, where $\ell(q)$ is the Euclidean length of the path q in the Euclidean vector space

$\mathbb{E}$. Thus

$$A(q) \geq \frac{1}{2T}\ell(q)^2.$$

Let t_* is a time for which $\|q(t)\|$ which realizes $\|q\| := sup_t \|q(t)\|$. Then $q(t_* + T) = e^{i\omega} q(t_*)$. Also $\ell(q) \geq |q(t_*) - q(t_* + T)|$ since Euclidean line segments realize the infimum of distance between any two points, here the points $q(t_*)$ and $q(t_* + T)$ in $\mathbb{E}$. But the angle between $q(t_*), q(t_* + T)$ is ω from which it follows that $\|q(t_*) - q(t_* + T)\| = 2R\sin(\omega/2)$ where $R = \|q\| = |q(t_*)|$. In other words

$$A(q) \geq \frac{1}{2T} 4\|q\|^2 \sin^2(\omega/2).$$

This implies coercivity as long as $\omega \neq 2\pi k$ for some integer k. We will deal with case $\omega = 2\pi k$ where k is a nonzero integer seperately, later on.

Equicontinuity. In order to establish the equicontinuity on the minimizing sequence we again only use the kinetic part of the action. We have that $q(s) - q(t_0) = \int_{t_0}^{s} \dot{q} dt$ because our paths q are assumed absolutely continuous and the coordinates of absolutely continuous paths have measurable derivatives and are equal to the integrals of these derivatives. ROYDEN REF. Then

$$\|q(s) - q(t_0)\| \leq \int_{t_0}^{s} \|\dot{q}(t)\| dt.$$

Using Cauchy-Schwartz again, we have

$$\int_{t_0}^{s} \|\dot{q}(t)\| dt \leq \sqrt{|s - t_0|} \int_{t_0}^{s} \|\dot{q}(t)\|^2 dt \leq \sqrt{|s - t_0|} 2A(q)$$

since $A(q) \geq \int_0^T K(\dot{q}) dt = \frac{1}{2}\int_0^T \|\dot{q}\|^2 dt \geq \frac{1}{2}\int_{t_0}^{s} \|\dot{q}\|^2 dt$. Putting these inequalities together yields

$$\|q(s) - q(t_0)\| \leq \sqrt{|s - t_0|} 2A(q)$$

(The inequality asserts that the q's are uniformly Hölder continuous of exponent 1/2.) Equicontinuity follows immediately. Given an ϵ with which we want to bound $\|q(s) - q(t_0)\|$ by, take the corresponding $\delta = \epsilon^2/C^2$ where C is any constant such that $2A(q^{(n)} \leq C$ for all n sufficiently large, so for example $C = 2.1A_*$ would work. Here we are using that we are only concerned with the tail of the sequence $q^{(n)}$ - we can throw out any finite number of inital $q^{(n)}$'s when applying Arzela-Ascoli.

Now we can invoke Arzela-Ascoli and get some subsequence, relabelled $q^{(n)}$, of our initial sequence, with $q^{(n)} \to q_*$ in the C^0 norm.

Step 3. The limiting path is collision-free. By the dominated convergence theorem $\liminf A(q^{(n)}) \geq A(q_*)$, so $\infty > A_* := \liminf A(q^{(n)}) \geq A(q_*)$ which implies that q_* is collision-free since $A(q_*) = +\infty$ for any path with collisions, by the lemma. *It is this step that is the hardest in the Newtonian* $1/r$ *case.*

Step 4. The limiting path is smooth and satisfies Newton's equations.

" Routine" - leave to appendix ??

This completes the proof of Poincaré's theorem as long as $\omega \neq 2\pi k$, k an integer. For that case we have not shown coercivity.

COERCIVITY IN THE CASE OF $\omega = 2\pi k$, k A NON-ZERO INTEGER. Look first at the problem of minimizing the length of plane curves $q : \mathbb{R} \to \mathbb{R}^2$ over all curves missing the origin and winding once around it. Without any additional constraint the infimum is 0 and is realized by a shrinking family of circles. But if we impose the constraint $\|q\| := sup_t |q(t)| = R$ with $R \neq 0$ then any minimizing sequence will tend to a line segment of length R, traversed twice, with one endpoint the origin, thus one for which $\ell(q) = 2R$. FIGURE. The Cauchy-Schwartz argument identical to that of the paragraph " $\omega \neq 0 \implies$ coercivity" in step 2 above gives

$$\frac{1}{T}\|q\|^2 = \frac{1}{T}R^2 \leq \frac{1}{2}\int_0^T |\dot{q}(s)|^2 ds.$$

Now suppose $\omega = 2\pi$ so $k = 1$ and apply this argument separately to each edge $q_{ab}(t) \in \mathbb{R}^2$ of our moving triangle and sum the inequalities with appropriate mass weights to get

$$\frac{1}{T}\|q\|^2 \leq \int_0^T K(\dot{q}(t))dt.$$

which is coercivity.

The infimum of the length among all curves in the plane which wind k times around the origin, $k \neq 0$, and have $\|q\| \geq R$ is the same: $2R$ and yields the same coercivity bound.

This completes Poincaré's theorem for all nonzero values of ω.

The proof we presented of Poincaré's theorem goes through verbatim for any coercive class if the action we are minimizing is that of a strong force potential. Hence we get:

Theorem 3.9. *For the planar N-body problem each of the following families of loops or relative loops are coercive, and hence each is realized by an action-minimizing solution for any power law potential U_α having $\alpha \geq 2$.*

- *i) any relative braid type provided we require edge* 12 *of the N-gon suffers a net rotation of $\omega \neq 0$ in one period*
- *ii) any "homologically tied" relative braid type*

We explain the classes and their coercivity. The first class (i) provides a direct generalization of Poincare's theorem from relative homology to relative homotopy. The homology of $\mathbb{C}^N \setminus \Delta$ is generated by $\binom{N}{2}$ loops, one for each edge $q_{ab} = q_a - q_b$, the loop being realized by masses a and b making one full counterclockwise revolution around each other while all the other masses are far away and unmoving. To directly generalize Poincaré, we could insist that q_{12} revolves by $2\pi\omega$ while all the other q_{ab} rotate by $2\pi\omega + 2\pi n_{ab}$ for some integers n_{ab}. The n_{ab} realize the 1st homology of the planar N-body shape space minus collisions. Then we would fix $\omega \neq 0$ and the $\binom{N}{2} - 1$ integers n_{ab}. Now there is a canonical map, the Hurwewicz map $P_N \to \mathbb{Z}^{\binom{N}{2}}$ and when we quotient by rotations it induces the map $P_N/Z \to \mathbb{Z}^{\binom{N}{2}-1}$. This map factors through the relative free homotopy classes. The coercivity of this class follows in an identical manner to Poincaré's class.

By taking $\omega = k$ a non-zero integer we get, above, winding number k for edge 12 and $k + n_{ab}$ for all the other edges. The projection of the braid onto the center is zero if and only if all OOPS ...

(ii) as per my orig paper

4.2. Interlude: Gordon's work. The key to Gordon's work is the realization that the action of a periodic solution to Kepler's problem only depends on its period.

Proposition 3.1. *Fix the period T. Then the action suffered by going once around any Keplerian solution having this period is the same, and equal to $2\pi\mu^{1/2}a^{1/2}$ where $H = -\frac{\mu}{2a}$ is the orbit's energy and $T^2 = \mu a^3$ relates the period and size a.*

ref?? Appendix? -RM

Proof. The Lagrangian and Hamiltonian are related by $L(q,\dot{q}) = p\dot{q} - H(q,p)$ provided p and $\dot{q}$ are related to each other by the Legendre transformation (A.1, **??**) The relation inducing the Legendre transformation can be understood as an equality of one-forms:

$$Ldt = pdq - Hdt$$

in an extended phase space, pulled back to solution curves. The form pdq denotes the canonical one-form on phase space (A). Integrate this relation over one period, using the fact that the energy and period are related by Kepler's third law and in particular that fixing the energy fixes the period. We get

$$\int_0^T Ldt = \int_c pdq - HT.$$

The integral of Ldt is the action. Now let c_1, c_2 be two solutions having the same energy, and hence the same period T. We can subtract the two integrals to get

$$A(c_2) - A(c_1) = \int_{c_2} pdq - \int_{c_1} pdq$$

Now the energy level set is path connected for Kepler's problem. Parameterize the two curves and connect their initial points $c_1(0), c_2(0)$ by a smooth curve $\gamma(s), 0 \le s \le 1$ which lies within the fixed energy level. Because $\omega(X_H, v) = dH(v)$ we have that $\omega(X_H, \dot{\gamma}) = 0$ But $\omega = dp \wedge dq$. Applying the Hamiltonian flow to all the points of the arc γ for one full period we sweep out an annulus whose boundary is $c_2 - c_1$. We have just shown that this annulus is ω-isotropic: the pull-back of $dp \wedge dq$ to this annulus is zero. It follows by Stoke's theorem that $\int_{c_2} pdq - \int_{c_1} pdq = 0$ or $A(c_2) = A(c_1)$.

To get the precise expression for the action evaluate it on the circular solution $q(t) = aexp(i\omega t)$. (Use the defining equation of Kepler's problem to see that $\omega^2 = \mu/a^3$.) QED

Proof of the second part of Gordon. We show any minimizing sequence tends to a collision-ejection solution. Let's first go twice around a Keplerian orbit. If the period of this orbit is P then to go twice around takes time $T = 2P$. On the other hand the action of this twice-circled orbit is $2a(P) = 2cP^{1/3}$. The action of the collision-ejection solution of period T is $cT^{1/3} = 2^{1/3}cP^{1/3}$ which is less than twice around our twice-circled orbit since $2^{1/3} < 2$. This proves that going twice around the classical solution will not minimizer the action. Why not?

Take the collision ejection solution and just back up a bit from collision and pause to concatenate with a tiny quick circle traversed twice. For example, if collision occurs when $t = 0$ then it occurs again at $t = T$. Using $r(t) = r(-t) = r(T-t)$, travel along the collision-ejection solution from $t = \epsilon/2$ to $t = T - \epsilon/2$. At that stopping time we are at a point a distance $a = c\epsilon^{2/3}$ from the origin. Now take the Keplerian cirlce of this radius. Its action, twice around is $4\pi\mu^{1/2}a^{1/2}$ which tends to zero with ϵ. Now, we've messed up the period condition but only by $O(\epsilon)$ which can be normalized back to T by a time rescaling at the risk of a vanishingly small change to the action.

Without much difficulty one can show the scenario just sketched is optimal: any minimizing sequence converges to the collision-ejection orbit. Moreover, the same is true no matter what winding number we take as long as it is not $0, 1$ or -1.

4.3. Hunting the Eight and the Hip-hop.

4.4. A workhorse : Marchal's lemma. Establishing the existence of the hip-hop and figure eight gave a proof of concept. Variational methods actually can be used to say something interesting about the gravitational N-body problem. Not long after hearing a talk on the figure eight Christian Marchall proved a lemma which became central to subsequent work.

Marchal lemma

LEMMA 3.3. *[Marchal] Consider the N-body problem in d dimensions, $d > 1$ with any power law pair potential, Newton's $\alpha = 1$ included. Let $q_0, q_1 \in \mathbb{E} = (\mathbb{R}^d)^N$ be fixed endpoints and $T > 0$ a fixed time. Then every action minimizer $q : [0,T] \to \mathbb{E}$ connecting q_0 to q_1 in time T has no interior collision points. In other words, if $q : [0,T] \to \mathbb{E}$ is such a minimizer then $q(t) \notin \Delta$ for all $t \in (0,T)$.*

REMARKS. The lemma allows either endpoint to be a collision point.

Since minimizers solve Newton's equations away from collision, the minimizers of Marchall's lemma are solutions to Newton's equations.

Marchall proved his lemma for the Newtonian case and $d = 3$. Chenciner and then Ferrrario and Terracini showed how to prove it for any d and any power law $1/r^\alpha$, $\alpha > 0$.

Before Marchall's lemma the only general facts available regarding action minimizers were that their collision locus was a closed set of measure zero and that upon restriction to any component of the complement to the collision locus the curve solved Newton's equations. These facts allow for the collision set to be a Cantor set in which case the "solution" afforded by the minimizer consisted of a countable set of solutions defined on small open intervals, concatenated to each other at elements of the Cantor set. The jump from the Cantor set all the way down to the empty set is a big one!

PROOF SKETCH.

We will give the proof under a simplifying assumption. For a full proof we recommend [**22**] or [**39**].

The proof is by contradiction. Assume our minimizer has interior collisions. Our simplifying assumption is that the collisions occuring are isolated binary collisions. Let $t_c \in (0,T)$ be one such collision time and δ_1 a time such no other binary collision arises within $(t_c - \delta_1, t_c + \delta_1)$. Shift time by t_c so as to center the collision time to $t = 0$ and so that our minimizer starts at q_0 at time $-t_c$ and ends at q_1 at time $q(T - t_c)$. For each point $s \in \mathbb{S}^2$ we will construct a perturbation $q_s(t)$ of $q(t)$ which agrees with $q(t)$ outside the critical interval $[-\delta_1, \delta_1]$ and has no collisions within the interval. We will not be able to compare the action $A(q_s)$ to $A(q)$ for any particular s to any advantage. Marchall's brilliance was to instead *average* $A(q_s)$ over the whole sphere and show that the average action decreases:

{Marchall_average}

$$\Delta A := \fint_{s \in \mathbb{S}^2} A(q_s) d\sigma(s) - A(q) < 0, \tag{86}$$

where the symbol $\fint_{s \in \mathbb{S}^2} \ldots ds$ means to take the average over the sphere with respect to the usual area measure of the sphere. It follows from inequality (86that for some $s_* \in \mathbb{S}^2$ we have that $A(q_{s_*}) < A(q)$ so that in particular , contradicting minimality.

The key to Marchall's proof of inequality (86) is potential theory. He identified the leading term in the averaged perturbed action with the potential due to a homogeneous gravitating spherical shell.

The particular family of perturbations of Marchall is simple but beautiful. Let us suppose that it is body 1 and 2 that are colliding at time $t = 0$. Let $f(t; \epsilon)$ be the

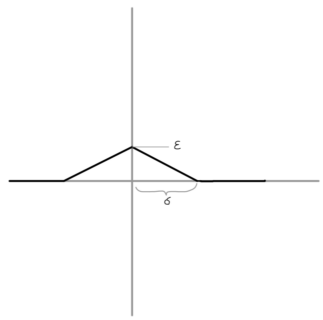


FIGURE 7. The function f defining the perturbation. fig: graph_f

continuous even non-negative piecewise linear scalar function equal to ϵ for $t = 0$ and having support on

$$I_c := [-\delta, \delta]$$

where $\delta \le \delta_1$ (A particular choice of δ will be made later, one guaranteed so that within I_c our alleged minimizer has well described asympotics.) See figure 7. Thus

$$f(t;\epsilon) = \epsilon(1 - t/\delta), 0 \le t \le \delta$$

Let μ_1, μ_2 be the usual 'reduced mass" coefficients so that $m_1\mu_1 - m_2\mu_2 = 0$ $\mu_1 + \mu_2 = 1$. Define the perturbation $q_s^\epsilon(t) = (q_{1,s}^\epsilon(t), q_{2,s}^\epsilon(t), \ldots, q_{N,s}^\epsilon(t))$ by

$$q_{1,s}^\epsilon(t) = q_1(t) + \mu_1 f(t;\epsilon)s,$$

$$q_{2,s}^\epsilon(t) = q_2 - \mu_2 f(t;\epsilon)s,$$

while

$$q_{a,s}^\epsilon(t) = q_a(t), a > 2.$$

Note that

$$m_1 q_{1,s}^\epsilon + m_2 q_{2,s}^\epsilon = m_1 q_1 + m_2 q_2$$

: the perturbation has not changed the 1-2 center of mass, nor the overall center of mass. Also note that

$$q_{1,s}^\epsilon - q_{2,s}^\epsilon = q_{12}(t) + f(t;\epsilon)s$$

where $q_{12}(t) = q_1(t) - q_2(t)$. The perturbation as destroyed the collision at $t = 0$, shifting it by the vector s.

Since q and q_s agree outside of $I_c = [-\delta, \delta]$ we have

$$A(q_s^\epsilon) - A(q) = \int_{I_c} (K(q_s(t)) - K(q(t))) + U(q_s(t)) - U(q(t))dt$$

for each s. The proof is achieved by averaging this action difference over $s \in \mathbb{S}^2$ and showing that it becomes negative for appropriate $\delta > 0$ and all ϵ sufficiently small.

To this end we will show that as $\epsilon \to 0$ the dominant term of the action difference comes from the 12 potential term:

$$\Delta A_{12} := \fint_{s \in \mathbb{S}^2} \int_{I_c} \frac{m_1 m_2}{|q_{12}(t) + f(t,\epsilon)s|} - \frac{m_1 m_2}{|q_{12}(t)|} ds$$

We estimate ΔA_{12} now. Switch the order of integration:

$$\Delta A_{12} := \int_{I_c} (\fint_{\mathbb{S}^2} \frac{m_1 m_2}{|q_{12}(t) + f(t,\epsilon)s|} - \frac{m_1 m_2}{|q_{12}(t)|} ds$$

At this point the promised potential theory enters, potential theory going back to Newton. Newton established that a uniform spherical shell of mass exerts no force inside, while outside it exerts the same force as would be exerted if all of its mass were concentrated at a point at its center. This reasoning implies that if $f > 0$ denotes a constant radius and if we define a function $W(Q)$ on $\mathbb{R}^3$ by

$$W(Q; f) = \fint_{s \in \mathbb{S}^2} \frac{ds}{|Q - fs|}$$

then

$$W(Q) = \begin{cases} \frac{1}{|Q|}, |Q| > f \\ \frac{1}{f}, |Q| \le f \end{cases} \tag{87}$$

Note also that $W(Q,f) = \fint_{s \in \mathbb{S}^2} \frac{ds}{|Q+fs|} ds$ since the integral is invariant under $s \mapsto -s$. Applying this realization regarding $W(Q)$ to the integrand inside our last expression for ΔA_{12}. Setting $Q = q_{12}(t)$, $f = f(t)$, multiplying by $m_1 m_2$ and using $|Q| = r_{12}$ we see that the two terms of this integrand cancel unless $r_{12} < f$ in which case we get the integrand to be $\frac{m_1 m_2}{f} - \frac{m_1 m_2}{r_{12}}$. Thus

$$\Delta A_{12}^{\epsilon} = m_1 m_2 \int_{t \in I_c : r_{12}(t) \le f(t;\epsilon)} (\frac{1}{f(t;\epsilon)} - \frac{1}{r_{12}(t)}) dt.$$

We are done with our use of potential theory.

Since $1/r_{12} > 1/f$ on the region of integration we have that $\Delta A_{12} < 0$. We will need the leading asymptotics in ϵ of ΔA_{12} to insure it dominates the rest of the average action, keeping it negative. To obtain this asymptotics we will need that of $r_{12}(t)$ as $t \to 0$:

{eq: r asymptotics}
$$r_{12}(t) \sim \gamma t^{2/3} + O(t), \gamma = (\frac{9}{2}(m_1 + m_2)G)^{1/3} \tag{88}$$

This estimate is valid for the separate solution arcs on either side of collision. The asymptotics (88) was supplied a century ago by Levi-Civita. See section 13 of [**81**]. We show how derive this in an exercise at the end of the chapter. **We choose δ independent of ϵ so that this asymptotics is valid on the interval $[-\delta, \delta]$ for both collision branches with fixed constant bounds**

include γ get it by either owrking it out by energy balance or look up in levi-civita and add eq ref; chenc says $\gamma = (9/2)^{1/3}$ for standard Kepler -RM

really? add some exercises? -RM

I claim that estimate (88) implies

{eq: Delta A}
$$\Delta A_{12}^{\epsilon} = -\frac{4 m_1 m_2}{\gamma^{3/2}} \sqrt{\epsilon} + O(\epsilon). \tag{89}$$

Focus on the branch leaving collision: $0 \le t$. We have $f(t)$ decreases monotonically from ϵ to 0 while r_{12} increases monotonically from 0. It follows that there is a unique first time $t_0 = t_0(\epsilon)$, close to zero such that $r_{12}(t_0) = f(t_0)$. See figure 8 We can estimate that time by replacing $r_{12}(t)$ by $\gamma t^{2/3}$ and solving $f(t) = \gamma t^{2/3}$ to

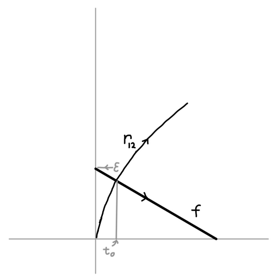


FIGURE 8. The point t_0 is where $r_{12}(t)$ first exceeds $f(t)$. fig: t0

leading order, using $f(t) = \epsilon(1 - t/\delta)$ which yields

$$t_0 = (\frac{\epsilon}{\gamma})^{3/2} + O(\epsilon^2).$$

The errors made by replacing r_{12} by $\gamma t^{2/3}$ in the integrand of ΔA_{12} change it by terms which are $O(\epsilon)$. The integral is now estimated by

$$m_1 m_2 (\int_0^{t_0(\epsilon)} \frac{dt}{f(t;\epsilon)} - \int_0^{t_0(\epsilon)} \frac{dt}{\gamma t^{2/3}}).$$

The second integral we do immediately to get $-\frac{3}{\gamma} t_0^{1/3}$. The first integral could be done but instead we estimate it by $\frac{t_0^{1/3}}{\gamma} \geq \int_0^{t_0} \frac{dt}{f(t)}$ as follows. The function f is monotone decreasing on $[0, t_0]$ so $1/f(t_0) \geq 1/f(t)$ on that interval. Thus $\frac{t_0}{f(t_0)} = \int_0^{t_0} \frac{1}{f(t_0)} dt \geq \int_0^{t_0} \frac{dt}{f(t)}$. Now use $f(t_0) = \gamma t_0^{2/3} + O(t_0)$. Summing, the two integrals we get $(1-3) m_1 m_2 \frac{t_0^{1/3}}{\gamma} +$ lower order terms $\geq \Delta A_{12}|_{t>0}$ or

$$-2 \frac{m_1 m_2}{\gamma^{3/2}} \epsilon^{1/2} \geq \Delta A_{12}|_{t>0}$$

where $\Delta A_{12}|_{t>0}$ denotes the contribution to the averaged action change obtained by integrating over for the interval $[0, t_0]$. The estimate for $\Delta A_{12}|_{t<0}$, the averaged 12 potential part of the action for the interval $[-\delta, 0]$ before collision proceeds identically and yields the same upper bound. Adding the two upper bounds yields the claimed result, inequality (89).

It remains to show that ΔA_{12} dominates the other terms in the averaged change in action. Expand out the total averaged change in action:

$$\Delta A = \Delta A_K + \Sigma_{j>2}(\Delta A_{1j} + \Delta A_{2j}) + \Delta A_{12}$$

where ΔA_K denotes the averaged change in the kinetic part of the action, and where ΔA_{ij} denote the differences in perturbed and unperturbed averaged action arising from the $1/r_{ij}$ term of the potential. There are no terms ΔA_{bc} for $b, c > 0$ since the motions of all the other bodies $a > 2$ were left unchanged by the perturbation. terms of the potential are zero. We will show that

$$\Delta A_K = O(\epsilon^2/\delta)$$

and

$$\Delta A_{1j} = O(\epsilon\delta), \Delta A_{2j} = O(\epsilon\delta).$$

(Recall δ is independent of ϵ and is chosen so that the collision asymptotics of both branches held.) These estimates show that the negativity of ΔA_{12} dominates the rest of the action as claimed, and will complete the proof.

(ΔA_K) The integrand for the difference in kinetic parts of the action is

$$\frac{1}{2}\Sigma m_a(|\dot{q}^{\epsilon}_{a,s}(t)|^2 - |\dot{q}_a(t)|^2)$$

Since we've only changed the trajectories for $a = 1, 2$ these are the only ones that contribute. Since

$$\dot{q}^{\epsilon}_{1,s} = \dot{q}_1 + \mu_1 \dot{f}(t;\epsilon)s$$

the kinetic difference term in the integrand for mass 1 is

$$\frac{1}{2}m_1(2\mu_1 \dot{f}_1 \dot{q}_1 \cdot s + \mu_1^2(\dot{f})^2)$$

Now

$$\dot{f}(t;\epsilon) = \epsilon\frac{\Theta(t;\delta)}{\delta}$$

where

$$\Theta(t,\delta) = \begin{cases} -1, 0 \le t \le \delta \\ 1, -\delta \le t < 0 \\ 0, |t| > \delta \end{cases}$$

From this expansion of the integrand it looks like the averaged kinetic is $O(\epsilon)$ right away. However the velocity $\dot{q}_1$ blows up as $t \to 0$ so we would need its asymptotics to estimate this term which is apparently linear in ϵ. The integral arising from this term is at least finite since the rate of blow up of the velocity is $t^{-1/3}$. But viewed as a function of s this first term is linear in s and linear functions on the sphere average to zero! It follows that the only term that contributes to the averaged kinetic action from mass 1 is the term $m_1\mu_1^2(\dot{f})^2$. The same argument holds for the kinetic term arising from mass 2. Adding the two, using $\dot{f}^2 = \epsilon^2/\delta^2$ on I_c and doing the integral we get :

$$\Delta A_K = \frac{\epsilon^2}{\delta}(m_1\mu_1^2 + m_2\mu_2^2)$$

exactly.

($\Delta A_{1j}, A_{2,j}, j > 2$) We use the fact that the collision is isolated. Thus, there is a positive constant C such that $r_{1j}, r_{2j} > C$ as long as $t \in I_c$ and $j > 2$. Write $q_{ij} = q_i - q_j$. One expands out $\frac{1}{r^{\epsilon}_{1j}} - \frac{1}{r_{1j}}$, using

$$(r^{\epsilon}_{1j})^2 = r^2_{1j}(1 + \frac{2}{r^2_{1j}}\mu_1 f(t,\epsilon) q_{1j} \cdot s + \frac{1}{r^2_{1j}} f(t,\epsilon)^2)$$

and the binomial expansion to find that

$$|\frac{1}{r^{\epsilon}_{1j}} - \frac{1}{r_{1j}}| \leq \frac{\epsilon}{2}\frac{\mu_1}{C^2} + O(\epsilon^2)$$

where we used that $|f| \leq \epsilon$ on I_c. An identical estimate holds for $\frac{1}{r^{\epsilon}_{2j}} - \frac{1}{r_{2j}}$. Since the integrands arising in $\Delta A_{1j}, A_{2,j}$are bounded by ϵ and their interval of integration is of size δ we get that $\Delta A_{1j}, \Delta A_{2,j} = O(\epsilon\delta)$ as claimed.

QED (for our proof of Marchal's lemma)

4.5. Isolating Collisions.

4.6. Variants of Marchal. One cannot directly apply Marchal to eliminate collisions for the minimizers obtained by applying the direct method to paths with Γ-symmetry. This is because Marchal's averaged variations are random in any direction and destroy the symmetry. Ferrario and Terracini [**39**] showed how to construct an equivariant version of Marchal's lemma provided the representation of Γ obeyed a property they called the rotating circle property. We refer to section 2.6 of [**?**] and section 9 of [**39**] for the precise statement of this property and a host of solutions to which it applies.

5. A ballet tour

We have already described the figure eight and hip-hop orbits. Here we describe a sampling of a few of the many other orbits obtained by combining the direct method with symmetry constraints.

