OSCILLATING ABOUT COPLANARITY IN THE 4 BODY PROBLEM.

RICHARD MONTGOMERY

Abstract. For the Newtonian 4-body problem in space we prove that any zero angular momentum bounded solution suffers infinitely many coplanar instants, that is, times at which all 4 bodies lie in the same plane. This result generalizes a known result for collinear instants ("syzygies") in the zero angular momentum planar 3-body problem, and extends to the $d+1$ body problem in d-space. The proof begins by identifying the translation-reduced configuration space with real $d \times d$ matrices, the degeneration locus (set of coplanar configurations when $d = 3$) with the set of matrices having determinant zero, and the mass metric with the Frobenius (standard Euclidean) norm. Let S denote the signed distance from a matrix to the hypersurface of matrices with determinant zero. The proof hinges on establishing a harmonic oscillator type ODE for S along solutions. Bounds on inter-body distances then yield an explicit lower bound ω for the frequency of this oscillator, guaranteeing a degeneration within every time interval of length π/ω. The non-negativity of the curvature of oriented shape space (the quotient of the translation-reduced configuration space by the rotation group) plays a crucial role in the proof.

1. Results.

Consider the Newtonian 4 body problem in Euclidean 3-space. Typically, the four point masses form the vertices of a tetrahedron. As the masses move about, at isolated instants the tetrahedron which they form might degenerate so that all 4 bodies lie on a single plane. Must such co-planar instants always occur?

A solution is called *bounded* if the interparticle distances r_{ab} between the four masses $m_a, a = 1, 2, 3, 4$ are bounded for all time in the solution's domain of definition.

Theorem 1.1. *For the 4 body problem in 3-space, any bounded zero angular momentum solution defined on an infinite time interval suffers infinitely many coplanar instants.*

Thus the theorem asserts that the coplanar configurations define something like a global slice for the spatial 4 body problem when the total angular momentum is zero.

Theorem 1.1 follows directly from the finite time interval oscillation result, Theorems 1.2 below, which in turn follows immediately from Theorem 1.4. Those results are stated more generally, for the $d+1$-body problem in d-dimensional Euclidean space. The results were inspired by [13] which concerns the case $d = 2$, which is to say the recurrence of collinearity in the case $d = 2$ of the zero angular momentum planar three-body problem .

Date: October 13, 2018.

Write $q_a \in \mathbb{R}^d, a = 1, \dots, d+1$ for the positions of the bodies. Typically, at each instant the q_a form the vertices of a $d+1$-simplex, meaning that their convex hull has nonzero d-dimensional volume. At special instants this volume may vanish by virtue of all bodies instantaneously lying on some affine hyperplane. We call these *degeneration instants.* Write $r_{ab} = |q_a - q_b|$ for the distances between bodies, $M = \Sigma m_a$ for the total mass and G for the universal gravitational constant. G is included to get our units straight: $GM/{r_{ab}}^3$ has the units of $1/(time)^2$, the units of a frequency squared.

Theorem 1.2. *Consider any zero angular momentum solution to the standard attracive ($1/r$ potential) Newton's equations for $d+1$ bodies in d-dimensional. Suppose that along this solution the inter-body distances satisfy the bound*

$$r_{ab} \leq c \tag{1}$$

Then, within every time interval of size $\frac{1}{\pi}(\frac{c^3}{GM})^{1/2}$, this solution has a degeneration instant.

Remark. Theorem 1.2 represents a quantitative improvement of the syzygy estimates found earlier in the case $d = 2$ described above.

Necessity of zero angular momentum in even dimensions. The regular simplex is a central configuration in all dimensions d. If the dimension d is even, say $d = 2k$, then one can uniformly rotate the simplex in a way consistent with a splitting of $\mathbb{R}^d$ into k two-planes to get a relative equilibrium solution to the $d+1$-body problem in $\mathbb{R}^d$ which has nonzero angular momentum and never degenerates. These even-dimensional analogues of the Lagrange rotating equilateral triangle illustrate that for even dimensions d the hypothesis that the angular momentum be zero is necessary in theorem 1.2.

Open Question. Does Theorem 1.1 hold if the zero angular momentum hypothesis is dropped?

Wintner's examples and another question. Wintner [21] describes a four-body solution of the following form. (See the footnote on p. 245 in section 325 of [21].) At each instant the four bodies lie in a plane. That plane spins about a fixed line in space. The angular momentum is not zero. One can turn Wintner's example around with the following question. Suppose the four bodies are coplanar at each instant *and* that their total angular momentum is zero. Must the instantaneous plane containing them be fixed, constant in inertial space, so that the solution is actually one to the planar four-body solution ? In lemma 3.3 below we answer this question "yes".

General two-body type potentials.. There is nothing special about the Newtonian $1/r$ potential in theorem 1.2. It is enough to have a sum of pair potentials of the form

$$V(q) = G\Sigma_{a \neq b} m_a m_b f_{ab}(r_{ab}(q)) \tag{2}$$

where the individual two-body potentials f_{ab} are attractive. (Choose the units so that the f_{ab} have units $1/(length)$.) Assume that

$$f'_{ab}(r) > 0, f''_{ab}(r) < 0, \text{ for } r > 0, \text{ and } \lim_{r \to \infty} \frac{f'_{ab}(r)}{r} = 0. \tag{3}$$

(These assumptions can probably be relaxed to some extent and still yield our results.)Examples include the standard Newtonian 3-dimensional gravitational potential $f_{ab}(r) = -1/r$ and the power law potentials $f_{ab}(r) = -k_{ab}/r^\alpha$ for positive exponent α and positive constants k_{ab}. Hypothesis (3) guarantees that the functions $f'_{ab}(r)/r$ are positive and strictly monotone decreasing so that for each $c > 0$ and pair ab we have that $r_{ab} \leq c \implies \frac{f'_{ab}(r_{ab})}{r_{ab}} \geq \delta_{ab} := \frac{f'_{ab}(c)}{c}$. Taking δ to be the minimum of these δ_{ab} over all pairs we get

$$r_{ab} \leq c \text{ for all pairs } ab \implies \frac{1}{r_{ab}} f'_{ab}(r_{ab}) \geq \delta > 0 \text{ for all pairs } ab. \tag{4}$$

Then, we have

Theorem 1.3. *Consider the zero angular momentum Newton's equations for $N = d+1$ bodies moving in Euclidean d-dimensional space under the influence of the attractive potential (2) whose 2-body potentials satisfy hypothesis (3). Suppose that along such a solution all its inter-body distances r_{ab} satisfy the bound $r_{ab} \leq c$. Then, in every time interval of size $(GM\delta)^{-1/2}/\pi$, this solution has a degeneration instant. Here δ is as in implication (4) above, and M the total mass.*

We now describe the key ingredients behind these Theorems.

Definition 1.1. *Σ is the degeneration locus within configuration space – the set of configurations for which the $d+1$ masses all lie on a single affine hyperplane.*

Σ is a singular hypersurface which cuts configuration space into two disjoint congruent halves, the simplices of positive volume, and those of negative volume. (The sign of the volume depends on the orientation of Euclidean space and the ordering of the masses, which we fix once and for all.) Write $sgn(det(q))$ for the sign of the volume, defined for $q \notin \Sigma$. For example, if $d = 3$, then $sgn(det(q))$ is the sign of the triple product $(q_2 - q_1) \cdot ((q_3 - q_1) \times (q_4 - q_1))$.

Definition 1.2. *The signed distance $S(q)$ of a configuration q of $d+1$ point masses in $\mathbb{R}^d$ is the signed distance from q to the degeneration locus relative to the mass inner product (described in subsection 3.1.1), that distance being assigned a sign according to that of the signed volume of q:*

$$S(q) = sign(det(q))dist(q, \Sigma),$$

with $S(q) = 0$ if and only if $q \in \Sigma$.

S is not smooth everywhere, however its singular locus has codimension 2, which means that most solutions miss this singular set, a fact essential to our proof. On the other hand, the singular set of the distance $|S|$ has codimension 1 since it contains the hypersurface Σ, and for this reason using $|S|$ instead of S would have complicated our proof. In Prop. 6.1 below we prove that this distance $|S(q)| = dist(q, \Sigma)$ is the smallest singular value of a $d \times d$ matrix representing q in the center-of-mass frame.

Theorem 1.4. *[Main computation.] If S is smooth along a zero angular momentum solution $q(t)$ to Newton's equations then $S(t) := S(q(t))$ evolves according to*

$$\ddot{S} = -Sg(q, \dot{q}), \text{ with } g > 0 \text{ everywhere }.$$

If, moreover, all interparticle distances r_{ab} satisfy $r_{ab} \leq c$ then $g \geq GM/c^3$ for the Newtonian ($f_{ab}(r) = -1/r$) potential case, and, more generally, $g \geq GM\delta$ for potentials of the form (2), with (3) in force and δ as per (4).

Acknowledgements: I would like to thank Alain Albouy, Gil Bor, Joseph Gerver, Connor Jackman, Adrian Mauricio Escobar Ruiz, Robert Littlejohn, and Rick Moeckel for useful discussions. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1440140 while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2018 semester.

2. Motivation and Main ideas.

Newton's N-body equations in d-space are invariant under the isometry group of the inertial Euclidean space, $\mathbb{R}^d$, so we can push them down to form a system of ODEs on "shape space", by which we mean the quotient space of the N-body configuration space by the isometry group of $\mathbb{R}^d$. There are actually two shape spaces, depending on whether or not we allow orientation reversing isometries. In the body of this paper we will work on the oriented shapes space which is the quotient formed by using the group $SE(d)$ of orientation-preserving isometries to idenify configurations. In the appendix we discuss the unoriented shape space and the map from oriented to unoriented shape space, which is a branched double cover.

We will speak of "downstairs" to mean we are working on the quotient and "upstairs" to mean we are working on the original configuration space. Upstairs, Newton's equations have the form $\ddot{q} = -\nabla V(q)$. Downstairs on shape space, the equations have precisely this same form **provided** that the total angular momentum is zero [1]. In writing down the downstairs zero-angular momentum Newton's equations, the acceleration $\ddot{q}$ is replaced by the covariant acceleration $\nabla_{\dot{q}}\dot{q}$ where ∇ is the Levi-Civita connection arising out of the the induced *shape metric* downstairs. This shape metric, induced by the flat kinetic energy metric upstairs, is curved.

Robert Littlejohn [9] pointed out to me that the shape space for 4 bodies in $\mathbb{R}^3$ is homeomorphic to the Euclidean space $\mathbb{R}^6$. I do not use this fact, but it is what inspired my belief that some version of the theorems stated here must hold. Littlejohn's observation generalizes to $d+1$ bodies in $\mathbb{R}^d$: its shape space is homeomorphic to $\mathbb{R}^{\binom{d}{2}}$, a fact known to some statisticians (eg [6] and [7]). This fact for the case $d = 2$ of the 3 body problem is explicit and central to all of my explorations into the planar three-body problem. We provide details regarding this topological fact for general d, and of the relation between the oriented and unoriented shape space in the appendix.

