# Analysis and Computation of Two Body Impact in Three Dimensions

**Yan-Bin Jia**
Department of Computer Science
Iowa State University
Ames, IA 50011, USA
Email: jia@iastate.edu

**Feifei Wang**
Department of Computer Science
Iowa State University
Ames, IA 50011, USA
Email: wangff@iastate.edu

*A formal impulse-based analysis is presented for the collision of two rigid bodies at single contact point under Coulomb friction in three dimensions (3D). The tangential impulse at the contact is known to be linear in the sliding velocity whose trajectory, parametrized with the normal impulse and referred to as the hodograph, is governed by a generally non-integrable ordinary differential equation (ODE). Evolution of the hodograph is bounded by rays in several invariant directions of sliding in the contact plane. Exact lower and upper bounds are derived for the number of such invariant directions, utilizing the established positive definiteness of the matrix defining the governing ODE. If the hodograph reaches the origin, it either terminates (i.e., the contact sticks) or continues in a new direction (i.e., the contact resumes sliding) whose existence and uniqueness, only assumed in the literature, are proven. Closed-form integration of the ODE becomes possible as soon as the sliding velocity turns zero or takes on an invariant direction. Assuming Stronge's energy-based restitution, a complete algorithm is described to combine fast numerical integration with a case-by-case closed-form analysis. A number of solved collision instances are presented. It remains open whether the modeled impact process will always terminate under Coulomb friction and Stronge's (or Poisson's) restitution hypothesis.*

## 1 Introduction

A collision between two rigid bodies is often modeled as an event happening at a single contact point with an infinitesimal duration. Despite the infinitely large impulsive force, the momentum of each involved body will undergo only a finite amount of change. Newton's second law in the integral form equates this change with the impulse exerted by the other body, which suggests that the impact as a process should be described in terms of the finite impulse rather than the infinitesimal time.

The impact problem cannot be solved under conservation of momentum alone. Three hypotheses were introduced by Newton (1686), Poisson (1827), and Stronge (1990) to respectively quantify the relationships between the pre- and post-impact velocities, between the impulses accumulated during the two impact phases: compression and restitution, and between the energies accumulated/released during these phases. Their measures, taking the forms of ratio and concerning quantities along the contact normal, are referred to as the kinematic, kinetic, and energetic coefficients of restitution, respectively.

During the collision, impulse also exists in the tangential direction at the contact due to friction and tangential compliance. Study has primarily focused on the effect of friction, although that of tangential compliance has begun to receive some attention (Stronge, 2000; Cross, 2010; Jia 2013).

To deal with friction, application of Newton's hypothesis previously relied on some theoretically unjustified ratios of tangential impulse to normal impulse (Brach 1989; Smith 1991). Meanwhile, algebraic approaches based on linear complementarity formulations (Glocker and Pfeiffer 1995; Stewart 2000) or resorting to the use of two restitution parameters (Chatterjee and Ruina 1998) did not capture the progressive nature of the physical process.

Most of the investigations on impact with friction have applied Poisson's hypothesis. In two dimensions (2D), Routh (1905) developed a graphical representation of the impulse accumulation. His approach was extended by Wang and Mason (1992) into a complete solution which considered all five possible sequences of contact and impact mode changes. The existence of a closed-form solution to the 2D problem is primarily attributed to the fact that contact slip can only take

place along either of the two opposite tangential directions at the contact.

In three dimensions (3D), however, slip may happen along any direction in the contact plane. Darboux (1880) and Keller (1986) found that the evolution trajectory of the sliding velocity, referred to as the hodograph, was governed by a first order differential equation (ODE) which is analytically non-integrable in general. The ODE, derived under Coulomb's law of friction, specifies the derivative of the sliding velocity with respect to normal impulse as a linear function of the velocity direction.

Although only the ending point of the hodograph is needed to determine the tangential impulse accumulation during the slip, generally there is no short cut to locate this point other than numerical integration. Stable (but not efficient) numerical integration methods were investigated by Zhao and Liu (2007), and Zhang and Sharf (2007). Slow integration hinders not just the modeling effort but also, in a broader sense, application of impact to real tasks where modeling may be executed as a subroutine many times by some higher level modeling or planning algorithm.

As this paper will demonstrate, considerable speedup in numerical integration can indeed be achieved by exploiting the geometry of the hodograph as a parametric plane curve. Another improvement will be based on the observation that closed-form integration is possible (and thus numerical integration should terminate) in a couple of situations which happen frequently. In the first situation, the growing hodograph may reach one of several directions, referred to as invariant directions (Bhatt and Koechling 1995; Batlle 1996; Elkaranshawy2007), in which the sliding velocity and its derivative are collinear. The second situation occurs when the sliding velocity turns zero. It is known (Keller 1986; Bhatt and Koechling 1995; Batlle 1996) that starting at this moment either stick will happen or slip will resume in a constant direction. Along this new direction the tangential contact acceleration would coincide with the prediction by the governing differential equation if the velocity assumes the same direction. Whether the contact mode turns into stick or resumed slip, the change in the tangential impulse will assume a linear form.

Most of the aforementioned works on frictional 3D impacts applied Poisson's hypothesis. Although evolution of the hodograph is not affected by the particular impact hypothesis adopted, where the curve will end is. Poisson's hypothesis, along with Newton's, was found to be energetically inconsistent with Coulomb's law of friction (Kane and Levinson 1985; Stronge 2000) if the direction of slip either reverses during a 2D collision or varies during a 3D collision. Thus, neither hypothesis is truly adequate for describing the 3D impact. Stronge's hypothesis, directly related to irreversible deformation at the contact [4, p. 47], is the only one consistent with energy conservation. In this paper, we will use Stronge's hypothesis to model the energy loss.

An efficient solution of the 3D impact problem should begin with fast numerical integration, and switch to closed-form evaluation as soon as sliding either takes on an invariant direction or (momentarily) reaches zero velocity. The impact phase—compression or restitution—is meanwhile tracked or predicted, based on the changing potential energy stored at the contact.

Section 2 will go over the setup of the impact equation, and establish the positive definiteness of the inverse inertia matrix which, when multiplied with the impulse, yields the change in contact velocity. It will also overview the commonly used solution scheme under Strong's hypothesis (easily adaptable to work with Poisson's hypothesis). Section 3 will derive the linear dependence of impulse on the sliding velocity. Their relationship was given in an integral form before (Keller 1986). Section 4 will study the governing ODE of the hodograph, bounding the number of invariant directions. (To keep the flow of reading, the proof of this result will appear in Appendix A.) Section 5 will be devoted to an analysis of the situation when the sliding velocity reaches zero, establishing the existence and uniqueness (proof given in Appendix B) of the centrifugal direction along which sliding will resume if sticking does not happen. Such existence and uniqueness were merely stated and assumed to be true without proof in the past. Section 6 will present a complete algorithm which takes into account all possible orderings of events where the sliding velocity becomes zero, the sliding direction becomes invariant, or compression ends. Numerical integration will be carried out using a step size adapted to the hodograph's differential geometry. Several solved instances will also be presented.

While a real collision always ends, termination of the modeled process, whether under Poisson's or Stronge's hypothesis, has so far been only assumed but not mathematically proven. This could be an ultimate check on the consistency between Coulomb's law and either hypothesis. To elaborate, termination of the modeled impact process hinges on endings of both compression and restitution in all possible impact scenarios. Ending of compression, for instance, depends on the normal contact velocity eventually increasing to zero. This velocity component, however, is generally affected by tangential impulse due to Coulomb friction. The first author has been able to prove termination of all 2D impacts (Jia et al. 2016). For a 3D impact, we will provide an explanation in Section 2.3 why it is conceivably much harder to establish impact termination. There is a good news though. Termination will be guaranteed under a mild condition, which is rarely violated given the positive definiteness of the inverse inertia matrix.

## 2 Impact Equations

Fig. 1 shows two bodies $\mathcal{B}_1$ and $\mathcal{B}_2$ colliding at a point $\mathbf{c}$. For convenience, we place the world frame $\mathcal{W}$ at $\mathbf{c}$ with the $z$-axis in the direction of the inward surface normal of $\mathcal{B}_1$ (and thus the $xy$-plane aligned with the tangent plane). Each $\mathcal{B}_i$ has a body frame $\mathcal{F}_i$, located at its center of mass $\mathbf{o}_i$ and defined by its principal axes, in which the angular inertia matrix $Q_i$ is diagonalized. The rotation of $\mathcal{F}_i$ from $\mathcal{W}$ is described by the matrix $R_i$, while the translation is $-\mathbf{r}_i$, where $\mathbf{r}_i = \mathbf{c} - \mathbf{o}_i$.

Let $\mathbf{V}_i$ be the velocity of the body $\mathcal{B}_i$. Also, let $\boldsymbol{\omega}_i$ be the

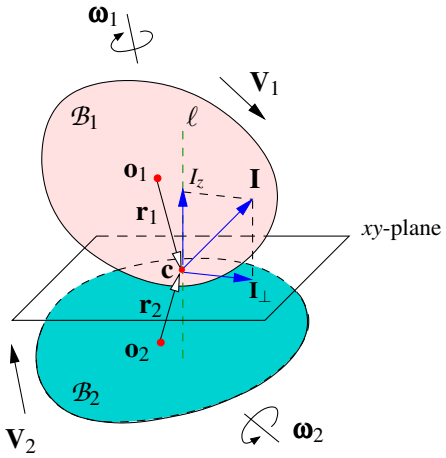



Fig. 1. Impact between two bodies. The impulse $\mathbf{I}$ consists of a component $I_z$ along their contact normal (aligned with the $z$-axis), and a tangential component $\mathbf{I}_\perp$ (in the $xy$ plane) due to contact friction. The line of impact $\ell$ through $\mathbf{c}$ is normal to the two contacting surfaces.

angular velocity of $\mathcal{B}_i$ which, by convention, is in terms of a fixed frame instantaneously coincident with $\mathcal{F}_i$.

The line $\ell$ through $\mathbf{c}$ and normal to the two contacting surfaces (and thus aligned with the $z$-axis) is called the *line of impact*. The impact is *central* if $\ell$ passes through $\mathbf{o}_1$ and $\mathbf{o}_2$, and *eccentric* otherwise [4, pp.2–3]. It is called a *direct* impact if both pre-impact velocities $\mathbf{V}_i^-$, $i = 1,2$, are along $\ell$, and an *oblique* impact otherwise. From now on, the superscripts '$-$' and '$+$' of a quantity will refer to its values just before and after the impact, respectively.