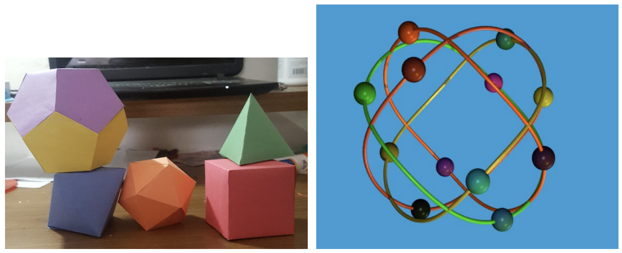


FIGURE 9. The symmetry groups of the five Platonic solid have various representations which are realized by solutions to the spatial N-body problem. by way of the direct method. For N you can take the order of the group. Courtesy of D. Ferrario.

platonics

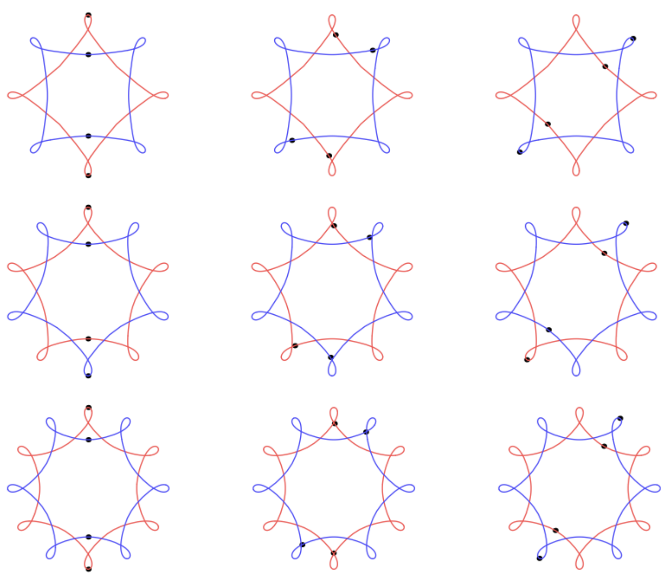


FIGURE 10. Chen established the existence of a countable family of solutions to the $2N$ body problem. Half the masses travel one choreography (blue) clockwise. The other half travel a rotate of the same choreography curve (red) counterclockwise. In the figure $2N = 4$. Courtesy of K.C. Chen.

fig: rosettes

double rosettes from C.U.P: so no permission necessary -RM

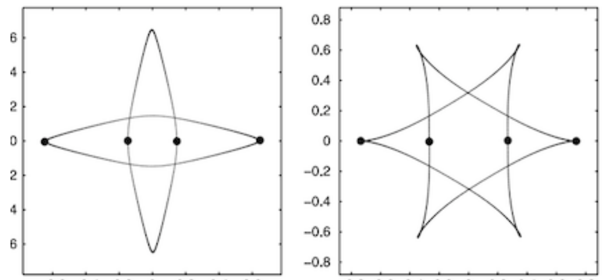


FIGURE 11. Chen Ouyang and Xie established the existence of various families of solutions by arranging points on two lines and using the angle between the lines and the arrangements of the points as inputs into the direct method. Courtesy of K.C. Chen.

fig: 4braid

picture from their Math. Res. Lett. paper -RM

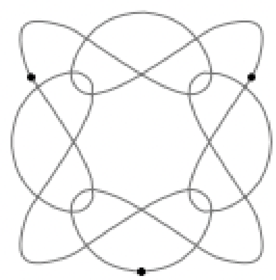


FIGURE 12. Steckles and Montaldi classified possible symmetry types of planar choreographies. All types are realized by strong force potentials. This one they call $D(3,4)$. Numerical evidence suggests it is realized gravitationally. Courtesy of J. Montaldi.

fig: D(3,4)

include constraints on Fourier coeffs from their paper ? -RM

dynamical method

6. Dynamical method.

We will use exponential notation to denote stutters. For example $1^2 2^3 = 11222$. We say that an eclipse sequence consists of stutter blocks in the range $[N_0, N]$ if it is of the form $\ldots a_0^{n_0} a_1^{n_1} \ldots a_i^{n_i}$ with $a_0, a_1, \ldots \in \{1,2,3\}$ and the integers $n_0, n_1, \ldots$ lying in the interval $[N_0, N]$.

thm:MoeckelMontgomery

THEOREM 3.10. *Consider the planar three body problem as in the previous theorem. There is a (possibly large) integer N_0 with the property that for every $N > N_0$ and every eclipse sequence, infinite or not, consisting of stutter blocks in the range $[N, N_0]$ there is a solution realizing that sequence. If the sequence is periodic then the realizing solution is relative periodic.*

PROOF OF THEOREM 3.11 FROM THEOREM 3.10 Given a relative braid type, there is a reduced periodic eclipse sequence $w = a_1 a_2 \ldots a_{2k}$ representing it. (Lemma **??**.) Replace each a_i in w with the stutter block $a_i^{n_i}$ with n_i odd and in the range $[N_0, N]$ to get a new word and repeat the new word so as to make it periodic. By theorem 3.11 this new periodic eclipse sequence is realized by a relative periodic solution. Since stutters cancel in pairs (subsection 1.1.2), this relative periodic curve, viewed in the shape sphere is freely homotopic to the relative braid type encoded by w. QED

meaning and depending of N on angular momentum? -RM

None of these guaranteed solutions are minimizers in their free homotopy classes. Indeed

LEMMA 3.4. *The action of any relative periodic solution which has a stutter can be decreased while remaining relative periodic and in the same relative free homotopy class.*

PROOF. Our proof is by "orbit surgery". Suppose q is the solution and the stutter is realized by two successive times t_0, t_1. Look at the image curve $s(t)$ in the shape sphere. The operation of reflecting the arc $s([t_0, t_1])$ about the equator while leaving the rest of $s(t)$ unchanged keeps its eclipse sequence the same. Some thought shows that this 'surgery operation' on the shape curve can be lifted. To do so, let ℓ_0 be the line through the three masses at time t_0 and ℓ_1 the line at t_1. Reflect the arc $q([t_0, t_1])$ about ℓ_0 while keeping $q([0, t_0])$ unchanged to arrive at a new arc whose shape projection on the interval $[t_0, t_1]$ is the reflection of $s([t_0, t_1])$. This new arc ends in a collinear configuration lying on the line $\tilde{\ell}_1$ obtained by applying the reflection to ℓ_1. Let g be the rotation taking ℓ_1 to $\tilde{\ell}_1$. Apply g to the final segment $q([t_1, T])$ of q and concatenate the result with the previous part of the curve. By this process we have made a continuous relative periodic curve $\tilde{q}$ with the same eclipse sequence as q and the same action.

Now, by way of contradiction, suppose q cannot be shortened in its homotopy class. Then q must satisfy the Euler Lagrange equations and in particular be analytic. But $\tilde{q}$ has the same action and represents the same homotopy class as q so it, also, cannot be shortened and so must be a solution. But our surgery process has destroyed the analyticity of $\tilde{q}$ at t_0 (and t_1). This contradicts minimality. QED

Theorem 3.10 is achieved using Moeckel's detailed knowledge of dynamics near triple collision. A cartoon of the solutions guaranteed is as folllows

write up "cartoon "? -RM

6.1. Route to a full answer? Theorem 3.11 *almost* solves the original open question 3.2 for $N = 3$. That question asks us to find, for each *unreduced* free homotopy class of loops in $\hat{\mathbb{E}}(2,3)$, a solution to Newton's three-body problem representing it. We have found solutions representing each *reduced* free homotopy class.

Free homotopy classes are in bijection with conjugacy classes in the pure braid group P_3 while reduced free homotopy classes are in bijection with conjugacy classes of the projective pure braid group, $P_3/\mathbb{Z}$. Thus, we have answered the Open Question modulo the center $\mathbb{Z}$ of the braid group. This center is generated by an overall rotation, so by travelling circuits of the Lagrange relative equilibrium. So, we would answer the original open question affirmatively if we could "tack on" any number of nearly Lagrange orbits to the dynamically realized solutions of theorem 3.11. We would also have to make sure the resulting orbits were periodic and not just relative periodic. Can this be done? We do not know. Following or extending our dynamical proof of theorem 3.10 through to this end would require keeping track of the overall rotations due to close encounters with triple collision in the solutions we constructed and this we have not done.

..

There is something unsatisfactory about the answer, at least for someone addicted to variational principles -include a finaly section here of "more approachable open questions ? – how to go from 123123 to $1^a2^a3^a1^a2^a3^a$ ... ? -RM

6.2. Danya Rose results ??

7. N=3. Almost solved.

eclipse_main

THEOREM 3.11 (Moeckel-Montgomery; 2015). *Every reduced free homotopy class of loops is realized by some periodic solution to the Newtonian three-body problem.*

This theorem almost solves the original open question 3.2 for $N = 3$. That question asks us to find, for each *unreduced* free homotopy class of loops in $\hat{\mathbb{E}}(2,3)$, a solution to Newton's three-body problem representing it. Instead we have found solutions representing each *reduced* free homotopy class. Recall that free homotopy classes are in bijection with conjugacy classes in the pure braid group P_3 while reduced free homotopy classes are in bijection with conjugacy classes of the projective pure braid group, P_3/Z. Thus, we have solved the problem modulo the center Z of the braid group. This center is generated by an overall rotation, so by travelling circuits of the Lagrange relative equilibrium. Powers of the Lagrange solution now generate all the "missing" central free homotopy classes! So why aren't we done with answering the open question? These central classes are all trivial as reduced free homotopy classes. It remains to represent all the non-central conjugacy classes in P_3. Each such class projects onto some reduced free homotopy class which we know is realized. But there are a $\mathbb{Z}$'s worth of unreduced free homotopy classes, which could be constructed one from the other by concatenating on powers of Lagrange's solution. Are all of these realized? Probably. But to follow our proof through to this end would require keeping track of overall rotations due to close encounters with triple collision in the solutions we constructed and this we have not done.

In contrast to theorem 3.4, the solutions established by theorem 3.11 all have nonzero angular momentum. The angular momentum is small, but non-zero. This "shortcoming" begs us to ask:

OPEN QUESTION 3.3. *Is every reduced free homotopy class of loops realized by some* **zero angular momentum** *relatively periodic solution to the Newtonian three-body problem?*

The proof of theorem 3.11 requires that the mass distributions m_1, m_2, m_3 lie in a large open set of mass parameter space which includes the case $m_1 = m_2 = m_3$ of equal masses. Whether or not every reduced free homotopy type can be realized when the masses lie in the complement of this set remains an open question. For example, does the theorem hold in the regime of a dominant primary mass, say $m_1 >> m_2, m_3$?

We proved the theorem 3.11 using dynamical methods pioneered by Moeckel, rather than variational methods. Moeckel's methods require analysis of the flow near the McGehee blown-up triple collision manifold REF CC Chapter and as result the solutions have repeated close approach to triple collision. See figure.

The eclipse sequences which the periodic solutions of theorem 3.11 generate may have many many stutters. Define a stutter block of size k to mean one of the three eclipse sequences of the form a^k with $a = 1, 2$ or 3 occurring exactly k times in a row. We say that an eclipse sequence has stutter range $[K, K']$ if it is a concatenation of stutter blocks of size k with $K \le k \le K'$. Then, for example, $1^3 2^8 3^3$ has stutter range $[3, 8]$ but does not have stutter range $[3, 7]$. Note though that $1^3 1^8$ is considered as having stutter range $[3, 8]$, since we do not forbid stutters of stutter blocks. Theorem 3.11 was proved as a corollary of:

eclipse_w_stutters

THEOREM 3.12. *Let $\mu = |J|^2|E|$ be the Dziobek constant. Then there is an open neighborhood U of the equal mass diagonal (m, m, m) in the three-dimensional mass parameter space with the following property. For the Newtonian 3-body problem with masses in U there exist integers K, K' with $K' = K'(\mu) > K$ such that every bi-infinite eclipse sequence in the stutter range $[K, K']$ is realized by a solution. These solutions come very close to triple collision when transitioning from one stutter block to the next. If the eclipse sequence is periodic then so is the solution. Moreover $K'(\mu) \to \infty$ as $\mu \to 0$.*

PROOF OF THEOREM 3.11 FROM THEOREM 3.12. Take masses in U. Let a reduced eclipse sequence be given, say 1213 for the sake of example. Choose an odd integer m in the range $[K, K']$ By theorem 3.12 $1^m 2^m 1^m 3^m$ (repeated periodically) is realized by a periodic solution. Cancel all the stutters until none are left to see that homotopically the loop generated by this solution is the one encoded by eclipse sequence 1213 (repeated). QED

Note that if we could take $K = 1$ in theorem 3.12 then we would have an affirmative answer to open question 3.2.

Non uniquenesses of representing solutions ? - many! many! - many eclipse sequences rep the same free homotopy class -RM

8. More questions.

1. Does the zero angular momentum three-body problem admit a solution which is contractible when viewed as a loop in collision-free configuration space? In other words, can one realize the trivial free homotopy class by a zero angular momentum periodic solution?

2. Is there a free homotopy class which is realized but for which every realizing solution is linearly unstable?

3. Moeckel and I showed that every free homotopy class is realized by a periodic orbit if we allow ourselves a bit of angular momentum and a large number of stutters. For some of classes, for example that of the eight, both the angular momentum and the number of stutters can be brought down to zero. Can you find a class for which the number of stutters in any realizing solution is not zero? Is greater than 10? Greater than 100?

4. Throughout this section we restricted ourselves to the planar problem, focusing on realizing a given free homotopy type, which is to say, conjugacy class in $\pi_1(\hat{\mathbb{E}}(2, N))$. For $d > 2$ the fundamental group is zero and its algebraic replacement is either π_{d-1} or H_{d-1}. By playing some variational or min-max game for the N body problem in d-dimensions, $d > 2$ using these higher classes, can you find any new or interesting classes of orbits?

5. For fixed negative energy are there an infinite number of choreographies modulo rotation for the equal mass-three body problem? Do they represent an infinite number of distinct eclipse sequences? If we also fix angular momentum does this number remain infinite?

9. Notes.

Wu-yi Hsiang introduced me to this problem for $N = 3$ in about 1997. He was insistent about using the Jacobi-Maupertuis metric approach to solve it. So far, that method has not panned out. But see REFS of his.

The variational method in mechanics has a long history, and is central to Feynman's approach to quantum mechanics. The Poincaré work described here is the first result we know of applying the direct method of the calculus of variations to an N-body problem. There are many works from Italian schools of analysis applying the direct method to various potential force problems and predating Venturelli's work on the hip-hop and our work on the figure eight. Often these start with a list of assumptions on the potential. These assumptions often exclude Newton's N-body problem. When they include it the "solutions" they yield are generalized. They allow for a collision set which could be as big as the Cantor set - some closed set of measure zero. Outside the collision set the curves thus obtained solve Newton's N-body problem. The various solution branches on the intervals forming the complement of the Cantor set are concatenated continuously and one does not know much more than that.

I was proud of theorem XXX when I proved it. I did not know about the earlier works of Poincaré or of Moore. Poincaré had come to grips with the difference between homology and homotopy before 1896. It is likely that he knew the result and just found it easier to write out his short paper the way he did, using the older more familiar language of winding numbers instead of homotopy. It seems certain that if told the result he could have proved it effortlessly. My ignorance aided my morale. Knowing how far others had already gone along lines I thought were new and was just starting off on could have demoralized me to the point of quiting the N-body world.

GORDON: In the 1970s W. B. Gordon re-discovered Poincaré's basic observation without ever discovering his two page paper. He placed it into a wider setting using something like a functional analyst's eye and began the first steps of extending it to the Newtonian problem.

Hip hop vs eight: , discoveries both occurred in 1999 and were published in 2000. "In the air"

Our early work was a kind of "by hands" getting rid of collisions. You can read a bit of the history around the eight, and its interaction with Moore's work in our paper REF and on my web site. Marchall's lemma provided the big breakthrough to let loose the flood of proofs establishing families of choreographic solutions.

Marchal. I met Christian Marchal in December of 1999 at the Don Saari birthday/ Northwestern retirement party. He gave an astounding talk that had nothing to do with celestial mechanics. He talked about how the U.N. had all their models wrong for human population growth. They universally over-predicted world population. Data showed that birth rates were actually declining precipitously around the world, in contradiction to all models. (He justified giving this talk in a celestial mechanics conference by relating, in a single slide, to the mis-guided epicycles of Ptolemy before Kepler.) For me, this was the best news I had heard in a decade! I was brought up in Northern California in the 70s and had an oceanography teacher, Mr. Sikora, in Junior High for whom the up-coming Population Bomb, predicted by the Ehrlich couple, was the worst disaster the earth was facing. For Christian, this birth rate data was horrible news! Who was going to pay for his retirement? He himself had seven children. What would Europe be coming to? Later, I was blessed with a few dinners with he and his wife. He lived in the apartment building he had grown up in, and could remember the bombs dropping and Nazi soldiers occupying Paris.

continue in this vein? stories on Simo, Chenciner, Moeckel, .. Gole? Hsiang ?? -RM

The story of how a comment by Carles Simo led to the proof of theorem REF by Moeckel and I is told in Ref and REF .

Danya Rose's thesis REF provides a large data set to explore for the zero-angular momentum equal mass three body problem. It contains information which can be translated to the syzygy type and XXX

****

TO NOTES ?? :

Notes on the proof set-up. One could use the monodromy relation to replace the domain $[0, T]$ of $C^0([0, T], \mathbb{E})$ with a circle, at the expense of changing the set up for the action by going into a frame rotating with constant frequence ω/T.

****

better change to a circle from $[0, T]$ -RM

However, those solutions found by variational means are often quite pleasing to the eye and enjoy remarkable properties. This branch of investigation has generated a large body of research over the last two decades.

Among the most well known of these variational orbits is the figure eight. See figure 3.3 in the 'Tour' of chapter 1. The eight has eclipse sequence 123123. .. PUT IN TOUR −¿ was discovered by Cris Moore using variational methods, and rediscovered by Chenciner and myself. We gave a proof of the eight's existence by combining the direct method with two other ingredients, additional symmetries which arise when all three masses are equal, and the geometry of the shape sphere. Our proof was explicity enough that Carles Simo could use it as a seed to find initial conditons yielding the eight.

state theorem w minimezer in the D_6-equivariance class ; hexagon...

This discovery led to a host of new solutions found by similar methods for the equal mass N body problems. We describe some of these in a later section.

****

There are two sources of non-compactness in the N-body problem: $r_{ab} \to 0$ and $r_{ab} \to \infty$.

Since Poincaré was essentially the inventer of the first homotopy group he certainly could have gotten the homotopy version of this theorem which we will describe momentarily.

To the best of our knowledge, Poincaré was the first one to get this method to work for three bodies. There is a large body of work by the "Italian school" in the 1970s-1990s but with little concrete to show for it. Cris Moore took this approach in a beautiful five page paper "Braids in Classical Dynamics" written in 1993 and implemented numerically to good effect.

****

However, as with many problems in celestial mechanics, someone had scooped Cris Moore's key idea by many decades.

******

Jacobi and Maupertuis showed how to reframe the N-body problem as a geodesic problem on a manifold that is Riemannian but not compact or even complete. Due to incompleteness and other problems the direct method seems to be quite hard to get to work for finding periodic orbits. In place of extremizing the Jacobi-Maupertuis arclength we can, instead, apply the direct method the action.

******

10. Exercises

EXERCISE 3.2. *Here we derive equation (eq: r asymptotics).*

Suppose that for some graviational N-body solution that q_1 *and* q_2 *engage in an isolated collision at time* $t = 0$*, so their mutual distance* $r_{12} \to 0$ *while all the other* r_{ab} *remain bounded away from zero. Write*

$$r(t) = r_{12}(t)$$

and H *for the total energy of the solution. The derivation rests on the fact that* $lim_{t\to 0} r(t)H = 0$.

a) Show that the motions of all the other bodies $q_a(t), a > 2$ *are* C^1 *near* $t = 0$ *so their contributions to the total kinetic energy is bounded.*

b) Show that the center of mass motion of q_1 *and* q_2 *is* C^1 *near* $t = 0$.

c) From (a) and (b) we have that $H = K - U = K_{12} - \frac{m_1 m_2}{r} +$ *bounded where* K_{12} *is the kinetic energy of the relative* $q_1 - q_2$ *part of 12 system. We have* $K_{12} = \frac{1}{2}\frac{m_1 m_2}{m_1+m_2}(\dot{r}^2 + r^2\dot{\theta}^2)$. *Use* $rH \to 0$ *to conclude that*

$$\frac{1}{2}\frac{m_1 m_2}{m_1 + m_2}\dot{r}^2 - m_1 m_2 = 0$$

are bounded.

EXERCISE 3.3. *Make the ansatz* $q(t) = At^b$ *for the one-dimensional Kepler's equation. Show that* $b = 2/3$ *yields a solution. Compute the action to collision for such a solution and see that it is finite over any finite interval* $[0, c]$ *containing the collision time* $t = 0$.

Any finite group sits in S_N for some N. Take the Monster group for example. It is represented on $[N]$ where $N = ...$ is the order of the Monster group. We will have to take the trivial representation on $\mathbb{T}$ and $\mathbb{R}^d$ since the Monster is simple. Is there an N-body solution realizing this Monster symmetry? Would anyone besides me care?

CHAPTER 4

Does a Scattered Beam have a Dense Image?

chapter: scattering
open:scattering

Open Question 4.1. *Is the image of a beam of N-body solutions under scattering open and dense in the sphere of outgoing directions in configuration space?*

This question concerns the positive energy N-body problem. There are no periodic or bounded orbits so that the problem concerns asymptotic behaviours. A beam is a family of solutions close to parallel to each other in the distant past. Their velocities limit to the same value as $t \to -\infty$. Computing their limiting velocities as $t \to +\infty$ defines the scattering map, a map to the sphere of outgoing directions. The question asks if the image of this map is open and dense. It will take effort to make all these notions precise. Before we take that effort we provide the classic motivation behind the question.

sec: Rutherford

1. Motivation. Rutherford and the discovery of nuclei

The discovery of the nucleus depended on Rutherford's analysis [**133**] of classical two-body scattering which includes defining the scattering map for the two-body problem. See figure 1. Rutherford, Geiger and Marsden had been aiming beams of alpha particles (Helium nuclei) at gold foil and measuring the directions of the

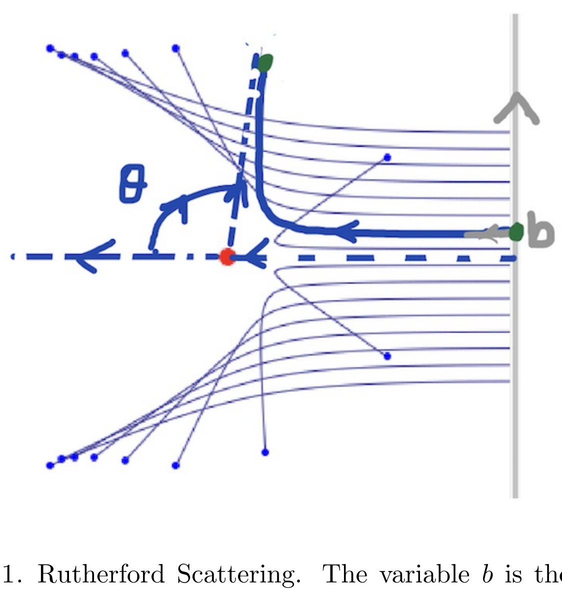


Figure 1. Rutherford Scattering. The variable b is the impact parameter and θ coordinatizes the direction of outgoing rays.

fig:Rutherford

outgoing particles scattered by the foil. They were surprised to find a measurable fraction (1/20,000) suffered nearly complete recoil. To explain their results, Rutherford posited replacing Thompson's 'plum pudding' model of the atom by one in which nearly all of its mass is concentrated in a point-like positive charge at the center (the nucleus) this center being surrounded by a diffuse cloud of very light negative charges (electrons). He assumed the alpha particles to be also positively charged and that in order to suffer their strong deviations they must have penetrated through the electron cloud so as to be strongly interacting with the repulsive Coulomb force of the gold nuclei. Ignoring the electrons, Rutherford was faced with a well known problem, the Coulomb problem which is just the Kepler problem (equation (28)) with the sign of the Kepler constant μ reversed so as to yield a repulsive instead of an attractive center.

Figure 1 indicates a beam of parallel lines coming in from infinity and, interacting with the Coulomb force at the origin, getting deflected, and retreating to infinity in a splay of directions. We parameterize the rays of the incoming beam by a real parameter b, a coordinate on a line orthogonal to the beam's direction. We take $b = 0$ to correspond to the ray heading dead straight in to the origin. This b is called the "impact parameter" as it measures how close the corresponding trajectory would come to impact with the nucleus if we were to turn off the Coulomb force. We label the outgoing trajectories by the angle θ made by their asymptotic direction and the original incoming beam direction. Coulomb dynamics then defines the scattering map $b \mapsto \theta = f(b)$. Rutherford computed

{Rutherford scattering}

$$\theta = f(b) = 2arctan(\frac{-\mu}{2Eb}) \tag{90}$$

Here E is the energy of the incoming beam and μ is the 'Kepler-Coulomb -constant' of $\ddot{q} = -\mu q/|q|^3$, so that $\mu < 0$ for Coulomb and $\mu > 0$ for Kepler. See [**70**], particularly p 286, eq (12.3.3), or [**76**] for this computation. *It is this map, extended to and made sense of in the context of the N-body problem, that the open question refers to.*

We pause to re-iterate that for the case of a Keplerian attractive center - the expression (90) for the scattering map continues to hold. See figure 2 for a picture of the attractive scattering map. It is remarkable that the differential cross sections REF ?? XXX for the attractive and repulsive cases are identical, with μ entering only through its square.

The map of equation (90) is a smooth invertible map

$$f : \mathbb{R} \setminus \{0\} \to S^1 \setminus \{0, \pi\}$$

In particular its image is open and dense,missing exactly two angles, $\theta = 0$ and $\theta = \pi$ and so, in the two-body case the answer to our question is 'yes'. Somewhat remarkably, the Coulomb scattering map extends smoothly to the closure $\mathbb{R} \cup \{\infty\} = \mathbb{RP}^1 = S^1$ of its domain. The extension sends $b = 0$ to $\theta = \pi$ corresponding to the ray reflecting off the nucleus and it sends $b = \infty$ to $\theta = 0$ corresponding to no deflection at all for infinitely distant trajectories. We will see below that this extension is really the one-dimensional version of stereographic projection.

Rutherford and crew built their experiment in three-space, not in the plane. By rotational symmetry, Coulomb scattering in three dimensions can be obtained from two-dimensional Coulomb scattering by spinning the picture (figure 2) for two-dimensional scattering about the beam axis. Take the beam axis to be $-e_3 =$

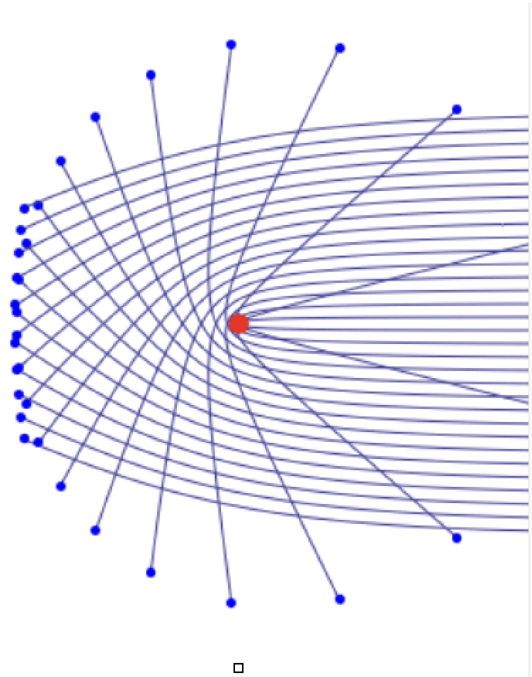


FIGURE 2. Kepler Scattering.

fig:RutherfordAttractive

$-(0,0,1)$. Then the incoming beam is parameterized by the xy plane and outgoing directions are parameterized by the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ so the relevant map is a map $\mathbb{R}^2 \to \mathbb{S}^2$. Use polar coordinates (b,θ) on xy plane and spherical coordinates (ϕ,θ) on the sphere. Fixing θ defines the plane of motion for the trajectory of the beam and this does not change during the motion. Then, by the rotational symmetry, the 3-dimensional scattering map is $(b,\theta) \mapsto (f(b),\theta)$. But $(r,\theta) \mapsto (2arctan(1/r),\theta)$ is the expression for the inverse to stereographic projection, $\mathbb{R}^2 \to \mathbb{S}^2$. Upon setting $r = \pm 2Eb/\mu$ we find that Coulomb scattering and stereographic projection are the same map! See Figure 3. It is remarkable that the scattering map analytically extends so as to yield a diffeomorphism $\mathbb{R}^2 \cup \{\infty\} \to \mathbb{S}^2$ covering the entire sphere of outgoing directions. This fact is one reason to suspect the answer to the Open Question is 'yes'.

1.1. Scattering Cross-section. Scattering experiments are inherently probabalistic. One does not have a single nucleus, but rather an enormous number of gold nuclei whose precise location are unknown and essentially unknowable. One thinks of the result of the experiment as a sampling of impact parameters.

What is taken to be measured, or approximated, is the "differential cross-section" which is the push-forward of Lebesgue measure db by the scattering map. Instead Rutherford was fix

scattering cross-section for Rutherford here ...? from Knauf.. -RM

2. Scattering in the N-body problem.

Classical scattering is based on the idea that if the forces between bodies decay with distance and the bodies are far apart and receding from each other then each

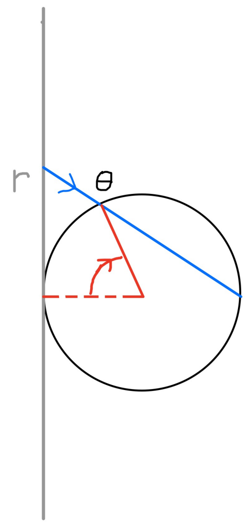


FIGURE 3. Stereographic Projection stereoProj

body ought to move asymptotically along a straight line. N straight lines in $\mathbb{R}^d$ yield a single line in the N-body configuration space $\mathbb{E}$, and conversely, such a single line defines N individual particle straight lines. A line $\ell(t)$ in $\mathbb{E}$ can be written as

{eq:lines} (91) $$\ell(t) = At + B, \qquad A, B \in \mathbb{E}, \qquad \|A\| = 1, \qquad B \perp A.$$

A represents the direction of the line and B is its impact parameter, the unique point on the line closest to the origin.

A is a point in the unit sphere $\mathbb{S}^{M-1} \subset \mathbb{E}$ of dimension $M-1$ where $M = dN$. Together $(A, B) \in T\mathbb{S}^{M-1}$, the tangent bundle of the sphere, so that we identify the space of lines in $\mathbb{E}$ with this tangent bundle. We may want to center the solutions by going to the center of mass frame in which case we replace $\mathbb{E}$ by $\mathbb{E}_0$, the centered configuration space and this dimension M becomes $M = d(N-1)$.

As a zeroth approximation, the scattering map which we will eventually define is a map on the space of oriented lines. It sends an incoming line representing the asymptotics of a solution at $t = -\infty$ to the corresponding outgoing line representing the same solution at $t = +\infty$. Identify $T\mathbb{S}^{M-1}$ with $T^*\mathbb{S}^{M-1}$ using the mass metric, so as to bring in ideas from symplectic geometry. The scattering map should be a symplectic map of the form

{full scattering} (92) $$SC : T^*\mathbb{S}^{M-1} \dashrightarrow T^*\mathbb{S}^{M-1}.$$

defined by the dynamics. The broken arrow notation for the map indicates that its domain will not be all of $T^*\mathbb{S}^{M-1}$ due to the incompleteness of the flow. Some incoming solutions will end in collision (or a non-collision singularity) and cannot be continued into the future. We will call the eventual map inspired by equation (92) "the full scattering map". The scattering map of the Open Question is built from the full scattering map by restricting this map to a single fiber representing a single incoming beam direction, and then projecting the resulting image to the base sphere of directions.

A *beam* is the space of all oriented lines having a fixed direction A in equation (91) so that a beam is identified with a single fiber $T^*_A\mathbb{S}^{M-1} = A^\perp$. The beam is parameterized by fixing A and varying B in equation 91. The map whose image is the subject of the Open Question 4.1 is the composition

add more complicated for N body see sec q CHAZY for details. -RM

{scat_map}

$$T^*_A\mathbb{S}^{M-1} \xrightarrow{i_A} T^*\mathbb{S}^{M-1} \overset{SC}{\dashrightarrow} T^*\mathbb{S}^{M-1} \xrightarrow{\pi} \mathbb{S}^{M-1} \tag{93}$$

where i_A is the inclusion $T^*_A\mathbb{S}^{M-1} \to T^*\mathbb{S}^{M-1}$ and π is the projection. This map takes all of the solutions which fill out a fixed incoming beam, integrates the flow along them from time $t = -\infty$ to $t = +\infty$ and then projects out the final limiting direction for each resulting outgoing solution. We call this composition the *beam-restricted scattering map* when we need to distinguish it from the full scattering map (92) defined earlier. The beam-restricted scattering map is a map $\mathbb{R}^{M-1} \dashrightarrow \mathbb{S}^{M-1}$. The open question asks if its image, like that of Coulomb scattering's image, is open and dense.