We will continue to write Σ for the degeneration locus, either upstairs or downstairs. Downstairs, in shape space, Σ is a totally geodesic hypersurface, at least at its smooth points. This total geodesy is strange since Σ is not totally geodesic upstairs. For example, when $d = 3$, the geodesic connecting a quadrilateral lying in one plane to another quadrilateral in an orthogonal planes consists of straight lines along which the vertices travel, and the resulting one-parameter family in shape space will have nonzero volume, i.e will not be planar, at most instants. To see that total geodesy holds downstairs, choose any orientation reversing isometry R, for example, in the case $d = 3$, a reflection about the xy plane. R induces an isometric involution downstairs, one whose fixed point set is precisely Σ. A general

[1]If the angular momentum is non-zero there are additional 'magnetic' terms" in the equations downstairs, meaning terms linear in velocities, and also additional equations involving 'internal variables" which represent instantaneous rigid body tumbling coupled to dynamics on the shape space, these internal variables lying in co-adjoint orbits for $SO(d)$.

theorem in Riemannian geometry asserts that the fixed point set of an isometric involution is totally geodesic implying the totally geodesic nature of Σ.

Heuristics behind the theorems. Theorem 1.4 asserts that the signed distance S from Σ behaves qualitatively like a one-dimensional harmonic oscillator, oscillating around $S = 0$. The physical intuition behind this phenomenon was pointed out to me by Mark Levi many years ago. The potential is invariant under isometries so descends to a function downstairs. How to interpret this potential downstairs? Write $\Sigma_{ab} \subset \Sigma$ for the binary collision locus $r_{ab} = 0$. One computes that $r_{ab}(s) = \mu_{ab} dist(s, \Sigma_{ab})$ where $dist(s, \Sigma_{ab})$ is the distance from s to Σ_{ab} and where $\mu_{ab} = \sqrt{M/m_a m_b}$. Consequently, re-interpreted downstairs, formula (2) for the potential asserts that a point s in shape space is subjected to the force of an attractive potential exerted by the $\binom{d}{2}$ sources Σ_{ab}, all of which lie in the "hyperplane" Σ. So, of course, the shape is always attracted to Σ! And as long as the shape's "vertical' kinetic energy is not too large, it will always return to cross Σ, oscillating forever back and forth across the attracting 'hyperplane' Σ.

Choice of S versus signed volume. In [13], in proving Theorem 1.1 for the case $d = 2$, I used a function "z" in place of the S of Theorem 1.4. This z equals the signed area of the oriented triangle normalized by divided the area by the moment of inertia I that the triangle *would* have if all masses were assigned the value 1. The obvious generalizations z_d of this z to $d > 2$, namely a normalized signed volume, did not work out. All my attempts at proving a version of Theorem 1.4 for such a function in place of S failed. The function z_2 satisfies a kind of monotonicity relation with respect to geodesics orthogonal to Σ which fails for $z_d, d > 2$ and this monotonicity was required to get positivity of g in Theorem 1.4. The need for such a relation led to introducing S. After the fact, one observes that the identity $z = S/\sqrt{I}$ holds for equal masses when $d = 2$, and fails for $d > 2$.)

Key ingredients to the proof. The proof of Theorem 1.4 relies on four key facts .

- Fact 1. S satisfies the Hamilton-Jacobi equation $\|\nabla S\| = 1$ wherever S is smooth. This fact implies that the integral curves of the gradient flow of S are geodesics.
- Fact 2. The shape metric is everywhere non-negatively curved.
- Fact 3. There is a close relationship between the sign of the second fundamental form of distance level sets (the $\{S = t\}$'s) from a totally geodesic submanifold ($\Sigma = \{S = 0\}$) and the sign of the curvature of the ambient space within which the level sets lie. This relation is detailed in section 2 of Gromov [5], see particularly p. 44 there, and recalled below as proposition 4.1.
- Fact 4. (Theorem 6.1). The singular locus of S has codimension 2. This locus, denoted $Sing(S)$ below, consists of all points at which S is not smooth.

3. Set-up and Reduction.

The proofs of all the theorems hinge on Theorem 1.4 which is a computation. We achieve the computation by exploiting the relations between Newton's equations at zero angular momentum as expressed upstairs on the usual configuration space and downstairs on shape space. The process of pushing the equations downstairs is referred to as "reduction". Our reduction procedure is a metric reduction, putting kinetic energy to the fore, as opposed to the oft-used symplectic reduction. The two reduction procedures are formally equivalent but the metric approach makes our computation tractable. In this section we go through the reduction for the case $d = 3$. At the end, in subsection 3.3, we describe the small changes needed for the set-up of reduction for higher d.

Write $M(k, m)$ for the space of $k \times m$ real matrices. The configuration space for the 4 body problem in $\mathbb{R}^3$ can be naturally identified with the space $M(3, 4)$. To do so, think of the four vectors $q_1, q_2, q_3, q_4 \in \mathbb{R}^3$ defining the positions of the four bodies as column vectors, and place them side-by-side to form the 3×4 matrix

$$q = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \end{pmatrix} \in M(3, 4). \tag{5}$$

The translation subgroup $\mathbb{R}^3$ acts on $M(3, 4)$ by $q_a \mapsto q_a + b$, $b \in \mathbb{R}^3$, which in matrix terms is

$$q \mapsto q + (b, b, b, b) \tag{6}$$

The quotient of $M(3, 4)$ by this action can be identified with the matrix space $M(3, 3)$. This identification depends on choosing a basis for the 3-dimensional subspace $x_1 + x_2 + x_3 + x_4 = 0$ of the mass label space $\mathbb{R}^4$. Such a choice is equivalent a choice of "Jacobi vectors" in Celestial Mechanics but the results that follow are independent of this choice. See the last two paragraphs of subsection 3.1.1 just below for some details.

Definition 3.1. *The rank of a configuration $q = (q_1, q_2, q_3, q_4)$ is the codimension of the smallest affine subspace in $\mathbb{R}^3$ which contains all 4 vertices q_1, q_2, q_3, q_4.*

Upon projection to $M(3, 3)$ the rank of a configuration equals the rank of the representing 3-by-3 matrix so that the degeneration locus is

$$\Sigma = \{q \in M(3, 3) : det(q) = 0\}.$$

FACT. Σ is a singular algebraic surface whose smooth locus consists of the matrices of rank 2.

For a sketch of a proof of this fact see the end of our Appendix and references therein.

3.1. **Oriented Shape Space.** Rotations of $\mathbb{R}^3$ act on both $M(3, 4)$ and its translation-quotient $M(3, 3)$ by

$$q \mapsto gq, g \in SO(3).$$

Definition 3.2. *The oriented shape space $Sh = Sh(3, 4)$ is the topological quotient space $M(3, 3)/SO(3)$. The quotient map $M(3, 3) \to Sh(3, 4)$ will be denoted by π. The projection of a configuration $q \in M(3, 3)$ will be called "the shape" of q.*

REMARK. The group of orientation-preserving isometries of $\mathbb{R}^3$, denoted $SE(3)$, is made up of the translations ($\mathbb{R}^3$) and the rotations ($SO(3)$). We identify $Sh(3, 4)$ with the quotient $M(3, 4)/SE(3)$, by using reduction in stages: we first quotient by translations $\mathbb{R}^3$ to get $M(3, 3)$ and then quotient it by rotations $SO(3)$ to get to $Sh(3, 4)$. The projection $M(3, 4) \to Sh(3, 4)$ will also be denoted by π.

Recall that we say that G acts freely on a set Q if whenever $gq = q$ for some $q \in Q$ we must have that $g = id \in G$. It is well-known (see for example Prop 4.1.23 of [1]) that when a compact Lie group acts freely on a smooth manifold Q then the quotient space Q/G is a smooth manifold in such a way that the quotient map $Q \to Q/G$ is a smooth submersion. $SO(3)$ does not act freely on $M(3, 3)$. However the restriction of the action to the open dense set of matrices having either rank 3 or rank 2 is free. The rank 3 matrices form the complement of Σ. The rank 2 matrices form the smooth points of Σ which itself a singular algebraic hypersurface, and are open and dense within Σ. The remaining points, being the matrices of rank 1 or 0, form the singular locus of Σ. Each such singular point q has continuous isotropy, i.e the subgroup of $SO(3)$ fixing q has positive dimension, being respectively a one-parameter group of rotations or the whole group. Hence we get

Proposition 3.1. *Let $M_{gen} \subset M(3,3)$ be the dense open subset consisting of those matrices whose rank is either 3 or 2. (The subscript 'gen' is for 'generic'.) The restriction of $\pi : M(3,3) \to Sh(3,4)$ to M_{gen} gives $\pi(M_{gen}) \subset Sh(3,4)$ a smooth structure in such a way that this restricted projection is a smooth submersion. Moreover this restricted projection has the structure of a principal $SO(3)$ bundle.*

Remark. In the appendix we show that the complement of M_{gen} has codimension 4 within $M(3,3)$.

3.1.1. *Newton's Equations.* To write down Newton's equations for the motion of the 4 bodies, we need the potential and the choice of masses. We have written down the potential (eq (2)). A choice of mass $m_a > 0$ for each body $a = 1,2,3,4$ defines an inner product $\langle \cdot , \cdot \rangle$ on $M(3,4)$ called the "mass metric" or "kinetic energy metric" according to

$$\frac{1}{2}\langle \dot{q}, \dot{q} \rangle = \frac{1}{2}\Sigma_{a=1}^{4} m_a |\dot{q}_a|^2.$$

We use the absolute value symbol for the usual norm in our Euclidean inertial $\mathbb{R}^3$. When we interpret $\dot{q} \in M(3,4)$ to represent the velocities of the four bodies then the above expression is the usual expression for the total kinetic energy. Newton's equations can now be written

$$\ddot{q} = -\nabla V(q) \tag{7}$$

where the gradient ∇V is computed using the mass inner product: $dV(q)(\delta q) = \langle \nabla V(q), \delta q \rangle$.

The mass inner product induces an inner product on $M(3,3)$ and an isometric embedding of $M(3,3)$ into $M(3,4)$. Indeed, we can identify $M(3,3)$ with the mass-orthogonal subspace to the translation subspace $(b,b,b,b), b \in \mathbb{R}^3$ of eq. (6). This orthogonal subspace equals those configurations $q : \mathbb{R}^4 \to \mathbb{R}^3$ whose center of mass is zero: $q_{cm} = 0$, where $q_{cm} := q(\vec{m}) = m_1 q_1 + m_2 q_2 + m_3 q_3 + m_4 q_4$, and where $\vec{m} = (m_1, m_2, m_3, m_4) \in \mathbb{R}^4$. By choosing a basis for the three-dimensional subspace $\{x : x_1 + x_2 + x_3 + x_4\} \subset \mathbb{R}^4$ we may coordinatize $M(3,3)$ as a space of 3-by-3 matrices in such that the induced mass inner product becomes

$$\langle q, q \rangle = Tr(q^t q),$$

namely, the inner product under which the matrix entries q_{ij} form an orthonormal linear coordinate system. Such a basis corresponds to a choice of *normalized Jacobi vectors*. Once done, Newton's equations on the translation-reduced space $M(3,3)$ have precisely the form as eq (7).