Let $\mathbf{F}$ be the contact force exerted on $\mathcal{B}_1$ by $\mathcal{B}_2$. Since the impact duration $\tau$ is very small (tends to zero, mathematically), $\mathbf{F}$ is significantly larger than the gravitational forces on the two objects. For this reason, we will ignore gravity in impact analysis. Newton's and Euler's equations on dynamics are respectively

$$\begin{aligned} \mathbf{F} &= m_1 \dot{\mathbf{V}}_1, \\ R_1^{-1}(\mathbf{r}_1 \times \mathbf{F}) &= Q_1 \dot{\boldsymbol{\omega}}_1 + \boldsymbol{\omega}_1 \times Q_1 \boldsymbol{\omega}_1, \end{aligned}$$

where the dot '.' denotes differentiation with respect to time. Integrate the above equations over the time period $[0,\tau]$, neglecting the integral of the finite term $\boldsymbol{\omega} \times Q_1 \boldsymbol{\omega}$ as $\tau \to 0$:

$$\begin{aligned} \mathbf{I} &= m_1 \Delta \mathbf{V}_1, \\ R_1^{-1}(\mathbf{r}_1 \times \mathbf{I}) &= Q_1 \Delta \boldsymbol{\omega}_1. \end{aligned}$$

Here, $\mathbf{I}$ is the impulse exerted by $\mathcal{B}_2$ onto $\mathcal{B}_1$, and $\Delta\mathbf{V}_1$ and $\Delta\boldsymbol{\omega}_1$ are the changes in $\mathbf{V}_1$ and $\boldsymbol{\omega}_1$, respectively. The impact equations for $\mathcal{B}_2$ can be set up similarly, and solved together with those for $\mathcal{B}_1$ for the velocity changes during the impact:

$$\begin{aligned} \Delta\mathbf{V}_1 &= \frac{1}{m_1}\mathbf{I} \quad \text{and} \quad \Delta\boldsymbol{\omega}_1 = Q_1^{-1}R_1^{-1}(\mathbf{r}_1 \times \mathbf{I}), \\ \Delta\mathbf{V}_2 &= -\frac{1}{m_2}\mathbf{I} \quad \text{and} \quad \Delta\boldsymbol{\omega}_2 = -Q_2^{-1}R_2^{-1}(\mathbf{r}_2 \times \mathbf{I}). \end{aligned} \tag{1}$$

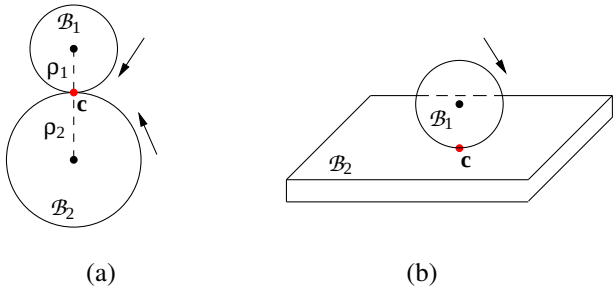



Fig. 2. Impacts (a) between two balls and (b) between a ball and a half-space with infinite mass.

The contact velocity (of $\mathcal{B}_1$ relative to $\mathcal{B}_2$) at $\mathbf{c}$ is

$$\mathbf{v} = (v_x, v_y, v_z)^T = \mathbf{V}_1 + (R_1\boldsymbol{\omega}_1) \times \mathbf{r}_1 - \mathbf{V}_2 - (R_2\boldsymbol{\omega}_2) \times \mathbf{r}_2. \tag{2}$$

Denote by $P_i$ the antisymmetric matrix such that $P_i\mathbf{u} = \mathbf{r}_i \times \mathbf{u}$ for all vectors $\mathbf{u} \in \mathbb{R}^3$, and by $U_3$ the $3 \times 3$ identity matrix $\mathrm{diag}(1,1,1)$. Let

$$S = P_1R_1Q_1^{-1}R_1^{-1}P_1 + P_2R_2Q_2^{-1}R_2^{-1}P_2. \tag{3}$$

During the impact, from (2) we see that $\mathbf{v}$ changes from its pre-impact value $\mathbf{v}^-$ by the following amount:

$$\begin{aligned} \Delta\mathbf{v} &= \Delta\mathbf{V}_1 + (R_1\Delta\boldsymbol{\omega}_1) \times \mathbf{r}_1 - \Delta\mathbf{V}_2 - (R_2\Delta\boldsymbol{\omega}_2) \times \mathbf{r}_2 \\ &= W\mathbf{I}, \end{aligned} \tag{4}$$

after substitutions of (1), where

$$W = \left(\frac{1}{m_1} + \frac{1}{m_2}\right) U_3 - S. \tag{5}$$

Fig. 2(a) shows a ball ($\mathcal{B}_1$) of radius $\rho_1$ impacting another ball ($\mathcal{B}_2$) of radius $\rho_2$. Both balls have uniform mass density. Let the body frames of $\mathcal{F}_1$ and $\mathcal{F}_2$ have the same orientations as that of the world frame $\mathcal{W}$. Namely, $R_1 = R_2 = U_3$. Also, $\mathbf{r}_1 = (0,0,-\rho_1)^T$ and $\mathbf{r}_2 = (0,0,\rho_2)^T$, inducing

$$P_1 = \begin{pmatrix} 0 & \rho_1 & 0 \\ -\rho_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0 & -\rho_2 & 0 \\ \rho_2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Each ball has the angular inertia matrix $Q_i = \frac{2}{5}m_i\rho_i^2U_3$. It is then straightforward to evaluate (3) and (5):

$$S = \sum_{i=1}^{2} \frac{5}{2m_i\rho_i^2} P_iP_i = -\frac{5}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$W = \left(\frac{1}{m_1} + \frac{1}{m_2}\right) \begin{pmatrix} \frac{7}{2} & 0 & 0 \\ 0 & \frac{7}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Suppose we keep $\mathcal{B}_1$ as a ball but consider $\mathcal{B}_2$ to be an immobilized half-space (with infinite mass and thus infinite moments of inertia about its principal axes). The impact is shown in Fig. 2(b) with $W = \frac{1}{2m_1}\text{diag}(7,7,2)$.

### 2.1 Inverse Inertia Matrix

The matrix $W$ defined in (5) is called the *inverse inertia matrix* [4, p. 66], since its product with the impulse generates the change in the velocity, as described by (4).

**Lemma 1.** *The matrix $S$ defined in (3) is symmetric. It is negative definite if $\mathbf{o}_1$, $\mathbf{o}_2$, and $\mathbf{c}$ are not collinear, and negative semidefinite otherwise.*

*Proof.* Since $P_i^T = -P_i$, $(Q_i^{-1})^T = Q_i^{-1}$, and $R_i^{-1} = R_i^T$, the symmetry of $S$ follows from that

$$\begin{aligned} S^T &= \sum_{i=1}^{2} \left(P_i R_i Q_i^{-1} R_i^{-1} P_i\right)^T = \sum_{i=1}^{2} (-P_i) R_i Q_i^{-1} R_i^T (-P_i) \\ &= \sum_{i=1}^{2} P_i R_i Q_i^{-1} R_i^T P_i \;=\; S. \end{aligned}$$

Suppose $\mathbf{u} \neq \mathbf{0}$. We have, for $i = 1,2$,

$$\begin{aligned} \mathbf{u}^T \left(P_i R_i Q_i^{-1} R_i^{-1} P_i\right) \mathbf{u} &= (\mathbf{u}^T P_i R_i) Q_i^{-1} (R_i^T P_i \mathbf{u}) \\ &= -(R_i^T P_i \mathbf{u})^T Q_i^{-1} (R_i^T P_i \mathbf{u}) \\ &\leq 0, \end{aligned} \tag{6}$$

because $Q_i^{-1}$ is positive definite. Hence $\mathbf{u}^T S \mathbf{u} \leq 0$.

In (6), the product is zero if and only if $R_i^T P_i \mathbf{u} = R_i^{-1}(\mathbf{r}_i \times \mathbf{u}) = \mathbf{0}$, which is equivalent to $\mathbf{r}_i \times \mathbf{u} = \mathbf{0}$ in the world frame. Therefore, by (3), $\mathbf{u}^T S \mathbf{u} = 0$ if and only if $\mathbf{r}_1 \times \mathbf{u} = \mathbf{r}_2 \times \mathbf{u} = \mathbf{0}$, or equivalently, if and only if $\mathbf{u}$ is collinear with the vectors from the origin to $\mathbf{o}_1$ and $\mathbf{o}_2$. Trivial reasoning from here establishes the second statement of the theorem.

By Lemma 1, the matrix $-S$ is positive semidefinite. From (5), $W$ is the sum of $-S$ with the positive definite matrix $\left(\frac{1}{m_1} + \frac{1}{m_2}\right) U_3$.

**Proposition 2.** *The inverse inertia matrix $W$ is symmetric and positive definite.*

### 2.2 Impact Solution Scheme

The contact velocity $\mathbf{v}$, the contact force $\mathbf{F}$, and the impulse $\mathbf{I}$ each has a tangential component (in the $x$-$y$ plane) and a normal component (along the $z$-axis):

$$\mathbf{v} = \mathbf{v}_\perp + v_z \hat{\mathbf{z}} = v_x \hat{\mathbf{x}} + v_y \hat{\mathbf{y}} + v_z \hat{\mathbf{z}}, \tag{7}$$

$$\mathbf{F} = \mathbf{F}_\perp + F_z \hat{\mathbf{z}} = F_x \hat{\mathbf{x}} + F_y \hat{\mathbf{y}} + F_z \hat{\mathbf{z}}, \tag{8}$$

$$\mathbf{I} = \mathbf{I}_\perp + I_z \hat{\mathbf{z}} = I_x \hat{\mathbf{x}} + I_y \hat{\mathbf{y}} + I_z \hat{\mathbf{z}}, \tag{9}$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are the unit vectors in the directions of $x$-, $y$-, and $z$-axes of the world frame $\mathcal{W}$. Here, the subscript '$\perp$' of a vector refers to its projection onto the $xy$-plane. The tangential contact force $\mathbf{F}_\perp$ is due to Coulomb friction, with $\mu$ being the coefficient of friction.

Impact consists of two phases: compression and restitution. During compression, kinetic energy is converted to potential energy $E$ at the contact as the normal contact velocity $v_z$ increases to zero from its pre-impact value $v_z^- < 0$. This velocity component is, via a substitution of (4),

$$v_z = \hat{\mathbf{z}}^T (\mathbf{v}^- + \Delta\mathbf{v}) = v_z^- + \hat{\mathbf{z}}^T W \mathbf{I}. \tag{10}$$

During restitution, the elastic portion of the stored potential energy is converted back to kinetic energy. This portion is $e^2$, where $e$ is called the *energetic coefficient of restitution.* The restitution phase ends with $E = 0$.