What makes this zeroth approximation inaccurate is that N-body solutions are not, in fact, asymptotic to lines in any traditional sense. Instead there is a $\log(|t|)$ occuring in their asymptotics (equation (94) below) and this makes rigorously defining the map more complicated than that of the strategy just outlined.

THE COULOMB CASE. We return to Rutherford's scattering, that is the case of Coulomb or Kepler scattering in the plane, figure 1, to illustrate the preceding formalism. Then we have $M = 2$ so that

$$SC : T^*\mathbb{S}^1 \to T^*\mathbb{S}^1.$$

$T^*\mathbb{S}^1 = \mathbb{S}^1 \times \mathbb{R}$ is a cylinder parameterized by $(\theta, b) \in \mathbb{S}^1 \times \mathbb{R}$. The angle $\theta \in S^1$ describes the direction of the oriented line and the "impact parameter" $b \in \mathbb{R}$ is as marked in the figure and measures the (signed) distance of the line to the origin when the force is turned off. We have $A = (\cos\theta, \sin\theta)$ and $B = bA^\perp = b(-sin\theta, \cos\theta)$ in equation (**??**). The impact parameter equals the angular momentum times a constant depending on mass and energy and hence is unchanged in scattering. Using its constancy and rotational symmetry it follows that SC must have the form $SC : (\theta, b) \mapsto (\theta + f(b), b)$ for some function $f(b)$. Scattering for any decaying central force problem will take this same form. The function $f(b)$ for Kepler is the one given above in equation (90).

subsec: scattered image

2.1. Defining the image of the scattering map. For scattering to make sense one needs a family of unbounded solutions, hence our requirement that the energy be positive in the Open Question. See lemma 0.4. Despite being unbounded, positive energy N-body solutions are **not** asymptotic to lines in the standard sense.

Equation (94) below describes their actual asymptotics. The presence of the unbounded $log(|t|)$ term in that equation, combined with the fact that typically A and $\nabla U(A)$ are linearly independent, implies that the solution asymptotically diverges from any single line. This lack of a standard asymptotics means that the above description of N-body scattering, although morally correct, is not accurate. For this reason it will be easier to define the *image* of the beam-restricted scattering map directly rather than to attempt to define the full scattering map SC. Later on we will define this full scattering map.

section label? -RM

DEFINITION 4.1. *A solution $q(t)$ is called* **backwards hyperbolic** *if it has positive energy and if the limit of $s(t) = q(t)/\|q(t)\|$ as $t \to -\infty$ exists and is collision-free. Let $s_- \in \mathbb{S}^{M-1} \setminus \Delta$ denote this limit and refer to it as the backward asymptotic shape of $q(t)$.*

The reader will be able to formulate the analogous definitions of **forward hyperbolic** and forward asymptotic shape, s_+.

DEFINITION 4.2. *A solution is* **hyperbolic** *if it is both forwards and backwards hyperbolic.*

A hyperbolic solution $q(t)$ has backward and forward asymptotic shapes, s_- and s_+. We say that the solution connects s_- to s_+.

def:beam

DEFINITION 4.3. *The* **incoming beam** *with direction s_- and energy E is the set of all backwards hyperbolic solutions having fixed backward asymptotic shape $s_- \in \mathbb{S}^{M-1} \setminus \Delta$ and fixed energy $E > 0$*

f: scattered image beam

DEFINITION 4.4. *The scattered image of the incoming beam with direction s_- is the set of all outgoing directions s_+'s connected to hyperbolic solutions lying in this beam.*

With these definitions in place, the terms appearing in the Open Question are now defined.

scattering known results

3. What's Known?

1. The scattered image has nonempty interior. See [**30**] for the proof. The open region guaranteed is swept out by solutions for which the distance between all bodies remains very large for all time: that is they 'stay near infinity" for all time. The proof relies on analysis at infinity and a perturbation argument.

2. On their way from $t = -\infty$ to $t = +\infty$ the trajectories of a beam will hit every point $q_0 \in \mathbb{E}$ of configuration space, including all collision points. This result is established in [**84**] where the authors prove it thorugh a deep analysis of the associated Hamilton-Jacobi equations using weak KAM ideas. Their rays $q(-\infty, 0]$ are global JM minimizers connecting the asymptotic beam direction at $t = -\infty$ to the desired $q_0 \in \mathbb{E}$.

3. Rays in the beam which encounter collisions cannot be continued to infinity. But by item 2 just above, we are guaranteed some beams hit collision. Consequently the beam-restricted scattering map *never* has domain all of $T^*_{s_-}\mathbb{S}^{M-1}$.

4. For power law potentials $U = 1/r^\alpha$ with $0 < \alpha < 1$ the scattering map in **not** onto for the two-body problem. Indeed, in the planar version of the problem, in

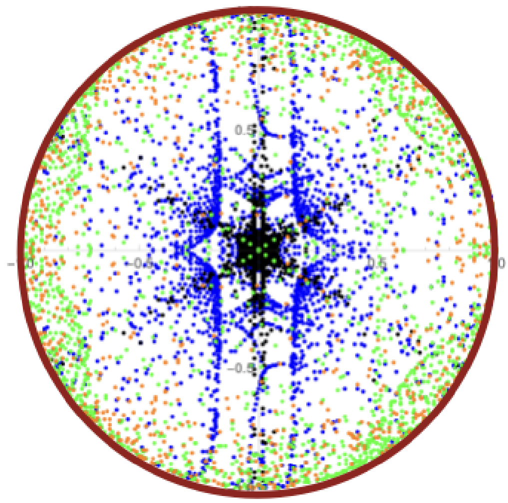


FIGURE 4. lThree-body scattering Scattering. fig:LagrangeScattering1

center-of-mass coordinates, the image of the map forms a circular arc of arclength $2\pi\frac{\alpha}{2-\alpha} < 2\pi$ centered on the beam direction. See figure 5 and 6. On the other hand, when $1 < \alpha < 2$ we find that the map is onto, with some outgoing directions covered more than once. As $\alpha \to 2$ the scattered image of the beam covers the circle more and more times, tending to cover each outgoing direction infinitely many times as $\alpha \to 2$.

These results can be found in the paper [**72**] where a surprising application of them is made. We will prove these results in Appendix E. Figure 5 describes the scattering of a half-beam for $\alpha = 1/2$, giving a sense of why the result holds and the mechanism behind it.

5. Numerical experiments done by Moeckel for an incoming Lagrange beam indicate that the scattering map for such a beam is onto, in keeping with a 'yes' answer to the open question. See figure 4.

As we defined it, for a centered beam the scattered image is a subset of $\mathbb{S}^3 \subset \mathbb{E}_0(2,3)$. What is depicted in the figure is the further projection of this image in $\mathbb{S}^3$ to the shape sphere $\mathbb{S}^2$, and from there, onto the shape disc $\mathbb{D}^2$. The shape disc is the quotient of the shape sphere by reflections so its points represent (unoriented) similarity classes of triangles. The reflection about any line in the plane generates $\mathbb{Z}_2 = O(2)/SO(2)$ which acts on shape space by $(w_1, w_2, w_3) \mapsto (w_1, w_2, -w_3)$. It follows that we can identify the space of congruence classes of triangles with the closed upper half space, $w_3 \geq 0$, and the space of similarity classes of triangles with the closed upper hemisphere of the shape sphere. Project out w_3 to get a convenient picture of the space of similarity classes of triangles as the *shape disc* : the unit disc in the $w_1 - w_2$ plane. The boundary of the disc represents collinear triangles and has the same 6 marked points as before, alternating binary and Euler points. In the disc's interior we have a single marked point representing the equilateral triangle.

When the masses are equal this point is at the center and the three isosceles circles become three equispaced diameters.

how to make gray scale version of ?? -RM

How was the figure made? Moeckel took a bunch of initial conditions lying in the unstable manifold of the Lagrange shape s (see the end of subsection 4.3 below) at zero angular momentum – so a Lagrange beam – and numerically evolved them forward into the distant future and plotted their final shapes on the shape disc to get the picture.

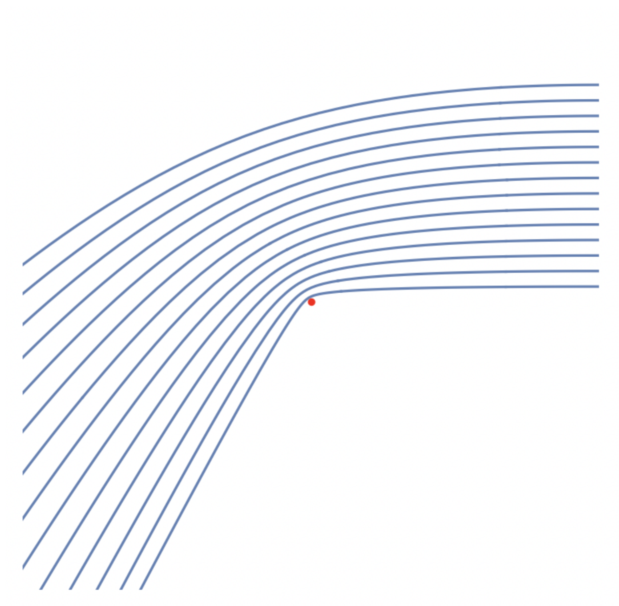


FIGURE 5. Scattering a half beam with power law $r^{-1/2}$

fig: twothirds_scattere

4. Building the Scattering Map.

We proceed to construct the scattering map. We begin by recalling Chazy's 1922 asymptotics for hyperbolic solutions. His asymptotics give us a definition for the full scattering map, but with a gap. That gap is filled by theorem 4.1. We prove the theorem by using a McGehee blow-up at infinity which adds a manifold at infinity together with equilibrium points at infinity which provide hyperbolic solutions a place to go and a a place to come from. Hyperbolic solutions then connect unstable equilibria at infinity to stable equilibria at infinity. The cotangent fibers $T^*_s\mathbb{S}$ arise as projectivizations of the tangent space to the unstable or stable manifolds at these equilibria. This blow-up perspective on scattering was essential in establishing that its image has non-empty interior.

4.1. Chazy asymptotics. Chazy [**18**] established the asymptotic expansion

{Chazy1]

$$q_+(t) = At - (\nabla U(A))log(t) + B + O(\frac{\log t}{t}), t \to \infty \tag{94}$$

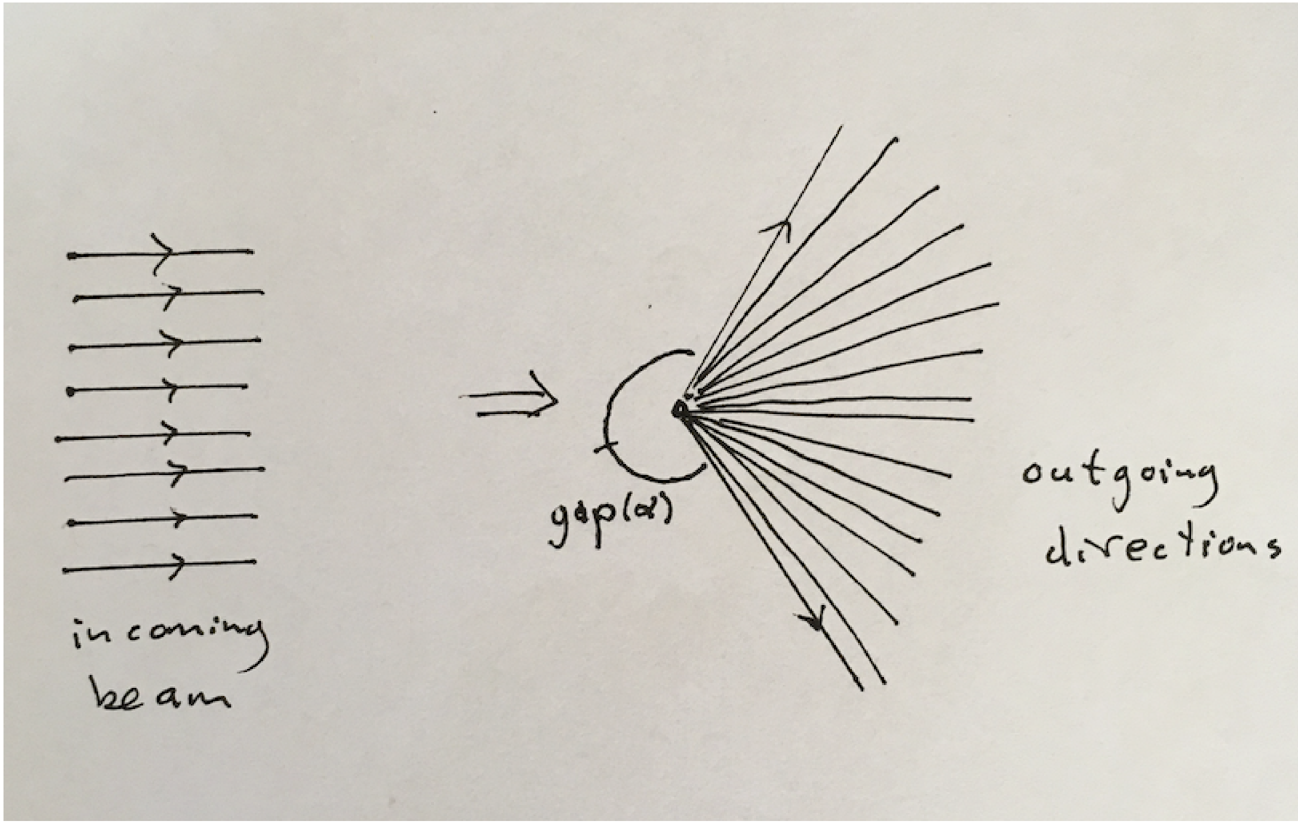


FIGURE 6. Gap in the scattering image for $\alpha = 1/2$ `fig: beam gap`

valid for any forward hyperbolic solutions $q_+(t)$. The dominant linear coefficient is related to the solution's energy $E > 0$ and asymptotic shape s_+ by

$$A := A_+ = \sqrt{2E}s_+.$$

The secondary $\log t$ term is needed for the expansion to solve Newton's equations to leading order $1/t^2$. The coefficient of this term is $\nabla U(A) = \mathbb{M}^{-1}F(A)$. Compare equation (35). There is no reasonable sense in which our solution is asymptotic to a line since its distance from any fixed line as $t \to \infty$ is typically unbounded. [1] Despite this lack of an asymptotic line we simply drop the $\log(t)$ term and let

$$B = B_+$$

play the role of the impact parameter of a line and so declare that the line associated to q_+ is the one whose parameters are $a = A_+, b = B_+(mod A_+)$ in equation (91). (Now this line has speed $\sqrt{2E}$ instead of speed 1. No matter.) *We call* (A, B) *the Chazy parameters of* $q_+(t)$. .

[1] If A and $\nabla U(A)$ are linearly independent, then q_+ and the line with parameters $a = A_+, b = B$ are asymptotic as unparameterized curves But this special case does not help us develop a general scattering map.

Similarly, if the solution $q = q_-(t)$ is backward hyperbolic with energy $E > 0$ and asymptotic shape s_- then Chazy's asymptotics in the backward time direction are

{Chazy2}
$$q_-(t) = A_- t - (\nabla U(A_-))\log(|t|) + B_- + O(\frac{\log|t|}{t}), t \to -\infty \tag{95}$$

with

{ChazyE}
$$A_- = -\sqrt{2E}s_-. \tag{96}$$

To justify the choice of minus sign here, note that equation (94) implies that

$$\lim_{t\to+\infty} \dot{q}_+(t) = A_+.$$

The minus sign relating A_- to s_- guarantees that

$$\lim_{t\to-\infty} \dot{q}_-(t) = A_-.$$

We call (A_-, B_-) the Chazy parameters of q_-.

If $q(t)$ is hyperbolic then both its forwards and backwards ends have Chazy parameters. The full scattering map takes one to the other:

{full scattering Chazy}
$$SC : (A_-, B_-^\perp) \mapsto (A_+, B_+^\perp) \tag{97}$$

where the $\perp$ is orthogonal projection to the respective normal space to the A's so that $B_\pm^\perp = B_\pm - \langle s_\pm, B_\pm \rangle s_\pm$.

4.2. The scattering map and Systems of Rays. In order for our formulation (97) of the scattering map to be well-defined we must know that a backward hyperbolic solution exists for any given $(A_-, B_-^\perp)$ and that this solution is unique modulo time translations. Once we know this, the domain of the scattering map is the set of (A_-, B_-) for which its solution extends indefinitely into the future, and the map is defined by taking this extension and working out its future asymptotics. The following theorem establishes the needed existence and uniqueness.

Chazy_uniqueness

THEOREM 4.1. *(I) Two trajectories which have the same Chazy parameters* (A_-, B_-) *are the same up to a time translation.*

Recall relation 96.

(II). The incoming beam with direction s_- *and energy* E *modulo time translation forms an affine space parameterized by the Chazy impact parameter* $B_- \in s_-^\perp$*. Any two trajectories in the beam remain a bounded distance apart as* $t \to -\infty$*.*

This theorem says that we can think of the Chazy parameters as initial conditions at infinity.

PROOF OF (II). Take two solutions $q, \tilde{q}$ in this incoming beam and subtract them. Because they have the same leading Chazy parameter $A = -\sqrt{2E}s_-$ their difference has asymptotic expansion $B_- - \tilde{B}_- + O(\log|t|/|t|)$ whose limit as $t \to -\infty$ is $B_- - \tilde{B}_-$, showing that the two trajectories remain a bounded distance apart and that we can use the impact parameter B_- as an affine parameter.

The affine parameter B_- is defined modulo A_-. To see this, consider a solution $q(t)$ in the the incoming beam and its time translate $q(t - t_0)$ which is also in the incoming beam and has the same impact parameter B_-. We have $q(t) - q(t - t_0) = At_0 - \nabla U(A)\log(|t|) + \nabla U(A)\log(|t - t_0|) = At_0 + O(1/t)$ where we have used that $\log|t| - \log(|t - t_0|) = log(|t|/|t - t_0|) = \log(1/(1 - t_0/t)) = O(1/|t|)$, $t << -1$. It follows that the "B_-"' for the time translate is $B_- - At_0$, so that modulo time translation the impact parameter is only defined modulo A_-.

ec: pf Chazy uniqueness

4.3. Proof of (I) of theorem 4.1 by Blowing up infinity. We are to show that for given Chazy parameters A_-, B_- there is *exactly one* backward hyperbolic solution, modulo time translation, having these Chazy parameters. There are several proofs available. We will follow [**30**]. Other approaches are Chazy's original approach [**18**] and one based the Dollard-Möller transformation that can be found in [**36**].

In the same spirit as McGehee blow-up, we add a manifold at infinity as a place for scattering solutions "to go". The flow, extended to infinity, has a manifold of fixed points which is normally hyperbolic and which splits in two parts according to incoming and outgoing beams. The unstable manifold of the incoming fixed point will correspond to an incoming beam.

To begin, introduce spherical coordinates $(r,s) \in (0,+\infty] \times \mathbb{S}$ on $\mathbb{E} \cong \mathbb{R}^{dN}$, as in the McGehee transformation:

$$q = rs, \text{ where } r = \|q\|, \;\; s = q/r \in \mathbb{S}.$$

Decompose the velocity vector $v := \dot{q}$ into radial and tangential or " spherical" components

$$v = \nu s + w \qquad \text{where } \nu = \langle s, v\rangle \text{ and } \langle s, w\rangle = 0.$$

Change the independent time variable according to

{eq_tau}

$$dt = r d\tau, \tag{98}$$

writing $' = \frac{d}{d\tau}$. Finally introduce

$$\rho = 1/r,$$

so that $\rho = 0$ corresponds to spatial infinity. Under these changes of variables, Newton's equations transform into

{eq_odex}

$$\begin{aligned} \rho' &= -\nu\rho \\ s' &= w \\ \nu' &= \|w\|^2 - \rho U(s) \\ w' &= \rho \tilde{\nabla} U(s) - \nu w - \|w\|^2 s \end{aligned} \tag{99}$$

where

$$\tilde{\nabla} U(s) = \nabla U(s) + U(s) s$$

is the tangential component of $\nabla U(s)$, since U is homogeneous of degree -1. In these new variables the energy H is given by

$$\frac{1}{2}\nu^2 + \frac{1}{2}\|w\|^2 - \rho U(s) = H.$$

The transformed equations extend analytically to $\rho = 0$. Further setting $w = 0$ yields all equilibria e of the extended flow. The equilibria e form a manifold $\mathcal{E}$ coordinatized by $\nu, s \in \mathbb{R} \times \mathbb{S}$, by writing $e = (0, \nu, s, 0)$. The dimension of $\mathcal{E}$ is $M = dim(\mathbb{E})$ which is half the dimension of the extended phase space $\mathcal{P}$.

The energy at infinity is $\frac{1}{2}\nu^2 + \frac{1}{2}\|w\|^2 = H$, so at equilibrium $H = \frac{1}{2}\nu^2$. We are only interested in the case $H > 0$ of positive energy solutions in which case the equilibrium manifold $\mathcal{E}$ falls into two (diffeomorphic) components $\mathcal{E}_-$ and $\mathcal{E}_+$ according to whether $\nu > 0$ or $\nu < 0$, where $\nu = \pm\sqrt{2H}$.

This equilibrium manifold $\mathcal{E}$ is normally hyperbolic. To be normally hyperbolic means that when we linearize the vector field at any equilibrium $e \in \mathcal{E}$ the resulting operator on the tangent space $T_e\mathcal{P}$ of phase space $\mathcal{P}$ splits into a two-by-two block diagonal form where one block is zero and corresponds to the tangent space to $\mathcal{E}$ and the other block corresponds to the generalized eigenspaces for the nonzero generalized eigenvalues. Moreover, and this is the 'hyperbolic part", all these nonzero generalized eigenvalues have nonzero real part. In our case there is only one nonzero generalized eigenvalue and it is $-\nu$ when $e = (0, \nu, s, 0) \in \mathcal{E}$.

A non-trivial simultaneous linearization theorem, theorem 3.1 of [**30**], now comes to play, and implies that the N-body flow near infinity is locally conjugate in a neighborhood of $\mathcal{E}_-$ to its linearization, with this conjuation varying analytically with e. In particular, this theorem yields, for each $e = (0, \nu, s, 0) \in \mathcal{E}_-$ an M-dimensional unstable manifold $W^u(e)$. The part of $W^u(e)$ lying in the interior $\rho > 0$ of $\mathcal{P}$ is the incoming beam with direction s and energy $\frac{1}{2}\nu^2$. When the parameters of the linearized data are compared with the Chazy parameters, the existence and uniqueness theorem follows. See theorem 4.1 of [**30**]. QED

Remark. There is a single generalized eigenvector associated to the eigenvalue $-\nu$. The corresponding nontrivial Jordan block in the linearization propagates to 'twist' the relations between the linearizing variables and the Chazy variables (see the equations in theorem 4.1 of [**30**]) and also yields the $log(|t|)$ term in the Chazy asymptotics.

Summary. We can summarize the proof just given as follows. The incoming beam with direction s and energy E is equal to $W^u(e) \cap \{\rho > 0\}$ where $e = (0, \nu, s, 0)$ as in the proof.

4.4. Systems of Rays. Arnol'd defines a system of rays as a Lagrangian submanifold of the cotangent bundle. He uses this idea and ideas from optics to formulate a number of interesting theorems and conjectures. Taking his perspective, we believe that the following facts may be of use.

thm: beam is Lag

Theorem 4.2. *The set of phase points $(q(t), p(t)) \in T^*\mathbb{E}_0$ swept out by the incoming beam with fixed direction s_- and energy E forms a Lagrangian submanifold $\mathcal{L}_-(s_-; E)$ lying within the given energy level set.*

The obvious analogues of theorems 4.1 and 4.2 hold for the **outgoing beam** for an $s_+ \in \mathbb{S} \setminus \Delta$. We write $\mathcal{L}_+(s_+; E)$ for the corresponding outgoing Lagrangian submanifold. If $\mathcal{L}_-(s_-; E)$ and $\mathcal{L}_+(s_+; E)$ intersect, then, since they are invariant under the backward and forward flows respectively, they must intersect along a union of energy E hyperbolic orbits. Thus s_- is connected to s_+ if and only if $\mathcal{L}_-(s_-; E) \cap \mathcal{L}_+(s_+; E) \neq \varnothing$ for some $E > 0$. (By scaling symmetry, whether they intersect or not does not depend on the energy E.) We have reformulated the Open Question of this chapter in terms of the popular subject of "Lagrangian intersection theory". Namely: given s_-, is it true that for an open dense set of s_+'s we have that $\mathcal{L}_-(s_-; E) \cap \mathcal{L}_+(s_+; E) \neq \varnothing$.

5. More questions

1. [This question represents a kind of marriage of the present chapter with the previous one]

The three bodies of hyperbolic solutions to the planar three body problem asymptote to lines, a triple of lines for the past, another triple for the future. Fix six such allowable lines, three for the past, three for the future. (For example, they could both represent the Lagrange shape– three lines through the origin each at $120^{\deg}$ with respect ot each other) The set of all collision free paths (not necessarily solutions!) having these past and future asymptotics has many different components, indeed they are labelled by reduced eclipse sequences, or free homotopy classes. Are all finite reduced sequences realized by hyperbolic solutions asymptotic to the given lines? Or, alternatively, is there some bound on the length of the sequence, say, in the equal mass case?

2. The scattering map for an incoming Lagrange beam at fixed positive energy is a map $SC : \mathbb{R}^3 \to \mathbb{S}^3$. Here the $\mathbb{R}^3$ represents the unstable manifold to that Lagrange , or, what is the same, the set of impact parameters b having energy E, and the $\mathbb{S}^3$ represents the space of outgoing directions. Let d^3b and $d^3\omega$ be the standard Lebesgue (= Riemannian) measures on $\mathbb{R}^3$ and $\mathbb{S}^3$. Compute the Radon-Nikodym derivative $d(Sc_* d^4b)/d^\omega$. In other words, if we push forward Lebesgue by SC we get a measure $SC_* d^3b = f(\omega)d^3\omega$. Compute, or at least say something about, the function $f(\omega)$. This function is known as the "differential cross section'. [The main open question of this chapter asks whether the support of f has full measure.]

3. Analysis of the central force problem for the other homogeneous potentials $1/r^\alpha$ show that the scattering map for the two-body problem is not onto, and that the set of accessible angles $\Delta\theta(\alpha)$ goes to zero with α. Work out the correct 'billiards at infinity' model for the associated broken geodesics for three-body scattering with these other potentials. Is it the "obvious" non-deterministic billiards, except instead of having all outgoing angles allowed as in the $1/r$ case, only angles in the sector $\Delta\theta(\alpha)$ are allowed?

6. Notes.

Andreas Knauf introduced me to this problem in 2012. He has numerous papers on scattering. I highly recommend XXX.

The book by REF is a kind of bible of classical scattering.

...

Part 3

Appendix

APPENDIX A

Geometric Mechanics

AppendixA

sec: formalisms

1. The Lagrangian and Hamiltonian Formalisms

We describe the elements of geometric mechanics as used in the book. This appendix can be taken as a brief introduction to the subject aimed at someone who knows differential geometry but has never had a course in classical mechanics. For more complete treatments we recommend [**10**], [**1**], or [**70**] We end the appendix with a derivation of the "particle in a magnetic field " version of the reduced planar N-body equations by using Poisson reduction.

The mathematical foundations of modern Classical mechanics are formed by the Lagrangian and Hamiltonian formalisms and the relations between them. We begin with a manifold Q called configuration space. Points of Q are thought of as representing the 'positions' of the interacting objects forming the mechanical system of study - in this book N point masses. The Hamiltonian formalism lives on the cotangent bundle of the configuration space. The Lagrangian formalism lives on its tangent bundle. The Hamiltonian formalism plays a central role in chapter two of this book while the Lagrangian formalism plays a central role in chapter three.

Definition A.1. *Let Q be a manifold representing the configuration space of a mechanical system. A Hamiltonian is a smooth function $H : T^*Q \to \mathbb{R}$ on the cotangent bundle of configuration space. A Lagrangian is a smooth function $L : TQ \to \mathbb{R}$ on the tangent bundle of configuration space. Either TQ or T^*Q are referred to as the phase space of the system and will sometimes be denoted by $\mathcal{P}$.*

Coordinates q^i on M induce fiber-linear coordinates v^i on TQ and p_i on T^*Q by writing a vector $v \in T_qQ$ as $v = \Sigma v^i \frac{\partial}{\partial q^i}$ and a co-vector $p \in T^*_qQ$ as $p = \Sigma p_i dq^i$. Then $p(v) = \Sigma p_i v^i$. (Here $q \in Q$ is any point covered by our coordinate neighborhood.) In terms of these coordinates the Hamiltonian and Lagrangian are functions of the forms $H = H(q^1, \dots, q^n, p_1, \dots p_n)$ and $L = L(q^1, \dots, q^n, v^1, \dots, v^n)$ where $n = dim(Q)$. Hamilton's equations, written in these coordinates, is the system of equations

{eq: Hamiltons}

$$
\begin{aligned}
\dot{q}^i &= \frac{\partial H}{\partial p_i} \\
\dot{p}_i &= -\frac{\partial H}{\partial q^i}.
\end{aligned} \tag{100}
$$

The Euler-Lagrange equations for L, written in the above coordinates, are

{eq: E-L}

$$
\frac{d}{dt}\left(\frac{\partial L}{\partial v^i}\right) = \frac{\partial L}{\partial q^i}. \tag{101}
$$

When performing the time derivative $\frac{d}{dt}$ on the left-hand side of the Euler-Lagrange equations in order to write them out as a system of second order ODEs, we use the chain rule, and when we are finished we impose the equation $\dot{q}^i = v^i$ which says that v is indeed the velocity associated to a curve $q^i(t)$ in M.