As an example of Jacobi vectors let p_a be any quadruple of positive numbers such that $p_1 + p_2 = p_3 + p_4 = 1$. Then the three vectors

$$J_1 = (1,-1,0,0), J_2 = (0,0,1,-1), \text{ and } J_3 = (-p_1, -p_2, p_3, p_4)$$

are a basis for the subspace $\{x_1 + x_2 + x_3 + x_4 = 0\} \subset \mathbb{R}^4$. We can form such p_a's from our masses m_a by the recipe $m_{12} = m_1 + m_2, m_{34} = m_3 + m_4, p_1 = m_1/m_{12}, p_2 =$

$m_2/m_{12}, p_3 = m_3/m_{34}, p_4 = m_4/m_{34}$ Then J_1, J_2, J_3 form an orthogonal basis for the subspace $\{x_1 + x_2 + x_3 + x_4 = 0\}$ of the mass label space $\mathbb{R}^4$, relative to the restriction of the metric $m_1 dx_1^2 + m_2 dx_2^2 + m_3 dx_3^2 + m_4 dx_4^2$. Normalize the J_i to obtain the normalized Jacobi vectors $E_i = \sqrt{\mu_i} J_i$ with normalization factors μ_i computed to be $\frac{1}{\mu_1} = \frac{1}{m_1} + \frac{1}{m_2}, \frac{1}{\mu_2} = \frac{1}{m_3} + \frac{1}{m_4}$, and $\frac{1}{\mu_3} = \frac{1}{m_{12}} + \frac{1}{m_{34}}$. Set $\rho_i = q(E_i) \in \mathbb{R}^3$, $i = 1, 2, 3$. For example, $\rho_1 = \sqrt{\mu_1}(q_1 - q_2)$. If $q_{cm} = 0$ then we compute that $\|q\|^2 = |\rho_1|^2 + |\rho_2|^2 + |\rho_3|^2$. (Compare [9], eqs (2.5), (2.6), pp. 2036).) To summarize, by identifying $M(3, 3)$ with the subspace $q_{cm} = 0$ and then expressing the resulting elements of $M(3, 3)$ in terms of the E_i as the 3-by-3 matrix whose columns are ρ_1, ρ_2, ρ_3 we have realized our translation reduced space $M(3, 3)$ as the space of real 3-by-3 matrices in such a way that the the squared norm for the mass inner product is equal to the usual matrix squared norm $Tr(q^t q)$.

Remark. Albouy and Chenciner [2] have developed a beautiful and deep linear algebra which they call the algebra of "dispositions' to explain and understand the choosing of Jacobi vectors, their relation to the subspace $\{x_1 + x_2 + x_3 + x_4\} \subset \mathbb{R}^4$, and to allow as much as possible of the reduction process to be achieved in a mass-independent and basis-independent manner. We do not need dispositions here, but they have steered our computations and our understanding. Also Moeckel [10] for an exceptionally clear exposition of dispositions.

3.2. **Reduced Newton's equations.** We push Newton's equations and the kinetic energy metric down to shape space. For this purpose it will be helpful to keep in mind the following generalities.

Metric projections and Riemannian submersions. Whenever we have a metric space M with distance function d_M and an onto map $\pi : M \to B$ we can *try* to define a metric d_B on B by $d_B(b_1, b_2) = d_M(\pi^{-1}(b_1), \pi^{-1}(b_2))$, or, in English, the distance between points downstairs is the distance between their corresponding fibers upstairs. When this construction works we say that $\pi : M \to B$ is a metric projection or *submetry*. If $B = M/G$ is the quotient of M by the action of a compact Lie group acting on M by isometries and π is the quotient projection then the construction always works. If, in addition, M is a manifold whose metric d_M comes from a Riemannian metric and if the G-action is free so that the quotient map π is a smooth submersion with smooth B, then the induced distance function d_B also arises as the distance function of a Riemannian metric on M. In this case $\pi : M \to B$ is a *Riemannian submersion* which has the following infinitesimal meaning. The 'vertical space" $V_q \subset T_q M$ through $q \in M$ is defined to be the kernel of $d\pi_q$; equivalently, it is the tangent space at q to the fiber $\pi^{-1}(s) = Gq$ through q. Define the "horizontal space" H_q to be the orthogonal complement to the vertical: $H_q = V_q^{\perp}$. Then the restriction of $d\pi_q$ to H_q is a linear isomorphism. Declaring this linear isomorphism $H_q \to T_s B$ to be an isometry induces an inner product on $T_s B$, and this inner product is independent of the point $q \in \pi^{-1}(s)$ since G acts isometrically on M. Distance minimizers between fibers upstairs are geodesics in M orthogonal to the fibers. From this follows the well-known fact that geodesics orthogonal to fibers at one point are orthogonal at every point, and that the geodesics downstairs in B are precisely the projections of horizontal geodesics upstairs.

In this way, starting from the mass metric on $M(3,4)$ or $M(3,3)$, we get a metric on $Sh = Sh(3,4)$ which is Riemannian at the generic shapes (those of rank 2 or 3) and over these points is such that $\pi : M(3,3) \to Sh$ is a Riemannian submersion.

To push Newton's equations down to Sh we must understand the *dynamical meaning* of being horizontal in $M(3,3)$. In [12] (or [16]) I compute that $\dot{q}$ is orthogonal to the $SO(3)$ orbit through q if and only if the total angular momentum $J(q,\dot{q})$ of the pair $(q,\dot{q})$ is zero. The expression for J as a function on $TM(3,4) = M(3,4) \times M(3,4)$ is $J(q,\dot{q}) = \Sigma m_a q_a \wedge \dot{q}_a$ for $M(3,4)$, and is the same when restricted to $TM(3,3)$ viewed as subspace of $TM(3,4)$. To reiterate:

$$V_q^{\perp} := H_q = \{v \in T_q M(3,3) : J(q,v) = 0\} \tag{8}$$

Recall that J is conserved for any potential of the form of eq (2), that is to say $J(q(t),\dot{q}(t)) = J(q(0),\dot{q}(0))$ along solutions $q(t)$ to Newton's equations. Now let ∇ be the Levi-Civita connection for the shape metric. Observe that since the potential is $SE(3)$ invariant it also defines a projection on Sh. We will use the same symbol V for the potential upstairs and downstairs. We have

Lemma 3.1. *Any zero angular momentum solution to Newton's equations passing through generic (i.e. rank 2 and 3) points of $M(3,3)$ projects to a curve γ in shape space which satisfies*

$$\nabla_{\dot{\gamma}} \dot{\gamma} = -\nabla V(\gamma(t)).$$

Conversely, the horizontal lift of such a solution is a zero-angular momentum solution to Newton's equations upstairs.

Regarding 'horizontal lift" see, again [12] or chapter 13 of [16].

Proof. This theorem is a general fact, holding for any Hamiltonian of the form kinetic plus potential on any manifold endowed with the smooth free action of a Lie group which keeps both the kinetic (metric) and potentials invariant. For a proof see for example, [16].

The special case when $V = 0$ will be useful below.

Lemma 3.2. *Any zero angular momentum straight line $q + tv$ in $M(3,3)$ projects to a geodesic in Shape space $Sh(3,4)$. Conversely, the horizontal lift of any geodesic in $Sh(3,4)$ is a zero-angular momentum straight line in $M(3,3)$. The geodesic is parameterized by arc length if and only if $\|v\| = 1$.*

A planar aside. By using the fact that the constraint $J = 0$ of eq (8) defines a connection on the principal $SO(3)$ bundle $U \to \pi(U)$ we can prove the following

Lemma 3.3. *Let $q(t) = (q_1(t), q_2(t), q_3(t), q_4(t))$ be a solution to the 4-body problem in $\mathbb{R}^3$ and suppose that for each time t in some interval all 4 bodies lie in a (possibly moving) plane $\Pi(t) \subset \mathbb{R}^3$. Suppose moreover that the solution has angular momentum zero. Then the plane $\Pi(t)$ is in fact a constant plane Π and the solution is a solution to the 4-body problem restricted to this plane.*

Proof. Suppose that the interval is of the form $[0,\epsilon)$. For simplicity, let us assume that $q(0)$ is in fact planar, rather than collinear. Let $\Pi \subset \mathbb{R}^3$ be the plane containing the $q_a(0)$ and write $W(\Pi) \subset T_q M(3,3)$ for the linear subspace of all velocities tangent to Π, that is to say, the linear space of velocities $v = (v_1, v_2, v_3, v_3) \in M(3,3)$ for which each $v_a \in \Pi$. It is enough to show that

$$\text{CLAIM 1:} \qquad \dot{q}(0) \in W(\Pi), \tag{9}$$

since if both positions and velocities lie in the fixed plane Π, then the entire solution $q(t)$ lies in that plane.

Write $c = \pi(q(0)) \in Sh(3,4)$ for the projection of the initial configuration to shape space and $\dot{c} = d\pi_{q(0)}\dot{q}(0)$ for the initial shape velocity. Since $q(0)$ is planar we have that c is rank 2 and so forms a smooth point of $\pi(\Sigma)$. By the planarity assumption we know that $\dot{c} \in T_c\Sigma \subset T_cSh(3,4)$. Write $h_{q(0)} : T_cSh(3,4) \to H_{q(0)} \subset T_qM(3,3)$ for the horizontal lift operator. Recall that $h_{q(0)}$ is the inverse of the restriction of the linear operator $d\pi_q : T_qM(3,3) \to T_cSh(3,4)$ to the subspace $H_{q(0)}$. Then we have that $\dot{q} = h_{q(0)}\dot{c}$ since $J(q(0), \dot{q}(0)) = 0$. (Compare eq (8).) Thus $\dot{q}(0) \in h_{q(0)}(T_c\Sigma)$. It follows that we have established our claim (9) once we show that

$$\text{(10)} \qquad \text{CLAIM 2:} \qquad h_{q(0)}(T_c\Sigma) = W(\Pi) \cap H_{q(0)}$$

.

At this point the reader may find it instructive to compute the dimensions of both sides of eq (10) to verify that they are both 5 dimensional hypersurfaces within the 6 dimensional $H_{q(0)}$.

To establish the claim of eq (10) , note that any vector $v \in W(\Pi) \cap H_{q(0)}$ represents a deformation $q_a \mapsto q_a + \epsilon v_a$ of $q(0)$ *within* the plane Π and hence projects to a vector $\delta c = d\pi_{q(0)}v \in T_c\Sigma$. Moreover, any vector $\delta c \in T_c\Sigma$ can be so realized as $d\pi_{q(0)}v$ for such a v since any planar deformation of $q(0)$ can be rotated so as to be realized within Π, and since $H_{q(0)} = ker(d\pi_{q(0)})^{\perp}$, so that intersecting with $H_{q(0)}$ is just a way of getting rid of this rotational ambiguity in the representation. Thus $d\pi_{q(0)}(W(\Pi) \cap H_{q(0)}) = T_c\Sigma$. Apply the inverse $h_{q(0)}$ of $d\pi_{q(0)}$ to both sides of this last equality to obtain eq (10).