Since the normal contact force $F_z$ is greater than zero during the impact, its integral, the normal impulse $I_z$, increases monotonically. It replaces time as the variable. From now on, we will let '$'$' denote differentiation with respect to $I_z$. Assuming linear normal stiffness, it is easy to show that $E' = -v_z$ [6], and hence, by (4),

$$\begin{aligned} E' &= -v_z^- - \hat{\mathbf{z}}^T W \mathbf{I} \\ &= -v_z^- - \left(\frac{1}{m_1} + \frac{1}{m_2}\right) I_z + \hat{\mathbf{z}}^T S \mathbf{I}. \end{aligned} \tag{11}$$

For the convenience of description, we introduce the following notation:

$$\begin{aligned} I_{zc}&: \text{ value of } I_z \text{ at which compression ends } (v_z = 0); \\ I_{zr}&: \text{ (positive) value of } I_z \text{ at which restitution ends } (E = 0); \\ \mathbf{I}_c&: \text{ value of } \mathbf{I} \text{ when } I_z = I_{zc}; \\ \mathbf{I}_r&: \text{ value of } \mathbf{I} \text{ when } I_z = I_{zr}, \text{ i.e., the total impulse;} \\ E_c&: \text{ value of } E \text{ when } I_z = I_{zc}, \text{ i.e., maximum value of } E. \end{aligned} \tag{12}$$

The impact problem is solved via increasing $I_z$ from zero, tracking the contact mode (stick or slip) to obtain the tangential impulse $\mathbf{I}_\perp$ as a function of $I_z$, as well as tracking the impact phase. The following steps are carried out:

1. Solve $E' = -v_z = 0$ for $I_{zc}$ with (11) and $\mathbf{I}(I_z)$ plugged in. At $E' = 0$, the impulse $\mathbf{I}$ lies in the *plane of compression* $\mathcal{P}_c$ defined by $\hat{\mathbf{z}}^T W \mathbf{I} + v_z^- = 0$ in the 3D impulse space.
2. Obtain $E_c$ at the end of compression via integrating (11) over $[0, I_{zc}]$ with the initial value 0.
3. After compression, the energy dissipates by a factor of $1 - e^2$. Obtain the energy form during restitution from integrating (11) over $[I_{zc}, I_z]$ with the initial value $e^2 E_c$.
4. Solve $E = 0$ to obtain $I_{zr}$ and the total impulse $\mathbf{I}_r$.
5. For $i = 1,2$, evaluate $\Delta\mathbf{V}_i$ and $\Delta\boldsymbol{\omega}_i$ using (1) with $\mathbf{I} = \mathbf{I}_r$. The post-impact velocities are $\mathbf{V}_i^+ = \mathbf{V}_i^- + \Delta\mathbf{V}_i$ and $\boldsymbol{\omega}_i^+ = \boldsymbol{\omega}_i^- + \Delta\boldsymbol{\omega}_i$, $i = 1,2$.

The main effort is to keep track of the values of $I_z$ at which the contact mode switches or the impact phase changes. The tangential impulse $\mathbf{I}_\perp$ will change its form with every contact mode switch, subsequently affecting the energy form by (11).

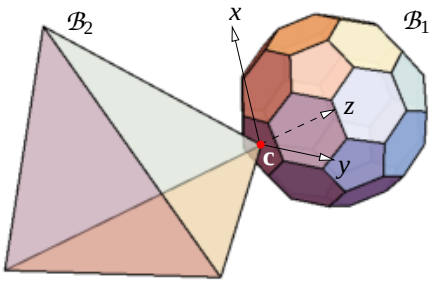



Fig. 3. Configuration of impact ($\mu = 0.8$ and $e = 0.95$) between an icosahedron $\mathcal{B}_1$ and a tetrahedron $\mathcal{B}_2$, both with uniform mass density, respectively. The $x$-$y$ plane coincides with a pentagon face of $\mathcal{B}_1$, whose center coincides with a vertex of $\mathcal{B}_2$ at $\mathbf{c}$. All pentagons and hexagons on $\mathcal{B}_1$ have side length 0.1. The vector from $\mathbf{c}$ to the vertex of the contacting pentagon with the largest $y$-coordinate rotates from the $y$-axis about the $z$-axis (through $\mathcal{B}_1$'s center of geometry) through $\pi/10$. The remaining three vertices of the tetrahedron are located at $(0,0,-0.75)^T$, $(0.5,0,-0.5)^T$, and $(0,0.5,-0.5)^T$. Other geometric and physical parameters include: $\mathbf{o}_1 = (0,0,0.232744)^T$, $\mathbf{o}_2 = (0.125,0.125,-0.4375)^T$, $m_1 = 3.0$, $m_2 = 1.0$, $Q_1 = 0.0671673\,U_3$, and $Q_2 = \mathrm{diag}(0.017239, 0.022813, 0.027135)$.

Fig. 3 shows the impact between an icosahedron and a tetrahedron. The shapes, selected for their appeal, do not increase technical complexity.[1] The inverse inertia matrix is computed as

$$W = \begin{pmatrix} 11.5984 & -0.910367 & 2.44236 \\ -0.910367 & 9.90134 & 1.95747 \\ 2.44235 & 1.95746 & 2.59042 \end{pmatrix}. \tag{13}$$

The pre-impact velocities are

$$\begin{aligned} \mathbf{V}_1^- &= (-0.1, 0.5, 0.95)^T, & \boldsymbol{\omega}_1^- &= (0.1, 0.1, 0.1)^T, \\ \mathbf{V}_2^- &= (0.1, 0.1, 1)^T, & \boldsymbol{\omega}_2^- &= (0.1, -0.1, 0.1)^T. \end{aligned} \tag{14}$$

The total impulse, computed by Algorithm 1 to be described in Section 6, is

$$\mathbf{I}_r = (-0.00326657, -0.0592263, 0.1007)^T. \tag{15}$$

It yields the following post-impact velocities:

$$\begin{aligned} \mathbf{V}_1^+ &= (-0.101089, 0.480258, 0.983567)^T, \\ \boldsymbol{\omega}_1^+ &= (0.138764, -0.100369, 0.0755903)^T, \\ \mathbf{V}_2^+ &= (0.103267, 0.159226, 0.8993)^T, \\ \boldsymbol{\omega}_2^+ &= (-0.281687, 0.402075, 0.589565)^T. \end{aligned}$$

[1] The inverse inertia matrix $W$ in (5) governing impact dynamics (4) depends on the angular inertia matrices of the two bodies, and the relative positions of the contact point to the two centers of mass. It does not depend on specific geometry of the two involved bodies.

### 2.3 Sufficient Condition for Impact Solution

Impact happens because $v_z^- < 0$. The solution scheme described in Section 2.2 works only if $v_z$ will eventually increase to zero to end compression, and after that, the energy $E$ will eventually decrease to zero to end restitution. Both events have been assumed to happen in previous works on impact. But, is it possible that one of them will not happen? For instance, what if $v_z$ never increases to zero, or even decreases so the two objects will "penetrate" into each other infinitely?

To understand how $v_z$ varies, we look at its derivative $v_z'$ obtained below from (4) and (9):

$$v_z' = \frac{d}{dI_z}\left(v_z^- + \hat{\mathbf{z}}^T \Delta\mathbf{v}\right) = \hat{\mathbf{z}}^T W \mathbf{I}' = \hat{\mathbf{z}}^T W \hat{\mathbf{z}} + \hat{\mathbf{z}}^T W \mathbf{I}_\perp'. \tag{16}$$

Meanwhile, it follows that

$$\mathbf{I}_\perp' = d\mathbf{I}_\perp / dI_z = \dot{\mathbf{I}}_\perp / \dot{I}_z = \mathbf{F}_\perp / F_z. \tag{17}$$

Thus, under Coulomb's law of friction, $\mathbf{I}_\perp' = -\mu \mathbf{v}_\perp / \|\mathbf{v}_\perp\|$ if $\mathbf{v}_\perp \neq \mathbf{0}$. When $\mathbf{v}_\perp = \mathbf{0}$, it will be shown in Section 5 that $\mathbf{I}_\perp'$ has a constant value. To summarize, $\mathbf{I}_\perp'$ is determined by $\mathbf{v}_\perp$.

But how will $\mathbf{v}_\perp$ change? It follows from $\mathbf{v}_\perp = (U_3 - \hat{\mathbf{z}}\hat{\mathbf{z}}^T)\mathbf{v}$ and (4) that $\mathbf{v}_\perp' = (U_3 - \hat{\mathbf{z}}\hat{\mathbf{z}}^T) W \mathbf{I}'$. Thus, $\mathbf{v}_\perp'$ depends on $\mathbf{I}_\perp'$, and essentially on $\mathbf{v}_\perp$.

That $v_z' < 0$ can happen. For example, consider $\mathbf{v}_\perp$ to be non-zero in a direction $\hat{\mathbf{s}}$ such that $\hat{\mathbf{z}}^T W \hat{\mathbf{s}} > 0$. Thus, $\mathbf{I}_\perp' = -\mu\hat{\mathbf{s}}$. By (16), a large enough value of $\mu$ will result in $v_z' < 0$.

When $v_z' < 0$ happens, will $v_z$ continue to decrease or eventually increase (and finally reach zero to finish compression)? This question was never raised in the previous works on impact. For a 2D impact, it can be shown that the normal contact velocity will eventually increase at a constant rate to end the impact (Jia et al. 2016). The proof relies on that a slip may happen in only two (opposite) directions. In a 3D impact, a slip may happen in any direction in the contact plane, and evolution of $\mathbf{v}_\perp$ is governed by a differential equation (to be derived in Section 4), which is generally non-integrable. Another complication is the consideration of all possible topologically distinct sequences of contact mode switches and the end of compression (as ordered by the $I_z$ value). To answer the aforementioned question would be too complex (if not impossible) a task.

**Theorem 3.** *Suppose the following condition holds:*

$$\hat{\mathbf{z}}^T W \hat{\mathbf{z}} - \mu \hat{\mathbf{z}}^T W \hat{\mathbf{s}} > 0, \quad \textit{for any unit tangent vector } \hat{\mathbf{s}}. \tag{18}$$

*Then the impact process as modeled in Section 2.2 will end.*

*Proof.* Suppose that (18) holds. Let $b > 0$ be the minimum value of $\hat{\mathbf{z}}^T W \hat{\mathbf{z}} - \mu \hat{\mathbf{z}}^T W \hat{\mathbf{s}}$ for all tangent vectors $\hat{\mathbf{s}}$. Since $\|\mathbf{I}_\perp'\| \leq \mu$ under Coulomb's law, we infer from (16) that $v_z' \geq b$. Thus, $v_z$ will increase to zero monotonically to end compression. After compression, $E$ will decrease at the rate of $v_z$, which will continue to increase at a rate at least $b$. Thus, $E$ will decrease to zero to end restitution.

The condition (18) is called the *impact solution condition*. It is a sufficient condition for us to solve the impact problem.[2] The condition is often satisfied given the positive definiteness of $W$ and the orthogonality of $\hat{\mathbf{s}}$ and $\hat{\mathbf{z}}$.[3] From now on, the impact configuration will be assumed to satisfy the condition (18) for the impact problem to be solvable.[4]

### 2.4 Solution of Frictionless Impact

When $\mu = 0$, the impact solution condition (18) trivially holds given the positive definiteness of $W$. The tangential impulse $\mathbf{I}_\perp$ does not exist, namely, $\mathbf{I}_\perp = \mathbf{0}$. With $\mathbf{I} = I_z\hat{\mathbf{z}}$, equation (10) simplifies to $v_z = v_z^- + \frac{1}{m}I_z$, where $m = (\hat{\mathbf{z}}^T W\hat{\mathbf{z}})^{-1} > 0$ since $W$ is positive definite. Letting $v_z = 0$, we obtain $I_{zc} = -mv_z^-$ from the above equation.

The energy at this point, from integrating $E' = -v_z$, has the maximum value as follows:

$$E_c = \left(-v_z^- I_z - \frac{1}{2m}I_z^2\right)\Bigg|_0^{I_{zc}} = \frac{m(v_z^-)^2}{2}.$$

During restitution, the energy assumes the following expression after dissipation of $(1-e^2)E_c$:

$$E = e^2E_c - \left(v_z^- I_z + \frac{1}{2m}I_z^2\right)\Bigg|_{I_{zc}}^{I_z} = \frac{m}{2}\left(e^2-1\right)v_z^{-2} - v_z^- I_z - \frac{1}{2m}I_z^2.$$

The total normal impulse is the root of the quadratic equation $E = 0$ that is greater than $I_{zc}$: $I_{zr} = -mv_z^-(1+e)$. Hence, we have the total impulse: $\mathbf{I}_r = I_{zr}\hat{\mathbf{z}} = -mv_z^-(1+e)\hat{\mathbf{z}}$.