To relate the Lagrangian and Hamiltonian systems set $p^i = \frac{\partial L}{\partial v^i}$ within the left-hand side of the Euler-Lagrange equations, thus defining a map $(q,v) \mapsto (q,p) = (q, p(q,v)$ called the Legendre transform. The Legendre transformation is an intrinsically defined bundle map, (generically not linear)

$$FL : TQ \to T^*Q$$

with the following coordinate-free definition as a fiber derivative: if $v \in T_qQ$ then $p = FL(q,v) \in T_q^*Q$ is the linear functional $p : T_qQ \to \mathbb{R}$ given by $p(w) = \frac{d}{dt}|_{t=0} L(q, v + tw)$. When the Legendre transformation is invertible, we can write its inverse as $(q,p) \mapsto (q, v(q,p))$ and using this inverse we can view the function

$$H = p(v) - L(q,v)$$

as a function of q and p. The resulting function $H : T^*Q \to \mathbb{R}$ is called the Legendre transform of L. The Legendre transform turns the Euler-Lagrange equations for L into Hamilton's equations for H.

write notes -RM

notes. Dirac. L-C Young. In the case that the Legendre transform is not invertible

exer: euclidean newton

Exercise A.1. *1. Take $Q = \mathbb{R}$ and write $x \in \mathbb{R}$ for the coordinate. The standard Lagrangian of 1-dimensional mechanics is $L(q,v) = \frac{1}{2}mv^2 - V(x)$ where the constant $m > 0$ is the particle's mass. Verify that its Euler-Lagrange equations are Newton's equations: $m\ddot{x} = -V'(x)$. Verify that its Legendre transform is $(x,v) \mapsto (x,p) = (x, mv)$. Verify that its Hamiltonian is $H = \frac{1}{2}\frac{1}{m}p^2 + V(x)$ and that Hamilton's equations are equivalent to Newton's equations.*

2. Take $Q = \mathbb{R}^n$ be Euclidean space with its standard inner product and let $V : \mathbb{R}^n \to \mathbb{R}$ be a smooth function called the potential. Let M be a positive parameter corresponding to mass. The corresponding Newton's equations are $M\ddot{q} = -\nabla V(q)$. Verify that the Euler-Lagrange equations of the Lagrangian $L(q,v) = \frac{1}{2}M|v|^2 - V(q)$ are Newton's equations. Verify that the Legendre transformation $p = Mv$ when we use the standard inner product to identify $(\mathbb{R}^n)^*$ with $\mathbb{R}^n$, and so both TQ and $T*Q$ with $\mathbb{R}^n \times \mathbb{R}^n$. Verify that the Hamiltonian is then by $H(q,p) = \frac{1}{2}\frac{1}{M}|p|^2 + V(q)$ and that Hamilton's equations are also equivalent to Newton's equations.

Exercise A.2. *Modify the 2nd part of the previous exercise so as to work for the N-body problem, thus expressing the N-body problem as an Euler-Lagrange equations and as a Hamilton's equations. Restrict to the non-collision configurations so derivatives of the potential are defined.*

1.1. Natural Mechanical Systems. If our configuration space Q comes along with a Riemannian metric ds^2_Q and a 'potential function' $V : Q \to \mathbb{R}$ then we call it a "natural mechanical system". If a Lie group G acts on Q by isometries of the metric which leave the potential invariant then we say we have a "natural mechanical system with symmetry". The N-body problem is such a system.

Write $\langle \cdot, \cdot \rangle_q$ for the fiber inner product on T_qQ defined by the metric on Q and $\langle \cdot, \cdot \rangle_q^*$ for the associated dual metric on T_q^*Q. Write $K(q,v) = \frac{1}{2}\langle v, v \rangle_q$ and $K^*(q,p) = \frac{1}{2}\langle p, p \rangle_q^*$ for the kinetic energies. We take for the Lagrangian $L(q,v) =$

$K(q,v) - V(q)$ while the Hamiltonian is $H(q,p) = \frac{1}{2}K^*(q,p) + V(q)$. Coordinate expressions for K and K^* may be useful. If $ds_Q^2 = \Sigma g_{ij}(q)dq^i dq^j$ expresses the metric in coordinates then $K = \frac{1}{2}\Sigma g_{ij}(q)v^i v^j)$ while $K^* = \frac{1}{2}\Sigma g^{ij}(q)p_i p_j$ where g^{ij} is the inverse matrix to g_{ij}.

EXERCISE A.3. *Verify that the Legendre transform for the L just described is the operation of "lowering indices" which takes $v = T_q Q$ to the covector $p = FL(q,v)$ with $p(w) = \langle v, w\rangle_q$. In coordinates: $p_i = \Sigma g_{ij}(q)v^j$. Verify that the Legendre transform of L is the Hamiltonian H above.*

EXERCISE A.4. *Verify that both Hamilton's and Lagrange's equations of motion for a natural mechanical system are equivalent to their Newtonian form $\nabla_{\dot q}\dot q = -\nabla V(q)$ where ∇ is the Levi-Civita metric for the Riemannian metric on Q.*

2. Symplectic structure: the Hamiltonian side.

The Hamiltonian side of mechanics enjoys the benefits of symplectic geometry, a geometry associated to a particular type of two-form on a manifold known as a *symplectic form*. The symplectic form for mechanics is called the *canonical two-form* on T^*Q and is given in our local coordinates by

$$\omega = \Sigma dq^i \wedge dp_i.$$

EXERCISE A.5. *Verify that ω is well-defined, independent of the choice coordinates q^i on Q used to induce the coordinates q^i, p_i for T^*Q.*
Verify that ω is closed, i.e. that $d\omega = 0$.
Verify that ω is non-degenerate.

To say that ω is non-degenerate means that the fiber-linear 'index lowering' map $v \mapsto i_v\omega := \omega(v,\cdot)$ from the tangent bundle to the cotangent bundle is invertible. Equivalently, any two-form ω has the local coordinate expression $\omega = \Sigma\omega_{ij}dx^i \wedge dx^j$ relative to coordinates x^i on the manifold. 'Non-degeneracy' of ω is equivalent to the skew-symmetric matrix ω_{ij} being invertible.

DEFINITION A.2. *A symplectic form is a closed non-degenerate two-form on a manifold. A manifold P with a symplectic form ω is called a symplectic manifolds.*

Symplectic manifolds must have even dimensions since skew-symmetric matrices in odd dimensions always have kernels. Moreover, any two symplectic manifolds of the same dimension are locally diffeomorphic as symplectic manifolds by the following lemma.

LEMMA A.1 (Darboux). *About any point of an symplectic manifold there exist coordinates, called "Darboux" or "canonical" coordinates, wuch that the expression for the two-form in these coordinates is the one given above.*

The matrix of our ω in terms of the canonical coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ is

$$\mathbb{J} = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

where I_n is the $n \times n$ identity matrix and $n = dim(Q)$. One has $\mathbb{J}^2 = -I_{2n}$ hence the matrix of ω is invertible and the canonical two-form is indeed non-degenerate.

EXERCISE A.6. *Write Θ for the one-form on T^*Q given in local coordinates as $\Theta = \Sigma p_i dq^i$. Then $\omega = -d\Theta$. Verify that Θ and hence ω are globally defined forms on T^*Q, independent of the choice of coordinates q^i used to induce the system of canonical coordinates (q^i, p_i) on T^*Q.*

See [**1**] for a coordinate-free definition of Θ.

Given a function H on such a symplectic manifold P we define the associated Hamiltonian vector field X_H on P by solving

$$\omega(X_H, \cdot) = dH. \tag{102}$$

for the vector field X_H. The non-degeneracy of ω implies the solvability of this linear equation. We also write $i_{X_H}\omega$ for the one-form $\omega(X_H, \cdot)$.

DEFINITION A.3. *The vector field X_H just defined is called the Hamiltonian vector field of the Hamiltonian H.*

EXERCISE A.7. *Verify that Hamilton's equations (100) are the ODEs associated to the Hamiltonian vector field X_H as just defined on $P = T^*Q$, relative to the canonical symplectic form.*

Cartan's magic formula asserts that $L_X\alpha = di_X\alpha + i_X d\alpha$, where L_X is the Lie derivative with respect to a vector field X and where α is any k-form. Applied to $X = X_H$ and $\alpha = \omega$, Cartan's formula yields $L_{X_H}\omega = 0$. Thus the flow Φ_t^H of any Hamiltonian vector field X_H is a flow through symplectomorphisms: diffeomorphisms which preserve ω. Such transformations are also refered to as "canonical transformations".

DEFINITION A.4. *A symplectic transformation, also known as a canonical transformation, is any diffeomorphism of P preserving the symplectic form. We write $Symp(P)$ for the (infinite-dimensional) group of symplectic transformations.*

The huge and varied nature of the canonical transformations give the symplectic formalism its power. We can use canonical transformations to put the Hamiltonian, and hence the dynamics, into some normal form, perhaps approximate, or to better understand the properties of its flow.

$Diff(Q) \subset Symp(T^*Q)$. The diffeomorphism group of Q embeds in the symplectomorphism group of T^*Q as follows. Given $\psi : Q \to Q$ a diffeomorphism, define $\psi_* : T^*Q \to T^*Q$ by $\psi_*(q,p) = (\psi(q), (d\psi_q)^{-1,T}(p))$. This induced map is called the "cotangent lift" of ψ and is easily verified to be a canonical transformation. The process of cotangent lift defines an injective homomorphism $Diff(Q) \to Symp(T^*Q)$.

At the level of Lie algebras, cotangent lift induces an inclusion of the space of all vector fields on Q into the space of Hamiltonian vector fields on $T*Q$ as follows.

EXERCISE A.8. *If Y is a vector field on Q, define $P_Y : T^*Q \to \mathbb{R}$ by $P_Y(q,p) = p(Y(q))$. Show that the Hamiltonian vector field X_H of $H = P_Y$ projects to Y by the projection $T^*Q \to Q$. Write $exp(tY) : Q \to Q$ for the flow of Y. Show that the Hamiltonian flow defined by P_Y is the cotangent lift of the flow of Y.*

ex: momentum functions

In canonical coordinates $P_Y = \Sigma p_i Y^i(q)$ where $Y = \Sigma Y^i(q)\frac{\partial}{\partial q^i}$. In other words, Hamiltonians linear in momentum variable correspond to the vector fields on configuration space.

DEFINITION A.5. *We call P_Y the momentum function , or simply the momentum of the vector field Y.*

def: momentum function

This last observation allows us to quantify the vast gulf between diffeomorphisms of Q and symplectomorphisms of T^*Q. At a formal level, we can view the space $C^\infty(T^*Q)$ of all Hamiltonians H as forming the Lie algebra of the group of symplectomorphisms $Symp(T^*Q)$, with the exponential map $Lie(G) \to G$ taking H to the time t flow of X_H. Now Taylor expand a given H in terms of the fiber variable $p \in T_q^*Q$:

{H} (103) $$H(q,p) = H_0(q) + H_1(q)(p) + H_2(q)(p,p) + \dots$$

where H_0 is independent of p, $H_1 = H_1(q)(p)$ is linear in p and thus has the local form $H_1(q)(p) = \Sigma Y^i(q)p_i$, $H_2(q)(p,p) = \frac{1}{2}\Sigma g^{ij}(q)p_i p_j$ is quadratic in p, etc. We have seen that H corresponds to a diffeomorphism of the configuration space if and only if $H = H_1$, which is to say, if and only if $H_0 = H_2 = H_3 = \dots = 0$. We have just characterized the image of the embedding of $Diff(Q)$ in $Symp(T^*Q)$ as having infinite codimension, requiring as it does infinitely many equalities to hold.

Return to natural mechanics. A natural mechanical system is generated by a Hamiltonian which truncates at the second order and has no first order piece: $H = H_0 + H_2$. We identify H_0 with the potential and H_2 with the kinetic energy. When a first order piece H_1 is present we may associate it to a vector field, or, dually (via H_2 assuming that it is non-degenerate) a one-form "vector potential" as in electromagnetism.

3. Variational structure: the Lagrangian side.

The Lagrangian defines a function on paths in configuration space called the action.

Definition A.6. *The action A associated to the Lagrangian $L : TQ \to \mathbb{R}$ is the function*

$$c \mapsto A(c) := \int_I L(c(t), \dot{c}(t))dt$$

defined on the space of absolutely continuous paths $c : I \to Q$. Here $I \subset \mathbb{R}$ is any closed bounded interval.

Recall that a path is absolutely continuous if it is differentiable a.e. and if the fundamental theorem of calculus holds locally: if $c(t) = (q^1(t), \dots, q^n(t))$ is the coordinate expression for c in some neighborhood then $q^i(t) = q^i(a) + \int_0^t \dot{q}^i(s)ds$. (See [**132**].) Being absolutely continuous does not depend on the choice of local charts.

In calculus of variations one differentiates the action with respect ot sufficiently smooth paths c. If we restrict the action to the space of all smooth paths joining two fixed points of Q in some fixed interval I of time, then the Euler-Lagrange equations are the necessary conditions [**?**], [**88**] for a path to be a critical point of the restricted action. We now verify this assertion by hand for a natural mechanical system on a Euclidean space $Q = \mathbb{E}$ as is relevant for the N-body problem. To do so we make explicit the differentiation process.

Definition A.7. *By a compact perturbation of a path $q : (a,b) \to \mathbb{E}$ we will mean a family $q_\epsilon(t) = q(t) + \epsilon h(t)$ of curves where the support of h is some compact sub-interval J of (a,b).*

Theorem A.1. *Consider the case of the action $A = \int Ldt$ for a natural mechanical system $L = K - V$ on the Euclidean vector space $\mathbb{E}$, where K corresponds* *to half the inner product. Suppose that the path q is twice-differentiable and that the derivative of A is zero at q for all compact twice differentiable perturbations $q_\epsilon = q + \epsilon h$ of q. Then q solves Newton's equations.*

PROOF. Differentiate $A(q + \epsilon h)$ with respect to ϵ. Use $\dot{q}_\epsilon = \dot{q}(t) + \epsilon \dot{h}(t)$ to expand out and find

$$K(\dot{q}_\epsilon(t)) = K(\dot{q}(t)) + \epsilon\langle \dot{q}(t), \dot{h}(t)\rangle + \epsilon^2 K(\dot{h}(t)).$$

Also

$$V(q_\epsilon(t)) = V(q(t)) + \epsilon\langle \nabla V(q(t)), h(t)\rangle + O(\epsilon^2)$$

where we have assumed that V is C^2 along the path $q(t)$. It follows that

$$A(q_\epsilon) = A(q) + \epsilon(\int \langle \dot{q}, \dot{h}\rangle - \langle \nabla V(q), h\rangle)dt + O(\epsilon^2)$$

so that

$$\frac{d}{d\epsilon}|_{\epsilon=0} A = \int_J \langle \dot{q}, \dot{h}\rangle - \langle \nabla V(q), h\rangle)dt.$$

where J is the closed interval supporting h. We note that $\frac{d}{dt}(\langle \dot{q}, h\rangle = \langle \ddot{q}, h\rangle + \langle \dot{q}, \dot{h}\rangle$ or

$$\langle \dot{q}, \dot{h}\rangle = \langle -\ddot{q}, h\rangle + \frac{d}{dt}(\langle \dot{q}, h\rangle)$$

It follows that

$$\frac{d}{d\epsilon}|_{\epsilon=0} A = \int_J \langle -\ddot{q} - \nabla V(q), h\rangle dt + (\langle \dot{q}, h\rangle)|_c^d$$

where $J = [c, d]$. The final boundary term is zero since h has compact support within the open interval J and so $h = 0$ at the endpoints c, d of J. So, we have shown that

$$\frac{d}{d\epsilon} A = \int_J \langle -\ddot{q} - \nabla V(q), h\rangle)dt.$$

Now assume the hypothesis of the theorem, that this derivative is zero for all twice differentiable compactly supported h. Since the set of such h is dense in $L_2 = L_2(J, \mathbb{E})$ for this interval J, where the L_2 pairing is $(w, h) \mapsto \int_J \langle w(s), h(t)\rangle ds$ we must have $w = -\ddot{q} - \nabla V(q)$ is zero a.e.. But $q(t)$ is assumed C^2 so we have that Newton's equations $\ddot{q} = -\nabla V(q)$ is satisfied on J. Since $J \subset I$ is arbitrary, w have that q satisfies Newton's equations on all of I.

QED

With extra work we can weaken the hypothesis of the theorem to :

THEOREM A.2. *Suppose that the path $q : I = [a, b] \to \mathbb{E}$ is absolutely continuous and V is C^2 in a neighborhood of $q((a, b))$. Then A is Frechet-differentiable at q and the derivative is zero at q if and only if $q(t)$ is twice-differentiable on the interior of I and satisfies Newton's equations there.*

We will prove this theorem in the next appendix, which is an appendix on the direct method.

yes? if you promise it, do it! or, simply give a reference to .. -RM

Extremal path need not exist. For example, take $Q = \mathbb{R}$ with standard coordinate q and $L = \dot{q} + q$. Then the Euler Lagrange equations for L read $\frac{d}{dt}(1) = 1$ or $0 = 1$. There is no solution. Correspondingly, the action of a path $q : [0, T] \to \mathbb{R}$ is $\int_0^T (\dot{q}(t) + q(t)dt = q(T) - q(0) + \int_0^T q(s)ds$. This function has no extremals for fixed values of T, $q(0)$ and $q(T)$. Indeed this action functional is linear so has no extremal and in particular no minimum or maximum.

In physics texts one reads the adage " physical trajectories minimize the action" not "extremize the action". The effort to minimize the action initiated the calculus of variations and is one of the most powerful methods available to this calculus. But, as we just saw, the minima may not exist. With this in mind, the next proposition is important in chapter 3.

Proposition A.1. *For natural mechanical systems action minimizers exist between sufficiently close points. Specifically, if we restrict the action to all absolutely continuous paths joining two points q_0, q_1 in a time T and if the points are sufficiently close and the time sufficiently short, then the minimum exists, is smooth, and solves the Euler-Lagrange equations. Conversely, any sufficiently short subarc of a solution to the Euler-Lagrange equations minimizes the action among all paths connecting the endpoints of this subarc.*

For a proof of a somewhat more general version of this proposition we refer to what Mañe calls "Weierstrass's theorem" on p.24-6 of his book.

The existence of minima for arbitrary points and times is closely linked to the completeness of the Euler-Lagrange flow. In the case of Riemannian geometry, which is to say, natural mechanical systems with potential $V = 0$, this fact is well-known and goes under the name of the Hopf-Rinow theorem: the geodesic flow is complete if and only if any two points can be connected by a length-minimizing geodesic. In the case of the N-body problem, the collisions get in the way of completeness of the flow. Despite this fact, by the miracle discovered by C. Marchal **??**, collision-free action minimizers always exist for the gravitational two-point boundary value problem, provided the endpoints are not collision points and the dimension of space is greater than 1. See XX

put Marchall in , further on -RM

4. Poisson Bracket Formalism and Reduction

At various points in this book we work with the N-body problem after modding out by its symmetry group G. is the group of rigid motions G acting on N-body phase space $\mathcal{P}$. There are two primary methods for forming the quotient of phase space by the group action. The most direct way and computationally effective method is to just do it: form the quotient phase space $\mathcal{P}/G$. However, the quotient phase space is no longer a symplectic manifold and another geometric formalism enters in if we are to understand what we are doing down there on the quotient. That formalism is Poisson geometry. Instead of inheriting a symplectic structure or variational structure from $\mathcal{P}$, this quotient space inherits a "Poisson structure", hence this section.

The other method of forming the quotient is known as "symplectic reduction". Poisson manifolds admit intrinsic (singular) foliations whose leaves are symplectic manifolds. These leaves can be identified with the symplectic reduced spaces of $\mathcal{P}$ and these will be described in a later section.

To begin our brief survey of Poisson geometry, return to symplectic geometry, so a manifold $\mathcal{P}$ endowed with a symplectic form ω. We define the Poisson bracket on $\mathcal{P}$ by

$$\{F, H\} = \omega(X_F, X_H)$$

where X_F, X_H are the Hamiltonian functions of the smooth functions F, H on $\mathcal{P}$. We could also write $\{F, H\} = dF(X_H) = -dH(X_F)$. This bracket defines a bilinear skew-symmetric operation on functions on $\mathcal{P}$. In terms of our canonical coordinates q^i, p_i we have

$$\{F, H\} = \Sigma \frac{\partial F}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q^i}$$

Exercise A.9. *Verify that when ω is the canonical symplectic form as above, we have the bracket relations, also known as the canonical commutation relations:*

{eq: CCRs} (104)
$$\{q^i, q^j\} = 0, \{q^i, p_j\} = \delta^i_j, \{p_i, p_j\} = 0$$

Exercise A.10. *If $\mathcal{P} = T^*Q$, let $X \mapsto P_X$ be the momentum function defined earlier. Verify that $\{P_X, P_Y\} = -P_{[X,Y]}$ where $[X, Y]$ is the Lie bracket of the vector fields X, Y on Q. Also verify that $\{\pi^* f, P_X\} = \pi^*(X[f])$ where $\pi : T^*Q \to Q$ is the projection and $\pi^* : C^\infty(Q) \to C^\infty(T^*Q)$ is the pull-back operation.*

We formalize Poisson brackets as follows

Definition A.8. *A Poisson bracket on a manifold $\mathcal{P}$ is a Lie algebra structure $F, H \mapsto \{F, H\}$ on its smooth functions which, in addition, is Liebnitz:*

$$\{F, GH\} = G\{F, , H\} + \{F, G\}H$$

There is a second Liebnitz identity relative to the first slot "F" of the Poisson bracket which follows automatically from the skew-symmetry of the bracket. The reader may wish to verify that the bracket previously defined on a symplectic manifold is actually a Poisson bracket, the main issue being the Jacobi identity.

A Poisson manifold is, by definition, a manifold with a Poisson bracket operation on its smooth functions. The Liebnitz identity allows us to associate a tensor $B \in \Gamma(\Lambda^2 T\mathcal{P})$ to a Poisson manifold. B is called the "Poisson tensor". If x^i are local coordinates on $\mathcal{P}$ then we write $B^{ij}(x) = \{x^i, x^j\}(x)$ and then $B = \Sigma_{i<j} B^{ij}(x)\frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j}$. Then $\{F, H\} = B(dF, dH) = \Sigma_{ij} B^{ij}(x)\frac{\partial F}{\partial x^i}\frac{\partial H}{\partial x^j}$ where we use $B^{ji} = -B^{ij}$.

Exercise A.11. *If $\omega = \Sigma \omega_{ij} dx^i \wedge dx^j$ is the coordinate expression for a symplectic form ω on a symplectic manifold, verify that the Poisson tensor corresponding to its Poisson bracket is given by $B = \Sigma \omega^{ij}\frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j}$ where ω^{ij} is the inverse matrix to ω_{ij}.*

4.1. Dynamics. As in symplectic geometry, there is a Hamiltonian vector field X_H associated to each smooth function H on a Poisson manifold. We define X_H by its action on functions: $X_H[f] = \{f, H\}$. This agrees with our previous definition of X_H in the symplectic case. Write Φ^H_t for the flow of X_H. The Jacobi identity implies that this flow preserves the Poisson bracket. If we set $f_t = f \circ \Phi^H_t$ for $f \in C^\infty(\mathcal{P})$ then we have that:

$$\frac{d}{dt} f = \{f, H\}$$

which provides us with another equivalent way to express Hamilton's equations when $\mathcal{P} = T^*Q$. Just let f vary over q^i, p_i and write out the above equation to recover our original version of Hamilton's equations (100).

4.2. New Poisson manifolds by quotient. If $\mathcal{P}$ is a Poisson manifold and if a compact Lie group acts freely on $\mathcal{P}$ by Poisson automorphisms then $\mathcal{P}/G$ inherits the structure of a Poisson manifold. Write the action of pulling a function back by the action of g as $F \mapsto g^*F$. Then to say that G acts by Poisson automorphisms is to say that for all $g \in G, F, H \in C^\infty(\mathcal{P})$ we have that $\{g^*F, g^*H\} = g^*\{F, H\}$. Since G is compact and acts freely, the quotient space $\mathcal{P}/G$ is automatically a smooth manifold. We can identify the ring of smooth functions on $\mathcal{P}/G$ with $C^\infty(\mathcal{P})^G$ - the ring of G-invariant smooth functions on $\mathcal{P}$. The fact that G acts by Poisson automorphisms implies that the Poisson bracket of two G-invariant functions is again G-invariant. Thus we have defined a Poisson structure on the quotient space $\mathcal{P}/G$.

In case $\mathcal{P} = T^*Q$ and the compact Lie group G acts freely on Q , then its cotangent lift acts freely on T^*Q by symplectic, and hence Poisson automorphisms. It follows that $(T^*Q)/G$, is a Poisson manifold. This class of examples arises in the N-body problem. We will describe its details in the next section.

Take the case of G acting on itself by left translation: $L_g(h) = gh$, $L_g : G \to G$. This is a free action and we can identify $T^*G/G = T^*_eG = \mathfrak{g}^*$ with the dual of the Lie algebra by left-translating covectors back to the identity. Now if X is a left-invariant vector field then the momentum function P_X is a left-invariant function on T^*G. By the previous exercise $\{P_X, P_Y\} = -P_{[X,Y]}$. Viewed on T^*_eG, we have that $P_X : \mathfrak{g}^* \to \mathfrak{g}$ is given by $P_X(\mu) = \mu(X(e))$. We identify $\mathfrak{g}$ with left-invariant vector fields. In this way, $X \in \mathfrak{g}$ becomes an element of $\mathfrak{g}^{**}$ -i.e. a linear function on $\mathfrak{g}^*$. Take a basis E_i of $\mathfrak{g}$ and let c^k_{ij} be the corresponding structure constants. Write $x_1, \ldots, x_n$ for the coordinate system on $\mathfrak{g}^*$ corresponding to this basis. We have just shown

$$\{x_i, x_j\} = -\Sigma c^k_{ij} x_k.$$

This Poisson bracket was found by Lie, and rediscovered by Kostant, Kirrilov and Souriau. See [**157**] for a bit of the history.

EXERCISE A.12. *If the action of G is instead by right translation verify that the induced quotient bracket on $\mathfrak{g}^*$ is $\{x_i, x_j\} = +\Sigma c^k_{ij} x_k$.*

When I was a graduate student Alan Weinstein published a beautiful paper called "the Local structure of Poisson Manifolds" One of the central observations of that paper is that every Poisson manifold admits a canonically defined, typically singular foliation by what he christened "symplectic leaves". To form the symplectic leaf through the point $p \in \mathcal{P}$ form the collection of all Hamiltonian vector fields $X_H = \{\cdot, H\}$ and integrate them, starting from p. The Jacobi identity shows us that this collection of vector fields is involutive, so, by an extension of the Frobenius integrability theorem due to Sussman, the collection of all endpoints of such vector fields is an locally embedded submanifold of $\mathcal{P}$. Its tangent space is spanned by the X_H's. The symplectic form on the leaf can be obtained by writing $\omega(X_F, X_G) = \{F, G\}$. some of this personal stuff To NOTES? -RM

EXERCISE A.13. *Suppose that G is connected. Then the symplectic leaves of the Lie-Poisson bracket on $\mathfrak{g}^*$ are the co-adjoint orbits: the orbits of $\mathfrak{g}^*$ under the dual of the adjoint action.*

5. Reduction by the circle

The formalism of Poisson reduction works out to be simple and explicit in the case relevant to the *planar* N-body problem. Here G is the circle group $\mathbb{S}^1$ acting by cotangent lift on the phase space of a natural mechanical system. The goal is then to identify the quotient Poisson manifold $(T^*Q)/G$ in the case where $G = \mathbb{S}^1$. We are interested in the case where $\mathbb{S}^1$ acts on the configuration space of a natural mechanical system Q so that it leaves both the metric and potential invariant. Then this $\mathbb{S}^1$ acts on T^*Q by cotangent lift so as to leave the Hamiltonian of the system invariant and preserve the Poisson brackets. This is the situation which arises for the planar N-body problem. The grand finale and main point of this section is the realization of "Wong's equations" (108) - the equations of a particle in a magnetic field on the shape space $S = Q/G$ - as being the reduced version of the planar N-body problem.

In order to proceed with the computation we will need the notion of a connection on a circle bundle which we will review as we come to it. We suppose, as is the case for the planar N-body problem away from total collision, that our $\mathbb{S}^1$ action is free. Write $\frac{\partial}{\partial\theta}$ for the infinitesimal generator of the $\mathbb{S}^1$-action. Thus, $\frac{\partial}{\partial\theta}$ is the vector field $q \mapsto \frac{\partial}{\partial\theta}(q) := \frac{d}{d\theta}|_{\theta=0}e^{i\theta}(q)$ on Q. Set

$$V = span\frac{\partial}{\partial\theta} \subset TQ$$

We call V the vertical distribution and vectors lying in V are called "vertical". V is a trivial line sub-bundle of TQ. Also form the momentum function:

$$J = P_{\frac{\partial}{\partial\theta}} : T^*Q \to \mathbb{R}.$$

We call J the momentum map for the circle action. Being $\mathbb{S}^1$-invariant, J can be viewed as a function

$$J : (T^*Q)/\mathbb{S}^1 \to \mathbb{R}$$

EXERCISE A.14. *Verify that J agrees with the angular momentum in the case of the planar N-body problem, with $\mathbb{S}^1 = SO(2)$ the standard rotational action.*

.

Write V^0 for the annihilator of V, that is, the space of all covectors $p \in T^*Q$ such that $p = 0$ for all $v \in V$.

$$V^0 = J^{-1}(0) \subset T^*Q$$

Indeed, $J(q,p) = p(\frac{\partial}{\partial\theta}_q)$.

The advantage of assuming that $\mathbb{S}^1$ acts freely on Q is that this implies that the quotient space $B = Q/\mathbb{S}^1$ is a smooth manifold and that the quotient projection

$$\pi : Q \to Q/\mathbb{S}^1 = B$$

gives Q the structure of a principal circle bundle over the base space B. We might also want to call the base B the "shape space" in honor of the planar three-body problem. Observe that

$$V = ker(d\pi).$$

Circle bundles admit local trivializations so that, if $x^i, i = 1, \dots, n-1$ are coordinates on B defined within a locally trivializing neighborhood and if θ is the standard circle variable, we get induced coordinates x^i, θ on Q such that in these coordinates $\pi(x^i, \theta) = (x^1, \dots, x^{n-1})$ and such that the infinitesimal generator $\frac{\partial}{\partial\theta}$ is the coordinate $\frac{\partial}{\partial\theta}$. These coordinates x^i induce the usual momentum coordinates p_i on T^*B and the x^i, θ induce coodinates p_i, p_θ on T^*Q. In terms of these coordinates on T^*Q we have that $J(x^i, \theta, p_i, p_\theta) = p_\theta$, the momentum dual to rotation.