The case where $q(0)$ is collinear rather than planar takes more work and we will not delve into it.

3.3. **Set-up for general dimension d.** GOING FROM $d = 3$ TO GENERAL d. The configuration space for N bodies in $\mathbb{R}^d$ is the space $M(d,N)$ of $d \times N$ real matrices. Its quotient by the translation group of $\mathbb{R}^d$ forms an $M(d, N-1)$ once a basis for the hypersurface $x_1 + x_2 + \ldots + x_N = 0$ of the 'zero centered' mass-label space is chosen. This basis can be viewed as a choice of Jacobi vectors. When $N = d+1$ we get the translation-reduced configuration space to be the space $M(d,d)$ of square matrices with the degeneration locus $\Sigma = \{q : det(q) = 0\}$. Shape space is $Sh(d, d+1) = M(d,d)/SO(d) = M(d, d+1)/SO(d)$. Proposition 3.1 holds with 'rank 2 and 3' replaced by 'rank $d-1$ and rank d.

Introducing masses puts an inner product on the mass label space, and so on $M(d, d+1)$ and on its translation quotient $M(d,d)$. The masses also allow us to identify $M(d,d)$ as a linear subspace, rather than a quotient space, of $M(d, d+1)$, namely as the subspace of center-of-mass zero configurations. An orthonormal basis for the hypersurface $\Sigma x_i = 0$ is equivalent to a choice of normalized Jacobi vectors $q_1, \ldots, q_d$ and relative to the coordinates whose components are the coordinates of these normalized Jacobi vectors ($q_{ij} := (q_j)_i$) the mass-induced inner product structure on $M(d,d)$ becomes standard : $\langle q, q\rangle = tr(q^tq) = \Sigma_{i,j}q_{ij}^2$. The rotation group $SO(d)$ acts isometrically on $M(d,d)$ by left multiplication. The action also leaves invariant so that both the metric and the zero-angular momentum Newton's equation push down to the quotient shape space $Sh(d, d+1) = M(d,d)/SO(d)$.

The reduction lemmas 3.1 and 3.2 for this reduced dynamics hold as stated, upon replacing '3' by 'd' in the obvious places.

4. Proving Theorem 1.4: signed distance as an oscillator.

We proceed to differentiate S along a solution arc which does not pass through any singular point of S. We have

$$\dot{S} = \langle \nabla S, \dot{\gamma} \rangle$$

so that

$$\ddot{S} = \langle \nabla S, \nabla_{\dot{\gamma}} \dot{\gamma} \rangle + \langle \nabla_{\dot{\gamma}} \nabla S, \dot{\gamma} \rangle \tag{11}$$

$$= \langle \nabla S, -\nabla V \rangle + \langle \nabla_{\dot{\gamma}} \nabla S, \dot{\gamma} \rangle \tag{12}$$

We estimate each term of this last equation separately, showing that each term has the form $-Sg$ with $g \geq 0$. We verify that the 'g' for the first term is always positive and satisfies the stated bounds when $r_{ab} \leq c$.

First term, $\langle \nabla S, -\nabla V \rangle$. The vector field ∇S exists and is smooth in a neighborhood of any smooth point p of S. The integral curve of ∇S through such a p is the geodesic orthogonal to the level set $\{S = S(p)\}$. We will implement this fact along $\Sigma = \{S = 0\}$. All rank 2 points of $M(3,3)$, i.e the planar configurations of Σ, are smooth points of S, so that at such a point the integral curve of ∇S is a geodesic orthogonal to Σ, which itself is smooth at p. (See Theorem 6.1 below. Note that at a rank 2 point of Σ the smallest singular value, namely $|S|$, is zero, while the second smallest is nonzero.) In particular these geodesics are orthogonal to the level set $S = 0$, the degeneration locus Σ. These facts regarding the relationship between S, its smooth points, geodesics, and ∇S hold true generally for the signed distance function S from a hypersurface on any Riemannian manifold, and are closely related to the fact that signed distance satisfies the Hamilton -Jacobi equation: $\|\nabla S\| = 1$.

We proceed in the special case of $d = 3$ for this paragraph, for simplicity. The geodesics in $M(3,3)$, or in shape space, are the projections of straight lines $q + tv$ in $M(3,4)$ for which $(q,v) \in M(3,4) \times M(3,4)$ has zero total angular momentum and zero total linear momentum. See lemma 3.2 above. (Zero linear momentum arises by identifying $M(3,3)$ with the zero center-of-mass configurations. Alternatively, having zero linear momentum is equivalent to the assertion that the velocity v is orthogonal to the translation action.) The parameter t is arclength provided $\langle v, v \rangle = 1$. Now the smooth points q of the degeneration locus Σ are the planar points. In order for a geodesic to be perpindicular to Σ at such a q we must have that v is perpindicular to all $\delta q \in T_q\Sigma$. By rotating, we may assume that the 4 vertices of q lie in the xy plane which we will denote by "$\mathbb{R}^2$". Then any variation $\delta q = (\delta q_1, \delta q_2, \delta q_3, \delta q_4)$ with $\delta q_a \in \mathbb{R}^2$ represents a planar variation of q and hence a tangent vector to Σ at q. Since

$$\langle \delta q, v \rangle = \Sigma m_a (\delta q_a) \cdot v_a$$

and since the δq_a are arbitrary vectors in $\mathbb{R}^2$, we see that our tangent vector v must have all 4 of its component vectors v_a perpindicular to $\mathbb{R}^2$, which is to say, along the z-axis. But then, along our geodesic the squared inter-body distances are

$$r_{ab}^2 = |q_a + tv_a) - (q_b + tv_b)|^2 = r_{ab}(0)^2 + t^2 |v_a - v_b|^2 \tag{13}$$

where the cross term is zero since q_a, q_b lie in $\mathbb{R}^2$ while v_a, v_b are orthogonal to $\mathbb{R}^2$.

For general d, equation (13) continues to hold for a geodesic orthogonal to the degeneration locus. Indeed, the only real difference between the proof above for $d = 3$ and the proof for $d > 3$ is notational. Now the q_a, representing a point on the degeneration locus, can be taken to all lie in a fixed affine hyperplane of $\mathbb{R}^d$ so that the variations δq_a, $a = 1, \ldots N = d + 1$ can be taken to be arbitrary vectors tangent to the correspoding linear hyperplane $\mathbb{R}^{d-1}$. As a consequence the v_a all lie in the one-dimensional orthogonal to this $\mathbb{R}^{d-1}$ and the computation is the same.

Now look at the negative of the potential in the gravitational case:

$$U = -V = G\Sigma \frac{m_a m_b}{r_{ab}}$$

along our geodesic. Each individual term $\frac{m_a m_b}{r_{ab}}$ is *strictly decreasing* or constant in t^2. Indeed $\frac{d}{dt}\frac{1}{r_{ab}(t)} = -\frac{t|v_{ab}|^2}{r_{ab}(t)^3} = -S\frac{|v_{ab}|^2}{r_{ab}(t)^3}$, since $S = t$ as long as the geodesic is the unique minimizer to the degeneration locus. Summing, we obtain

$$\langle \nabla S, -\nabla V \rangle = \langle \nabla S, \nabla U \rangle = -S g_1$$

with

$$g_1 = G\Sigma m_a m_b \frac{|v_{ab}|^2}{r_{ab}(t)^3} > 0$$

as desired.

If each r_{ab} is bounded above by c, we have that $g_1 \geq \frac{G}{c^3}\Sigma m_a m_b |v_{ab}|^2$. But, if $\Sigma m_a v_a = 0$, we find that $\|v\|^2 = \Sigma m_a m_b |v_{ab}|^2 / M$ ("Lagrange's identity") and since we have that $\|v\|^2 = 1$ (since t is arclength) it follows that $\Sigma m_a m_b |v_{ab}|^2 = M$ which yields $g_1 \geq GM/c^3$, which completes the proof for the gravitational case.

In the case of a general potential satisfying hypothesis (2), (3) we get that $\frac{d}{dt} f_{ab}(r_{ab}) = f'_{ab}(r_{ab})\frac{d}{dt}(r_{ab}(t)) = f'_{ab}(r_{ab})(t v_{ab}^2)/r_{ab} = S(\frac{f'(r_{ab})}{r_{ab}})(v_{ab})^2$. Summing, we get $\langle \nabla S, -\nabla V \rangle = -S g_1$ with $g_1 = G\Sigma m_a m_b (\frac{f'(r_{ab})}{r_{ab}})(v_{ab})^2 > 0$. Under the boundedness assumption, eq (4) yields that $\frac{f'(r_{ab})}{r_{ab}} > \delta$ for all pairs a, b and the lower bound for g_1 proceeds exactly as in the previous paragraph.

QED for Term 1.

Second term,$\langle \nabla_v \nabla S, v \rangle$. For a fixed shape p, $p \notin Sing(S)$

$$v \mapsto Q_p(v, v) := \langle \nabla_v \nabla S, v \rangle, v \in T_p Sh$$

is a quadratic form on the tangent space $T_p Sh$. **We will show that** $Q_p(v, v) = -S(p)H_p(v, v)$ **where** $H_p \geq 0$ **is a positive semi-definite quadratic form.** The trick for achieving this inequality is to recognize the quadratic form Q_p as being essentially the second fundamental form of the equidistant hypersurface Σ_t from Σ which passes through p, namely

$$\Sigma_t := \{S = t\}; \text{ where } t = S(p)$$

and then to use a relation between the sign of such second fundamental forms and the sign of the ambient curvature.

Take $v = \nabla S$ in $Q_p(v, v)$. Differentiate the identity $\langle \nabla S, \nabla S \rangle = 1$ with respect to v to see that $\langle \nabla_v \nabla S, v \rangle = 0$, so that $Q_p(v, v) = 0$.

Take $v \perp \nabla S$. Then v is tangent to Σ_t while ∇S is the unit normal N to Σ_t. Recall that second fundamental form to a hypersurface V with unit normal vector field N is the quadratic form $\Pi(v, v) = v \mapsto \langle \nabla_v N, v \rangle$ defined for vectors v tangent to V. It follows that for $Q_p(v,v) = \Pi_p(v,v)$ for $v \perp \nabla S$ is the second fundamental form Π_p of the hypersurface Σ_t at the point $p \in \Sigma_t$. Summarizing:

$$Q_p(v,v) = \begin{cases} 0 \text{ for } v \parallel \nabla S \\ \Pi_p(v,v) \text{ for } v \perp \nabla S \end{cases}$$

We recal some facts about the second fundamental form Π of a hypersurface.