The icosahedron-tetrahedron impact from Fig. 3, with $\mu$ reset to zero, generates the the final impulse $\mathbf{I}_r = (0,0,0.058)^T$ (cf. the value in (15) with friction).

## 3 Sliding Velocity

While the normal contact velocity $v_z$ decides the phase of impact, the tangential contact velocity $\mathbf{v}_\perp$ affects the mode of contact. It is more convenient to consider the tangential contact velocity as a 2-tuple: $\boldsymbol{\gamma} = (v_x, v_y)^T$, and refer to it as the *sliding velocity*. We obtain

$$\begin{aligned}
\boldsymbol{\gamma} &= (\hat{\mathbf{x}},\hat{\mathbf{y}})^T\mathbf{v} \\
&= (\hat{\mathbf{x}},\hat{\mathbf{y}})^T(\mathbf{v}^- + W\mathbf{I}) \qquad \text{(by (4))} \\
&= \boldsymbol{\gamma}^- + (\hat{\mathbf{x}},\hat{\mathbf{y}})^T W\mathbf{I} \qquad (19) \\
&= \boldsymbol{\gamma}^- + (\hat{\mathbf{x}},\hat{\mathbf{y}})^T W(I_x\hat{\mathbf{x}} + I_y\hat{\mathbf{y}} + I_z\hat{\mathbf{z}}) \\
&= \boldsymbol{\gamma}^- + B\begin{pmatrix} I_x \\ I_y \end{pmatrix} + I_z\mathbf{d}, \qquad (20)
\end{aligned}$$

where

$$B = (\hat{\mathbf{x}},\hat{\mathbf{y}})^T W(\hat{\mathbf{x}},\hat{\mathbf{y}}), \qquad (21)$$
$$\mathbf{d} = (\hat{\mathbf{x}},\hat{\mathbf{y}})^T W\hat{\mathbf{z}}. \qquad (22)$$

For the impact configuration in Fig. 3, we obtain

$$B = \begin{pmatrix} 11.5985 & -0.910369 \\ -0.910369 & 9.90134 \end{pmatrix} \text{ and } \mathbf{d} = \begin{pmatrix} 2.44236 \\ 1.95747 \end{pmatrix}. \qquad (23)$$

For a central impact (like the ball-ball and ball-halfspace impacts in Fig. 2), $\mathbf{r}_i \times \hat{\mathbf{z}} = \mathbf{0}$. Therefore,

$$R_i^T(\mathbf{r}_i \times \hat{\mathbf{z}}) = \mathbf{0} \Leftrightarrow R_i^T P_i\hat{\mathbf{z}} = \mathbf{0} \Rightarrow P_i R_i Q_i^{-1} R_i^T P_i\hat{\mathbf{z}} = \mathbf{0},$$

for $i = 1,2$. Thus, $S\hat{\mathbf{z}} = \mathbf{0}$ by (3), which implies $\hat{\mathbf{x}}^T S\hat{\mathbf{z}} = \hat{\mathbf{y}}^T S\hat{\mathbf{z}} = \mathbf{0}$, and consequently $\mathbf{d} = \mathbf{0}$.

**Proposition 4.** *The $2\times 2$ matrix $B$ is symmetric and positive definite.*

*Proof.* The symmetry of $B$ follows directly from (21) and the symmetry of $W$. Given any $\mathbf{u} \in \mathbb{R}^2$ and $\mathbf{u} \neq \mathbf{0}$, we have

$$\mathbf{u}^T(\hat{\mathbf{x}},\hat{\mathbf{y}})^T W(\hat{\mathbf{x}},\hat{\mathbf{y}})\mathbf{u} = \Big((\hat{\mathbf{x}},\hat{\mathbf{y}})\mathbf{u}\Big)^T W\Big((\hat{\mathbf{x}},\hat{\mathbf{y}})\mathbf{u}\Big) > 0,$$

because $(\hat{\mathbf{x}},\hat{\mathbf{y}})\mathbf{u} \neq \mathbf{0}$ and $W$ is positive definite.

Under Proposition 4, the inverse $B^{-1}$ exists. From (20) we obtain the tangential impulse in terms of $\boldsymbol{\gamma}$ and the normal impulse $I_z$:

$$\begin{pmatrix} I_x \\ I_y \end{pmatrix} = B^{-1}\left(\boldsymbol{\gamma} - \boldsymbol{\gamma}^- - I_z\mathbf{d}\right). \qquad (24)$$

If $\boldsymbol{\gamma}$ is fixed at $\mathbf{0}$, then all possible impulses $\mathbf{I}$ to realize it form a line (parametrized with $I_z$), referred to as the *line of sticking*, in the impulse space:[5]

$$\mathcal{L}_s:\ \begin{pmatrix} I_x \\ I_y \end{pmatrix} + B^{-1}\left(\boldsymbol{\gamma}^- + I_z\mathbf{d}\right) = \mathbf{0}. \qquad (25)$$

We can write the impulse and its derivative as follows:

$$\mathbf{I} = I_z\hat{\mathbf{z}} + (\hat{\mathbf{x}}\hat{\mathbf{y}})B^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^- - I_z\mathbf{d}), \qquad (26)$$
$$\mathbf{I}' = \hat{\mathbf{z}} + (\hat{\mathbf{x}},\hat{\mathbf{y}})B^{-1}(\boldsymbol{\gamma}' - \mathbf{d}). \qquad (27)$$

[2] Even if $\hat{\mathbf{z}}^T W\hat{\mathbf{z}} - \mu\hat{\mathbf{z}}^T W\hat{\mathbf{s}} \leq 0$ for some $\hat{\mathbf{s}}$, we may still be able to solve the impact problem using the above procedure as long as $\hat{\mathbf{z}}^T W\hat{\mathbf{z}} + \hat{\mathbf{z}}^T W\mathbf{I}'_\perp > 0$ is satisfied throughout the impact.

[3] In (18), the product $\hat{\mathbf{z}}^T W\hat{\mathbf{s}}$ has the most positive value when $\hat{\mathbf{s}}$ is in the direction of $W\hat{\mathbf{z}} - \hat{\mathbf{z}}\hat{\mathbf{z}}^T W\hat{\mathbf{z}}$, the projection of $W\hat{\mathbf{z}}$ onto the tangent plane.

[4] No violations have happened in our simulation of dozens of impact instances.

[5] When $\mathbf{I}$ reaches the line, sticking only happens if $\boldsymbol{\gamma}$ stays zero.

## 4 Hodograph and Invariant Directions

The contact slides when $\boldsymbol{\gamma} \neq \mathbf{0}$. It follows from (17) that

$$\begin{pmatrix} I_x' \\ I_y' \end{pmatrix} = \begin{pmatrix} F_x \\ F_y \end{pmatrix} \Big/ F_z = -\mu \hat{\boldsymbol{\gamma}}, \tag{28}$$

where $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}/\|\boldsymbol{\gamma}\|$. From (28) we easily obtain the derivative of the impulse during sliding:

$$\mathbf{I}' = \hat{\mathbf{z}} - \mu(\hat{\mathbf{x}}, \hat{\mathbf{y}})\hat{\boldsymbol{\gamma}}. \tag{29}$$

Let us differentiate (20) and then substitute (28) in:

$$\boldsymbol{\gamma}' = B\begin{pmatrix} I_x' \\ I_y' \end{pmatrix} + \mathbf{d} \tag{30}$$

$$= -\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d}. \tag{31}$$

The trajectory of the sliding velocity $\boldsymbol{\gamma}$, referred to as the *hodograph*, is a plane curve parametrized with $I_z$. The curve has its velocity $\boldsymbol{\gamma}'$ at a point $\boldsymbol{\gamma}(I_z)$ completely determined by the direction $\hat{\boldsymbol{\gamma}}$ from the origin to the point.

Fig. 4(a) plots the hodograph of the impact instance in Fig. 3. In the figure, $\mathbf{g}_c$ and $\mathbf{g}_r$ are the values of $\boldsymbol{\gamma}$ at the end of compression and restitution, respectively.

In the case of a central impact, the differential equation (31) is simplified to $\boldsymbol{\gamma}' = -\mu B\hat{\boldsymbol{\gamma}}$ under $\mathbf{d} = \mathbf{0}$. The magnitude $\|\boldsymbol{\gamma}\|$ of the sliding velocity decreases monotonically because $\left(\boldsymbol{\gamma}^T\boldsymbol{\gamma}\right)' = -2\mu\boldsymbol{\gamma}^T B\boldsymbol{\gamma}/\|\boldsymbol{\gamma}\| < 0$, given the positive definiteness of $B$.

As the hodograph of $\boldsymbol{\gamma}$ evolves under (31) with $I_z$ increasing, it may reach a point $\mathbf{g}_l$ at $I_z = I_{zl}$ where $\boldsymbol{\gamma}$ and $-\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d}$ become collinear, namely, $\boldsymbol{\gamma} \times (-\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d}) = \mathbf{0}$.[6] The case $\boldsymbol{\gamma} = \mathbf{0}$ will be considered separately in Section 5. Suppose $\mathbf{g}_l \neq \mathbf{0}$ so it has the direction $\hat{\mathbf{g}}_l$. It follows that $-\mu B\hat{\mathbf{g}}_l + \mathbf{d} = \lambda\hat{\mathbf{g}}_l$ for some $\lambda$. For $I_z \geq I_{zl}$, equation (31), reduced to $\boldsymbol{\gamma}' = \lambda\hat{\mathbf{g}}_l$, can be integrated until $\boldsymbol{\gamma} = \mathbf{0}$ (if this condition ever holds):

$$\boldsymbol{\gamma} = \mathbf{g}_l + \lambda(I_z - I_{zl})\hat{\mathbf{g}}_l. \tag{32}$$

Starting at $\mathbf{g}_l$, $\boldsymbol{\gamma}$ will be moving on a line in the direction $\hat{\mathbf{g}}_l$. There are three cases:

1. $\lambda < 0$. The sliding speed $\|\boldsymbol{\gamma}\|$ will decrease.
2. $\lambda = 0$. This means that $\boldsymbol{\gamma}' = -\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d} = \mathbf{0}$. Hence, $\boldsymbol{\gamma}$ will not change in the rest duration of the impact.
3. $\lambda > 0$. The sliding speed $\|\boldsymbol{\gamma}\|$ will increase.

The direction $\hat{\mathbf{g}}_l$ is called *invariant* since the sliding velocity $\boldsymbol{\gamma}$ starting at $\mathbf{g}_l$ will keep its direction until either $\boldsymbol{\gamma} = \mathbf{0}$ or the impact ends. Every invariant direction is identified with a unit vector $\hat{\mathbf{s}} \in \mathbb{R}^2$ satisfying the following equation:

$$\hat{\mathbf{s}} \times (-\mu B\hat{\mathbf{s}} + \mathbf{d}) = \mathbf{0}. \tag{33}$$

Such $\hat{\mathbf{s}}$ is called *centripetal invariant* [20] if $\hat{\mathbf{s}}^T(-\mu B\hat{\mathbf{s}} + \mathbf{d}) \leq 0$, and *centrifugal invariant* if $\hat{\mathbf{s}}^T(-\mu B\hat{\mathbf{s}} + \mathbf{d}) > 0$. Invariant directions depend on the impact configuration and the coefficient of friction but not on the pre-impact velocities.