LEMMA A.2. *$J^{-1}(0)/\mathbb{S}^1 = T^*B$ in a canonical way.*

PROOF. Locally, in the above coordinates, we have $J^{-1}(0) = \{p_\theta = 0\}$ and so x^i, θ, p_i coordinatize $J^{-1}(0)$. Forming the quotient by $\mathbb{S}^1$ gets rid of θ so that the quotient map $J^{-1}(0)/\mathbb{S}^1$ is, in these coordinates, $(x^i, \theta, p_i, 0) \mapsto (x^i, p_i)$ yielding canonical coordinates on T^*B. Some thought about how coordinates and local trivializations fit together shows this computation yields a global map and finishes the proof.

However, it may be worth understanding the diffeomorphism $J^{-1}(0)/\mathbb{S}^1 = T^*B$ at a coordinate-free level. Take a covector $\alpha \in T_b^*B$. Then $\pi^*\alpha \in T^*Q$ is a covector which annihilates V thus $\pi^*\alpha \in J^{-1}(0)$. There is an ambiguity as to where $\pi^*\alpha$ lives. To make it a covector attached to a particular $q \in \pi^{-1}(b)$ we set $p = (\pi^*\alpha)_q = d\pi_q^*\alpha$. Now p annihilates $V_q = ker(d\pi_q)$ so that $J(q,p) = 0$. The ambiguity in where p is attached washes out upon forming the quotient, since the $\mathbb{S}^1$ orbit of any point is the whole π-fiber. It follows that $(b, \alpha) \mapsto (q, \pi^*\alpha_q) mod \mathbb{S}^1$ is a well-defined map $T^*B \to J^{-1}(0)/\mathbb{S}^1$. To go the other way, take any $\beta_q \in J^{-1}(0) \cap T_q^*Q$. We define a covector $\alpha = [\beta_q] \in T_b^*B$ as follows. I need to tell you what α does to a $w \in T_bB$. In order to do so, choose any $W_q \in T_qQ$ such that $d\pi_q W_q = w$ and set $\alpha(w) = \beta_q(W_q)$. Because $\beta_q(V_q) = 0$ the result is well defined independent of the choice of W_q since such a choice is well-defined modula a vertical vector $\lambda\frac{\partial}{\partial\theta} \in V_q$.

QED

Circle bundles admit connections. A connection is an $\mathbb{S}^1$ invariant complement H to V, so that

$$TQ = H \oplus V.$$

(Apologies for the double use of the symbol H: for horizontal and for Hamiltonian.) Thus, for each $q \in Q$ we have a hyperplane $H_q \subset T_qQ$. Equivalently, a connection is an $\mathbb{S}^1$-invariant one-form $A \in \Omega^1(Q)$ normalized according to

$$A(\frac{\partial}{\partial\theta}) = 1.$$

A determines H by $H = ker(A)$. , H plus the normalization determines A. Connections have curvatures:

$$dA = \pi^*F$$

where F is a two-form on B. In terms of our locally trivializing coordinates x^i, θ we have

$$A = d\theta + \Sigma A_i(x)dx^i$$

and

$$F = d(\Sigma A_i(x)dx^i.$$

thm: circle reduction

THEOREM A.3. *A choice of connection for the circle bundle $Q \to B = Q/\mathbb{S}^1$ induces a diffeomorphism*

$$(T^*Q)/\mathbb{S}^1 \cong T^*B \times \mathbb{R},$$

under which the momentum map $J : T^*Q/\mathbb{S}^1 \to \mathbb{R}$ is projection onto $\mathbb{R}$. The induced Poisson structure on $T^*B \times \mathbb{R}$ is characterized by

$$\{x^i, x^j\} = 0, \{x^i, p_j\} = \delta^i_j, \{p_i, p_j\} = JF_{ij}$$

together with

$$\{x^i, J\} = \{p_i, J\} = 0.$$

Here x^i, p_i are a standard set of canonical coordinates on T^*B induced by choosing coordinates x^i on B, and $F_{ij} = F(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$ are the components of the curvature F of the connection with respect to the x^i.

Proof. A connection A is a one-form, so for each $q \in Q$ a momentum $A_q \in T_q^*Q$. The normalization condition $A_q(\frac{\partial}{\partial \theta}) = 1$ is equivalent to $J(q, A_q) = 1$. Now $J(q,p)$ is linear in p so it follows that $J(q, p + \lambda A_q) = J(q,p) + \lambda$. Setting $\lambda = -J(q,p)$ allows us to defines a map $T^*Q \to J^{-1}(0)$ according to $(q,p) \mapsto (q, p - J(q,p)A_q)$. This map is $\mathbb{S}^1$-equivariant since A and J are $\mathbb{S}^1$-invariant. By adding back in J we arrive at an $\mathbb{S}^1$-equivariant bundle isomorphism $T^*Q \cong V^0 \times \mathbb{R} = J^{-1}(0) \times \mathbb{R}$ of vector bundles over Q , namely

$$(q,p) \mapsto ((q, p - J(q,p)A_q), J(q,p) \in J^{-1}(0) \times \mathbb{R}.$$

($\mathbb{S}^1$ acts trivially on the $\mathbb{R}$-factor.) Recall from the lemma above that $J^{-1}(0)/\mathbb{S}^1 = T^*B$. It follows that this isomorphism, upon forming its quotient by $\mathbb{S}^1$, induces an isomorphism $T^*Q/\mathbb{S}^1 \to T^*B \times \mathbb{R}$ of vector bundles over B.

Look at this isomorphism in the local section induced coordinates $x^i, \theta, p_i, p_\theta$ on T^*Q as described above, and the coordinates $x^i, \theta, p^i, p_\theta$ that they induce on $T^*Q/\mathbb{S}^1$. The coordinates x^i on B also induce coordinates on T^*Q, which, for the purposes of the proof, we write as $x^i = X^i, P_i$. In terms of these coordinates our isomorphism is given by

$$X^i = x^i$$
$$P_i = p_i - p_\theta A_i(x)$$
$$J = p_\theta.$$

To finish the proof, we compute Poisson brackets of these new functions. That $\{X_i, X_j\} = 0$ and $\{X_i, P_j\} = \delta^i_j$ is almost immediate. The only tricky bracket to compute is $\{P_i, P_j\} = \{p_i - p_\theta A_i, p_j - p_\theta A_j\} = -p_\theta\{A_i, p_j\} - p_\theta\{p_i, A_j\}$ Since $\{f, p_j\} = \frac{\partial f}{\partial x^j}$ for any function f of the x's alone this last expression equals $p_\theta(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j})$. But $F = dA = \Sigma(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_j}{\partial x^i})dx^i \wedge dx^j$ which leads to $\{P_i, P_j\} = JF_{ij}$. Since X^i, P_i are the coordinates refered to as x^i, p_i in the statement of the theorem, this completes the proof.

QED

5.1. Application to Natural Mechanical Systems. A natural mechanical system with $\mathbb{S}^1$ symmetry admits a canonical connection with horizontal spaces

$$H = V^\perp,$$

where '$\perp$ is defined using the $\mathbb{S}^1$-invariant metric ds_Q^2. Since the circle acts by isometries with respect to ds_Q^2 this choice of horizontal is $\mathbb{S}^1$-invariant and hence defines a connection.

Definition A.9. *The connection associated to this choice of metric-induced horizontal is called the "natural mechanical connection"*

Since $V_q = ker(d\pi_q)$, the restriction of $d\pi_q$ to H_q is a linear isomorphism. Declare it to be an isometry $d\pi_q : H_q \to T_bB, b = \pi(q)$, where we endow H_q with the restriction of ds_Q^2 as its inner product. This process induces an inner product on T_bB , $b = \pi(q)$, and hence a Riemannian metric ds^2_B on the base space B which, in deference to the N-body problem, we refer to as the "shape metric". To see that this inner product on T_bB is independent of the choice of $q \in \pi^{-1}(b)$ use the fact that $\mathbb{S}^1$ acts by isometries of ds^2_Q and that the $\mathbb{S}^1$-orbits are the fibers of π. We also have an $\mathbb{S}^1$ invariant metric on the fibers, summarized by the squared length of the infinitesimal generator:

$$I(q) = \langle \frac{\partial}{\partial\theta}(q), \frac{\partial}{\partial\theta}(q) \rangle_q.$$

We can summarize the above discussion by saying that we have an $\mathbb{S}^1$-invariant splitting:

{eq: split1} (105) $$TQ = \pi^* TB \oplus \mathbb{R}$$

with corresponding metric split as

{eq: split2} (106) $$ds^2_Q = \pi^* ds^2_B \oplus I d\theta^2.$$

A word or two is in order regarding π^*TB. π^*TB is the vector bundle over Q which assigns to $q \in Q$ the vector space $T_{\pi(q)}B$. Then $H \cong \pi^*TB$ with isomorhism $d\pi_q$ restricted to H_q. We write the inverse of this linear isomorphism as

$$h_q : T_bB \to Hor_q$$

and we call it the horizontal lift operator.

An alternative expression for the conserved angular momentum J, valid upon using the Legendre transformatio to identify TQ and T^*Q is

{eq: tangent J} (107) $$J(q, v) = \langle v, \frac{\partial}{\partial\theta} \rangle_q$$

J is an $\mathbb{S}^1$ invariant, so can be thought of as a function on the quotient space $(T^*Q)/\mathbb{S}^1$. In this way $J_q = J(q, \cdot)$ becomes an $\mathbb{S}^1$-invariant one-form on Q.

Lemma A.3. *The connection one-form for the natural mechanical connection is given by*

$$A(q)(v) = \frac{1}{I(q)} J(q, v)$$

where J denotes the angular momentum as per equation (107).

Proof. Since $H = V^\perp$ and V is spanned by $\frac{\partial}{\partial\theta}$ it is clear that $ker J_q = H$. It follows that A_q and J_q are multiples of each other: $A_q = f(q)J_q$ for some non-vanishing scalar function f The multiplier f is determined by the normalization $A(\frac{\partial}{\partial\theta}) = 1$. We have $J(q, \frac{\partial}{\partial\theta}) = I(q)$ which yields $f = \frac{1}{I}$.

QED

The dual of our metric splitting (equations (105, 106)) yields the $\mathbb{S}^1$-invariant vector bundle splitting

$$T^*Q \cong \pi^* T^* B \oplus \mathbb{R}.$$

with corresponding dual quadratic form splitting as

$$K = K^*_B \oplus \frac{1}{2}\frac{J^2}{I}.$$

The total Hamiltonian has the form $H = K + V$. Using the kinetic energy splitting we can rewrite $H = K_B + \frac{1}{2}\frac{J^2}{I} + V := K_B + V_{eff}(x)$ where we have written

$$V_{eff}(x) = V_{eff,J}(x) := \frac{1}{2}\frac{J^2}{I(x)} + V(x),$$

The new function V_{eff} is called the "effective potential", hence its subscript. In our local coordinates then

$$H(x,p,J) = K_B + V_{eff,J}(x), \qquad K_B = \frac{1}{2}g^{ij}p_i p_j$$

We now use the Poisson bracket formalism to compute the equations of motion: namely, we compute $\dot{f} = \{f, H\}$ for $f = x^i, p_i, J$. Now $\{x^i, V_{eff,J}(x)\} = 0$ since $\{x^i, f(x)\} = 0$ for any function of the x^j's alone. So

$$\dot{x}^i = \{x^i, K_B\} = g^{ij}(x)p_j.$$

We move on to the momentum evolution equations:

$$\dot{p}_i = \{p_i, H\} = \{p_i, K_B\} + \{p_i, V_{eff}(x)\},$$

$$\dot{J} = 0$$

Now, $\{p_i, V_{eff}(x)\} = -\frac{\partial V_{eff}}{\partial x^i}$ while two terms arise out of $\{p_i, K_B\}$, one from the fact that $\{p_i, x^j\} = -\delta_i^j$ and the other from $\{p_i, p_j\} = JF_{ij}$. We compute that $\{p_i, K_B\} = -(\frac{\partial}{\partial x^i}\frac{1}{2}g^{km})p_k p_m + g^{km}p_k JF_{im}$. From the $\dot{x}^i$ equation, this last term equals $JF_{ik}\dot{x}^k$. Thus

$$\dot{p}_i = -(\frac{\partial}{\partial x^i}\frac{1}{2}g^{km}(x))p_k p_m + JF_{ik}\dot{x}^k - \frac{\partial V_{eff}}{\partial x^i}.$$

If we set the last two terms in the last equation to zero what remains is the geodesic equations for the shape metric. The deviation from geodesy is given by these last two terms which we may interpret as forces. These forces are the coordinate expression of the one-form $JF(\dot{b}, 0) - dV(b)$. Raising indices using the metric and recalling that the contravariant version of the geodesic equation is $\nabla_{\dot{b}}\dot{b} = 0$ we see that the full system of equations is equivalent to

{eq: Wongs}

$$\nabla_{\dot{b}}\dot{b} = -J(F(\dot{b}, \cdot)^{\#} - \nabla V_{eff,J}(b). \tag{108}$$

where J is a constant and where $\# : T^*Q \to TQ$ is the metric-induced index raising map (the inverse of the Legendre transformation).

In the case of the planar N-body problem, once we center the bodies and delete total collision we get $Q = \mathbb{E}_0 \setminus \{0\} = \mathbb{C}^{N-1} \setminus \{0\} = \mathbb{S}^{2N-3} \times \mathbb{R}^+$, $Q/\mathbb{S}^1 = \mathbb{CP}^{N-2} \times \mathbb{R}^+$. The $\mathbb{R}^+$ factor corresponds to radial dilations, is coordinatized by $r = \sqrt{I}$ and its generating vector field -dilations - has zero angular momentum. The quotient map $Q \to Q/\mathbb{S}^1$ is the Hopf fibration $\mathbb{S}^{2N-3} \to \mathbb{CP}^{N-2}$ trivially extende to the $\mathbb{R}^+$ factor by having it be the identity. The connection form induced by K and the circle action is the standard connection for the Hopf fibration. The curvature form F is the Fubini-study Kahler form on $\mathbb{CP}^{N-2}$, trivially extended to the $\mathbb{R}^+$ factor. In the case $N = 3$ of the planar N-body problem, under our shape space identifications, this two-form F is a constant multiple of the solid angle form on $\mathbb{R}^3 \setminus \{0\} \cong \mathbb{CP}^1 \times \mathbb{R}^+$. A vector interpretation of this solid angle form and determination of the constants leads immediately to the reduced 3-body equations (51) of the introduction.

as per p 165 of Tour, pf of Wong's eqns -RM

6. Noether, Symmetries, and the Momentum map.

If our Lagrangian is independent of one of the position coordinates, say q^1, then $\frac{\partial L}{\partial q^1} = 0$ and the first Euler-Lagrange equation (101) reads

$$\frac{d}{dt}(\frac{\partial L}{\partial v^1}) = 0.$$

Thus symmetry of L with respect to translation of the q^1 coordinate yields conservation of the corresponding dual momentum coordinate $p_1 = \frac{\partial L}{\partial v^1}$. This is the simplest instance of "Noether's theorem" relating symmetries to conservation laws. Recall that under the Legendre transform $\frac{\partial L}{\partial v^1} = p_1$ is in fact the momentum variable dual to q^1. A quick computation shows that our Hamiltonian $H = H(q,p)$ is also independent of q_1. Hamilton's equations (100) yield this same conservation law $\dot{p}_1 = 0$. The classical language for this situation is to say that q^1 is a "cyclic variable" for our system.

The variable p_1 plays another role here by way of its Hamiltonian vector field. Recall that $\{f, p_1\} = \frac{\partial f}{\partial q^1}$ so that the Hamiltonian vector field of the conserved coordinate yields translation of q_1, providing a basic link between the conserved quantity and the symmetry it is associated to, namely , translation of q^1.

The situation just described arises when the Lagrangian is invariant under the action of a circle $\mathbb{S}^1$ acting on configuration space. At points where the circle orbits do not degenerate to points, we can take the cyclic variable to be the angular coordinate $q^1 = \theta \in \mathbb{S}^1$ coordinatizing the circle, and supplement it with the needed number of additional coordinates $q^2, q^3, \ldots, q^n$ transverse to the circle which label the nearby circle orbits. In this way the circle action becomes $(\theta_0, q^2, \ldots, q^n) \mapsto (\theta_0 + \theta, q^2, \ldots, q^n)$. The corresponding conservation law for circular invariance of L is the momentum $p_1 = p_\theta$ and is called the "angular momentum" or "momentum map" associated to the action.

The discussion above generalizes to a general Lie group G acting on Q, so that G replaces $\mathbb{S}^1$. We suppose that this G-action leaves the Lagrangian L invariant: $L(q,v) = L(gq, dg_q v)$. Here we have written $g \in G$ and have written the corresponding transformation as $q \mapsto gq$. This transformation induces the transformation $dg_q v$ of velocities where $dg : TQ \to TQ$ denotes the differential with respect to q of the transformation $q \mapsto gq$. Under the Legendre transformation, the Hamiltonian is invariant under the corresponding cotangent lifted action $g(q,p) = (gq, (dg)_q^{-1T} p)$.

Noether's theorem gives us r independent conservation laws where $r = dim(G)$. To express these laws, let $\mathfrak{g}$ denote the Lie algebra of G, an r-dimensional vector space. We will write one conservation law for each $\xi \in \mathfrak{g}$. Write $\xi_Q(q) = \frac{d}{d\epsilon}|_{\epsilon=0}(exp(\epsilon\xi)q$ for the infinitesimal generator of the G-action on Q. Here $exp(\epsilon\xi) \in G$ is the one-parameter subgroup corresponding to $\xi \in \mathfrak{g}$. Then we set

$$J^\xi = P_{\xi_Q}$$

in other words, $J^\xi(q,p) = p(\xi_Q(q))$. This expression is linear in ξ so we can view J as a map

{eq: mom map}

$$J : T^*Q \to \mathfrak{g}^*. \tag{109}$$

DEFINITION A.10. *The map of equation (109) just defined is called the momentum map for the given G-action.*

Choosing a basis $E_1, \dots, E_r$ for $\mathfrak{g}$ gives us linear coordinates on $\mathfrak{g}^*$ and expresses J as an r-vector $J = (J^1, \dots, J^r)$ with $J^a = J^{E_a}$.

REMARK. In the previous section we had $G = \mathbb{S}^1$ and wrote J for the above p_θ, the single component of J. Now J is a vector-valued.

PROPOSITION A.2 (Noether). *Suppose, as above, that the Lagrangian is G-invariant and that its Legendre map is invertible so that the associated Hamiltonian H is defined on all of T^*Q. Then the J^ξ, $\xi \in \mathfrak{g}$ are all conserved: $\{J^\xi, H\} = 0$.*

complete! -RM

PROOF. Recall the momentum functions of ... . If ψ_t is the cotangent lift of the flow of the vector field Y we showed that $\frac{d}{dt}\psi_t^* F = \{F, P_Y\}$. Now from $J^\xi = P_{\xi_Q}$ we see that $\frac{d}{dt} g_t^* F = \{F, J^\xi\}$ where $g_t : T^*Q \to T^*Q$ is the cotangent lift of the action of $exp(t\xi)$ on Q. It follows that if F is G-invariant then $\{F, J^\xi\} = 0$ for all $\xi \in \mathfrak{g}$. But H is G-invariant since L is, and $\{H, J^\xi\} = -\{J^\xi, H\}$. QED

In the setting of a natural mechanical system, we suppose that the Lie group G acts on Q so as to preserve the metric ds^Q and the potential V. In this case the lifted G-actions preserve both L and H. Recall that the Legendre transformation sends $v \in T_q Q$ to the linear functional $\langle v, \cdot \rangle_q \in T_q^* Q$. It follows the momentum map, when viewed on the tangent bundle side by using the inverse of the Legendre transformation has the form

$$J^\xi(q, v) = \langle v, \xi_Q(q) \rangle_q.$$

EXERCISE A.15. *Let $G = SE(d) = \mathbb{R}^d \times SO(d)$ be the group of rigid motions acting on $Q = \mathbb{E}(d, N)$, the configuration space for N-bodies in d-space. Put the mass metric on $\mathbb{E}$ so as to define a metric ds_Q^2. Verify that the momentum map $\mathcal{P} = T^*\mathbb{E} \to \mathfrak{g}^*$ for the G-action on the N-body phase space $\mathcal{P}$ encodes the linear and angular momenta as its $\mathbb{R}^{d*}$ and $\mathfrak{so}(d)^*$ compenents. Here $\mathfrak{g} = \mathbb{R}^d \oplus \mathfrak{so}(d)$. Discuss the various identifications required to make this translation.*

The following is important when we get to symplectic reduction in the next section

EXERCISE A.16. *The group G has a representention on its dual Lie algebra given by $g \cdot \mu = Ad^*_{g^{-1}} \mu$ where Ad_g is the usual adjoint action (conjugation for matrix groups). Upon using this action show that the momentum map as defined above is G-equivariant $J(gp) = g \cdot J(p)$.*

`sec: symp reduc`

7. Symplectic Reduction

The momentum map (equation (109)) allows us to construct a refined version of Poisson reduction known as symplectic reduction. The symplectic reduced spaces are parameterized by $\mu \in \mathfrak{g}^*$ and can be identified with the symplectic leaves of the Poisson reduced space, at least when G is compact and acts freely. The reduced Poisson dynamics thus restricts to each reduced space, giving us a family of dynamical systems parameterized by μ.

Here is the definition. Fix $\mu \in \mathfrak{g}^*$. Form

$$J^{-1}(\mu) \subset \mathcal{P}.$$

Since J is equivariant, the stabilizer $G_\mu = \{g \in G : Ad_g^* \mu = \mu\}$ leaves $J^{-1}(\mu)$ invariant.

DEFINITION A.11. *The symplectic reduced space for the value μ is the quotient space $(T^*Q)_\mu := J^{-1}(\mu)/G_\mu$.*

There is no reason, a priori, for the symplectic reduced space to be a manifold, let alone a symplectic manifold. However, we have

PROPOSITION A.3. *If G is compact and acts freely on T^*Q then each symplectic reduced space $J^{-1}(\mu)/G_\mu$ is a smooth manifold and inherits a symplectic form ω_μ from the canonical form on T^*Q.*

We will go through the details leading to this proposition in the remainder of this section.

7.1. Linear Symplectic Reduction. A symplectic vector space $\mathbb{V}$ is a real vector space endowed with a constant symplectic form ω_0. Thus our phase space is a symplectic vector space, as is the tangent space to any symplectic manifold. If $\mathbb{L} \subset \mathbb{V}$ is a linear subspace, then by its reduction we mean the quotient space $\mathbb{L}/ker(\omega_{\mathbb{L}})$ where $\omega_{\mathbb{L}}$ is the restriction of ω to $\mathbb{L}$ and its kernel consists of all $v \in \mathbb{L}$ such that $\omega(v, w) = 0$ for all $w \in \mathbb{L}$. Equivalently, $ker(\omega_{\mathbb{L}}) = \mathbb{L} \cap \mathbb{L}^\perp$ where by $\mathbb{L}^\perp \subset \mathbb{V}$ we mean the ω-perpindicular to $\mathbb{L}$, that is, the set of all $v \in \mathbb{V}$ such that $\omega(v, w) = 0$ for all $w \in \mathbb{L}$. As per the inner product situation, $\mathbb{L}^\perp$ is a subspace of $\mathbb{V}$ whose dimension is complementary to $\mathbb{L}$'s. The form ω induces a linear symplectic form on the quotient in the standard way.

ssec: local reduction

7.2. Local nonlinear symplectic reduction. Moving on to the non-linear world, if $(\mathcal{P}, \omega)$ is a symplectic manifold and $S \subset \mathcal{P}$ a submanifold, then $\omega_S = j^*\omega$ is a closed two-form on S whose rank may vary. Here $j : S \to P$ is the inclusion. At points where this rank is locally constant the kernel has constant rank and defines an involutive distribution, so we can form the local leaf space $S/ker\omega_S$. This local quotient space is the local symplectic reduction of S.

Words are in order regarding the meaning of the "local quotient" of a manifold M by an involutive distribution D. The Frobenius normal theorem asserts the following normal form for D: every point of M is contained in a coordinate neighborhood $U \cong \mathbb{R}^k \times \mathbb{R}^\ell$ where k is the rank of D and under which the distribution D corresponds to the $\mathbb{R}^k$ directions. Thus, there are coordinates $u^a, v_i, a = 1, \ldots, k, i = 1, \ldots, \ell$ with respect to which D is defined by $dv_i = 0, i = 1, \ldots, \ell$. The plaques (local leaves) are the integral manifolds $v_i = const_i$ for $D|_U$, which is to say the image of any one of the k-planes $\mathbb{R}^k \times pt$ under the Frobenius normal form diffeomorphism. Whenever a leaf of D intersects U it does so in a union of these plaques. The plaques themselves are the connected components of the sets obtained by intersecting a leaf with U. In particular the quotient $U/D|_U$ consists of the space of plaques in U and is diffeomorphic to $\mathbb{R}^\ell$ as coordinatized by the v_i. Any one of these spaces is what we mean by a local quotient of M by D.

As a primary example of this reduction process, suppose that H is a smooth function on the phase space $\mathcal{P}$ and that $c \in \mathbb{R}$ is a regular value for H. Then the set $S := \{H = c\} \subset \mathcal{P}$ is a smooth embedded hypersurface and kernel of ω_S is spanned by X_H. A local symplectic reduction of $\{H = c\}$ is a space whose points are local integral curves for X_H which lie within $\{H = c\}$. To form the local quotient invoke the the symplectic version of the straightening lemma, which yields canonical coordinates $q^1, p_1, q^2, p_2, \ldots, q^n, p_n$ such that $H = p_1$ and those under which the local integral curves of X_H are the q^1-curves. It follows that $q^2, p_2, \dots, q^n, p_n$ coordinatize the local symplectic quotient.

More viscerally, choose any disc $D \subset S$ transverse to X_H. This disc must be symplectic: the restriction of ω to the disc is symplectic, since $ker(\omega_S)$ is spanned by X_H. Use the flow of X_H to form a tubular neighborhood of D diffeomorphic to $D \times (-\epsilon, \epsilon)$ with the $(-\epsilon, \epsilon)$ direction corresponding to the X_H direction. This process realizes the local symplectic quotient $S/ker(\omega_S)$ as D. Our description of the symplectic slice for realizing the Poincaré map can be viewed as an instantiation of this construction.

7.3. Symplectic reduction, redux. Return to our definition of the symplectic reduced space. First let us unwind the relations between $J^{-1}(\mu)$ being a manifold and the G action being free. A group action of a Lie group G on a manifold $\mathcal{P}$ is called *locally free* at $z \in \mathcal{P}$ if the infinitesimal generator map $\mathfrak{g} \to T_x\mathcal{P}$ is injective.

Exercise A.17. *Verify that "free" implies "locally free.*

Verify that μ is a regular value of the momentum map J if and only the G-action on $\mathcal{P}$ is locally free at points $z \in J^{-1}(\mu)$.

Thus, if G acts freely on $\mathcal{P}$ then each value $\mu \in \mathfrak{g}^*$ is a regular value for J and hence each $J^{-1}(\mu)$ is a smooth embedded submanifold of $\mathcal{P}$. Write $i_\mu : J^{-1}(\mu) \to \mathcal{P}$ for the inclusion.

Exercise A.18. *Verify that the leaves of the kernel of the two-form $i_\mu^*\omega$ on $J^{-1}(\mu)$ consists of the connected components of the orbits of the stabilizer G_μ.*

Write $\pi_\mu : J^{-1}(\mu) \to J^{-1}(\mu)/G_\mu$ for the projection, assuming that G_μ acts in such a way that the quotient space is indeed a manifold and the quotient projection a smooth submersion. Then $J^{-1}(\mu)/G_\mu$ is a manifold whose charts realize the local symplectic quotienst of subsection 7.2. It comes with a symplectic form ω_μ characterized by $i_\mu^*\omega = \pi_\mu^*\omega_\mu$.

If G is compact and acts freely then the same is true of G_μ and automatically the hypothesis of the above paragraph hold. The symplectic reduced space $J^{-1}(\mu)/G_\mu$ is indeed a smooth symplectic manifold with symplectic form ω_μ.

Finally, we asserted that the reduced space can be identified with a symplectic leaf of the Poisson reduced space $\mathcal{P}/G$. In order to do so consider the following diagram

$$\begin{array}{ccccc}
J^{-1}(\mu) & \longrightarrow & GJ^{-1}(\mu) = J^{-1}(G\mu) & \longrightarrow & \mathcal{P} \\
\downarrow{\scriptstyle G_\mu} & & \downarrow{\scriptstyle G} & & \downarrow{\scriptstyle G} \\
J^{-1}(\mu)/G_\mu & \longrightarrow & J^{-1}(G\mu)/G & \longrightarrow & \mathcal{P}/G
\end{array}$$

The horizontal arrows are inclusions. The vertical arrows are quotients, first by G_μ and then after that by G. It is not a hard job to show that the composite bottom arrow $J^{-1}(\mu)/G_\mu \to J^{-1}(G\mu)/G$ is injective and its image is a symplectic leaf, at least in the case when the G-action is free. The subsets $G\mu \subset \mathfrak{g}^*$ are the co-adjoint orbits, and known to be the symplectic leaves for $\mathfrak{g}^*$ with its standard Poisson structure (either left or right).

This diagram becomes particularly simple for the circle since it is an Abelian group. In this case $G\mu = \mu$ and $G_\mu = G = \mathbb{S}^1$. We return to our earlier picture as described in theorem A.3 of reduction by the circle. The symplectic reduced spaces can be identified with $T^*(Q/G)$, but with the standard symplectic structure "twisted" by J times the curvature form as per the earlier description of the Poisson manifold there.

When the G action is not free, both the Poisson and symplectic reduced spaces will typically be singular algebraic varieties and all kinds of interesting complications come in to play.

8. Notes

IN To NOTES These leaves are known as the symplectic reduced spaces, or Marsden-Weinstein-reduced spaces, or Meyer-Marsden-Weinsten reduced spaces.

make notes, organize... -RM

For us, "reduction" means the process of working with these quotient flows and understanding the various structures on the quotient space which are induced by the natural mechanical system on Q. We will describe details of reduction when $G = \mathbb{S}^1$ as is appropriate for the planar N-body problem. At the end we make some remarks regarding the case $d > 2$ of $SO(d)$ appropriate for the N-body problem in d-space.

put in Notes to this Appendix ! -RM

Symplectic and then Poisson reduction was my bread and butter as a graduate student. Here, I will not take these tacks, tacks that I was taught through chapter 4 of Abraham-Marsden [**?**] and subsequent papers by Marsden, Weinstein, Guilemin and Sternberg REFS ??. Instead I opt for a decidedly dynamical approach, pushing symplectic and Poisson structures off to the side for a while.