- (1) A hypersurface is *totally geodesic* if and only if $\Pi = 0$.
- (2)Replacing the choice of unit normal N to the hypersurface by its negative $-N$ replaces Π by its negative $-\Pi$.

Our hypersurface Σ is totally geodesic, as mentioned earlier in 'heuristics'. Indeed, Σ is the fixed point set of an isometric involution $i : Sh \to Sh$ and fixed point sets of isometric involutions are always totally geodesic. This isometric involution i, called "reflection about Σ", is implemented by the nontrivial element of the two-element group $O(d)/SO(d)$. Any orientation reversing orthogonal transformation $R \in O(d)$ realizes this nontrivial element and acts on shape space by sending the shape $s = \pi(q)$ to $i(s) = \pi(Rq)$. Now $i^*S = -S$ from which it follows that $i_*\nabla S = -\nabla S$, and thus, using item (2) above, that $i^*Q = -Q$. It follows that we can write $Q = -SH$ where $i^*H = H$. It remains to show that H is positive semi-definite.

A necessary detour into curvatures.

Definition 4.1. *A hypersurface is convex relative to the choice of normal N if $\Pi \geq 0$ for this choice of normal, and concave relative to N if $\Pi \leq 0$ for this choice of normal.*

Example 4.1. *The boundary of a convex domain having smooth boundary in Euclidean space is convex in the above sense provided we use the outward pointing normal.*

Let M be a Riemannian manifold and $V \subset M$ a hypersurface in M, together with a choice of unit normal N along the hypersurface. Then close to V we have the family $V_s, -\epsilon < s < \epsilon$ of nearby equidistant hypersurfaces formed by travelling along the geodesics tangent to the unit normal N for a distance s. By flowing along these geodesics we also have diffeomorphisms

$$\phi_s : V \to V_s.$$

Write Π_0 for the second fundamental form of V relative to N and Π_s for that of the equidistant V_s. Recall that we say that M is "non-negatively curved" if its sectional curvatures are all positive or zero, and "non-positively curved" if all of its sectional curvatures are all negative or zero. The following basic relationship between extrinsic and intrinsic curvature is found in section 2 of Gromov [5], particularly p. 44 there.

Proposition 4.1. *(See figures 1). If the ambient curvature of the Riemannian manifold M is non-negative and if the hypersurface $V \subset M$ is concave with respect to the choice of unit normal N for V, then its positive equidistants $V_s, s > 0$ are at least as concave as V: $\phi_s^*\Pi_s \leq \Pi_0 \leq 0$ for $s > 0$.*

If the ambient curvature of M is non-positive and if the hypersurface $V \subset M$ is convex with respect to N, then its positive equidistants V_s, $s > 0$ are at least as convex as V: $\phi_s^*\Pi_s \geq \Pi_0 \geq 0$ for $s > 0$.

End of proof for the 2nd term. By the O'Neill formula for curvature (see Cor. 1, eq (3), p. 466 of [18]) Riemannian submersions can only *increase* sectional curvatures. This fact implies that the base space B of a Riemannian submersion is non-negatively curved provided that its total space Q is non-negatively curved. Applying this formula to $\pi : M(d,d) \to Sh$ we get that Sh is a non-negatively curved manifold at all smooth points. (Indeed the sectional curvature of a two-plane in T_pSh which is spanned by orthonormal vectors $v, w \in T_pSh$ is $\sigma = \frac{3}{4}\|F_p(v,w)\|^2$ where F is curvature of the Riemannian submersion when viewed as a principal $SO(d)$-bundle. See also lemma 2, p.461 of [18] which describes the relation between a certain tensor A defined in his corollary 1, eq (3), and the curvature.)

By proposition 4.1 and the fact that $\Sigma = \Sigma_0$ is totally geodesic at each smooth point, we have that each $\Sigma_s, s > 0$ is concave relative to ∇S, which is to say that $Q_p \leq 0$ for $S > 0$. It follows that $H \geq 0$ and by symmetry, as above, we have that $Q = -SH$ with H a positive semi-definite form, as desired.

QED for the second term and the proof of Theorem 1.4.

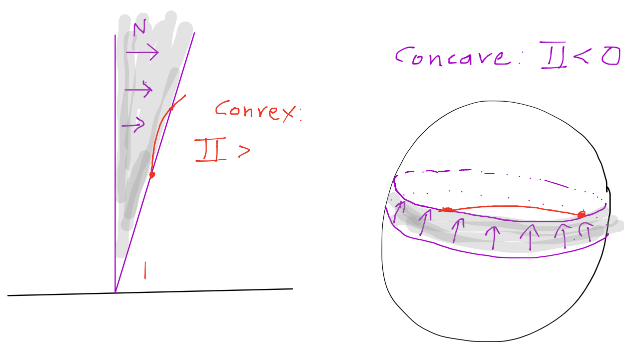

Figure 1. The relation between the sign of the sectional curvatures and convexity of equidistant hypersurfaces to a totally geodesic submanifold. The left figure depicts an equidistant from a geodesic in the hyperbolic plane (ambient curvature -1. The right figure pictures an equidistant from a geodesic on the sphere (ambient intrinsic curvature 1). The first equidistant is convex relative to the normal while the second is concave.

5. Proofs of Theorems

We prove theorems 1.1, 1.2 and 1.3 by strengthening Theorem 1.4:

Proposition 5.1. *Regardless of whether or not S is smooth along the zero angular solution γ to Newton's equations, the composition $S \circ \gamma$ is a convex function of t for $S > 0$ and a concave function for $S < 0$. If γ is bounded with bounds $r_{ab} \leq c$ then $S \circ \gamma(t) = 0$ for at least one time t in each time interval of length $\Delta t = \pi(c^3/GM)^{1/2}$, in the Newtonian potential case and length $\pi(GM\delta)^{-1/2}$ for the general potential case as per hypothesis (2), (3) and (4).*

Proof of Proposition 5.1 .

We first consider the case when S is smooth along γ, treating the general case as a limit of the smooth case.

If S is smooth along γ then Theorem 1.4 asserts that $\ddot{S} = -Sg$ with $g > 0$ and smooth. The convex/concave properties of $S \circ \gamma$ follow immediately. In case the bounds on the r_{ab} are in force then we know that and $g \geq \omega^2 = GM/\delta$ with δ as per hypothesis (2), (3) and (4) in the case of general two-body potential and $\delta = 1/c^3$ in the particular case of the Newtonian potential. Compare our differential equation for S to the oscillator equation $\ddot{S} = -S\omega^2$. The solutions of the later, being $S = Asin(\omega(t - t_0))$, have successive zeros $t_0, t_1, \ldots$ spaced regularly at increments of length π/ω. By the Sturm comparison theorem, between any two of these zeros lies a zero of our S. Since $1/\omega = \sqrt{\delta/GM}$ this yields the result for the smooth case.

For the general case, it will suffice to know that set of points at which S fails to be smooth has codimension 2. We call this the singular set of S and denote it by $Sing(S)$. This assertion regarding the codimension of $Sing(S)$ is theorem 6.1 of the next section.

Assuming the validity of this codimension theorem 6.1, let γ be a zero angular momentum solution to Newton's equations. Then, by using the smooth dependence of solutions on initial conditions, we can find a family of solutions γ_ϵ in $M(d, d)$ which avoids $Sing(S)$ and converges in the uniform (C^0) topology (or even C^k topology for any k) to γ on compact time intervals as $\epsilon \to 0$. By lemma 1, each $S \circ \gamma_\epsilon$ is convex wherever it is positive and concave wherever negative. The properties of being convex or concave are closed in the C^0-topology, i.e. the uniform limit of convex functions is convex. Since $S \circ \gamma_\epsilon \to S \circ \gamma$ in the C^0 topology our result for the convexity / concavity of $S \circ \gamma$ follows. We proceed to the boundedness implications. If the original γ satisfies $r_{ab}(\gamma(t)) \leq c$, then its approximating curves γ_ϵ almost satisfy this bound, namely, they satisfy $r_{ab}(\gamma_\epsilon(t)) \leq c + o(1)$ as $\epsilon \to 0$, since they C^0-converge to γ. Thus, by the preceding paragraph, each $S \circ \gamma_\epsilon$ has a zero in any interval of length $\Delta t = \pi/\omega(\epsilon)$ with $\omega(\epsilon) = GM/\delta(\epsilon)$ and $\delta(\epsilon) = \delta(c + o(1))$ as per eq (4) above. (Explicitly for the Newtonian case, $\delta(c + o(1)) = 1/(c + o(1))^3$.) Since $S \circ \gamma_\epsilon \to S \circ \gamma$ we see that $S \circ \gamma$ must have a zero in every time interval of length $\pi/\omega = \pi\sqrt{\delta/GM}$.

QED

Proof of Theorems. Theorems 1.2 and 1.3 follow immediately from proposition 5.1. Theorem 1.1 is the case $d = 3$ of theorem 1.2.

All that remains to do now in the way of proofs is to establish that the codimension of $Sing(S)$ is 2.

6. Singular set of S and Singular Value Decomposition.

We compute the codimension of $Sing(S) \subset M(d, d)$ (theorem 6.1) and relate the value of $S(m)$ to the singular value decomposition of the matrix $m \in M(d, d)$ (therem 6.1).

The signed distance function S enjoys a larger symmetry group than Newton's equations, symmetries crucial for identifying $Sing(S)$. We saw in subsection 3.3 that the translation-reduced configuration space is the space of square matrices

$M(d,d)$, that $\Sigma \subset M(d,d)$ is given by $det(q) = 0$ and that by choosing an appropriate basis ("Jacobi vectors") for "mass label space" we can insure that the mass inner product agrees with the standard Euclidean inner product so that the norm squared of a matrix is $tr(q^t q) = \Sigma_{i,j} q_{ij}^2$. By inspection, the action

$$q \mapsto g_1 q g_2^t, g_1, g_2 \in O(d). \tag{14}$$

of $O(d) \times O(d)$ on $M(d,d)$ is an isometric action which preserves Σ. It follows that $|S(q)|$, which is the distance from q to Σ is invariant under this group action. The action does not quite preserve our signed distance, since the $O(d)$'s can reverse orientation. Indeed

$$S(g_1 q g_2^t) = \pm S(q); \text{ where } \pm = det(g_1)det(g_1).$$

from which it follows that the action of $SO(d) \times SO(d)$ preserves S.

The first, or left $O(d)$ action (g_1, in eq (14)) is the usual action of rotations on the configuration space of the N body problem in $\mathbb{R}^d$. The second, or right $O(d)$ (g_2, in eq (14)) is not a symmetry of Newton's equations. Littlejohn and Reinsch [9] refer to this second $O(d)$ as the "democracy group" since its action on the matrix space corresponds to choosing new basis for the mass label space, and, at least in the case of equal masses, contains the permutation group which acts by interchanging masses.