**Theorem 5.** *Let $\lambda_1, \lambda_2 > 0$ be the two eigenvalues of $B$. The following statements hold:*

*(i) If $\lambda_1 = \lambda_2$ and $\mathbf{d} = \mathbf{0}$, every unit vector $\hat{\mathbf{s}} \in \mathbb{R}^2$ is an invariant direction.*
*(ii) Otherwise, there exist two to four invariant directions.*
*(iii) If $\|B^{-1}\mathbf{d}\| \leq \mu$, every invariant direction is centripetal.*
*(iv) If $\|B^{-1}\mathbf{d}\| > \mu$, there exists at least one centripetal invariant direction and exactly one centrifugal invariant direction.*

*Proof.* Since $B$ is symmetric, it has a spectral decomposition $B = U\Sigma U^T$, where $U$ is an orthogonal matrix and $\Sigma$ is the diagonal matrix $\mathrm{diag}(\lambda_1, \lambda_2)$. The following equivalences hold for a unit vector $\hat{\mathbf{s}}$:

$$\begin{aligned}
\hat{\mathbf{s}} \times (-\mu B\hat{\mathbf{s}} + \mathbf{d}) = \mathbf{0} &\Leftrightarrow \hat{\mathbf{s}} \times \left(-\mu U\Sigma U^T\hat{\mathbf{s}} + \mathbf{d}\right) = \mathbf{0} \\
&\Leftrightarrow U^T\hat{\mathbf{s}} \times \left(-\mu\Sigma U^T\hat{\mathbf{s}} + U^T\mathbf{d}\right) = \mathbf{0} \\
&\Leftrightarrow \hat{\mathbf{u}} \times \left(-\mu\Sigma\hat{\mathbf{u}} + U^T\mathbf{d}\right) = \mathbf{0} \quad (\hat{\mathbf{u}} = U^T\hat{\mathbf{s}}) \\
&\Leftrightarrow \hat{\mathbf{u}} \times \left(-\Sigma\hat{\mathbf{u}} + \mu^{-1}U^T\mathbf{d}\right) = \mathbf{0} \\
&\Leftrightarrow \hat{\mathbf{u}} \times \left(\Sigma\hat{\mathbf{u}} - \mu^{-1}U^T\mathbf{d}\right) = \mathbf{0}.
\end{aligned}$$

In the above, $U^T$ acts as a rotation. Also, $\hat{\mathbf{s}}$ and $-\mu B\hat{\mathbf{s}} + \mathbf{d}$ are in the same direction if and only if $\hat{\mathbf{u}}$ and $\Sigma\hat{\mathbf{u}} - \mu^{-1}U^T\mathbf{d}$ are in opposite directions. The proof of part (i) is then completed by applying part (i) of Theorem 8 in Appendix A, and that of part (ii) by applying parts (i)–(iv) of the same theorem.

Suppose $\|B^{-1}\mathbf{d}\| \leq \mu$. Hence, we infer

$$\begin{aligned}
\|\Sigma^{-1}(-\mu^{-1}U^T\mathbf{d})\| &= \mu^{-1}\|\Sigma^{-1}U^T\mathbf{d}\| \\
&= \mu^{-1}\|U\Sigma^{-1}U^T\mathbf{d}\| \\
&= \mu^{-1}\|B^{-1}\mathbf{d}\| \\
&\leq 1.
\end{aligned}$$

Applying part (v) of Theorem 8, every $\hat{\mathbf{u}}$ collinear with $\Sigma\hat{\mathbf{u}} - \mu^{-1}U^T\mathbf{d}$ must be in its direction, namely, in the opposite direction of $-\Sigma\hat{\mathbf{u}} + \mu^{-1}U^T\mathbf{d}$. Equivalently, every $\hat{\mathbf{s}}$ satisfying (33) is centripetal. This establishes (iii).

To establish part (iv), for convenience we introduce the notation $\sim$ such that $\mathbf{u} \sim \mathbf{w}$ means that the two vectors $\mathbf{u}$ and $\mathbf{w}$ are non-zero and in the same direction. From the sequence of equivalences earlier in the proof, we see that, for unit vectors $\hat{\mathbf{s}}$ and $\hat{\mathbf{u}} = U^T\hat{\mathbf{s}}$,

$$\begin{aligned}
\hat{\mathbf{s}} \sim -\mu B\hat{\mathbf{s}} + \mathbf{d} &\Leftrightarrow \hat{\mathbf{u}} \sim -(\Sigma\hat{\mathbf{u}} - \mu^{-1}U^T\mathbf{d}), \\
\hat{\mathbf{s}} \sim \mu B\hat{\mathbf{s}} - \mathbf{d} &\Leftrightarrow \hat{\mathbf{u}} \sim \Sigma\hat{\mathbf{u}} - \mu^{-1}U^T\mathbf{d}.
\end{aligned}$$

[6] The cross product of two tuples (representing two vectors in the $x$-$y$ plane) is always along the $z$-axis, and thus treated as a scalar for convenience.

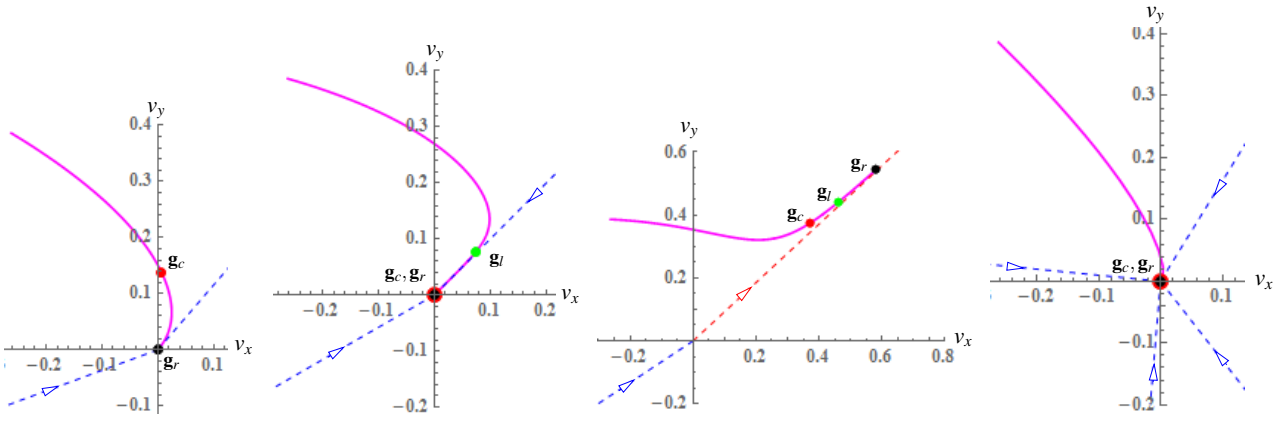



Fig. 4. Hodographs of four instances of the icosahedron-tetrahedron impact in Fig. 3 generated by changing the values of some physical parameters listed in its caption and (14): (a) no change; (b) $\mu = 0.4$ and $V_{1z}^- = 0.2$; (c) $\mu = 0.25$ and $V_{1z}^- = 0.2$; and (d) $\mu = 3.0$ and $V_{1z}^- = 0.5$. Here, $V_{1z}^-$ is the $z$-component of the pre-impact velocity $\mathbf{V}_1^-$ of the icosahedron. The direction out of the origin along a dashed (blue) line represents a centripetal invariant direction. The line is labeled with an arrow pointing towards the origin to indicate the magnitude change of the sliding velocity in this direction. The direction out of the origin along a dashed (red) line represents a centrifugal invariant direction. This line is labeled with an arrow pointing away from the origin.

That $\|B^{-1}\mathbf{d}\| > \mu$ implies $\|\Sigma^{-1}(-\mu^{-1}U^T\mathbf{d})\| > 1$. Applying part (vi) of Theorem 8 from Appendix A, there exists at least one unit vector $\hat{\mathbf{u}}$ such that $\hat{\mathbf{u}} \sim \Sigma\hat{\mathbf{u}} - \mu^{-1}U^T\mathbf{d}$, and exactly one such that $\hat{\mathbf{u}} \sim -\Sigma\hat{\mathbf{u}} + \mu^{-1}U^T\mathbf{d}$.

To find all the invariant directions, we first check if $(-1,0)^T$ is one. Then, let the unit vector $\hat{\mathbf{s}}$ represent a direction. Utilize the one-to-one correspondence $\hat{\mathbf{s}} = ((1-\eta^2)/(1+\eta^2), 2\eta/(1+\eta^2))^T$ from $(-\infty,\infty)$ to the set of all unit vectors except $(-1,0)^T$. Substitute this expression into (33). After some cleanup, we end up with a quartic polynomial equation in $\eta$, which has roots in analytic forms. Each real root of $\eta$ leads to a distinct invariant direction. If $\|B^{-1}\mathbf{d}\| \le \mu$, the found directions are all centripetal invariant. Otherwise, we can easily check the sign of the cross product $\hat{\mathbf{s}} \times (-\mu B\hat{\mathbf{s}} + \mathbf{d})$ to determine the unique centrifugal invariant direction.

Generally, $\lambda_1 \neq \lambda_2$ or $\mathbf{d} \neq \mathbf{0}$. The invariant directions $\hat{\mathbf{s}}_k$, $2 \le k \le 4$, partition the plane into up to four sectors. The hodograph $\boldsymbol{\gamma}(I_z)$ must stay within one sector and not cross either of its two bounding rays. Once $\boldsymbol{\gamma}(I_z)$ reaches such a bounding ray, it will stay on the ray until $\boldsymbol{\gamma} = \mathbf{0}$ or the end of the impact otherwise.

For the icosahedron-tetrahedron impact in Fig. 3, we calculate $\|B^{-1}\mathbf{d}\| = 0.3157$ from (23). Fig. 4(b)–(d) shows the hodographs of three modified instances of the impact by changing only the values of $\mu$ and $\mathbf{V}_1^-$. The one in Fig. 4(b), satisfying $\|B^{-1}\mathbf{d}\| < \mu$, has two centripetal directions. At $\mathbf{g}_l$, the hodograph reaches one of them and then moves along this direction to reach the origin. The instance in Fig. 4(c), under $\|B^{-1}\mathbf{d}\| > \mu$, has a centripetal direction as well as a centrifugal direction. At $\mathbf{g}_l$, the hodograph moves away from the origin in the centrifugal direction. Fig. 4(d) displays a situation with four centripetal invariant directions but no centrifugal invariant direction since $\|B^{-1}\mathbf{d}\| < \mu$.