APPENDIX B

The direct method of the calculus of variations

We can exploit the Lagrangian formalism by looking for minimizers of the action over some class of paths. By carefully specifying the class we can sometimes achieve solutions with desired properties. This method of obtaining solutions is known as the direct method in the calculus of variations.

The simplest version of the direct method is to the fixed end point problem.

PROBLEM B.1. *Fix* $q_0, q_1 \in \mathbb{E}$ *and a length of time* T. *Minimize the action* $A(q(\cdot))$ *over all paths* $q : [0,T] \to \mathbb{E}$ *such that* $q(0) = q_0$ *and* $q_1 = q(T)$.

Let us walk through the method for the fixed endpoint problem.

A bit of functional analysis is at the heart of the method. Define the function space Ω to be the space of all absolutely continuous paths $q : [0,T] \to \mathbb{E}$ satisfying the endpoint conditions $q(0) = q_0$ and $q(T) = q_1$ and having finite action: $A(q) < \infty$. Set

$$a = \inf_{c \in \Omega} A(c).$$

By the definition of infimum there exists a sequence $c_n \in \Omega$ such that $\lim_{n\to\infty} A(c_n) = a$. We call such a sequence is called a "minimizing sequence". The *direct method of the calculus of variations* now proceeds by completing the following steps.

STEP 1. Show that a subsequence of the sequence c_n converges to some continuous curve c_*.

STEP 2. Show that $c_* \in \Omega$.

STEP 3. Show that $A(c_*) = a$.

STEP 4. Show that A restricted to Ω is differentiable at c_* with derivative 0.

STEP 5. Conclude from step 4 and some version of the "fundamental lemma of the Calculus of variations" that c_* satisfies Newton's equations.

Fundamental lemma HERE..

The typical element of Ω does not have a continuous derivatives so Newton's equations does not make sense for such an element. The fact that the critical points automatically are twice-differentiable can be viewed as a very primitive version of "elliptic regularity" in PDE. Ω is not a vector space (unless $q_0 = q_1 = 0$). Rather it is an affine space.

To complete these steps we need some assumptions on the potential V. To make our life simple we will assume that V is smooth and that $-V(x) \geq 0$. Later on we must relax the 'smooth' assumption to allow for the N-body potential which blows up along Δ.

STEP 1.

Since $-V \leq 0$ we have that

{actionbd}
$$\frac{1}{2}\|\dot{q}\|_{L_2}^2 := \frac{1}{2}\int_0^T |\dot{q}(u)|^2 du \leq A(q). \tag{110}$$

Now absolutely continuous have derivatives a.e. Indeed, they are precisely those functions $c : I \to \mathbb{E}$ for which the fundamental theorem of calculus holds:

$$c(t) - c(s) = \int_s^t \dot{c}(u)du$$

[See eg Royden p number ...!] Use the Cauchy-Schwartz inequality in the form $\int fg \leq \sqrt{\int f^2}\sqrt{\int g^2}$ with $f = \|\dot{c}\|$ and $g = 1$ to get :

$$\int_s^t \|\dot{c}\|dt \leq \sqrt{|t-s|}\sqrt{\int_s^t \|\dot{c}(u)\|^2 du}$$

Now use $\|c(t) - c(s)\| = \|\int_s^t \dot{c}du\| \leq \int_s^t \|\dot{c}(u)\|du$ and the action bound eq [110], to conclude our basic inequality:

{star}
$$\|c(t) - c(s)\| \leq \sqrt{|t-s|}\sqrt{2A(c)}, \tag{111}$$

This inequality asserts that any family of curves c with bounded action is equicontinuous. Now if $c \in \Omega$ we have that $c(0) = q_0$ and so, by taking $s = 0$ and $t \leq T$ in the basic inequality we find that

{positionbd}
$$\|c(t) - q_0\| \leq \sqrt{T}\sqrt{2A(c)} \tag{112}$$

which asserts that the set of curves $c \in \Omega$ for which $A(c) \leq M$ is also a bounded family. Thus this set of curves is a bounded equicontinuous family and hence one to which the Arzela-Ascoli theorem applies. In particular, we can take $M = 2a$ for our minimizing sequence to see that every minimizing sequence has a C^0-convergent subsequence. We would like to say that this c_* is the limit of Step 1. But there is some more work to do.

H^1 and weak convergence. By H^1 we will mean the space of all absolutely continuous paths $q : [0,T] \to \mathbb{E}$ such that $\|\dot{q}\|_{L_2}^2 < \infty$. Now $N(q) = \|\dot{q}\|_{L_2}$ is not a norm on H^1 since any constant path has $N(q) = 0$. The standard norm on H^1 is the one whose square is

$$\|q\|_{H^1}^2 = \int |\dot{q}(s)|^2 ds + \int |q(s)|^2 ds = \|\dot{q}\|_{L_2}^2 + \|q\|_{L_2}^2.$$

EXERCISE B.1. *A) Show that the standard H^1 norm is Lipschitz equivalent to the norm whose square is $(\int_0^T |\dot{q}(s)|^2 ds) + |q(0)|^2$. B) Show that $q \in H^1$ if and only if q is absolutely continuous and $A(q) < \infty$.*

We now recall the Banach-Alaoglu theorem in the context of Hilbert spaces. any norm-bounded sequence in a Hilbert space H has a *weakly* convergent subsequence. We also recall the meaning of "weakly convergent", We say that a sequence $v_n \in H$ weakly converges to v, written $v_n \rightharpoonup v$. if for any continuous linear functional $\ell : H \to \mathbb{R}$ we have $\ell(v_n) \to \ell(v)$.

Basic example. Let $\{e_n\}$ be an orthonormal basis for the separable Hilbert space H. Then $e_n \rightharpoonup 0$, since any linear function ℓ has the form $\ell(v) = \langle u, v\rangle$ for some $u \in H$ and in our case $\ell(e_n) = u_n$ is the nth coordinate of u relative to our basis. Since $\|u\|^2 = \Sigma u_n^2$ we have that $u_n \to 0$. Note that we do not have that $e_n \to 0$ in the standard topology of H, since $\|e_n - 0\| = 1$!

By our basic bound, eq [111] and the bound of eq [112] we see that any $c \in \Omega$ we have that

$$\|q\|_{H^1}^2 \leq 2A(q) + TB^2; B = |q_0| + \sqrt{2A(q)T}.$$

In particular, our minimizing sequence is eventually H_1 bounded, and so has a weakly convergent subsequence.

LEMMA B.1. *The inclusion $H^1 \to C^0$ is weakly continuous.*

Proof. Fix $t \in [0, T]$. Then the evaluation map $q \mapsto q(t)$ is a linear map $H^1 \to \mathbb{E}$. It suffices to show that it is a continuous linear functional with a fixed bound, independent of t. Indeed, from our basic inequality

$$|q(t)| = |q(t) - q(0) + q(0)| \leq |q(t) - q(0)| + |q(0)| \leq \|\dot{q}\|_{L_2} + |q_0|$$

Now from $a + b \leq \sqrt{2(a^2 + b^2)}$, for $a, b > 0$ and Exercise A we have that $|q(t)| \leq \sqrt{2}L\|q\|_{H_1}$ which is continuity of the linear map.

Returning to our minimizing sequence we have that a subsequence of $\{c_n\}$, relabeled so as to be called c_n, satisfies:

$$c_n \rightharpoonup c_* (\text{ weak-}H^1), \text{ and } c_n \to c_* \ (C^0).$$

STEP 2. $c_* \in \Omega$. Indeed $c_* \in H^1$, and since $c_n \to c_*$ in the continuous topology and $c_n(0) = q_0, c_n(T) = q_1$ for all n, we have that $c_*(0) = q_0, c_*(T) = q_1$.

STEP 3. $A(c_*) = a$.

Since $c_* \in \Omega$ and since a is the infimum of A over Ω we have that $A(c_*) \geq a$. We show that $A(c_*) \leq a$. For this it suffices to show that $\lim_n A(c_n) \geq A(c_*)$, since $a = \lim_n A(c_n)$.

Since V is continuous and $c_n \to c_*$ in C^0 we have that the potential terms converge: $-\int_0^T V(c_n(s))ds \to -\int_0^T V(c_*(s))ds$. To deal with the kinetic term we need to use weak convergence. With this in mind, we set:

$$\langle\langle v, w\rangle\rangle = \int_0^T \langle \dot{v}(t), \dot{w}(t)\rangle dt, v, w \in H^1.$$

so that the integral of the kinetic term in the Lagrangian is

$$\int K(\dot{c}(t))dt = \frac{1}{2}\langle\langle c, c\rangle\rangle.$$

Now $\langle\langle \cdot, \cdot\rangle\rangle$ is "half" of the inner product (either one) which yields the H^1 norm squared. As a result, any $c \in H^1$ defines a continuous linear functional $h \mapsto \langle\langle c, h\rangle\rangle$ on H^1 (with the norm of this functional bounded by $\|c\|_{H^1}$). Take $c = c_*$ and use $c_n \rightharpoonup c_*$ to conclude that

{starstar} (113)
$$\langle\langle c_*, c_n\rangle\rangle \to \langle\langle c_*, c_*\rangle\rangle. \tag{113}$$

Now the quadratic form $\langle\langle \cdot, \cdot\rangle\rangle$ is nonnegative so that

$$0 \le \langle\langle c_* - c_n, c_* - c_n\rangle\rangle = \langle\langle c_*, c_*\rangle\rangle - 2\langle\langle c_* *, c_n\rangle\rangle + \langle\langle c_n, c_n\rangle\rangle$$

Taking lim infs and using equation (113) we get that

$$0 \le -\langle\langle c_*, c_*\rangle\rangle + \langle\langle c_n, c_n\rangle\rangle$$

or

$$\langle\langle c_*, c_*\rangle\rangle \le \lim_n \langle\langle c_n, c_n\rangle\rangle$$

It follows that the kinetic part of the action satisfies

$$\int K(\dot{c}_*(t)) \le \int K(\dot{c}_n(t))$$

Combined with the result $-\int_0^T V(c_n(s))ds \to -\int_0^T V(c_*(s))ds$. we have proven that $A(c_*) \le a = \lim_n A(c_n)$.

QED for step 3.

STEP 4. The differential of a function A at a point c_* in the direction h is defined as

$$dA(c_*)(h) = \frac{\partial}{\partial \epsilon} A(c_* + \epsilon h)|_{\epsilon=0}.$$

In our case

$$A(c_* + \epsilon h) = \frac{1}{2}\langle\langle c_* + \epsilon h, c_* + \epsilon h\rangle\rangle - \int_0^T V(c_*(t) + \epsilon h(t))dt$$

This first kinetic term is a bilinear expression and so the standard "differentiate a dot product" rule holds and yields $\langle\langle c_*, h\rangle\rangle$ for derivative. Since the expression $c_* + \epsilon h$ is smooth in ϵ and since V is C^1 we can use the chain rule to differentiate the second term, arriving at its differential is $-\int_0^T dV(c_*(t))h(t)dt = -\int_0^T \langle \nabla V(c_*(t)), h(t)\rangle dt$. In sum:

$$dA(c_*)(h) = \langle\langle c_*, h\rangle\rangle - \int_0^T \langle \nabla V(c_*(t)), h(t)\rangle dt.$$

showing that A is indeed differentiable at any path $c_* \in H^1$, and that this derivative is continuous as a function from H^1 to its dual space H^1 (use the weak topology on the dual!).

Now set

$$\Omega_0 = \{h \in H^1 : h(0) = h(T) = 0\}.$$

This is a closed linear subspace of H^1 and is the tangent space of the (affine) space Ω. In particular, since $c_* \in \Omega$ whenever $h \in \Omega_0$ we have that $c_* + \epsilon h \in \Omega$.

We now argue by contradiction that $dA(c_*)(h) = 0$ for all $h \in \Omega_0$. Suppose not. Then there is an $h \in \Omega_0$ such that $dA(c_*)(h) \ne 0$. By switching the sign of h and scaling if necessary, we can assume that $dA(c_*)(h) = -1$. Taylor expanding $A(c_* + \epsilon h)$ with respect to ϵ yields $A(c_* + \epsilon h) = A(c_*) - \epsilon + O(\epsilon^2) = a - \epsilon + O(\epsilon^2)$

which shows that we can lower the value of $A(c)$ to below a while staying in Ω. This contradicts the fact that $A(c_*) = a$ is the infimum of A over Ω.

Remark. This step is the standard argument from first quarter calculus!

STEP 5. If $dA(c_*)(h) = 0$ for all $h \in \Omega_0$ then c_* satisfies Newton's equation and in particular c_* is C^2. (!)

Remark. The "Euler-Lagrange equations" for L, which are precisely Newton's equations, express the condition $dA(c_*)(h) = 0$ for all $h \in \Omega_0$ *assuming* that c_* is C^2. This existence of continuous 2nd derivatives is used to integrate [**??**] by parts and so derive the Euler Lagrange equations. The main point in step 5 is that we do not, a priori, know that the minimizer is C^2 , so this integration by parts step is illegal.

Remark. This last step is "elliptic regularity" in its simplest setting.

The proof of the validity of step 5 is based on the

THEOREM B.1 (Fundamental Lemma of the Calculus of Variations). *Write* $\langle\langle v, w\rangle\rangle = \int_0^T \langle \dot{v}(t), \dot{w}(t)\rangle dt$ *for* $v, w : [0,T] \to \mathbb{E}$. *If* $c_* \in H^1$ *and* $\langle\langle c_*, h\rangle\rangle = 0$ *for all* $h \in \Omega_0$ *then* $\dot{c}_*$ *is a constant path, and so* $\ddot{c}_* = 0$.

For alternate proofs and an enlightened, beautiful and exceptionally clear statement around the Fundamental Lemma, please see p. 18 of the text by LC Young.

PROOF OF THE FUNDAMENTAL LEMMA. We use Fourier series. Since $\dot{c}_*$ is square integrable we can expand it as an L_2 convergent cosine series:

$$\dot{c}_* = A_0 + \Sigma_{k=1}^{\infty} a_k \cos(2\pi k/T); A_0, a_k \in \mathbb{E}$$

with

$$\langle\langle c_*, c_*\rangle\rangle = T|A_0|^2 + (T/2)\Sigma|a_k|^2$$

The functions $\sin(2\pi k/T)$, $k = 1, 2, \ldots$ all vanish at 0 and T, so, for any fixed $e \in \mathbb{E}$, we have that $h(t) = \sin(2\pi kt/T)e$ is a function in $h \in \Omega_0$. Now $\dot{h} = 2\pi k/T \cos(2\pi kt/T)e$ and, so, by the orthogonality of the cosines we have that for this h:

$$\langle\langle c_*, h\rangle\rangle = \pi\langle a_k, e\rangle$$

But by assumption, this latter is zero for all $k = 1, 2, \ldots$ and all $e \in \mathbb{E}$. It follows that all the nonzero Fourier series a_k of $\dot{c}_*$ are zero and hence $\dot{c}_* = A_0 = const.$, as claimed.

Proof of Validity of Step 5. We explain the logic, working backwards. If we integrate Newton's equations $\ddot{q} = -\nabla V(q)$ we get $\dot{q}(t) = -\int_0^t \nabla V(q(s))ds + v_0$ where v_0 is a constant.

Set

$$G(t) = \int_0^t \nabla V(q(s))ds.$$

so that $dG/dt = \nabla V(q(t))$. Use $\frac{d}{dt}\langle G(t), h(t)\rangle = +\langle \nabla V(q(t)), h(t)\rangle + \langle G(t), \dot{h}(t)\rangle$ Integrate by parts, to obtain that

$$\int \langle G(t), \dot{h}\rangle(t)dt = -\int \langle \nabla V(q(t)), h(t)\rangle dt$$

where the boundary term vanishes upon integration by parts because $h(0) = 0 = h(T)$. Set

$$\dot{\beta} = \dot{c}_* + G(t)$$

and observe that $\beta \in H_1$. The condition $dA(c_*)(h) = 0$ for all $h \in \Omega_0$ now reads $\langle\langle \beta, h\rangle\rangle = 0$ for all $h \in \Omega_0$. By the Fundamental lemma, we have that $\dot{\beta} = const. = v_0$, which means that $\dot{c}_* = -\int_0^t \nabla V(c_*(s))ds + v_0$ where v_0 is a constant vector. Now c_* is absolutely continuous, and hence so is $\nabla V(c_*)$ and thus the integral $G(t)$ is differentiable with continuous derivative. This shows that $\dot{c}_*$ is continuously differentiable, with derivative equal to $-\nabla V(c_*(s))$, concluding the proof.

(!! Yay!!)

APPENDIX C

Braids, Homotopy and Homology

sec:topology

Two closed curves in a space X are said to be freely homotopic if you can deform one into the other while staying within the space X. When defining the fundamental group of X we fix a base-point $x_0 \in X$ and we insist that all loops pass through x_0 at time $t = 0$. An element of the fundamental group is a homotopy classes of such a based loops. In free homotopy neither the loops nor the homotopies need pass through the base point. The set of all free homotopy classes of loops $S^1 \to X$ is denoted $[S^1, X]$ and can be canonically identified with the space of conjugacy classes of the group $\pi_1(X)$ when X is path-connected. This identification is achieved by taking a free loop $c : S^1 \to X$ and adding to it a leg connecting the base point to $c(0)$, travelling c, and then returning to the base point by travelling backward along the leg. [See figure]. In this way we get a composition of surjective maps

relegate this topological section to an appendix. (??) -RM

$$\pi_1(X) \to [S^1, X] \to H_1(X).$$

' The last space is the 1st homology group of X with integer coefficients. The composition of these two maps $\pi_1(X) \to H_1(X)$ is a surjective group homomorphism sometimes called the "Hurewicz map". This map "Abelianizes" the fundamental group by sending every commutator $aba^{-1}b^{-1}$ to $0 \in H_1(X)$.

Take $X = \mathbb{C}^N \setminus \Delta$, the configuration space of the collision-free planar N-body problem. Then $P_N = \pi_1(X)$ is the pure braid group. The above sequence then becomes

$$P_N \to [S^1, \mathbb{C}^N \setminus \Delta] \to \mathbb{Z}^{\binom{N}{2}}$$

P_N has $\binom{N}{2}$ standard generators, one for each pair ab of bodies, and denoted A_{ab} by Birman. A "tight binary" - a loop in which masses a and b make one full counterclockwise revolution around each other while the other masses stay still and far away represents this generator. Equivalently, this is a small loop in $\mathbb{C}^n$ encircling the binary collision plane $\Delta_{ab} \subset \mathbb{C}^N$ once, while avoiding all the other binary collision planes of $\Delta = \cup_{ij} \Delta_{ij}$. Mapped to homology, these generators form a standard integer basis for $H_1(\mathbb{C}^N \setminus \Delta) \cong \mathbb{Z}^{\binom{N}{2}}$ and the winding numbers of edge $q_a - q_b$ realize the coefficients of the homology class of a loop $q(t)$ relative to this basis.

PICTURE

The usual braid group is related to the pure braid group via an exact sequence

$$1 \to P_N \to B_N \to S_N \to 1$$

where S_N is the symmetric group on N letters. S_N acts freeling on $\mathbb{C}^N \setminus \Delta$ by interchanging mass labels and the quotient $(\mathbb{C}^N \setminus \Delta)/S_N$ is the configuration space of N indistinguishable but non-colliding particles. We have that $B_N = \pi_1(\mathbb{C}^N \setminus \Delta)/S_N)$. Generators of B_N can be obtained by taking the square roots σ_{ab} of the standard generators A_{ab}. The σ_{ab} map to the element $(ab) \in S_N$ which transposes a

and b. Not all are needed to generate. In the usual representation of the braid group one just takes $N-1$ of these $\binom{N}{2}$ generators, for example $\sigma_i = \sigma_{i,i+1}, i = 1\ldots, N-1$.

Instead of forming the quotient of $\hat{\mathbb{E}}_0 = \mathbb{C}^N \setminus \Delta$ by the group S_N of particle permutations we can form the quotient by the group of isometries and get shape space, which we have seen deformation retracts onto $\mathbb{CP}^{N-2} \setminus \Delta)$ The rotation part of the group of isometries generates the center of the braid group. In this way the sequence of spaces $S^1 \to \hat{\mathbb{E}}_0 = \mathbb{C}^{N-1} \setminus \Delta \to$ *Shapespace* leads to the sequence of groups $\mathbb{Z} \to P_N \to \mathbb{P}P_N := P_N/\mathbb{Z}$ where $\mathbb{Z} = \pi_1(S^1)$ is the center of P_N. This sequence of spaces deformation retracts to the Hopf fibration $S^{2N-3} \to \mathbb{CP}^{N-2}$, with collisions deleted. Deleting any one of the binary collision planes allows us to trivialize the Hopf fibration so that $\mathbb{S}^{2N-3} \setminus \Delta \cong \mathbb{S}^1 \times (\mathbb{CP}^{N-2} \setminus \Delta)$. It follows that

$$P_N = \mathbb{P}P_N \times \mathbb{Z}.$$

See corollary 3.4 of **??** USE: THE CENTER OF SOME BRAID GROUPS AND THE FARRELL COHOMOLOGY OF CERTAIN PURE MAPPING CLASS GROUPS YU QING CHEN, HENRY H. GLOVER AND CRAIG A. JENSEN. for an alternative proof of this fact regarding splitting the pure braid group.

Rotate a regular N gon one full revolution. Its vertices sweep out a pure braid which generates the center of the braid group on N strands.

Definition C.1. *The projective pure braid group, denoted $\mathbb{P}P_N$ is the quotient $P_N/\mathbb{Z}$of the pure braid group P_N by its center $\mathbb{Z}$, this center being generated by applying a full rotation to the initial configuration of N points.*

The Lagrange solution, projected to the shape sphere is trivial homotopically: it is a constant loop. A loop in the shape sphere S^2 represents a *relative* periodic curve: a curve which closes up modulo a rotation. If that loop avoids the three binary collision points Δ then the curve upstairs in configuration space is collision free. More generally we have

Lemma C.1. *The projective pure braid group $\mathbb{P}P_N$ is the fundamental group of $\mathbb{CP}^{N-2} \setminus \Delta$ - the collision-free shape space for the planar N-body problem.*

equivariant braids here? - RM

******* to rewrite or delete:

for $N = 3$

The explicit form of this map can be expressed by realizing that the 'tight binaries' - the loops where masses i and j circle each other once while all other masses stay still - form a generator for P_N. We will call these generators (ij). (Birman [**14**] writes them as A_{ij}.) For $N = 3$ they are subject to the single relation $(12)(23)(31) = Id$. Now write an eclipse sequence in terms of generators. For example $123123 = (12)(31)(23)$. To see where it maps to in homology, let the generators commute: $(12)(31)(23) = (12)(23)(31)$ which is trivial, being the class of the identity which is zero in homology. This computation shows that the figure eight's class, represented by 123123, is homologically trivial, a fact which can also be seen by drawing it on the shape sphere.

APPENDIX D

The Jacobi-Maupertuis metric

App: JM

The Jacobi-Maupertuis [JM] metric is a family of Riemannian metrics parameterized by energy E whose geodesics are *almost* in bijection with solutions to Newton's equations having energy E. We describe the metrics and their properties here. We will focus some attention on the "almost" nature of this bijection which concerns brake orbits and collisions and is rarely if ever treated in textbooks.

The JM trick works for any natural mechanical system (definition **??**). Such a system is specified by a Riemannian metric $\langle \cdot, \cdot \rangle$ and a function V on a manifold Q. Newton's equations for the system are

$$\nabla_{\dot{q}} \dot{q} = -\nabla V(q)$$

where ∇ is the Levi-Civita connection of the Riemannian metric. Newton's equations are equivalent to Euler-Lagrange equations for the Lagrangian $L = K - V$ and to Hamilton's equations for the Hamiltonian $H = K + V$ which is the Legendre transform of L. Here K is the kinetic energy for the metric is $K(q, v) = \frac{1}{2}\langle v, v \rangle_q$. The Legendre transform for L is the metric induced isomorphism $TQ \to T^*Q$. When we wrote

$$H(q, p) = K(q, p) + V(q))$$

as being the Legendre transform of L we introduced an abuse of notation by using the same symbol K for the quadratic form on T^*Q induced by the cometric to our metric. This other K is the Legendre-transform of the K original $K : TQ \to \mathbb{R}$. If we write the Riemannian metric in the traditional coordinate form $ds_K^2 = \Sigma g_{ij}(x) dx^i dx^j$ and if v^i are the usual coordinate-induced fiber-velocity coordinates according to $v = \Sigma v^i \frac{\partial}{\partial x^i}$ then $K = \frac{1}{2}\Sigma g_{ij}(x) v^i v^j = \frac{1}{2}\Sigma g^{ij}(q) p_i p_j$ where $g^{ij}(x)$ is the matrix inverse to $g_{ij}(x)$ and the Legendre transformation is $p_i = \Sigma g_{ij}(x) v^j$. Newton's equations for the N-body problem form the model example of a natural mechanical system. Then the metric is the mass metric and the potential is the usual N-body potential.

Energy is conserved along solutions to Newton's equations so that Hamilton's equations induces a flow on the constant energy hypersurface $\{(q, p) : H(q, p) = E\} \subset T^*Q$. Since $K \geq 0$ and $E = K + V = K - U$ the projection of this hypersurface to the configuration space Q forms the the Hill region, the domain

$$\Omega_E := \{q : E + U(q) \geq 0\} \text{ where } U = -V.$$

An energy E solution $q(t)$ to Newton's equations must lie in the Hill region.

def: Hill region

Definition D.1. *The region Ω_E is called the Hill region for energy E.*

The Hill boundary for energy E is the subset $\partial\Omega_E : \{q : E + U(q) = 0\} \subset \Omega_E$.

A brake orbit is an energy E solution $q(t)$ to Newton's equations which touches the Hill boundary at some instant.

We allow $U(q) = \infty$, calling these q's "collision points".

The interior of the Hill region Ω_E *is the the locus* $\{q : 0 < E + U(q) < \infty\} \subset \Omega_E$ *so coincides with the Hill region minus the Hill boundary and the collision locus.*

The JM *metric at energy* E *is the the Riemannian metric*

$$ds_E^2 = \lambda_E ds_K^2, \text{ with conformal factor } \lambda_E(q) := 2(E + U(q)).$$

and domain the interior of the Hill region.

PROPOSITION D.1. *Geodesics for the JM metric* ds_E^2 *at energy* E *are reparameterizations of collision-free, brake-free energy* E *solutions to Newton's equations. The reparameterizaton is given by* $ds = \lambda_E(q(t))dt$ *where* t *is Newtonian time* t *and* s *is is the JM arclength.*

The proof follows from a simple but powerful lemma in symplectic geometry.

LEMMA D.1. *Let* (P, ω) *be a symplectic manifold. Suppose that* $H, F : P \to \mathbb{R}$ *are two smooth functions on* P *which share a common level set:*

$$\{H = c_H\} = \{F = c_F\}$$

where the constants c_H, c_F *are regular values for the respective functions so that the shared level set is a smooth hypersurface in* P*. Then, on this hypersurface, the solutions to Hamilton's equations for* H *and for* F *are reparameterizations of each other.*

PROOF OF LEMMA. Recall that ω, being non-degenerate, defines a fiber linear isomorphism $T^*P \to TP$. The Hamiltonian vector field X_F of a function F on P is obtained by applying this isomorphism to the one-form dF, according to $\omega(X_F, \cdot) = dF$. It follows that if $dF(\zeta) = c dH(\zeta)$ at some point $\zeta \in P$ and for some constant c, then $X_F(\zeta) = cX_H(\zeta)$.

Write $\Sigma \subset P$ for the hypothesized common level set of the lemma. Since the constants are regular values Σ is a smooth hypersurface whose tangent space at $\zeta \in \Sigma$ is the kernel of $dF(\zeta)$ which is also the kernel of $dH(\zeta)$. It follows that $dF(\zeta) = \lambda dH(\zeta)$ for some nonzero constant λ. Letting ζ vary we get a smooth non-vanishing function $\lambda : \Sigma \to \mathbb{R}$ such that $dF = \lambda dH$ holds along Σ. It follows that the identity $X_F = \lambda X_H$ holds along Σ, which proves the lemma, with the function λ supplying the reparameterization. QED

PROOF OF PROPOSITION.

Apply the lemma with F being the Hamiltonian on T^*Q whose geodesic flow is that of the JM metric and H the Hamiltonian for our natural mechanical system. Since F is defined by the inverse metric to ds_{JM}^2 we have that $F = \frac{1}{\lambda_E} K$ or $F = \frac{1}{2}\frac{1}{\lambda_E}\|p\|^2$ where $K = \frac{1}{2}\|p\|^2$ denotes the original kinetic energy on T^*Q. (Strictly speaking we ought to write $\|p\|_q^2$ since the metric depends on q.) We now take $c_F = 1/2$ and $c_H = E$ and verify that the hypothesis of the lemma :

$$\begin{aligned}
F(q,p) = \frac{1}{2} \qquad &\iff \|p\|^2 = \lambda_E(q) \\
&\iff \|p\|^2 = 2(E - V(q)) \\
&\iff K(q,p) = E - V(q) \\
&\iff K(q,p) + V(q) = E \\
&\iff H(q,p) = E
\end{aligned}$$

It remains to verify that the conformal factor λ_E defines the reparameterization. So let $(q(t), p(t)) = \zeta(t)$ be an energy E solution for H and $(q(s), p(s)) = \zeta(s)$ the same solution curve reparameterized by arclength according to $F = 1/2$. According to what we just proved we have that $\frac{d}{dt}\zeta = f\frac{d}{ds}\zeta$ for some function f and the problem is to find f. From Hamilton's equations we have $\frac{d}{dt}q = g^{-1}p$ where $g^{-1} = g^{-1}(q)$ is the inverse of the kinetic energy metric. (In index notation $\dot{q}^i = g^{ij}p_j$.) On the other hand, from the geodesic equations, rewritten as Hamilton's equations for F, the Hamiltonian associated to the JM metric, we have that $\frac{d}{ds}q = \frac{1}{\lambda_E}g^{-1}p$. Comparing these two ODEs for q we see that $\frac{d}{ds} = \frac{1}{\lambda_E}\frac{d}{dt}$ or $\lambda_E dt = ds$.