The Singular Value Decomposition [SVD] from Matrix theory (eg see [19]) is a normal form theorem for this group action (14). This decomposition asserts that for any $q \in M(d,d)$ there is a *diagonal matrix* x and matrices $g_1, g_2 \in O(d)$ such that

$$q = g_1 x g_2^t, \qquad [SVD1] \tag{15}$$

Moreover the g_i can be chosen so as to force every nonzero entry of x to be positive, and the diagonal entries to be listed in descending order, thus:

$$x = diag(x_1, x_2, \ldots, x_d), x_1 \geq x_2 \geq \ldots \geq x_d \geq 0. \qquad [SVD2] \tag{16}$$

The diagonal x written in this form is unique. Its diagonal entries x_i are called the "ith principal values" of q. The x_i^2 are the eigenvalues of both $q^t q$ and qq^t.

Proposition 6.1. *The distance function $|S(q)|$ of $q \in M(d,d)$ to Σ is equal to x_d above, the dth (smallest) principal value of q.*

We prove this proposition in the next subsection, below.

If we impose the constraint that $(g_1, g_2) \in SO(d) \times SO(d)$ when performing the normal form computations, then we get the following 'specialized' version of the SVD called the "pseudo-singular value decomposition" by [7] (see p. 361).

Proposition 6.2 (PsSVD)**.** *Given any $q \in M(d,d)$ there is a pair $(g_1, g_2) \in SO(d) \times SO(d))$ and a unique diagonal $x = diag(x_1, x_2, \ldots, x_d)$ satisfying $x_1 \geq x_2 \geq \ldots \geq x_{d-1} \geq |x_d|$, and with $x_d < 0$ allowed, such that*

$$q = g_1 x g_2, g_i \in SO(d).$$

Then

$$S(q) = x_d$$

and $sign(x_d) = sign(det(q)) = sign(S(q))$.

In words, the signed distance S is the last 'signed' singular value of q in its pseudo-singular value decomposition.

Proof of Prop 6.2 assuming Prop 6.1. The value of $det(q)$ cannot be changed by acting on it by $(g_1, g_2) \in SO(d) \times SO(d))$ and is equal to $x_1 x_2 \dots x_d$ if $q = g_1 x g_2^t$ with $x = diag(x_1, \dots, x_d)$. Now use the SVD for q. If either one of the elements g_i of the SVD for q is in $O(d)$ but not in $SO(d)$ then we can premultiply that element by $diag(1, 1, \dots, 1, -1) \in O(d)$ to get a new $g_i \in SO(d)$ at the expense of perhaps changing x_d to $-x_d$. Keeping track of the signs of $det(q)$ and of S yields that $S(q) = x_d$, the last 'special' (or 'signed') singular value.

QED

Finally, here is the assertion we need to complete all our proofs.

Theorem 6.1. *The signed distance function $S : M(d,d) \to \mathbb{R}$ is smooth at any point q of $M(d,d)$ whose smallest two principal values are distinct. The complementary set, the singular locus of S, is the set of matrices q whose dth and $d-1$st singular values are equal: $x_{d-1} = |x_d|$. This locus is a semi-algebraic set of codimension 2 within $M(d,d)$.*

A slice. The fact underlying the proofs of the propositions and theorems just stated (theorem 6.1 etc) is that the linear subspace $D \subset M(d,d)$ of diagonal matrices is a *global slice* for our $O(d) \times O(d)$ action (eq (14)) on $M(d,d)$. Recall that the *orbit* of $q \in M(d,d)$ under this action is the set $\{g_1 q g_2^t : g_1, g_2 \in O(d)\} \subset M(d,d)$ and that, from basic manifold theory, the orbit is a smooth submanifold. The assertion that D is a slice for the action means a number of things

- (a) every $O(d) \times O(d)$ orbit intersects D
- (b) the orbit intersects D orthogonally
- (c) the intersection is transverse for generic orbit (i.e generic q)

Assertion (a) follows from the SVD.

Assertion (b) is a computation. Let ξ_1 and ξ_2 be skew symmetric matrices representing elements of the Lie algebra of our $O(d)$'s, understood to represent the derivatives of the g_i along curves passing through $g_i = Id$. Then the tangent space to the orbit through x for $x \in D$ of the orbit consists of all $d \times d$ matrices v of the form

$$v = \xi_1 x - x \xi_2. \tag{17}$$

One sees by direct computation that the diagonal entries of v are all zero, so that $v \perp D$.

Assertion (c) follows by taking "generic" matrix to mean one all of whose principal values are **distinct**, and then making a more detailed computation based on the orbit tangent space equation (17). If we take $\xi_2 = -\xi_1$ in that equation and set $\xi = \xi_1$ then we compute that v is skew-symmetric with entries $(x_i + x_j)\xi_{ij}$ where ξ_{ij} are the entries of ξ. On the other hand, if we take $\xi_2 = \xi_1 = \xi$ in eq(17) we obtain that v is a symmetric matrix with entries $(x_j - x_i)\xi_{ij}$. Now if the x_i are the distinct principal values, we have that $x_i \pm x_j \neq 0$ for all $i \neq j$ and it follows easily from this we can obtain any skew-symmetric matrix as a v as per eq (17), and that we can also obtain any symmetric matrix v which has zeros on its diagonal. Since any matrix at all is the sum of a symmetric and a skew-symmetric matrix we see that the tangent space to the orbit at a generic x consists of all matrices v with zero entries on the diagonal, which comprises the orthogonal complement to D.

Proof of Proposition 6.1. As noted just after we introduced the action in eq (14), the distance function $|S(q)|$ is invariant under the $O(d) \times O(d)$ action:

$$|S(g_1 x g_2^t)| = |S(x)|.$$

Now $det(x) = x_1 x_2 \dots x_d$ so that $\Sigma \cap D = \{x_1 x_2 \dots x_d = 0\}$ is the union of the d coordinate hyperplanes $x_i = 0$. The metric on $M(d,d)$ is Euclidean in the entries, and D is a d-dimensional linear space and in particular totally geodesic: any minimizing geodesic connecting points of D is a line segment within D. This implies that for $x \in D$ the $M(d,d)$-distance of x to Σ equals the D-distance of x to $\Sigma \cap D$, that is the distance as realized by line segments *within* D. It follows that the problem of computing that distance is a problem in Euclidean geometry.

To solve the problem, let us first fix attention to the case $d = 3$. Observe that x_1, x_2, x_3 are orthonormal linear coordinates on D. The Euclidean distance of (x_1, x_2, x_3) from the plane $x_1 = 0$ is $|x_1|$. Since $\Sigma \cap D$ is the union of the three planes $x_1 = 0$, $x_2 = 0$ and $x_3 = 0$, we have that

$$|S(x_1, x_2, x_3)| = min_i |x_i|.$$

But this minimum is the 3rd singular value of x, namely x_3 when the diagonal values are listed as per the SVD. The same logic works for general d and yields $S(x_1, \dots, x_d) = min_i |x_i|$, which is by definition the dth singular value of q. This proves proposition 6.1.

Proof of Theorem 6.1. *The case $d = 2$.* We begin with the case $d = 2$ for simplicity and intuition. The configuration space is $M(2,2)$. The degeneration locus $\Sigma = \{det(q) = 0\}$ is a quadratic cone of signature $(2,2)$ in the vector space $M(2,2)$. The group $O(2) \times O(2)$ acts isometrically on the matrix space and the diagonal matrices D form a global slice as described above. Write $q = diag(x,y) \in D$. Then $D \cap \Sigma$ forms the "cross" $xy = 0$. Within the plane the distance function is $S(x,y) = sign(xy) min(|x|, |y|)$. *See figure 6.* The non-smooth locus of S is the line $x = y$ and $x = -y$ corresponding to the matrices xI and xJ where I is the identity and $J = diag(1, -1)$ It now follows from symmetry that $Sing(S)$ is the union of two two-dimensional conical varieties intersecting at the origin, namely $\mathbb{R}SO(2)I$ and $\mathbb{R}SO(2)J$. Taken together this set is simply $\mathbb{R}O(2)$, since $J \in O(2)$ and $det(J) = -1$. If the masses are all equal then this singular locus corresponds to the Lagrange points (equilateral triangles) with one cone corresponding to the positively oriented Lagrange configurations (the north pole of the shape sphere) and the other cone to the negatively oriented Lagrange configurations.

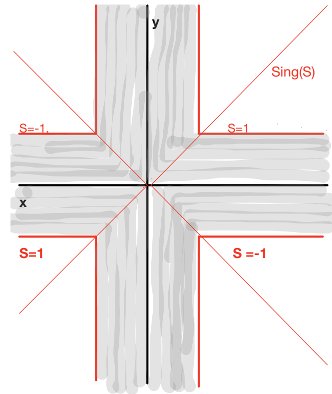

FIGURE 2. Equidistant curves to a cross $xy = 0$ have corners at which the distance function $|S|$ fails to be smooth. This picture models the contours of S restricted to the diagonal slice D for $d = 2$. The thin red diagonal lines indicate $Sing(S) \cap D$.

The case $d = 3$. The diagonals are still a slice for the $SO(3) \times SO(3)$ action, and S is invariant under this action. It follows that we can understand the singularity set of S by looking at its behaviour on the diagonal matrices $diag(x_1, x_2, x_3)$. First, suppose we are at a point where all $x_i > 0$, and such that $x_1 > x_2 > x_3 > 0$. Then, $S = x_3$ in a neighborhood of our point, which is clearly smooth. As we move from this point towards Σ along a geodesic orthogonal to Σ, the value of $x_3 = S$ steadily decreases until we hit $x_3 = 0$ at which point S continues to decrease, but smoothly. The equality $S = x_3$ continues into the region $x_3 < 0$ as long as $x_1 > x_2 > |x_3|$. This phenomenon is invariant under permutations of the coordinate indices. Indeed, restricted to D, we have that $S(x_1, x_2, x_3) = x_i$ where $|S(x_1, x_2, x_3)| = |x_i| := min_k |x_k|$. Thus the singular locus of S *restricted to* D lies on the locus where $|x_i| = |x_j|$ for some $i \neq j$. This locus is the union of 6 planes in D, so has dimension 2, or codimension 1, *within* D. (The singular locus of the restriction of S to D is a bit smaller that the union of these planes, since we do not need that all three principal values are distinct, but only the bottom, two, i.e we only need $x_2 \neq x_3$ if $x_1 \geq x_2 \geq x_3 \geq 0$ are the singular values.)

At first glance, one guesses that since the singular set has codimension 1 within D, then it has overall codimension 1 within $M(3, 3)$. *This logic is wrong.* Points on the singular set of S are *not* generic with respect to the $SO(3) \times SO(3)$ action: their symmetry type jumps. Orbits though points of $Sing(S)$ have dimension 5 or less, not 6 like the dimension of a generic point. (That the orbit through a generic point of D is 6-dimensional is item (b) of 'slice' above.) Since S is invariant under our group $SO(3) \times SO(3)$, so is its singular set, $Sing(S)$. Thus the singular set is the union of the orbits through the singular points of the restriction of S to D. $Sing(S) \cap D$ has dimension 2. If the orbit through any point of $Sing(S) \cap D$ has dimension 5 or less then the singular set itself has dimension at most $7 = 2 + 5$. Our space $M(3, 3)$ has dimension 9, which yields the claimed codimension of 2.