## 5 Vanishing of Sliding Velocity

Suppose that $\boldsymbol{\gamma}$ becomes $\mathbf{0}$ at some value $I_z = I_{zs}$ ($I_{zs} = 0$ if $\boldsymbol{\gamma}^- = \mathbf{0}$). Denote by $\mathbf{I}_s$ the value of $\mathbf{I}$ at the moment. For the contact to stick, $\boldsymbol{\gamma}' = \mathbf{0}$ must be maintained, which by (30) implies that the tangential impulse must change at a constant rate:

$$\begin{pmatrix} I_x' \\ I_y' \end{pmatrix} = -B^{-1}\mathbf{d}. \tag{34}$$

Under Coulomb friction, $\|\mathbf{I}_\perp'\| = \sqrt{I_x'^2 + I_y'^2} \le \mu$. We need to look at two cases: $\|B^{-1}\mathbf{d}\|$ exceeds $\mu$ or not. The matrix $B$ depends only on the impact configuration and the objects' mass and inertia properties. This means that we can predict at the start of impact what will happen after $\boldsymbol{\gamma}$ reaches zero — whether it would stick or continue sliding — without knowing if $\boldsymbol{\gamma} = \mathbf{0}$ will happen.

### 5.1 Stick

When $\|B^{-1}\mathbf{d}\| \le \mu$, friction will be enough to keep the contact stay in the sticking mode. Substituting $\boldsymbol{\gamma} = \mathbf{0}$ into (26), we obtain the impulse as a linear function of $I_z$ from now on until the impact ends:

$$\mathbf{I} = I_z\hat{\mathbf{z}} - (\hat{\mathbf{x}}, \hat{\mathbf{y}})B^{-1}(\boldsymbol{\gamma}^- + I_z\mathbf{d}), \qquad I_z \ge I_{zs}. \tag{35}$$

The above equation defines a line $\mathcal{L}_s$ in the impulse space. It is called the *line of sticking* [18].

For a central impact, $\mathbf{d} = \mathbf{0}$ and thus $B^{-1}\mathbf{d} = \mathbf{0}$. Sticking will immediately start when $\boldsymbol{\gamma} = \mathbf{0}$.

Fig. 5 shows the impulse curve for the impact instance in Fig. 3 with pre-impact velocities given in (14), and $B$ and $\mathbf{d}$ in (23). It generates the hodograph shown in Fig. 4(a).

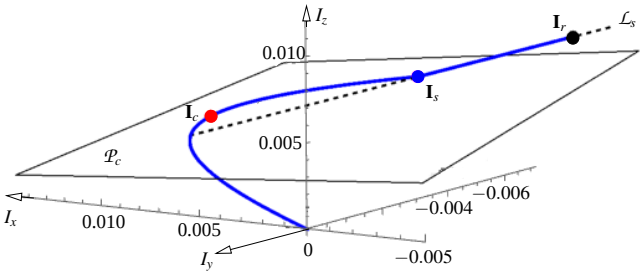



Fig. 5. Impulse curve for the icosahedron-tetrahedron impact. Also shown are the plane $\mathcal{P}_c$ of compression and the line $\mathcal{L}_s$ of sticking.

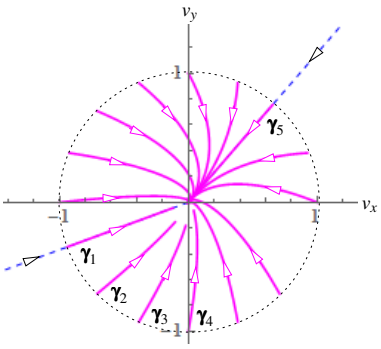



Fig. 6. Bundle of 15 hodographs resulting from unit $\boldsymbol{\gamma}^-$ values for the icosahedron-tetrahedron collision ($\mu = 0.8$ and $v_z = -0.2$).

Compression ends when the curve reaches $\mathbf{I}_c$. The sliding velocity becomes zero at $\mathbf{I}_s$. Since $\|B^{-1}\mathbf{d}\| = 0.3157 < \mu = 0.8$, the contact mode changes to stick. The impulse starts accumulating along the line of sticking $\mathcal{L}_s$ to reach the end point $\mathbf{I}_r$ given in (15). Sticking also happens with the impact instances in Fig. 4 (b) and (d).

Given $\mu$, the matrix $B$, and the vector $\mathbf{d}$, the hodograph defined by (31) has its shape completely determined by $\boldsymbol{\gamma}^-$, and its extension determined by the duration of impact which depends on $v_z^-$. Fig. 6 shows fifteen hodographs generated by different unit vectors $\boldsymbol{\gamma}^-$ for the icosahedron-tetrahedron impact in Fig. 3. There exists no centrifugal direction. All hodographs evolve towards the origin. Among them, $\boldsymbol{\gamma}_i$, $1 \le i \le 4$, will not reach the origin due to earlier ending of restitution. The remaining eleven hodographs will reach the origin and stay there with the contact sticking. The hodographs $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_5$, starting in the only two centripetal directions, are straight line segments.

### 5.2 Resumed Slip

When $\|B^{-1}\mathbf{d}\| > \mu$, the contact cannot stick (and sliding will continue). The velocity $\boldsymbol{\gamma}$ will grow out of zero at $I_z = I_{zs}$ in some direction $\hat{\mathbf{s}}$. This direction must agree with the direction of the acceleration at $I_{zs}$, which is $\boldsymbol{\gamma}' = -\mu B\hat{\mathbf{s}} + \mathbf{d}$ according to (31) with $\hat{\boldsymbol{\gamma}} = \hat{\mathbf{s}}$. Thus, the initial direction of resumed sliding must be centrifugal (and this sliding direction will be maintained). The corollary below then follows from Theorem 5(iv).

**Corollary 6.** *Suppose $\|B^{-1}\mathbf{d}\| > \mu$. Once the sliding velocity $\boldsymbol{\gamma}$ reaches zero, sliding will continue in the unique centrifugal direction.*

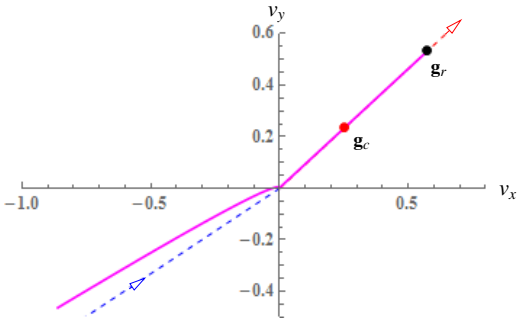



Fig. 7. Hodograph of the impact instance in Fig. 3 with the following changed parameter values: $\mu = 0.25$ and $\mathbf{V}_1^- = (-0.7, -0.35, -0.5)^T$.

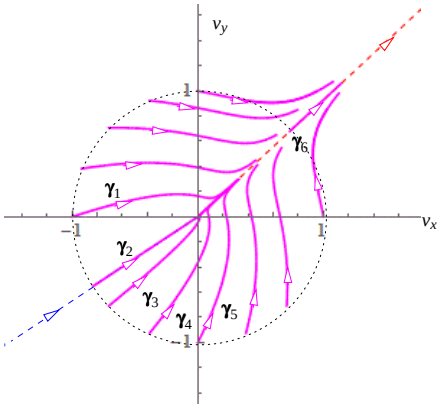



Fig. 8. Bundle of 13 hodographs resulting from unit $\boldsymbol{\gamma}^-$ values for the icosahedron-tetrahedron collision ($\mu = 0.25$ and $v_z = -0.8$).

Fig. 7 displays the hodograph of a modified instance of the icosahedron-tetrahedron impact. The sliding velocity $\boldsymbol{\gamma}$ becomes zero during compression, and afterwards, accelerates in a centrifugal invariant direction until the impact ends.

Fig. 8 displays thirteen hodographs resulting from different unit values of $\boldsymbol{\gamma}^-$ when $\|B^{-1}\mathbf{d}\| > d$. There exist one centrifugal and one centripetal directions. All the hodographs drift towards the centrifugal direction, although only six of them, $\boldsymbol{\gamma}_i$, $1 \le i \le 6$, finally reach it. Among these six hodographs, $\boldsymbol{\gamma}_6$ starts in the centrifugal direction and thus evolves out a straight trajectory; $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_4$, and $\boldsymbol{\gamma}_5$ reach the direction but not the origin; $\boldsymbol{\gamma}_2$ starts in the centripetal direction, reaches the origin, and switches to the centrifugal direction; and $\boldsymbol{\gamma}_3$ reaches the origin before continuing in the centrifugal direction.

After the sliding velocity $\boldsymbol{\gamma}$ becomes zero, the tangential impulse $\mathbf{I}_\perp$ will be accumulating along a constant direction. Since $-\mu B\hat{\mathbf{s}} + \mathbf{d} = \lambda\hat{\mathbf{s}}$ for some known $\lambda > 0$, we have $\boldsymbol{\gamma}' = \lambda\hat{\mathbf{s}}$ by (31) and $\mathbf{I}' = \hat{\mathbf{z}} + (\hat{\mathbf{x}}, \hat{\mathbf{y}})B^{-1}(\lambda\hat{\mathbf{s}} - \mathbf{d})$ by (27).

Suppose we know the value $I_z = I_{zs}$ at which $\boldsymbol{\gamma}$ will become zero. Whether sticking or resumed sliding will happen next, the impulse accumulation will start to take a linear form in $I_z$. Unfortunately, there is no way of determining $I_{zs}$ in general without integrating the differential equation (31).

## 6 Solution of the Impact Problem

This section will expand the solution scheme outlined in Section 2.2 into an algorithm. The normal impulse $I_z$, accumulating from zero until the impact ends, is the sole variable. As $I_z$ reaches certain value, the current impact phase will end or the current contact mode will change. The impulse $\mathbf{I}$ is tracked by integrating the system (31) over $I_z$, either numerically or analytically, and then evaluating (26). The energy, deciding when an impact phase ends, is tracked by integrating (11).

**Theorem 7.** *Given the impact configuration for two bodies $\mathcal{B}_1$ and $\mathcal{B}_2$, and their inertia properties, the total impulse $\mathbf{I}_r$ (and therefore the changes in their velocities according to (1)) depend on the pre-impact contact velocity $\mathbf{v}^-$ not on individual pre-impact velocities.*

*Proof.* The inverse inertia matrix $W$ defined in (5) is determined by the impact configuration and the bodies' inertia properties, so are the matrix $B$ in (21) and the vector $\mathbf{d}$ in (22) which govern the evolution of the sliding velocity $\boldsymbol{\gamma}$ in (31). During the impact, the impulse $\mathbf{I}$ is evaluated based on $\boldsymbol{\gamma}$ and the normal impulse $I_z$ according to (26). When the contact slides, $\boldsymbol{\gamma}$ evolves according to (31) which has a unique solution in terms of $I_z$. When the contact sticks, the impulse $\mathbf{I}$ has a simplified form (35) in terms of $I_z$. Slip or stick is initialized with $\mathbf{v}^-$, and determined by checking if $\boldsymbol{\gamma} = \mathbf{0}$, and, if true, by further checking if $\|B^{-1}\mathbf{d}\| \le \mu$. In summary, the growth in $\mathbf{I}$ is completely determined by the impact configuration and $\mathbf{v}^-$.

Meanwhile, the potential energy $E$ at the contact, evolving under (11), completely depends on $\mathbf{I}$ and $v_z^-$. So are the ends of compression ($E' = 0$) and restitution ($E = 0$). Thus, the value of $\mathbf{I}_r$ at the end of restitution is determined.