QED

Brakes and the Hill boundary. From $K = E + U$ we see that the conformal factor λ_E goes to zero along an energy E solution $q(t)$ precisely when $K(q(t), \dot{q}(t)) = 0$, which is to say, when $\dot{q}(t) = 0$. At such an instant t the energy E-solution $q(t)$ has instantaneously stopped, or "braked", hence the terminology "brake orbit". Since $p(t)$ is related to $\dot{q}(t)$ linearly we also have $p(t) = 0$ at the brake instant. If $t = 0$ is this brake instant then $q(-t) = q(t)$ since both curves $q(t)$ and $q(-t)$ solve Newton's equations and have the same initial conditions at $t = 0$. The brake solution instantaneously stops, pausing to turn around, and then retraces its path. Differentiation yields $\dot{q}(-t) = -\dot{q}(t)$ and so $p(-t) = -p(t)$. Newtonian orbits that touch the Hill boundary retrace their steps. A curve which retraces itself can never be a geodesic – a locally minimizing curve - for any length structure.

Collisions. Along collision solutions to the N-body problem we have $\lambda_E \to \infty$ since $U(q) \to +\infty$ precisely at collisions. Typically there is no consistent way to extend solutions through collisions. In special instances such as isolated binary collisions for the gravitational N-body problem we may be able to analytically extend the flow through collisions after forming a branched cover and reparameterizing time.

0.1. Metric completions for JM metrics. Associated to a Riemannian metric is a metric distance so we have a metric space structure on the interior of the Hill region. Metric spaces can be completed. The completion is unique up to isometry. What is the metric completion of the interior of the Hill region? Should a dynamicist even care?

The completion of the JM metric will contain Hill boundary points if there are any. It may contain collision points. It could contain added points "at infinity". We are most interested in the gravitational N-body problems and the power law N-body problems which contain them and focus our questions on these. We refer to the power law exponent as α. Thus $U = \Sigma m_a m_b / r_{ab}^{\alpha}$.

First recall the definition of the metric distance $d(A, B)$ between points A and B of a Riemannian manifold. We consider all paths joining A to B, take the infimum of their lengths and declare that to be the disrtance $d(A, B)$. If the manifold is connected the result is a metric in the standard sense. In this way, we get the JM distance function on the interior of the Hill region.

$E \geq 0$: No Hill boundary. When the energy E is positive or zero in power law N-body problems the corresponding Hill boundary is empty and the Hill region is all of configuration space $\mathbb{E}$ regardless of the exponent α.

Collisions. If the exponent $\alpha \geq 2$ then paths which hit the collision locus have infinite JM length, regardless of energy.

If $\alpha < 2$ then the Newtonian solutions ending in collision have finite JM length regardless of energy. Hence we must add in the collision locus in order to complete the metric space.

If $E > 0$ then paths heading out to infinity ($r \to \infty$) always have infinite length. It follows that no points needed to be added at infinity when completing. So for $E > 0$ and $\alpha \geq 2$ the Hill region minus collisions is complete relative to the JM metric.

$E = 0$- parabolic case. For $E = 0$ and $\alpha \geq 2$ the radial paths extended all the way to infinity have finite length since $\int dr/r^{\alpha/2}$ converges as $r_1 \to \infty$. Points at infinity need to be added at infinty to complete the space. We have not investigated the nature of this completion. For $\alpha < 2$ the length to infiinity is infinite. We do not need to add points at infiinity to complete.

$E < 0$: Dealing with the Hill boundary. When $E < 0$ the Hill boundary is the non-empty smooth non-compact hypersurface $\{U = -E\}$. This hypersurface is topologically isotopy to the boundary of a tubular neighborhood of the collision locus Δ. As pairs of points tend to the Hill boundary their distances converge to zero. We must add the Hill boundary to complete the metric space. But when we do , we must also collapse the entire boundary to a single point assuming it is connected. (In the case of the N-body problem in d-dimensions, $d > 1$, the Hill boundary is connected.) For the metric ds^2_E still makes sense on the boundary but then all paths travelling along the boundary have zero JM-length so that the JM distance function extended to the boundary yields that the distance between two points of the boundary is zero. To reiterate: the extended distance function to the Hill boundary is not a true distance fucntion but rather it is a *semi-distance* since any two points on the Hill boundary have zero distance apart. [1]

From a metric perspective there is only one way to proceed. Crunch the entire Hill boundary to a single point. We call this point the Hill point. This 'crunching" is the standard procedure used to produce a metric space out of a semi-metric space: identify all points a distance zero away from each other to a single point – divide out by the equivalence relation of "being a distance zero apart".

We now come to a rather surprising conclusion.

Proposition D.2. *Consider the gravitational N-body problem with energy $E < 0$. Then the completion of the JM metric in the interior of the Hill region forms a non-compact complete metric space structure on the entire Hill region with the Hill boundary collapsed to a point called the Hill point. All points are a finite distance D from the Hill point so that, in particular the space is bounded, no points being further than $2D$ apart.*

Question. When $N = 3$ is it true that this maximum distance D from the Hill point is realized by the triple collision point and that D is the JM length of the energy E Lagrange collision orbit ?

Remark. The same conclusion holds for negative energies and power law N-body problems with exponent α in the range $0 < \alpha < 2$ and with the bodies moving in a Euclidean space of dimension $d > 1$. For $\alpha \geq 2$ when we complete we exclude the collision locus since its points are infinitely JM-distant from any point in the interior of the Hill region.

[1] By a semi-distance on a set Ω we mean a function $d : \Omega \times \Omega \to \mathbb{R}$ which satisfies all the axioms of a metric except the one which says $d(A, B) = 0$ iff $A = B$.

.....

from the Hill region and the JM metric continues to be complete. .

In all cases we arrive at a complete metric space on the Hill region. When $E \geq 0$ the Hill boundary is empty so we do not have to worry about it. If $E < 0$ and $d > 1$, the Hill boundary forms a single component which we crunch to a point. If the power law has $\alpha < 2$, as is the case with gravity, then we must include the collision locus in order for the metric space to be complete.

complete in the zero or positive energy case... NOT in negative energy! For then any point can be connected to the Hill point by a finite length path whose length is bounded, roughly by the Kepler problems at infinity. So the entire metric space is finite diameter. A finite diameter locally compact complete metric space is compact. To make the negative energy space compact must attach two-discs at each end corresponding to the Kepler problem there -RM

Nope! not for negative energies! For neg energy have to add these Kepler discs at infinity -RM

0.2. Geodesics through collisions or the Hill boundary. Do geodesics extend *through* the added points? We will show that geodesics can be extended through the Hill point, but these extensions have essentially nothing to do with Newtonian dynamics. On the other hand we will see that geodesics cannot be extended through a collision, although by Levi-Civita the dynamics can be extended through isolated binary collisions.

We recall the notion of "geodesic" on a general metric space. See [**?**] for details. On any metric space (M, d) one can define the length $\ell(\gamma)$ of a continuous path $\gamma : [a, b] \to M$ by means of "polygonal approximations". The length may be infinite. If $A = \gamma(a)$ and $B = \gamma(b)$ then $d(A, B) \leq \ell(\gamma)$. We call γ a minimizing geodesic when equality holds: $d(A, B) = \ell(\gamma)$. We call M a geodesic length space if there is a minimizing geodesic joining any two points of M. The Hill region for the gravitational N-body problem, endowed with the JM metric, with boundary collapsed to a point when $E < 0$ has the structure of a geodesic length space [2] .

A curve can be a metric geodesic without being a minimizing geodesic. Think about great circles on the sphere. In order to define "geodesic" for general metric spaces we first talk about arclength parameterization. A curve is called rectifiable if its length is finite. Rectifiable curves can be reparameterized by arc length. When a minimizing geodesic is parameterized by arclength we find that it provides an isometric embedding of its domain into M, so that , in particular $d(\gamma(a), \gamma(b)) = b - a$. A geodesic is a curve in M parameterized by arclength for which all sufficiently short subarcs of the curve are minimizing geodesics. For example, if the domain of γ is all of $\mathbb{R}$ so that $\gamma : \mathbb{R} \to M$, then we say that γ is a geodesic if for every $t_0 \in \mathbb{R}$ there is an $\epsilon = \epsilon(t_0) > 0$ such that the restriction of γ to the interval $[t_0 - \epsilon, t_0 + \epsilon]$ is a minimizing geodesic between its endpoints: $d(\gamma(t_0 - \epsilon), \gamma(t_0 + \epsilon)) = 2\epsilon$.

It can happen that on a complete geodesic length space certain geodesics cannot be extended. For example a closed bounded strictly convex domain in $\mathbb{R}^n$ with its metric inherited from the Euclidean metric is a complete geodesic length space. Any geodesic (line segment) which hits the boundary cannot be extended beyond the boundary point as a curve in the domain and remain a metric geodesic.

JM Marchal theorem ... Lemma 1. [Jacobi-Maupertuis Marchal?s lemma] Given two points q_0 and q_1 in the Hill region for some energy E, a minimizing JM geodesic for this energy exists which connects the two points and which is collision free except possibly at one or the other of the endpoints. If this geodesic arc touches the Hill boundary at some point besides the endpoints then it hits the Hill point exactly once and is the the projection of the concatentation of two brake orbits and a zero length 'instantaneous' curve lying on the Hill boundary.

[2]One defines a metric "length space" by putting the length functional first, as in Riemannian geometry, and derives the distance function d out of the length functional. See [**?**] for an axiomatization of "length functional" and length space.

A JM version of the Marchal lemma holds. See **??**XUU. This means that JM geodesics ending in collision cannot be continued through collision and remain geodesics.

The celebrated Hopf-Rinow theorem asserts that a Riemannian manifold without boundary is complete in the metric space sense if and only if its geodesic flow is complete. The JM metric, extended to collisions, the Hill boundary (crunched to a point), and, in the negative energy case, to the 'Kepler boundaries" at infinity is complete, but fails to be Riemannian at the added points. We have seen that JM geodesics correspond to solutions to Newton's equations in the interior of Hill region. Do these geodesics continue as geodesics through all the points we've added when we complete the Hill region ?

The answer depends on what we take "geodesic" to mean. If we take the strict metric meaning of geodesic, which we will review momentarily, then the answer is "no" for collision points, and 'yes' for the Hill point. On the other hand, if we take the dynamical meaning of geodesic as solutions to a "geodesic equation" i.e. to Newton's equations (appropriately regularized), then the correct answer appears to be "yes" for isolated binary collision points in the gravitational N-body problem, "no" for all other collision points, and "yes" for the Hill boundary points. But the dynamically extended geodesics through the Hill point do **not** agree with the metric extended geodesics as we will see.

PROPOSITION D.3. *Any JM geodesic ending in collision cannot be extended beyond collision and remain a metric geodesic*

This proposition is a JM version of the Marchall lemma (3.3) so crucial for applying the action minimization techniques to the problem of chapter 3. For a proof see [**104**].

On the other hand, as far as hitting the Hill boundary we have

PROPOSITION D.4. *Every geodesic hitting the Hill boundary can be extended to remain a metric geodesic. There are a continuum of such extensions corresponding to all brake orbits distinct from this geodesic. None of these geodesic extensions solve Newton's equations. Only the original geodesic, viewed as a brake orbit, folding back on itself, satisfies Newton's equations and this unique Newtonian extension is not a metric geodesic.*

PROOF. Take such a geodesic $\gamma_1(t)$ with brake instant $t = 0$. Take any other brake orbit $\gamma_2(t)$ with brake instant $t = 0$ and distinct $\gamma_1(t)$. Being distinct, we have $\gamma_1(0) \neq \gamma_2(0)$ as points in configuration space. The concatenation of γ_1 for $t \leq 0$ with $\gamma_2(t)$ for $t \leq 0$ projects to a continuous curve $c(t)$ upon crunching the Hill boundary to a point, this point being $c(0)$. For ϵ small enough $c([-\epsilon, \epsilon])$ is a minimizing geodesic. The best (= JM shortest) way to get from $\gamma_1(-\epsilon)$ to $\gamma_2(\epsilon)$ is to head straight to the Hill boundary along γ_1, travel the boundary for no penalty – all paths have zero length on the boundary – to $\gamma_2(0)$, and then head straight back out to $\gamma_2(\epsilon)$ along γ_2. The unique dynamical extension of γ_1 is γ_1 itself - the brake orbit turns back on itself according to $\gamma_1(-t) = \gamma_1(t)$. A path which retraces itself cannot be locally minimizing in any neighborhood $(-\epsilon, \epsilon)$ of the turnaround point, i.e. the brake instant so that the dynamical extension fails to be a geodesic in the metric sense.

QED

blah blah blah:
a few words on the discontinuous solutions of Todhunter; related to what are called 'turnpike' optimal paths in economic theory. 'Turnback" and that other long Kepler paper [**154**], [**155**], [**118**]
a paragraph on seifert's paper and his trick [**?**], [**138**] -RM

APPENDIX E

One-degree of freedom and central scattering

App: Scattering

1. Radial motion as a one-degree of freedom system

In this appendix we gather well-known facts regarding one-degree of freedom motions and apply them to scattering for central potentials, which is to say, to two-body classical scattering problems.

By a one degree of freedom system we mean a one-dimensional Newton equation $\ddot{x} = -V'(x)$ where $x \in \mathbb{R}$. The associated energy $E = \frac{1}{2}\dot{x}^2 + V(x)$ is constant along solutions. In the phase space, whose coordinates are $x, \dot{x}$, solutions trace out contour level sets of energy.

One-degree of freedom systems arise within two-body problems where they have $x = r = r_{12}$ the distance between the two bodies. If we fix the angular momentum J and work in the center-of-mass frame then the two-body energy is

{eq: radial energy}

$$E = \frac{1}{2}\dot{r}^2 + \frac{1}{2}\frac{J^2}{r^2} - f(r) \tag{114}$$

where $f(r)$ is the negative of the two-body potential. The corresponding central force equation for the evolution of $q = q_1 - q_2$ is $\ddot{q} = f'(r)\frac{q}{r}$. See section 0.2. The function

q: effective potential}

$$V_J(r) := \frac{1}{2}\frac{J^2}{r^2} - f(r) \tag{115}$$

is called the effective potential and plays the role of the one-degree of freedom potential $V(x)$. In particular, r evolves by $\ddot{r} = -V_J'(r)$.

We can qualitatively understand everything about a one-degree of freedom system by simply graphing the potential $V(x)$ versus x. We then fix any energy value E of interest, draw the horizontal line $V = E$ and use the graph to mark out the intervals of the x-line where $V(x) \leq E$. We call these the Hill intervals associated to the energy E. Since $\dot{x}^2 \geq 0$ and E is constant, the solutions $x(t)$ having energy E are constrained to move within these Hill intervals. The endpoints of a Hill interval satisfy $V(x) = E$ which means that $\dot{x} = 0$ there. If we choose the time origin so that $t = 0$ when $\dot{x} = 0$ then one has $x(-t) = x(t)$ and $\dot{x}(-t) = -\dot{x}(t)$ and for this reason the Hill endpoints are also called "turning points". If a particular Hill interval is compact and both its endpoints are regular points for $V(x)$,meaning that $V'(x) \neq 0$, then the solution $x(t)$ shuttles back and forth between these endpoints. The solution is periodic, with the period equal to twice the time required to pass from one turning point to the other. When the Hill interval is a half-line $[a, \infty)$ with $V(a) = E$ and $V'(a) < 0$ then the solution come in from infinity, hits the turning point $x = a$ and turns around, retracing its steps to infinity.

Let us see how this one-degree-of-freedom analysis pans out for the Kepler problem. Then $V_J = \frac{1}{2}\frac{J^2}{r^2} - \frac{\mu}{r}$. The graph of V_J is shown in FIGURE 1with $J = 1$ and

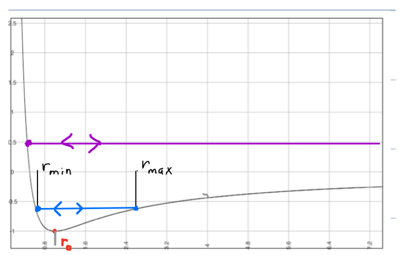


FIGURE 1. The graph of an effective potential arising in the Kepler problem. Here $V_J(r) = 1/r^2 - 2/r$. The black horizontal interval represents a chosen negative energy value with corresponding solution. The purple is a positive energy value with corresponding unbounded, or scattering orbit. fig: KeplerEffectiveB

$\mu = 2$. The effective potential has a global minimum when $J \neq 0$, this minimimum occuring at $r = r_0$. If $E < V_{min}$ then the Hill interval is empty. When $E = V_{min}$ the Hill interval is the single point r_0 which is a stable equilibrium for the one-dimensional dynamics and represents the circular Kepler orbit at that energy. For $V_{min} < E < 0$ the Hill interval is a bounded interval containing the circular radius r_0 and the turning points are the perihelion and apihelion radii r_{min} and r_{max} of the corresponding Kepler ellipse. When $E \geq 0$ the Hill interval is a half-line, representing parabolic or hyperbolic motion.

EXERCISE E.1. *Show that the radial motions for power law potentials $f(r) = 1/r^\alpha$ having $0 < \alpha < 2$ and $J \neq 0$ are qualitatively the same as those of Kepler. That is, when $J \neq 0$ the negative energy solutions $r(t)$ are oscillatory, with r shuttling back and forth between apicenter r_{min} and pericenter r_{max}, while when $E \geq 0$ the solutions 'scatter': they come in from infinity, hit an apicenter and return to infinity.*

For the strong force case of $\alpha > 2$ the situation is, in a way, the reverse of the previous $0 < \alpha < 2$ case. In figure 2 we graphed the effective potential for such a case,namely $\alpha = 4$ so that $f(r) = r^{-4}$, $\mu = .6$ and $J^2 = 2$. When $\alpha > 2$ and when $J \neq 0$ we that V_J now has a unique global maximum $V_{max} = V(r_0)$ instead of a unique global minimum and is unbounded below instead of above. Only for energy $E = V_{max}$ are the solutions bounded and collision free. Here the Hill interval is the single point r_0 which is an unstable fixed point for the one-dimensional dynamics, representing a circular orbit. If $E > V_{max}$ all solutions come in from infinity and hit total collision $r = 0$, or, are the time reversal of such an orbit. One can verify that they spiral infinitely often on the way to collision. If $E < 0$ then the Hill interval is of the form $(0, a]$ and all solutions end at total collision and reach a finite size corresponding to $a > 0$. For $0 < E < V_{max}$ there are two Hill intervals, one ending in collision at $r = 0$ and the other of scattering type, with a finite nonzero perihelion.

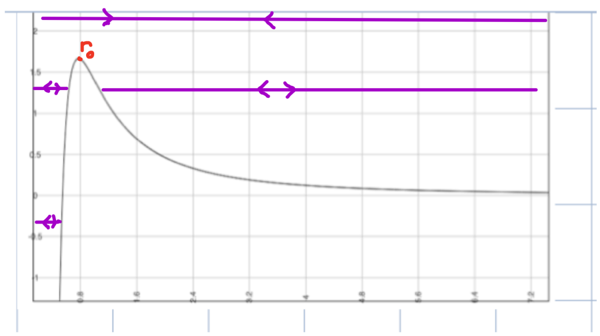


FIGURE 2. The graph of the effective potential $V_J(r) = 2/r^2 - .6/r^4$ arising in the $\alpha = 4$ strong central force problem when $\mu = .6$ and $J^2 = 2$. The purple horizontal lines represent various energy levels with corresponding solutions indicated. The top line goes from infinity to collision, or the reverse. The intermediate energy level splits into two Hill intervals, one ending in collision and bounded, the other coming in from infinity and returning to infinity. The bottom energy level is bounded and ends in collision.

fig: strongForceEffecti

The case $\alpha = 2$ is special among the power law potentials. Here the two terms of V_J balance. Indeed, by choosing $J^2 = \mu$ they cancel and we get a family of circular periodic orbits all having zero energy.

2. Scattering

Here we substantiate the 3rd claim of section 4.3 regarding 2-body scattering for power law potentials by using some one-degree-of-freedom tricks and the JM metrics for zero energy.

Proposition E.1. *Consider positive energy dynamics for a power law potential with exponent α satisfying $0 < \alpha < 1$. Then the image of a beam under the scattering map for the corresponding central force law $\ddot{q} = -\frac{q}{r^{\alpha+2}}$ in the plane is a circular arc of arclength $2\pi\alpha/(2-\alpha)$ which is centered on the 'undeflected' ray which travels in the beam direction. If $1 < \alpha < 2$ the scattered image of the beam covers the entire circle and then some: some directions are hit more than once. As $\alpha \to 2$ the circle is covered more and more times, so that in the limit $\alpha \to 2$ each direction is covered infinitely many times.*

The angular momentum for a central force problem, after normalizing units, satisfies $J = r^2\dot{\theta}$, so that

$$\frac{d\theta}{dt} = \frac{J}{r^2}$$

Solve the energy equation (114) for the radial velocity to get

$$\frac{dr}{dt} = \pm\sqrt{2(E - V_J(r)},$$

with the + sign valid for the branch of $r(t)$ receding from the origin and the − sign for the branch coming in towards the origin. Dividing the equation for dr/dt by that for $d\theta/dt$ we get an equation for $dr/d\theta$ valid along one branch or the another. Integrating this equation for $dr/d\theta$, and paying attention to signs we get, in the unbounded case,

$$\Delta\theta(E, J) = 2\int_{r_{min}}^{\infty} \frac{J/r^2}{\sqrt{2(E - V_J(r)}} dr$$

for the total change in angle $\Delta\theta = \Delta\theta(E, J)$ suffered in travelling along the trajectory labelled by E and J from $t = -\infty$ to $t = +\infty$. In the integral r_{min} is the pericenter or turn-around point, that unique value of r for which $E - V_J(r) = 0$. Note that $\Delta\theta$ is the angle between the two asympotic positions, $lim_{t\to-\infty}(q(t)/r(t))$ and $lim_{t\to+\infty}(q(t)/r(t))$. In particular, $\Delta\theta = \pi$ corresponds to straight line motion. In the Kepler case, this angle $\Delta\theta$ is the (signed) angle between the asymptotes of Keplerian hyperbola, with both asymptotes oriented to point outwards.

Recall from Rutherford scattering, 4.1 in chapter 4, the meaning of the impact parameter b. If we turn the force off , angular momentum is still conserved, and from this we see by a quick sketch or by an inspection of the relation $J = q \wedge \dot{q}$ as q tends to infinity along a line that

$$J = b.$$

Henceforth we use b in place of J. Choose the x-axis along the beam direction, with the rays of the travelling towards the origin, coming in from infinity , heading in the direction $(-1, 0)$ of the negative x-axis. Reflection about the x-axis maps solutions to solutions, preserving the beam direction and hence maps the entire beam's worth of trajectories to itself, acting on the impact parameter by $b \to -b$. Reflectional symmetry implies that $\Delta\theta(E,-b) = -\Delta\theta(E,b)$. As $b \to \infty$ the corresponding trajectories tend to straight lines, undeflected by the central force providing $f(r) \to 0$ sufficiently quickly as $r \to \infty$. The power laws for $\alpha > 0$ all decay sufficiently fast. Since $\Delta\theta = \pi$ for straight lines this yields $\lim_{b\to\pm\infty} \Delta\theta(E,b) = \pi$. Reflectional symmetry implies that the arc swept out by the scattering map is reflectionally symmetric, so of the form $(\pi - \Delta\psi, \pi + \Delta\psi)$. The midpoint of the arc is the angle π corresponding to the straight undeflected lines which correspond to $b = \pm\infty$.

The ray $b = 0$ corresponds to the solution with angular momentum zero and is the unique trajectory in the beam which ends in collision when $f(r) = r^{-\alpha}$ with $0 < \alpha < -2$. This collision trajectory associated to $b = 0$ splits the beam into two connected half-beams, one having $b > 0$ and the other having $b < 0$. Reflection symmetry shows that if the scattered image of one half beam is an arc $(\pi, \pi + \Delta\psi)$ then the the other half beam has scattered image equal to the arc $(\pi - \Delta\psi, \pi)$.

If, as is the case with power laws, the potential $f(r)$ is strictly monotone in r, then $\Delta\theta(E,b)$ is a strictly monotone function of b in either half-beam. It follows that the endpoints $\pi \pm \Delta\psi$ of the scattered arc are characterized by

$$\Delta\psi(E) = \lim_{b\to0^\pm} |\Delta\theta(E,b) - \pi|.$$

We will see momentarily that for power law potentials $\lim_{b\to0^+} \Delta\theta(E,b)$ is independent of the energy E as long as E is positive. Knauf and Krapf in [**72**] explicitly compute

{KKlimit}

$$\lim_{b\to0^+} \Delta\theta(E,b) = \frac{2\pi}{2-\alpha}, \qquad f(r) = r^{-\alpha} \tag{116}$$

by taking limits of integrals for power laws $f(r) = r^{-\alpha}$ of the integral expression (**??**) for $\Delta\theta(E,b)$ (See their equation (4.6).) As a consequence of their computation

{eq: limit scattering}

$$\Delta\psi(\alpha) = \pi\frac{\alpha}{2-\alpha}. \tag{117}$$

from which it follows that the overall size of the arc scattered into is $2\Delta\psi = 2\pi\frac{\alpha}{2-\alpha}$. as claimed.

Figure 3 summarizes the situation described in the last few paragraphs for the particular case of $\alpha = 1/2$ for which $\Delta\psi(1/2) = \pi/3$.

We will compute $\Delta\psi(\alpha)$ by a different means, taking advantage of the fact that the zero energy JM metric is conical for power-law potentials.

Scaling. In the first step of this computation we use scaling to trade $b \to 0$ with $E \to 0$. Recall the scaling symmety $q(t) \mapsto \lambda q(\lambda^{-\beta} t)$ where $\beta = 1 + \alpha/2$. (See 3.4.) Scaling a solution does not change its limiting direction but changes its energy and angular momentum scale according to $(E,J) = \lambda^{-\alpha}E, \lambda^{1-\alpha/2}J)$. It follows that

$$\Delta\theta(\lambda^{-\alpha}E, \lambda^c b) = \Delta\Theta(E,b), \qquad c = 1 - \frac{\alpha}{2}$$

It follows immediately from this scaling law that the image of the scattered beam does not depend on energy as long as the energy is positive. Scaling symmetry yields that $\Delta\theta(1,\lambda^c) = \Delta\theta(\lambda^\alpha, 1)$, or, more generally $\Delta\theta(E_*, \lambda^c b_*) = \Delta\theta(\lambda^\alpha E_*, b_*)$ for any positive constants E_*, b_*. Now let $\lambda \to 0$ and use that $\alpha, c > 0$ to conclude that

$$lim_{E\to0^+} \Delta\theta(E, b_*) = \lim_{b\to0} \Delta\theta(E_*, b)$$

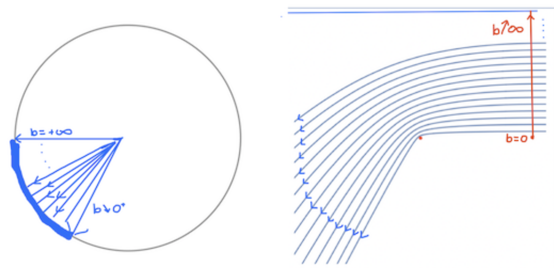


FIGURE 3. The right panel shows a half beam being scattered by an attractive central force power law $f(r) = r^{-1/2}$. The left panel shows the arc of the circle resulting from the scattering map applied to this half beam, its image being half of the scattered image of the full beam.

fig: halfbeam

for any $E_*, b_* > 0$. This common limit is the desired angle of equation (116).

Now in the scaling limit $E \to 0$ our entire beam (now with spatial scale λ growing according to the relation $\lambda^{-\alpha} = E \to 0$) limits to a beam of trajectories for the zero energy problem. At zero energy all solutions except the collision solution $b = 0$ suffer the same overall deflection and this angle is again the desired limiting angle $\lim_{b\to 0} \Delta\theta(E, b)$ which we want to compute.

STEP TWO. Reduction to conical geometry.

LEMMA E.1. *The Jacobi-Maupertuis metric at zero energy for the power law potential whose negative is $U(q) = f(r) = r^{-\alpha}$ is isometric to the cone metric $d\rho^2 + c^2\rho^2 d\theta^2$ where $c = 1 - \alpha/2$. This result is valid as long as $c \neq 0$.*

PROOF. In standard r, θ polar coordinates the JM metric is

$$ds^2_{JM} = r^{-\alpha}(dr^2 + r^2 d\theta^2) = r^{-\alpha} dr^2 + r^{2-\alpha} d\theta^2.$$

We can put this metric into the form $d\rho^2 + f(\rho)^2 d\theta^2$ by a change of variables $r \to \rho$ defined by imposing $r^{-\alpha} dr^2 = d\rho^2$. Thus $d\rho = r^{-\alpha/2} dr$. The change

$$\rho = \frac{1}{c} r^c$$

does the trick, with c as given in the lemma. Note that $\rho^2 = \frac{1}{c^2} r^{2c} = \frac{1}{c^2} r^{2-\alpha}$ or $r^{2-\alpha} = c^2\rho^2$. This last is our "$f(\rho)^2$". We have put the metric into the form

$$ds^2_{JM} = d\rho^2 + c^2\rho^2 d\theta^2,$$

that of the cone over a circle of circumference $2\pi c$ as claimed. QED

The further change of variables

$$\psi = c\theta$$

puts the metric into the form:

$$ds^2_{JM} = d\rho^2 + \rho^2 d\psi^2$$

which is that of the flat Euclidean metric on the plane if (ρ, ψ) are standard polar coordinates on that plane.

We summarize our change of variables:

$$\frac{1}{c}r^c = \rho \tag{118}$$

$$c\theta = \psi \tag{119}$$

$$c = 1 - \alpha/2 \tag{120}$$

This change can be summarized by $w = \frac{1}{c}z^c$ if we set $z = re^{i\theta}$ and $w = \rho e^{i\psi}$. Henceforth we will refer to the flat model as the "w-plane.