It remains to establish that the orbits through points $x \in Sing(S) \cap D$ have dimension 5 or less. The dimension of an orbit of Lie group action is the dimension of the group minus the dimension of the isotropy subgroup of that point. Our group has dimension 6. We show that the isotropy group at such a point x has dimension at least one. Write $x = diag(x_1, \lambda, \lambda)$ for such a singular point. Let $g(t)$ be the rotation about the 1st axis by t radians, and $g(-t)$ its inverse. Clearly $g(t)xg(-t) = x$, establishing that the isotropy group is *at least* one-dimensional, and hence the orbit has dimension $5 = 6 - 1$ *or less.* (A linear algebra computation, following equation (17) , shows that this dimension is exactly 5 as long as $x_1 \neq \lambda$, but is unneccessary here since all we need is that the codimension of $Sing(S)$ is at least 2.) In case $x = diag(x_1, \lambda, -\lambda)$ with $S(x_1, \lambda, -\lambda) = \lambda$ so that $|x_1| \geq |\lambda| \geq 0$, use left multiplication by the matrix $g_1 = diag(-1, 1, -1) \in SO(3)$ to replace this x by $x = (-x_1, \lambda, \lambda)$ which lies on the same orbit as the original x but now has the form of the computation just made. Since the orbit is homogeneous its dimension does not depend on where on the orbit we choose to compute dimension, and we arrive again at the fact that its dimension is 5 or less.

The case $d > 3$. The proof is nearly identical to the case $d = 3$. $Sing(S) \cap D$ has codimension 1, being contained in the union of the hyperplanes where $x_i = \pm x_j$. At a generic point of D, which is to say, off of these hyperplanes, the $SO(d) \times SO(d)$ action is "almost free" : the orbit's dimenison equals that of $SO(d) \times SO(d)$, as per item (b) of being a slice above. At a typical point on one of these hyperplanes the isotropy algebra is again one-dimensional , consisting of rotations of the double eigenvalue plane. (An "atypical" singular point would be one for which the three smallest singular values are all equal and here the the isotropy algebra has dimension at least 3.) Hence the *codimension* of $Sing(S)$ is $1 + 1$: 1 for the codimension within D and 1 for the extra continuous symmetry dimension (isotropy) associated to each such double "eigenvalue" diagonal matrix. (Sign discrepancies such as $x_i = -x_j \neq 0$ are at first bothersome, but the trick we used in the previous paragraph of multiplying by an element of $SO(d)$ with ± 1's to change the entries to $x_i = x_j$ works as before.)

QED

7. Dynamical Vistas and Open Questions

Planar precursor.

The planar case of theorem 1.2 or 1.3 asserts that any bounded solution to the planar three-body problem defined on the whole time line will suffer infinitely colinear instants. Colinear instances are also called "syzygies". Non-collision syzygies come in three flavors, 1, 2, and 3, depending on the mass in the middle. See figure 3. We can thus associate a syzygy sequence to such a solution. *What syzygy sequences are realized?* This question, still largely open, has motivated much work. See for example [11], and also the closely related work in which braids (equivalent to "stutter-reduced" syzygy sequences) rather than syzygy sequences are used for the symbolic encoding [17], [20], [8] and references therein.

Theorem 1.1 asserts that that any bounded solution to the spatial four-body problem defined on the whole time line will suffer infinitely coplanar instants. The generic coplanar configurations divide into 7 types as shown in figure 4. (We have excluded as "non-generic" configurations for which three of the masses are collinear. Binary collision configurations are thus excluded.) We now have a seven letter alphabet for potential symbol sequences, in analogy with the syzygy sequences of planar three-body dynamics.

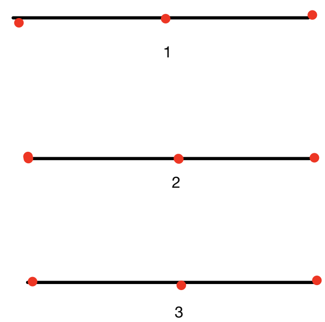

FIGURE 3. The 3 types of generic collinear 3 body shapes.

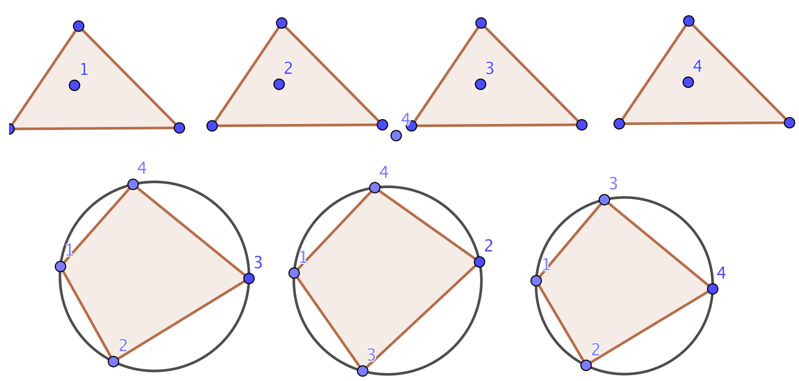

FIGURE 4. The 7 types of generic planar 4 body shapes
.

OPEN QUESTIONS FOR THE FOUR-BODY PROBLEM IN SPACE.

Q1. Are all possible symbol sequences in this 7-letter alphabet realized by a bounded solution having zero angular momentum?

Energy and angular momentum considerations Bounded solutions for the Newtonian N-body problem necessarily have negative energy. (As soon as $N > 2$ there are negative energy solutions which are unbounded.) Hence the following theorem (see [14]) represents a strengthening of 1.3for the case $d = 2$. **Theorem:** every zero angular momentum negative energy solution to the planar three-body problem which does not end and begin in triple collision hits the collinear locus infinitely often.

We do not know a single bounded or negative energy solution of the 4-body problem in space which never suffers co-planarities.

More Questions.

Q2. Do there exist negative energy collision-free solutions of the spatial four body problem which are defined over the whole time line and which never suffer coplanar instants ?

Q3. If the answer to Q2 is 'yes' then are any of these never-coplanar solutions bounded?

Q4. If the answer to Q2 is 'yes' do any of these never-coplanar solutions have zero angular momentum?

In regards to these last two questions, Joseph Gerver has pointed out that there are negative energy collision-free solutions which have no coplanar instants and are defined over a time ray $0 \leq t < \infty$. These solutions have nonzero angular momentum. Take the rotating Lagrange equilateral solution for three of the bodies. Now take the 4th body to be moving away from this triple along the line perpindiular to the plane of the rotating triangle. If the three masses are equal then the situation is symmetric and the 4th body will stay on this orthogonal line as it moves out. The energy of the bound triple can be taken to be sufficiently small so that the overall energy is negative while the 4th escaping body escapes hyperbolically to infinity.

Q5 **A reconstruction question.** What is the angle between the planes of successive coplanar configurations for a solution of the zero angular momentum 4 body problem?

The analogous $d = 2$ question asks for the angle between lines of succsessive collinear configurations in a solution to the zero angular momentum planar three-body problem. This question has a nice answer (see for example [4]): the angle is the spherical area in shape space enclosed by projection of that solution curve to the sphere and the requisite part of the collinear equator needed to close up this curve.

Returning to the $d = 3$ case, the condition of zero angular momentum defines a connection on the principal $SO(3)$ bundle $M_{gen}(3,3) \to Sh_{gen}(3,4)$. (See Proposition 3.1). The overall rotation relating one plane to the other can be computed in principal as the holonomy of this connection. Since the trace of a rotation matrix is $1 + 2\cos(\theta)$ where θ our reconstruction question is one of finding a useable expression for trace of the holonomy of this connection.

The answer in the planar case is tractable because the group is Abelian. Despite the nonAbelian nature of the connection in the $d = 3$ case it is conceivable that this question regarding the trace of the nonAbelian holonomy has a similarly nice answer.

Appendix A. Topologies of Unoriented and Oriented Shape Space

We describe the topology of the unoriented and oriented shape space in our case of $N = d + 1$ bodies in $\mathbb{R}^d$ and the relation between these spaces. We sketch the proof that the unoriented shape space is homeomorphic to the positive semi-definite cone - the closed convex cone of symmetric positive semi-definite matrices - within the vector space $Sym(d) \cong \mathbb{R}^{\binom{d}{2}}$ of real symmetric d by d matrices while oriented shape space is homeomorphic to this vector space itself. We also discuss the smoothness of the shape space in relation to the configuration space.

The configuration space for the N body problem in $\mathbb{R}^d$ is $Hom(\mathbb{R}^N, \mathbb{R}^d) = M(d, N)$. Let $E(d)$ and $SE(d)$ denote the group of isometries of $\mathbb{R}^d$ and its subgroup, the group of orientations preserving isometries. The unoriented shape space is the quotient $M(N, d)/E(d)$ while the oriented shape space is the quotient $M(N, d)/SE(d)$.

We form the quotient spaces in stages, first by translations, then by rotations. As described in section 3.3 and subsection 3.1.1, the quotient of $M(d, N)$ by the translations is isometric to the space $M(d, N-1)$. The remaining rotations acts on this space by left multiplication. In case $N = d+1$ we arrive at the space of real square d-by-d matrices. with the action of the orthogonal group of space being left multiplication by elements of $O(d)$. Thus unoriented shape space is $M(d, d)/O(d)$ while oriented shape space is $Sh(d, d+1) = M(d, d)/SO(d)$.

Unoriented shape space. The map

$$\pi_{un} : M(d, d) \to Sym(d); q \mapsto \pi_{un}(q) := q^t q$$

realizes the $O(d)$ quotient: given any pair $q', q \in M(d, d)$ there exists a $g \in O(d)$ such that $q' = gq$ if and only if $(q')^t q' = q^t q$. This fact follows from a basic theorem from representation theory. The matrix $q^t q$ is sometimes called the 'Gram matrix', being a matrix of inner products of the position vectors q_a. Any matrix of form $q^t q$ is positive semi-definite, and any positive semi-definite matrix s can be expresses as $s = q^t q$ for some $q \in M(d, d)$. (Take $q = \sqrt{s} \in Sym(d)$ for example.) *Thus we identify the unoriented shape space with the image of* π*, which is the positive semi-definite cone* $Sym_+(d) \subset Sym(d)$ *referred to above.* The boundary of this cone consists of those non-negative symmetric matrices whose rank is not full and so corresponds to our degeneration locus Σ.