Our impact solution will make use of events which represent certain moments during the impact. An event $\mathfrak{u}$ is measured in terms of the value of the normal impulse $I_z$ at which the event happens. For two events $\mathfrak{u}1$ and $\mathfrak{u}2$, we write $\mathfrak{u}_1 \prec \mathfrak{u}_2$ if $\mathfrak{u}_1$ occurs before $\mathfrak{u}_2$, and $\mathfrak{u}_1 \preceq \mathfrak{u}_2$ if $\mathfrak{u}_1$ occurs before or simultaneously with $\mathfrak{u}_2$. There are five possible events:

1. Event 0 refers to the start of the impact.
2. Event $c$ refers to the end of compression.
3. Event $r$ refers to the end of restitution.
4. Event $s$ happens when the sliding velocity $\boldsymbol{\gamma}$ first becomes zero. At the event, either the contact mode changes from sliding to sticking or sliding will resume along a new constant direction.
5. Event $l$ happens when $\boldsymbol{\gamma} \times (-\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d})$ first becomes zero but $\boldsymbol{\gamma}$ has been non-zero until this moment (inclusive). Afterward, the contact will be sliding along a line.

On the hodograph, events 0 and $s$ correspond to the origin, while events $c$, $r$, and $l$ are marked by the points $\mathbf{g}_c$, $\mathbf{g}_r$, and $\mathbf{g}_l$, respectively.

Events $c$ and $r$ always happen sequentially unless the coefficient of restitution $e$ is zero; namely, $c \preceq r$. The definitions of events $s$ and $l$ imply the following: a) neither event happens more than once; b) event $l$, if happens, must precede event $s$, namely, $l \prec s$ must hold; c) event $s$ may or may not happen. The event $s$ may coincide with either $c$ or $r$, leading to three possibilities: $s \preceq c$, $c \prec s \preceq r$, and no occurrence of $s$. Event $l$ also has the same three possibilities relative to events $c$ and $r$.

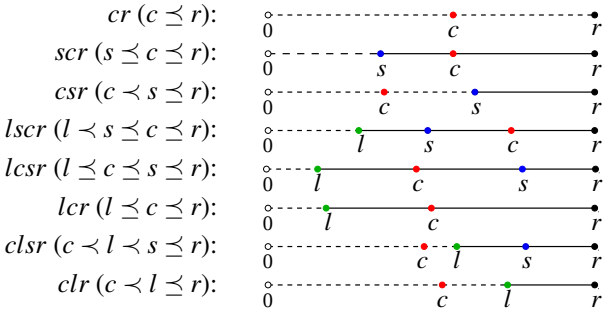



Fig. 9. Eight possible event sequences. A dashed line segment connecting two events represents a period over which the impulse growth has to be computed via numerical integration, while a solid line segment represents a period during which the impulse growth has a closed form.

Based on the above relationships, easy reasoning concludes with eight possible event sequences as shown in Fig. 9. The sequence *cr* represents no occurrence of event $s$ or $l$. The next two sequences *scr* and *csr* have the occurrence of event $s$ but not event $l$. The last five sequences all include event $l$, but differ in whether event $s$ also occurs (after $l$ if so) and in whether the events $l$ and $s$ occur during compression (i.e., $l, s \preceq c$) or restitution (i.e., $c \prec l, s$). Each sequence describes a different impact scenario.

Impact solution begins with numerical integration until either event $s$ or event $l$ happens.[7], or until event $r$ occurs otherwise. In the case of event $s$ or $l$, additional accumulation of the impulse $\mathbf{I}$ can be evaluated using one or multiple closed-forms until the impact ends.

Section 6.1 will describe how to efficiently conduct numerical integration, as this is the only way to solve the impact scenario *cr* with no occurrence of event $s$ or $l$. Scenarios *scr* and *csr*, with no occurrence of event $l$, will be treated in Sections 6.2 , while scenarios *lscr*, *lcsr*, *lcr*, *clsr*, and *clr*, all with occurrence of event $l$, will be treated in in 6.3.

### 6.1 Fast Numerical Integration

Before event $s$ or $l$ occurs, impact has to be simulated via numerical integration. The Euler method is carried out on (31) over the normal impulse $I_z$ to update the tangential contact velocity $\boldsymbol{\gamma}$, and then the impulse $\mathbf{I}$ according to (26). Within the same integration step, the potential energy $E$ is updated according to (11).

Viewing the hodograph as a parametric curve $\boldsymbol{\gamma}(I_z)$, we adjust the step size of the normal impulse $I_z$ according to the curve speed $\|\boldsymbol{\gamma}'\|$ and absolute curvature $|\kappa|$.

1. If we were only to consider constant increment in the arc length, say, $h_1$, then $I_z$ would follow a (varying) step size of $h_1/\|\boldsymbol{\gamma}'\|$.

[7] Each of the two events may happen at the start.

2. If we were only to consider constant increment in the tangential angle, say, $h_2$, then $I_z$ would follow a step size of $h_2/(|\kappa|\|\boldsymbol{\gamma}'\|)$. Here, we make use of the definition of curvature as the derivative of the tangential angle with respect to arc length [23, p. 29], to adjust the arc length increment $\Delta s$ as $h_2/|\kappa|$.

Combining the above two factors, we set the step size for numerical integration as below:

$$h = \varepsilon \frac{h_1}{\|\boldsymbol{\gamma}'\|} + (1-\varepsilon)\frac{h_2}{|\kappa|\|\boldsymbol{\gamma}'\|}, \qquad \text{for } 0 < \varepsilon < 1. \tag{36}$$

From (31), all high order derivatives of $\boldsymbol{\gamma}$ can be expressed in terms of $\boldsymbol{\gamma}$. The curvature is thus derived as

$$\begin{aligned}\kappa &= \frac{\boldsymbol{\gamma}' \times \boldsymbol{\gamma}''}{\|\boldsymbol{\gamma}'\|^3} \\ &= \frac{(-\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d}) \times \left[-\mu B\hat{\boldsymbol{\gamma}} \times \left((-\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d}) \times \hat{\boldsymbol{\gamma}}\right)\right]}{\|\boldsymbol{\gamma}'\| \cdot \|-\mu B\hat{\boldsymbol{\gamma}} + \mathbf{d}\|^3}.\end{aligned} \tag{37}$$

If $\boldsymbol{\gamma}$ locally curves a lot, i.e., if $|\kappa|$ is large, then the first summand in (36) dominates the second one. In this case, the step size is influenced more by the speed — the higher the speed the smaller the step size. If $\boldsymbol{\gamma}$ is locally straight, i.e, $|\kappa|$ is small, then the second summand will dominate the first one. The step size is influenced more by the curvature — the higher the curvature the smaller the step size. The coefficient $\varepsilon$ in (36) may be fixed or adjusted dynamically during the integration.

### 6.2 Event $s$ (Zero Sliding Velocity)

Suppose that event $s$ happens at $I_z = I_{zs}$.[8] The values of $\mathbf{I}$ and $E$ are respectively $\mathbf{I}_s$ and $E_s$ at this moment. First, substitute $\boldsymbol{\gamma} = \mathbf{0}$ and $I_z = I_{zs}$ into (26) to polish the value of $\mathbf{I}_s$ obtained through numerical integration. Combine (27) for sticking (with $\boldsymbol{\gamma}' = \mathbf{0}$) and (29) for resumed sliding (with $\hat{\boldsymbol{\gamma}}$ set to the unique centrifugal direction $\hat{\mathbf{s}}$) into one equation:

$$\mathbf{I}' = \boldsymbol{\sigma} = \begin{cases} \hat{\mathbf{z}} - (\hat{\mathbf{x}}\,\hat{\mathbf{y}})B^{-1}\mathbf{d}, \text{ if } \|B^{-1}\mathbf{d}\| \le \mu \text{ (sticking);} \\ \hat{\mathbf{z}} - \mu(\hat{\mathbf{x}}\,\hat{\mathbf{y}})\hat{\mathbf{s}}, \quad \text{if } \|B^{-1}\mathbf{d}\| > \mu \text{ (resumed sliding).} \end{cases} \tag{38}$$

The impulse will take the new form:

$$\mathbf{I} = \mathbf{I}_s + (I_z - I_{zs})\boldsymbol{\sigma}, \qquad \text{for } I_z \ge I_{zs}, \tag{39}$$

with the ending value $\mathbf{I}_r = \mathbf{I}_s + (I_{zr} - I_{zs})\boldsymbol{\sigma}$. Hence, the task becomes one to determine the normal impulse value $I_{zr}$ at the end of restitution.

Check the sign of $E'$ at $I_{zs}$ according to (11). If $E' \ge 0$, we infer $I_{zs} \le I_{zc}$ , the value of $I_z$ when compression ends, i.e., $s \preceq c$; otherwise, $I_{zs} > I_{zc}$, i.e., $c \prec s$. We substitute (39) into (11) to obtain the energy derivative after event $s$ (i.e., for $I_z \ge I_{zs}$):

$$\begin{aligned}E' &= -v_z^- - \left(\frac{1}{m_1} + \frac{1}{m_2}\right) I_z + \hat{\mathbf{z}}^T S\Big(\mathbf{I}_s + (I_z - I_{zs})\boldsymbol{\sigma}\Big) \\ &= -v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_s - I_{zs}\boldsymbol{\sigma}) - \frac{1}{m_\sigma} I_z,\end{aligned} \tag{40}$$

where $m_\sigma = (\frac{1}{m_1} + \frac{1}{m_2} - \hat{\mathbf{z}}^T S\boldsymbol{\sigma})^{-1}$. The above derivative has the following indefinite integral (with the constant term ignored):

$$\Phi_1(I_z) \equiv \left(-v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_s - I_{zs}\boldsymbol{\sigma})\right) I_z - \frac{1}{2m_\sigma} I_z^2. \tag{41}$$

Two cases below arise:

**Case 1** Compression has not ended by $I_z = I_{zs}$, yielding the event sequence *scr*. Example hodographs are shown in Figs. 4(d) and 7. (Note that event $s$ is represented by the origin.) Solving the equation $E' = 0$ after a substitution of (40), we first obtain $I_{zc} = m_\sigma\left(-v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_s - I_{zs}\boldsymbol{\sigma})\right)$. The maximum energy (achieved at the end of compression) is $E_c = E_s + \Phi_1(I_{zc}) - \Phi_1(I_{zs})$ . During restitution, the energy takes the form $E = e^2 E_c + \Phi_1(I_z) - \Phi_1(I_{zc})$. Letting $E = 0$ and substituting (41) in, we end up with a quadratic equation: $a_2 I_z^2 + a_1 I_z + a_0 = 0$, where $a_2 = -\frac{1}{2m_\sigma}$, $a_1 = -v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_s - I_{zs}\boldsymbol{\sigma})$, and $a_0 = e^2 E_c - \Phi_1(I_{zc})$.

**Case 2** Compression has ended by $I_z = I_{zs}$, yielding the event sequence *csr*. An example is Fig. 4(a). Since it is during restitution, the energy has the form $E = E_s + \Phi_1(I_z) - \Phi_1(I_{zs})$, for $I_z \ge I_{zs}$. Vanishing of $E$ yields the quadratic equation $a_2 I_z^2 + a_1 I_z + b_0 = 0$, where $b_0 = E_s - \Phi_1(I_{zs})$.

In either case, we solve the resulting quadratic equation to obtain $I_{zr}$ as the smallest root greater than $I_{zs}$.