As θ varies from 0 to 2π, the angle ψ varies from 0 to $2c\pi$. Thus the JM metric is isometric to that of a cone for which the circle of radius $\rho = 1$ about the cone point is a circle whose circumference is $2c\pi$. Such a circle can be constructed by a "paper and scissors" construction from the Euclidean plane by cutting out the interior of an angle with the complementary opening $2\pi - 2c\pi$ and gluing the bounding rays togethter . SEE FIGURE ...

Geodesics and Scattering.

In figure 4 we have drawn the flat w-plane with a geodesic - a line - and several congruent 'fundamental domains" corresponding to the sectors of angle $2c\pi$. The w-plane metric is flat so its geodesics are standard lines. Appropriately interpreted, these lines project to the geodesics of our JM cone metric. Choose k congruent sectors sharing rays so as to cover a half-plane containing the line, so roughly $k \sim 1/c$ sectors will do, one if $c \geq 1/2$. The metric cone is then the quotient of the large obtuse sector made from the k sectors by identifying points in different sectors by the appropriate multiple of the rotation $w \sim e^{ic2\pi}w$. In this way, we see the whole line and can visualize its wrappings around the cone, rather in the style that one investigates trajectories by square billiard table problems by tiling the plane.

These w-lines, being geodesics correspond to the zero energy solutions for the $r^{-\alpha}$ central force problem. Any line in the w- plane which does not pass through the origin represents a JM geodesic for which $b \neq 0$. The scattering angle for any such line is π. Indeed, take the line $X = t, Y = y_0 \neq 0$ where $w = X + iY$ and look at the change in angle ψ as suffered in travelling from $t = -\infty$ to $t = +\infty$ is π. (Rotational invariance implies that this angle does not depend on the choice of line.) Remember this is the overall ψ-angle suffered.

But the scattering angle described in equation (116) is the overall change in the angle θ suffered when travelling from $t = -\infty$ to $t = +\infty$, while what we just described is the overall change in ψ. Since $\theta = \psi/c$ we have that the overall angle suffered by θ is π/c which is equal to $2\pi/(2 - \alpha)$, thus establishing equation (116).

QED

CASE OF KEPLER, $\alpha = 1$ underlined next...

....

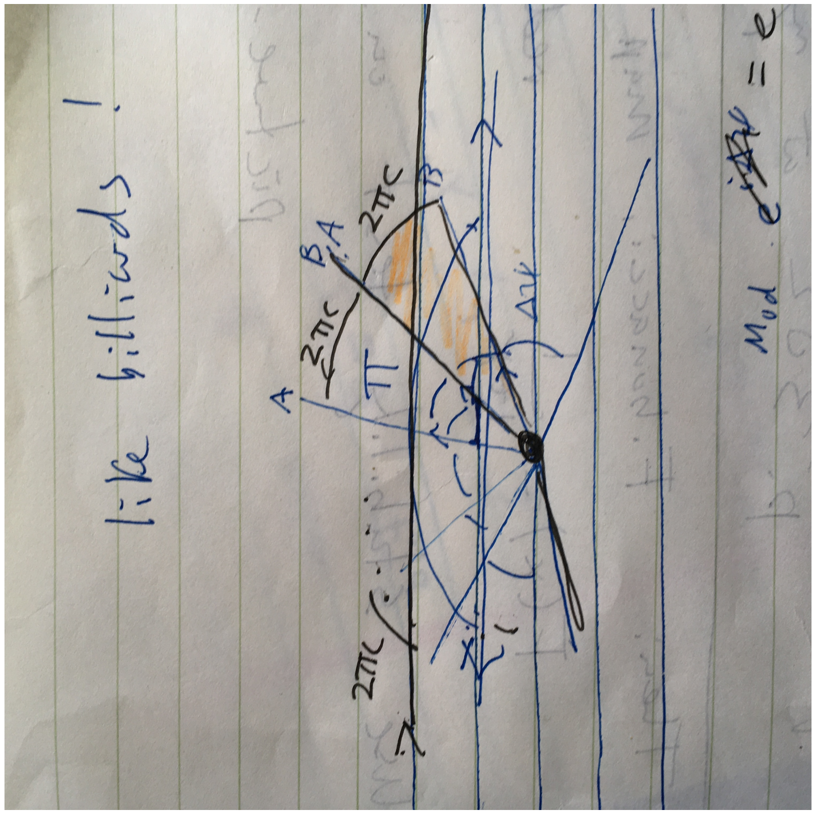


FIGURE 4. The cone over a circle of circumference $2\pi c$ is realizes as a quotient of any k sectors glued by rotation by the angle $2\pi c$. We choose k large enough so that the sectors cover a half-plane containing a straight line which induces a geodesic on the cone.

fig: alphaConemetric

Part 4

Bibliography

APPENDIX F

Bibliography

1. Bibliography

Bibliography

[1] R. Abraham and J. E. Marsden, **Foundations of Mechanics**, 2nd ed. Benjamin/Cummings, (1978).

[2] Albouy, A. (1996), *The symmetric central configurations of four equal masses*, in Hamiltonian Dy- namics and Celestial Mechanics, Contemp. Math **198**.

[3] Albouy, A. and Dullin R., 2020: Relative Equilibria of the 3-body Problem in $\mathbb{R}^4$ Journal of Geometric Mechanics, 12 (2020), 323-341 `http://www.aimsciences.org/article/doi/10.3934/jgm.2020012` `https://arxiv.org/abs/2002.00649`

[4]

[5] Albouy, A., Cabral H. and Santos A. (2012) *Some problems on the classical n-body problem* Celestial Mechanics and Dynamical Astronomy, 113 (2012), 369-375 `http://link.springer.com/article/10.1007%2Fs10569-012-9431-1` `http://arxiv.org/abs/1305.3191`

[6] Albouy, A., and Kaloshin V. (2012) *Finiteness of central configurations of five bodies in the plane* Annals of Mathematics **176**, 535-588 `http://annals.math.princeton.edu/2012/176-1/p10`

[7] Albouy, A. and Chenciner, A., (1998): *Le probléme des N corps et les distances mutuelles* Inventiones **131** , 151-184 `http://link.springer.com/article/10.1007%2Fs002220050200`

[8]

[9] V.I. Arnold, V.V. Kozlov, and A.I. Neishtatdt, **Dynamical Systems III**, v.3, Encyclopaedia of Mathematical Sciences, Springer- Verlag, (1988).

[10] Arnol'd, V.I., (1989), **Mathematical Methods in Classical Mechanics**, Springer-Verlag, 2nd ed..
Arnold, V. I.
1963a: Proof of A. N. Kolmogorov's theorem on the preservation of quasiperiodic motions under small perturbations of the Hamiltonian, *Russian Math. Surveys*, 18(5), 9–36.
1963b: Small divisor problems in classical and celestial mechanics, *Russian Mathematical Surveys*, 18(6), 85–192.
1964: Instability of dynamical systems with several degrees of freedom, *Sov. Math. Dokl.*, 5, 581–585.

[11] Benettin, G. , Fassò, F., and Guzzo, M. *Nekhoroshev-stability of L4 and L5 in the spatial restricted three-body problem* Regul. Chaotic Dyn., 3 no. 3, 56?72, (1998).

[12] Bierstone, E. , chapter 2 of *The Structure of Orbit Spaces and the Singularities of Equivariant Mappings*, Instituto de Mateática Pura e Aplicada, Rio de Janeiro, (1980).

[13] Birkhoff, G. D, (1927), **Dynamical Systems**, Colloq. 9, Amer. Math. Soc., Providence, R.I. USA.

[14] Birman, J. *Braids, Links, and Mapping Class Groups*, Princeton U. Press., (1974).

[15] Bost, J-B., (1986), *Tores invariants des systèmes dynamiques hamiltoniens* Astérisque, **133-134** Séminaire Bourbaki, no 639, 113-157

[16] Broucke R. (1975), *On relative periodic solutions of the planar general three-body problem*, Celest. Mech. **12** 439?62

[17] Calleja, R, Doedel, E. and Garcia-Azpeitia C, *Symmetries and choreographies in families bifurcating from the polygonal relative equilibrium of the n-body problem* Celestial Mechanics and Dynamical Astronomy **130**, no 48 (2018)

[18] Chazy, J., *Sur l'allure du mouvement dans le problème des trois corps quand le temps croît indéfiniment*, Ann. Sci. École Norm. Sup. (3), 29–130, (1922). `http://www.numdam.org/item?id=ASENS_1922_3_39__29_0`

[19] Chen, K-C., (2003),*Variational methods on periodic and quasi-periodic solutions for the N-body problem* Ergod. Th. & Dynam. Sys. **23**, 1691?1715

[20] Chen, K-C., Ouyang, T., and Xia, Z., (2012)*Action-Minimizing Periodic and quasi-periodic solutions in the n-body problem* Math. Res. Lett. **19** , no. 02, 483?497

[21] Chenciner, A. Féjoz, J. and Montgomery, R., *Rotating eights. I. The three Γ_i families*, Nonlinearity **18** , no. 3, 1407–1424, (2005) .

[22] Chenciner, A. *Action minimizing solutions of the Newtonian n-body problem: from homology to symmetry* International Congress of Mathematicians (2002).

[23] Chenciner, A. `https://perso.imcce.fr/alain-chenciner/Tsinghua%202019.pdf`

[24] Chenciner, A. and Montgomery, R. (2000) *A remarkable periodic solution of the three body problem in the case of equal masses* , Annals of Math, **152**, 881-901

[25] Chenciner A., and A. Venturelli, A. *Minima de l'intégrale d'action du problème de 4 corps de masses égales dans $\mathbf{R}^3$: orbites hip-hop* Cele Mech. Dyn. Astronom., 77, 139-152, (2000).

[26] Chierchia, L. and Pinzari, G. , *The planetary N-body problem: symplectic foliation, reductions and invariant tori.* Invent. Math. 186, no. 1, 1?77, (2011).

[27] Deprit A. and Delie, A. *Régularisation du problème des trois corps*, Arch. for Rat. Mech., **12**, 325-353, (1962).

[28] Deprit A. and Delie, A. *Courbure de la métrique riemannienne du probème plan des trois corps* Bull. Cl. Sc. Acad. Belg. , no. 5, **48**, 652-658, (1962).

maybe delete first two Deprit refs; unused -RM

[29] A. Deprit and A. Deprit-Bartholome, *Stability of the triangular lagrangian points* The Astronomical Journal, 72, 173–179, (1967).

[30] N. Duignan and H. Dullin *On the $C^{8/3}$ -regularisation of simultaneous binary collisions in the planar four-body problem* Nonlinearity, 34, (7), 4944-4982(2021).

[31] Duistermaat, J.J., *Symplectic Geometry*, Course Notes, Summer Course, 2004.

[32] H.Dullin and J. Scheurle, *Symmetry reduction of the 3-body problem in $\mathbb{R}^4$*, Journal of Geometric Mechanics, (in print, accepted 20 Nov 2019), arXiv:1908.04496.b

[33] Einstein, A. (1945), **The Meaning of Relativity**, (2nd ed.). Princeton, N.J.: Princeton University Press.

[34] L. Euler, *De motu rectilineo trium corporum se mutuo attrahentium*, (1765).

[35] L. Euler, *Considerations sur le probléme de trois corps*, Memoires de l?Académie de Sciences de Berlin, 19 194-220, (1770).

[36] Féjoz, J, Knauf, A. and Montgomery, R., *Classical n-body scattering with long-range potentials* Nonlinearity, **34** , Issue 11, 8017-8054

[37] Féjoz, J. *Démonstration du "théorème d'Arnol'd" sur la stabilité du système planétaire (d'après Herman)* Erg. Th. and Dyn. Systems, 24 (5):1521–1582, (2004).

[38] Féjoz, J., (2002), *Quasi Periodic Solutions in the Planar Three-Body Problem*, J. Differential Equations, vol.183, no. 2, pp. 303?341.

[39] Ferrario D. and Terracini S. *On the Existence of Collisionless equivariant minimizers for the classical n-body problem*, Invent. Math., 155 (2) 305-362, (2004).

[40] D. Ferrario *Transitive description of symmetry groups for the n-body problem*, arxiv:0603684 (2006).

[41] Feynman, R. and Wheeler, J. (1949), *Classical Electrodynamics in Terms of Direct Interparticle Action*, Rev. Modern Physics, **21**, n0.. 3, 425–433.

[42] Gelfand, I. and Fomin S., (2000), **Calculus of Variations**, Dover.

[43] Gerver, J. , `https://arxiv.org/abs/2202.08534`

[44] V. Ginzburg, *Hamiltonian dynamical systems without periodic orbits*, Northern California Symplectic Geometry Seminar, 35–48, Amer. Math. Soc. Transl. Ser. 2, 196, Amer. Math. Soc., Providence, RI, 1999.

[45] V. Ginzburg, *On closed trajectories of a charge in a magnetic field. An application of symplectic geometry.* Contact and symplectic geometry (Cambridge, 1994), 131–148, Publ. Newton Inst., 8, Cambridge Univ. Press, Cambridge, 1996.

[46] W. Gordon, *A Minimizing Property of Keplerian Orbits*, Am. J. Math., **99**, no. 5, 961-971, (1977).

[47] W. Gordon, *Conservative Dynamical Systems Involving Strong Forces*, Trans AMS, Volume 204, 113-135, (1975)

[48] Fusco, G., Gronchi, G.F, and Negrini P., (2011) *Platonic polyhedra, topological constraints and periodic solutions of the classical N-body problem, Invent. Math.* **185** 283-332.

[49] Guichardet, A. *On rotation and vibration motions of molecules* Ann. Inst. H. Poincare, Phys. Theor., **40** no.3 329-342 (1984).

[50] V. Guillemin and S. Sternberg, **Symplectic Techniques in Physics**, Cambridge U. Press, Cambridge, 1984.

[51] V. Guillemin and S. Sternberg, **Geometric Asymptotics**, Math. Surveys, **14**, AMS, Providence, R.I.,1977.

[52] Hampton, M and Moeckel R (2006) *Finiteness of relative equilibria of the four-body problem* Inventiones Mathematicae **163**, 289?312.

[53] Heggie, D. *A new outcome of binary-binary scattering*, Mon. Not. R. Astron. Soc., 318, 61-63, (2000).

[54] Hilbert D. and Courant R., **Methods of mathematical physics** 1st English ed., Interscience Pub., (1962).

[55] Heggie, D. and Hut, P. (1993), *Binary-Single-Star Scattering. IV. Analytic Approximations and Fitting Formulae for cross-sections and reaction rates*, Astrophys J., supplement series, **85**, 347–409.

[56] Hénon M (1974), *Families of periodic orbits in the planar three-body problem* Celest. Mech. **10** 375-388

[57] Hénon M (1976), *A family of periodic solutions of the planar three-body problem, and their stability* Celest. Mech. **13** 267?285

[58] M. Herman, Some open problems in dynamical systems, *Proc. Int. Congress of Math.*, Documenta Mathematica J. DMV, Extra volume ICM II, (1998) 797-808.

[59] Wu-Yi Hsiang, *Geometric Study of the Three-Body Problem I*, report -620, Center for Pure and Applied Math, University California, Berkeley, (1994).

[60] Wu-Yi Hsiang and E. Straume, , *Kinematic Geometric of triangles with given mass distribution*,report PAM-636, Center for Pure and Applied Math, University California, Berkeley, (1995).

[61] Wu-Yi Hsiang, *Kinematic geometry of mass-triangles and reduction of Schrödinger's equation of three-body systems to partial differential equations solely defined on triangular parameters.* Proc. Nat. Acad. Sci. U.S.A. **94**, no. 17, 8936–8938. (1997).

[62] T. Iwai, A gauge theory for the quantum planar three-body problem. *J. Math. Phys.* **26** (1987) 964-974.

[63] T. Iwai, *A gauge theory for the quantum planar three-body system*, J. Math. Phys. ,**28**, 1315-1326, (1987). WHICH 'IWAI' Richard? -RM

[64] T. Iwai, *A geometric setting for internal motions of the quantum three- body system*, J. Math. Phys. **28**,1315-1326, (1987).

[65] T. Iwai, *A geometric setting for classical molecular dynamics*, Ann. Inst. Henri Poincaire, Phys. Th., **47**, no. 2, 199-219, (1987).

[66] T. Kapela, C. Simó, (2007) Computer assisted proofs for nonsymmetric planar choreographies and for stability of the Eight, Nonlinearity, **20**, 1241-1255.

[67] Kapela,T. , Simó, S. (2017) *Rigorous KAM Results Around Arbitrary Periodic Orbits for Hamiltonian Systems* , Nonlinearity, **30** , 965-988. `https://arxiv.org/pdf/1105.3235.pdf`

[68] D. G. Kendall, Shape Manifolds, Procrustean Metrics, and Complex Projective Space, *Bull. London Math. Soc.* **16** (1984) 81-121.

[69] , Kobayashi, S. and Nomizu, N. (1969) **Foundations of Differential Geometry, vol. II**, John Wiley and Sons, New York.

[70] Knauf, A. **Mathematical Physics: Classical Mechanics**, Springer, (2012),.

[71] Klein, M. and Knauf, A., **Classical Planar Scattering by Coulombic Potentials**, LNP. 13, Springer-Verlag, (1992).

[72] Knauf, A. and Krapf, M. *The non-trapping degree of scattering*, Nonlinearity 21 2023-2041, (2008)

[73] Knauf, A. *Qualitative Aspects of Classical Potential Scattering*, Reg. and Chaotic Dyn. Sys., 4, 3-21, (1999)

[74] J. Lagrange, *Méchanique analitique*, first edition, Veuve Desaint, Paris, 1788.

[75] J.-L. Lagrange, Essai sur le Probléme des Trois Corps, in *Prix de l'Académie Royale des Sciences de Paris*, **IX**, 1772.

[76] Landau, I. and Lifshitz, E. (1976), **Mechanics**, Pergamon Press.

[77] Laskar, J., (2014), *Michel Hénon and the Stability of the Solar System* `https://arxiv.org/pdf/1411.4930.pdf`

[78] H. Le, *On geodesics in Euclidean Shape Spaces*, J. London Math. Soc. , no. 2 , **44**, 360-372, (1991).

[79] G. Lemaitre, em Coordonnées symétriques dans le problème des trois corps, Bull. Cl. Sc. Acad. Belg. no. 5, **38**, 582-592, 1218-1234, (1952).

[80] G. Lemaitre, em Regularization of the Three Body Problem, in **Vistas in Astronomy**, **1**, 207-215, (1955).

[81] T Levi-Civita, *Sur la régularisation du problème des trois corps* Acta Math. 42 , no. 1, 99?144, (1920).

[82] R. G. Littlejohn and M. Reinsch, *Gauge fields in the separation of rotations and internal motions in the n-body problem*, Rev. Modern Phys., **69**, no. 1, 213–275, (1997).

[83] Lobachevsky, N. *The new foundations of geometry with full theory of parallels* [in Russian], 1835-1838, In Collected Works, V. 2, GITTL, Moscow, 1949, p. 159.

[84] Maderna, E. and Venturelli, A. (2020) *Viscosity solutions and hyperbolic motions: a new PDE method for the N-body problem*, Annals of Mathematics, **192**, 499–550.

[85] Montaldi, J and Steckles K. *Classification of symmetry groups for planar n-body choreographies*, U. of Warwick MIMS EPrint: 2013.26

[86] Moser, J. (1973), **Stable and Random Motion**, Princeton, N.J.: Princeton University Press.

[87] Moser, J. (1978),*Is the Solar System Stable?* The Mathematical Intelligencer **1** 65–71.

[88] Mañe, R (199) Global Variational Methods in Conservative Dynamics, Colóquio Brasileiro de Matem
'atica

[89] Marchal, C. (2012), *Progress in Celestial Mechanics* Reg and Chaotic Dyn., , Vol. 17, No. 1, pp. 6?23.

[90] Marchal, C.

[91] Marchal, C. (1990), **The Three-Body Problem**, Amsterdam: Elsevier.

[92] J. Marsden and T. Ratiu, **Introduction to Mechanics and Symmetry**, Springer-Verlag, (1995).

[93] J. Marsden and A. Weinstein, *Reduction of Symplectic Manifolds with Symmetry*, Rep. Math. Phys., **5** (1974), 121–130.

[94] K. Meyer, **Periodic Solutions of the N-body Problem**, Lecture Notes in Math. 1719, Springer, (1999).

[95] K. Meyer, G. Hall, **Introduction to Hamiltonian Dynamical Systems and the N-Body Problem**, Applied Math. Sciences series 90, 1st edition, Springer-Verlag, (1991).

[96] K. Meyer, D. Offin, **Introduction to Hamiltonian Dynamical Systems and the N-Body Problem**, Applied Math. Sciences , 3rd edition, Springer (2017).

[97] K. Meyer, D. Schmidt, *The stability of the Lagrange triangular point and a theorem of Arnol'd* J. Differential Equations **62**, no. 2, 222?236 (1986).

[98] Milnor, J. (1973) **Morse Theory**, Ann. of Math Studies, **51**, Princeton U. Press.

[99] Moczurad, M., Zgliczyński, P. (2019), *Central configurations in planar n-body problem for n = 5,6,7 with equal masses*, Celest. Mech. Dyn. Astron. **131**, 46-74, `https://arxiv.org/abs/1812.07279`.

[100] Moeckel, R. (1988), *Some qualitative features of the three-body problem* , Contemp. Math.,**81**, Amer. Math. Soc., Providence, RI, 1–21.

[101] Moeckel, R., *Lecture Notes on Central Configurations*, `http://www-users.math.umn.edu/~rmoeckel/notes/Notes.html`

[102] Moeckel, R.,(2005) *A Variational Proof of the Existence of Transit Orbits in the Restricted Three-Body Problem*, Dynamical Systems, **20**, 45-58.

[103] Moeckel, R. and Montgomery, R., (2015) *Realizing All Reduced Syzygy Sequences in the Planar ThreeBody Problem*, Nonlinearity, **28**, 1919–1935.

[104] Moeckel, R. and Montgomery, R., Venturelli, A. (2012) *From Brake to Syzygy*, Archive for Rational Mechanics and Analysis; 204(3):1009-1060.

[105] Montaldi, J. *Existence of symmetric central configurations* arxiv 1408.5854

[106] Montaldi, J. and Steckles, K. (2013), *Classification of symmetry groups for planar n-body choreographies*, Forum of Mathematics, Sigma (online), also: arXiv:1305.0470.

[107] R. Montgomery, *How Much Does a Rigid Body Rotate?*, Am. J. Physics,v. 59, no. 5, 394-398, (1991).

[108] R. Montgomery , *Canonical Formulations of a Particle in a Yang-Mills Field*, Lett. Math. Phys. **8**, 59-67, (1984).
[109] R. Montgomery, *Gauge Theory and Control Theory*,pp. 343-378, in **Nonholonomic Motion Planning** J. Canny and Z. Li ed., Kluwer Acad. Press,Norwell, MA., USA, (1993).
[110] R. Montgomery, *Gauge Theory and the Falling Cat*, pp 193-218, in **Dynamics and Control of Mechanical Systems**, M. Enos, ed., Fields Inst. Communications series, vol 1 , AMS pub., Providence, RI. (1993).
[111] R. Montgomery, *The N-body problem, the braid group, and action-minimizing periodic orbit*, Nonlinearity, **11**, no. 2, 363–376, (1998).
[112] R. Montgomery, *The Geometric Phase of the Three-Body Problem*, Nonlinearity, **9** , no. 5, 1341-1360, (1996).
[113] R. Montgomery, *Infinitely many syzygies*, Archives for Rational Mechanics and Analysis, v. 164, no. 4, 311-340, (2002).
[114] Montgomery, R. *Fitting hyperbolic pants to a three-body problem*, Ergodic Theory Dynam. Systems v. 25 no. 3, 921–947, (2005).
[115] Montgomery, R. *Hyperbolic Pants fit a three-body problem* Ergodic Theory and Dynamical Systems, vol.25, 921-947, (2005).
[116] Montgomery, R. *The Three-Body Problem and the Shape Sphere*, Amer. Math. Monthly, v 122, no. 4, pp 299-321 , 2015; arXiv:1402.0841.
[117] R. Montgomery, R. **A tour of subriemannian geometries , their geodesics, and applications**, Mathematical Surveys and Monographs, vol. 91, American Math. Society, Providence, Rhode Island, 2002.
[118] Montgomery, R. (2020) , *Minimizers for the Kepler Problem*, Qualitative Theory of Dynamical Systems, vol. 19, no. 31.
[119] Moore,C. *Braids in Classical Gravity*, Phys. Rev. Lett. , 70, 3675-3679, (1993).
[120] Moulton, F.R. (1910) *The Straight Line Solutions of the Problem of n Bodies*, Ann. of Math. **12**, 1-17.
[121] Noether, E. (1918), *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse. 1235?257.
[122] B. O'Neill, *The fundamental equations of a subbmersion*, Mich. Math. J., **13**, 459-469, (1966).
[123] Palais, R. (1993), *The principle of symmetric criticality* Comm Math Phys **69**, 19-30.
[124] Palmore J. (1973), Classifying relative equilibria, I, Bull. AMS 79 (1973) 904-908; ..., II AMS 81 (1975) 489-491; ..., III, Lett.Math.Phys. 1 (1975) 71-73.
[125] , Peterson, I., (1993), **Newton's Clock**, W.H. Freeman and Co., NY
[126] Poincaré, H. **Les Methodes Nouvelles de la Mécanique Céleste**, v. 1, ch. 3, Paris, Gauthier-Villars et fils, (1892)
see also
introduction to English translation by D. Goroff, p. I 42. AIP, History of Modern Physics and Astronomy v. 13, (1993).
[127] Pöschel, J. , (1993), *Nekhoroshev estimates for quasi-convex Hamiltonian systems*, Math. Z., **213**, no. 2, 187–216.
[128] M. Reinsch, *Yang-Mills field equations for the Coriolis tensor in the n-body problem* J. Math. Phys. v. 39, no. 2, 1223-1235, (1998).
[129] G. Roberts, *Linear stability analysis of the figure-eight orbit in the three-body problem*, Erg. Th. & Dyn. Sys., v. 27, 1947-1963, (2007).
[130] G. Roberts, *A continuum of relative equilibria in the five-body problem*, Phys. D, v. 127, no. 3-4, 141-145, (1999).
[131] Rose, D (2015) *Geometric phase and periodic orbits of the equal-mass, planar three-body problem with vanishing angular momentum* PhD thesis, University of Sydney.
[132] , H.L. Royden **Real Analysis**, 2nd ed. Macmillan pub. (1968).
[133] Rutherford, E. (1911) *The Scattering of α and β Particles by Matter and the Structure of the Atom*, Philosophical Magazine. Series 6, vol. 21.
[134] M Saito and K. Tanikawa *Collinear Three-body problem with non-equal masses by symbolic dynamics*, in Order and Chaos in Stellar and Planetary Systems, ASP Conf series, Byrd et al editors, v 316, (2004).
[135] D. Saari, *From rotation and inclination to zero configurational velocity surfaces, I, a natural rotating coordinate system,* Celestial Mechanics, **33,** 299-318, (1984).

[136] D. Schmidt, *The Motion of the Moon*, Appendix A, pp 230-233, in Classical and Celestial Mechanics, the Recife Lectures, ed. H. Cabral and F Diacu, Princeton U. Press, (2002).
[137] [Sch] J. Schubart, J. (1956), *Nulerische Aufsuchung periodischer Lösungen im Dreikorperproblem*, Astronomische Nachriften, **283**, pp. 17-22.
[138] Seifert, H. translated by Bill McCain, *Periodic Motions of Mechanical Systems*, `https://people.ucsc.edu/~rmont/papers/periodicMcCain.pdf`
[139] A. Shapere and F. Wilczek, *Self-propulsion at low Reynolds number*, Physical Review Letters, vol. 58, 2051-2054, (1987).
[140] A. Shapere and F. Wilczek, *Geometric Phases in Physics*, World Sci., Singapore, (1989).
[141] A. Shapere, **Gauge Mechanics of Deformable Bodies**, PhD. thesis, Physics, Princeton, (1989).
[142] A. Shapere and F. Wilczek, *Geometry of Self-propulsion at low Reynolds number*, J. Fluid Mech. , **198**, 557-585, (1989).
[143] M. Shub, M. (1970) *Diagonals and relative equilibria, Appendix to Smale?s paper*, in Manifolds- Amsterdam, Springer Lecture Notes in Math. **197** 199-201.
[144] C. Simó, *Dynamical properties of the figure eight solution of the three-body problem*, in Proc. of the Cele. Mech. Conference dedicated to D. Saari for his 60th birthday, pp 209?228, Contemporary Mathematics v. 292, AMS, (2000).
[145] C. Simó, *private communications*, 1999-2021.
[146] Sitnikov FIND REF.
[147] Singer, S, (2003), **Symmetry in Mechanics: A Gentle, Modern Introduction**,
[148] Smale, S. (1970) *Topology and Mechanics, II*, Inventiones, **10**, 305- 331
[149] Smale, S. (1970), *Topology and Mechanics, II* Inventiones, **11**, 45-64
[150] Smale, S. (1998) *Mathematical Problems for the Next Century* Mathematical Intelligencer. **20** (2) 7?15
[151] Strömgren, E, (1933), *Connaissance actuelle des orbites dans le problm̀e des trois corps.* Bull. Astronomique, 9:87-130
[152] Szebehely V. and Peters, C. (1967) *Complete Solution of a General Problem of Three Bodies*, The Astronomical Journal, **72**, 876-882, `https://articles.adsabs.harvard.edu/pdf/1967AJ.....72..876S`
[153] Tanikawa K. and S. Mikkola S. , (2000) *One-Dimensional Three-body problem via symbolic dynamics*, Chaos, vol 10. no. 3, (2000).
[154] Todhunter, M. A., **Researches in the Calculus of Variations, principally on the theory of discontinuous solutions**, Cambridge, Macmillan and Co., 1871
[155] Todhunter, M. A. **History of the Calculus of Variations**, Cambridge, Macmillan and Co., 1871, `https://archive.org/details/histroyofthecalc033379mbp/page/n6`
[156] Weinstein, A. (1977) *Lectures on symplectic manifolds*, CBMS Regional Conference Series in Math., **29**, (1977).
[157] A. Weinstein, *The Local Structure of Poisson Manifolds*, J. Diff. Geom., **18**, 523-557, (1984).
[158] L.C. Young, **Lectures on the Calculus of Variations and Optimal Control Theory**, Chelsea, (1980).
[159] X. Yu *On the Stability of Lagrange Relative Equilibrium in the Planar Three-body Problem*, arXiv: 1911.12269.

Index