The map between shape spaces. When a topological group G acts properly on space X and a proper subgroup $H \subset G$ is selected, we obtain a projection $X/H \to X/G$ between quotient spaces whose typical fiber will be G/H if the G action is generically free, meaning that its principal orbit type has isotropy the identity. So, in our case we get the map labelled $\pi_{\mathbb{Z}_2}$ which completes the following diagram and whose typical fiber is $O(d)/SO(d) \cong \mathbb{Z}_2$, the two-element group:

$$\begin{tikzcd} & M(d,d) \arrow[dl, "\pi"'] \arrow[dr, "\pi_{un}"] & \\ Sh(d,d+1) \arrow[rr, "\pi_{\mathbb{Z}_2}"] & & Sym_+(d) \end{tikzcd}$$

The map $\pi_{\mathbb{Z}_2}$ between shape spaces is to be viewed as the map of forgetting orientations.

The atypical fibers are the projections to $Sh(d, d+1)$ of those points of $M(d, d)$ for which the $O(d)$ action is not free , i.e the isotropy is larger than the identity.

These are precisely the elements of Σ. It follows that our projection is a 2:1 cover branched along Σ. Indeed, an unoriented nondegenerate simplex shape has precisely two oriented representative shapes, one with positive volume, the other with negative volume, while degenerate shapes in oriented shape space have precisely one representative in the unoriented shape space.

Gluing cones. Take two identical copies of a closed convex cone with nonempty interior in any finite dimensional real vector space. Glue one copy to the other along the boundary, using the identity map of the boundary as gluing map. Because any such cone is homeomorphic to the closed half space (using a homogeneous degree zero map), one sees that the result is homeomorphic to the original vector space: we have 'blown up' or desingularize the boundary of the cone, turning the cone into the vector space. These considerations yield the theorem that oriented shape space is indeed *homeomorphic* to the Euclidean space $Sym(d)$.

Smoothness and Codimension considerations.

In subsection 3.1 we recalled that if a compact group G acts freely on a manifold Q then the quotient space Q/G inherits the structure of a smooth manifold from Q in such a way that the quotient map $Q \to Q/G$ is a smooth submersion. Moreover, this map gives Q the structure of a smooth principal G-bundle over the quotient. Our group $G = SO(d)$ does not act freely on $Q = M(d,d)$. However, $SO(d)$ does act freely upon restriction to the open dense subset of matrices M_{gen} which is the direct generalization of the set given the same symbol in Proposition 3.1 above.

Definition A.1. *Let M_{gen} denote the subset of $M(d,d)$ consisting of matrices which are either of full rank or rank $d-1$.*

Then $SO(d)$ acts freely on M_{gen}. It follows that the open dense subset $Sh_{gen} \subset Sh(d,d+1)$ of generic shapes within $Sh(d,d+1)$ inherits the structure of a smooth manifold for which the restricted quotient projection M_{gen} onto Sh_{gen} is a smooth submersion.

Proposition A.1. *The complement of M_{gen} is an algebraic subvariety of codimension 4.*

Case $d=3$. Think of $M(3,3)$ as $\mathbb{R}^3 \otimes \mathbb{R}^{3*}$. Then we can write any rank 1 matrix in the form $m = v \otimes \alpha$ with $v \in \mathbb{R}^3$ a non-vector and $\alpha \in \mathbb{R}^{3*}$ a nonzero covector. Thus we have a map $(\mathbb{R}^3 \setminus 0) \times (\mathbb{R}^{3*} \setminus 0)$ onto the rank 1 matrices. Since $(\lambda v) \otimes (\frac{1}{\lambda}\alpha) = v \otimes \alpha$ for all $\lambda \neq 0$ the fiber of the map is one-dimensional, showing that the image, the rank 1 matrices has dimension $6-1=5$. Finally, $4 = 9-5$.

General d. Write $M_r \subset M(d,d)$ for the space of matrices of rank r. On pages 28 to 30 of the book [3] one finds a proof that M_r is a smooth submanifold of $M(d,d)$ whose codimension is $(d-r)^2$, that is, the square of the corank. It is easy to see that the closure of M_r consists of the union $\cup_{j \leq r} M_j$. Now $M = \cap_r M_r$ while $M_{gen} = M_d \cap M_{d-1}$, so that $M(d,d) \setminus M_{gen} = \cap_{r<d-1} M_r$ is the union of the matrices of rank *less* than $r-1$, and as such is a singular subvariety, the union of $r-2$ smooth submanifolds, the largest having dimension $2^2 = 4$ and its closure containing all the lower strata.

For completeness, we sketch the argument from this book that M_r is smooth and has the stated codimension. The orbit (but not the closure of the orbit!) of the action of a smooth Lie group acting on a smooth manifold is a smooth immersed submanifold. $GL(d) \times GL(d)$ acts on on $M(d,d)$ by $(g_1, g_2) \cdot m = g_1 m g_2^{-1}$. Indeed, this action corresponds to changing bases in both the domain and the range of

$m : \mathbb{R}^d \to \mathbb{R}^d$. Given $m \in M_r$, by a standard argument from linear algebra we can choose these bases in such a way as to put m into the normal form

$$m = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where I_r is the r-by-r identity matrix, which shows that M_r is a single orbit, and thus is a smoothly immersed as a submanifold. To get the codimension perturb m in its normal form:

$$m(\epsilon) = \begin{pmatrix} I_r + \epsilon a & \epsilon b \\ \epsilon c & \epsilon d \end{pmatrix}$$

and ask that the perturbation remain of rank r. Here ϵ is small, and a, b, c, d are matrices of the appropriate size to fit in the blocks. For ϵ small enough, and a, b, c, d fixed, $m(\epsilon)$ always has rank *at least* r. Thus $m(\epsilon)$ is of rank r for small ϵ if and only if all of its $(r+1) \times (r+1)$ minors vanish. Recall that a minor is constructed by choosing $r+1$ row indices and $r+1$ column indices to extract a square $(r+1) \times (r+1)$ submatrix from $m(\epsilon)$. If two or more of column indices or the row indices lie outside the $r \times r$ upper block then the resulting minor vanishes to order at least ϵ^2. Thus the only minors vanishing to order ϵ come from selecting a row index , say i, and a single column index j, outside the block, which is to say, with $i, j > r$, in which case the resulting minor is $\epsilon d_{ij} + O(\epsilon^2)$. Consequently, at the linearized level, the equations asserting that our perturbation $m(\epsilon)$ has rank r is the matrix equation $d = 0$ which consists of $(n-r)^2$ scalar equations.

Finally, to see that $M \setminus M_{gen}$ is algebraic observe that it can be defined by the vanishing of all minors of $m \in M(d,d)$ size $(d-2) \times (d-2)$.

Semi-algebraic structure of quotient. We conclude this appendix by remarking that the shape space $Sh = Sh(d, d+1)$ can be identified with a semi-algebraic subvariety of some $\mathbb{R}^N$. We have $Sh(d, d+1) = M(d,d)/SO(d) = \mathbb{V}/G$ is the quotient of a real finite dimensional vector space $\mathbb{V}$ by the orthogonal action of a smooth Lie group G. The G-invariant polynomials P^G separate G-orbits of $\mathbb{V}$. By Hilbert's basis theorem P^G is finitely generated. Choose a set of N generators yields a polyomial map $\mathbb{V} = M(d,d) \to \mathbb{R}^N$ which realizes the quotient. Finally, by the Tarski-Seidenberg theorem the image of this map is a semi-algebraic set. This image is Sh.

It is not clear how useful this remark is. We make it with the hope that if someone needs to 'do calculus" at singular points of Sh it might be of use.

References

[1] R. Abraham and J.E. Marsden, **Foundations of Mechanics**, [1978], 2nd ed., Addison-Wesley, Reading, MA.

[2] A. Albouy and A. Chenciner, *Le probléme des n corps et les distances mutuelles*, Inventiones **131** (1998), 151-184.

[3] V.I. Arnol'd, S.M. Gusein-Zade, and A.N. Varchenko, **Singularities of Differentiable Maps, volume I**, [1985], Birkhauser, (see esp. p.28-30).

[4] A. Chenciner and R. Montgomery, *A remarkable periodic solution of the three-body problem in the case of equal masses*, with Alain Chenciner, Annals of Mathematics, **152**, 881-901 2000.

[5] M. Gromov, **Sign and Meaning of Curvature**, Rendiconti del Seminario Matematico e Fisico di Milano **61**, issue 1, (1991), 9-123., (see esp. p. 34-37).

[6] D. G. Kendall, D. Barden, T.K. Carne, and H. Le, **Shape and shape theory**, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, [1999].

[7] H. Le, *On Geodesics in Euclidean Shape Spaces*, J. London Math. Soc., (2), **44**, (1991) 360-372.
[8] X. Li and S. Liao, *Collisionless periodic orbits in the free-fall three-body problem*, arxiv:1805.07980.
[9] R. Littlejohn and M. Reinsch, *Internal or Shape coordinates in the n-body problem*, Phys. Rev. A , **52**, no. 3, (1995),2035-2051.
[10] R. Moeckel, *Lecture Notes on Central Configurations*, `http://www-users.math.umn.edu/~rmoeckel/notes/Notes.html`.
[11] R. Moeckel and R. Montgomery. *Realizing All Reduced Syzygy Sequences in the Planar Three-Body Problem*, Nonlinearity, **28** (2015), 1919-1935.
[12] R. Montgomery, *The isoholonomic problem and some of its applications*, Comm. Math. Phys., **128**, (1990), 565-592.
[13] R. Montgomery, *Infinitely Many Syzygies Archives for Rational Mechanics*, **164**, no. 4, (2002), 311–340. (DOI) 10.1007/s00205-002-0211-z.
[14] R. Montgomery, *The zero angular momentum three-body problem: all but one solution has syzygies*, Erg. Th. and Dyn. Systems, **27**, no. 6, (2007), 1933-1946.
[15] R. Montgomery, *The Three-Body Problem and the Shape Sphere*, Amer. Math. Monthly, **122**, no. 4,(2015), 299-321.
[16] R. Montgomery, **A tour of sub-Riemannian Geometry**, Mathematical Surveys and Monographs, **91**, American Math. Society, Providence, Rhode Island, [2002]. (see esp. p. 204).
[17] C. Moore, *Braids in Classical Gravity*, Phys. Rev. Lett., **70**, (1993), 3675-3679.
[18] B. O'Neil, *The fundamental equations of a submersion*, Michigan Math. J. , **13**, no. 4, (1966) 459–469, (see esp. Cor. 1, eq (3), p. 466). doi:10.1307/mmj/1028999604.
[19] G. Strang, **Introduction to Linear Algebra**, (2nd. ed.), [1976], Wellesley-Cambridge Press, ISBN 0-9614088-5-5, (esp. section 3.4, 141-142).
[20] M. Suvakov and V. Dmitrasinovic, *Three Classes of Newtonian Three-Body Planar Periodic Orbits*, Phys. Rev. Lett., **110**, (2013),114301-114305. arxiv:1303.0181.
[21] A. Wintner, **The Analytical Foundations of Celestial Mechanics**, Princeton Univ. Press, 1941.

Mathematics Department, University of California, Santa Cruz, Santa Cruz CA 95064
E-mail address: `rmont@ucsc.edu`