### 6.3 Event $l$ (Straight Sliding)

Suppose that event $l$ happens at $I_z = I_{zl}$ with $\boldsymbol{\gamma}$ at $\mathbf{g}_l$. Denote by $\mathbf{I}_l$ and $E_l$ the current values of $\mathbf{I}$ and $E$, respectively. Since $\mathbf{g}_l \ne \mathbf{0}$, we let $\hat{\mathbf{g}}_l$ be its direction. For $I_z \ge I_{zl}$ and as long as $\boldsymbol{\gamma} \ne \mathbf{0}$, $\boldsymbol{\gamma}' = \lambda\hat{\mathbf{g}}_l$ for some $\lambda$. This implies (32), from which we obtain the corresponding normal impulse $I_{zs} = I_{zl} + \Delta I_z$ at which event $s$ would happen:

$$\Delta I_z = \begin{cases} -\|\mathbf{g}_l\|/\lambda, \text{ if } \lambda < 0, \text{ i.e., } \hat{\mathbf{g}}_l \text{ is centripetal,} \\ \infty, \qquad\qquad \text{otherwise.} \end{cases} \tag{42}$$

Meanwhile, the impulse is, from integrating (29),

$$\mathbf{I} = \mathbf{I}_l + (I_z - I_{zl})\boldsymbol{\delta}, \tag{43}$$

where

$$\boldsymbol{\delta} = \hat{\mathbf{z}} - \mu(\hat{\mathbf{x}}, \hat{\mathbf{y}})\hat{\mathbf{g}}_l. \tag{44}$$

[8] This includes the case $\boldsymbol{\gamma}^- = \mathbf{0}$ where $I_{zs} = 0$.

Substitute (43) into the energy derivative (11):

$$E' = -v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_l - I_{zl}\boldsymbol{\delta}) - \frac{1}{m_\delta} I_z, \tag{45}$$

where $m_\delta = (\frac{1}{m_1} + \frac{1}{m_2} - \hat{\mathbf{z}}^T S\boldsymbol{\delta})^{-1}$, and integrate:

$$\Phi_2(I_z) = \left(-v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_l - I_{zl}\boldsymbol{\delta})\right) I_z - \frac{1}{2m_\delta} I_z^2. \tag{46}$$

If $l \prec c$, solution of the equation $E' = 0$ yields the normal impulse value $\zeta_c = m_\delta \left(-v_z^- + \hat{\mathbf{z}}^T S(\mathbf{I}_l - I_{zl}\boldsymbol{\delta})\right)$ at which compression would end if $\boldsymbol{\gamma} \neq \mathbf{0}$ is to be maintained. Two event sequences arise from comparing $I_{zs}$ with $\zeta_c$.

**Case 1 ($I_{zs} \leq \zeta_c$)** The sequence is *lscr*. Here, $\boldsymbol{\gamma}$ will have decreased to $\mathbf{0}$ by the end of compression. An example instance is shown in Fig. 4(b). Set $E_s = \Phi_2(I_{zs})$, and

$$\mathbf{I}_s = \mathbf{I}_l + (I_{zs} - I_{zl})\boldsymbol{\delta}. \tag{47}$$

From this point on, handle the case exactly the same as the event sequence *scr* in Case 1 in Section 6.2.

**Case 2 ($I_{zs} > \zeta_c$)** In this case, sliding will continue into restitution. Set $I_{zc} = \zeta_c$ and $E_c = E_l + \Phi_2(I_{zc}) - \Phi_2(I_{zl})$. Before $\boldsymbol{\gamma}$ becomes zero, the energy is $E = e^2 E_c + \Phi_2(I_z) - \Phi_2(I_{zc})$. Again, by setting $E = 0$ we end up with a quadratic equation: $\Phi_2(I_z) + e^2 E_c - \Phi_2(I_{zc}) = 0$. Let $\zeta_r$ be the smallest root greater than $I_{zc}$. This is the value of $I_z$ at the end of restitution if sliding continues until then. There are two further subcases:

**Case 2a ($I_{zs} \leq \zeta_r$)** This leads to the event sequence *lcsr*. An instance is shown in Fig. 10(a). Let $E_s = e^2 E_c + \Phi_2(I_{zs}) - \Phi_2(I_{zc})$ and set $\mathbf{I}_s$ according to (47). Handle the period $I_z > I_s$ exactly as for the sequence *csr* described as Case 2 in Section 6.2.

**Case 2b ($I_{zs} > \zeta_r$)** This is the event sequence *lcr*. The sliding velocity will never decrease to zero. Fig. 10(b) shows an example. We let $I_{zr} = \zeta_r$ and $\mathbf{I}_r = \mathbf{I}_l + (\zeta_r - I_{zl})\boldsymbol{\delta}$.

If compression has ended by $I_z = I_{zl}$, the energy takes the form $E = E_l + \Phi_2(I_z) - \Phi_2(I_{zl})$ for $I_z \geq I_{zl}$. By setting $E = 0$ and substituting (46) in, we obtain the quadratic equation $\Phi_2(I_z) + E_l - \Phi_2(I_{zl}) = 0$. Solve the equation for $\zeta_r$, the smallest root greater than $I_{zc}$. Again, two cases arise:

**Case 3 ($I_{zs} \leq \zeta_r$)** This yields the event sequence *clsr*. See Fig. 10(c) for an impact instance. Let $E_s = E_l + \Phi_2(I_{zs}) - \Phi_2(I_{zl})$ and set $\mathbf{I}_s$ according to (47). Determine $\mathbf{I}_r$ as in Case 2 in Section 6.2.

**Case 4 ($I_{zs} > \zeta_r$)** This yields the event sequence *clr* in Fig. 9(h). The sliding velocity will never decrease to zero, as with the impact instance in Fig. 4(c). Let $I_{zr} = \zeta_r$ and $\mathbf{I}_r = \mathbf{I}_l + (\zeta_r - I_{zl})\boldsymbol{\delta}$.

**Algorithm 1** Impulse Computation

```
1:  evaluate v^- according to (2)
2:  I ← 0
3:  evaluate B and d according to (21) and (22)
4:  while (I_z = 0 or E ≠ 0) and neither of events s and l happens do
5:      reset the step size h for I_z according to (36)
6:      update γ via one-step integration of (31) over I_z
7:      update I according to (26)
8:      update v_z according to (10)
9:      update E via one-step integration of (11)
10:     if v_z = 0 (compression ends) then
11:         I_zc ← I_z
12:         E_c ← E
13:         E ← e^2 E
14:     end if
15: end while
16: if E = 0 then
17:     return I
18: else
19:     set σ according to (38)
20:     if event s occurs then
21:         I_s ← I_z ẑ − (x̂, ŷ) B^{-1} (γ^- + I_z d)
22:         determine I_zr as in Section 6.2
23:         return I_s + (I_zr − I_zs) σ
24:     else
25:         determine I_zs, I_s (if needed) and I_zr as in Section 6.3
26:         if event s occurs then
27:             return I_s + (I_zr − I_zs) σ
28:         else
29:             set δ according to (44)
30:             return I_l + (I_zr − I_zl) δ
31:         end if
32:     end if
33: end if
```

### 6.4 Impact Algorithm and Execution Examples

Algorithm 1 computes the total impulse $\mathbf{I}_r$. The while loop of lines 4–15 performs numerical integration of (31) to track the sliding velocity $\boldsymbol{\gamma}$. Lines 21–23 are executed when event $s$ occurs. Lines 25–31 are executed when event $l$ occurs.

The algorithm computes the icosahedron-tetrahedron impact in Fig. 3 with the pre-impact velocities (14), and another impact between a pin and a ball used in bowling (hypothetically without the floor support). The second impact, with its configuration specified in Fig. 11, has the inverse inertia matrix

$$W = \begin{pmatrix} 1.55465 & 0.00238675 & -0.360808 \\ 0.00238676 & 5.15191 & -0.00313296 \\ -0.360808 & -0.00313296 & 1.105 \end{pmatrix}.$$

The pre-impact velocities of the pin and ball are

$$\mathbf{V}_1^- = (-0.1, 0.2, -0.5)^T, \quad \boldsymbol{\omega}_1^- = (0.1, 0.1, 0.1)^T,$$
$$\mathbf{V}_2^- = (0.1, 0.1, 1)^T, \quad \boldsymbol{\omega}_2^- = (0.1, -0.1, 0.1)^T.$$

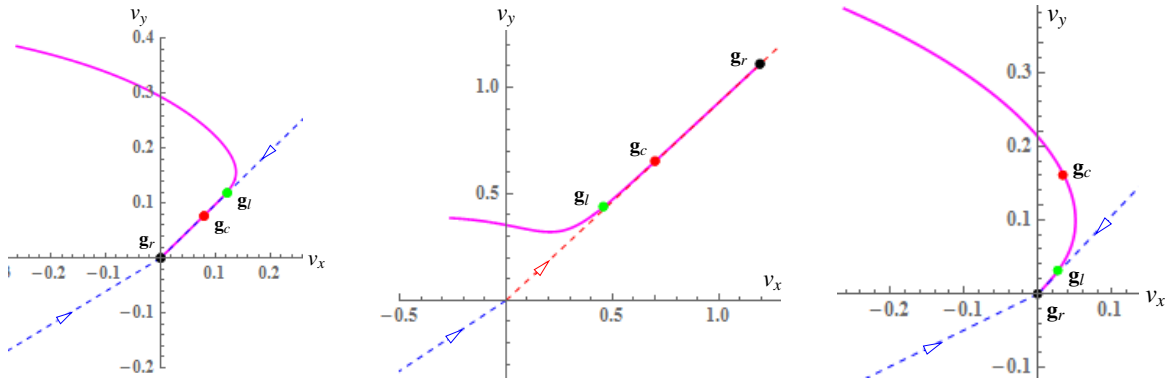



Fig. 10. Hodographs of the icosahedron-tetrahedron impact in Fig. 3 with the following changed parameter values: (a) $\mu = 0.35$ and $V_{1z}^- = 0.2$; (b) $\mu = 0.25$ and $V_{1z}^- = -1$; and (c) $\mu = 0.55$ and $V_{1z}^- = 0.9$.

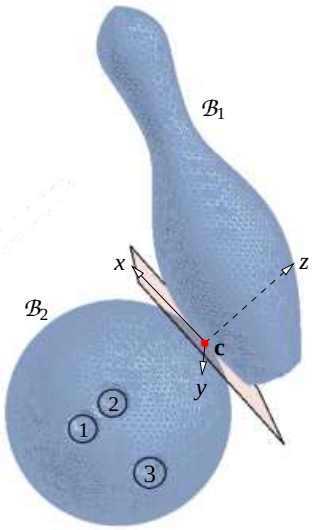



Fig. 11. Configuration of impact ($\mu = 0.7$ and $e = 0.7$) between a bowling pin ($\mathcal{B}_1$) and a ball ($\mathcal{B}_2$), both with uniform mass distribution. Here, $\mathbf{o}_1 = (0.068436, -0.00000242, 0.073735)^T$, $\mathbf{o}_2 = (-0.000061, -0.000890, -0.108458)^T$, $m_1 = 1.63293$, $m_2 = 6.35029$, $Q_1 = \text{diag}\ (0.0021186, 0.012092, 0.014029)^T$, and $Q_2 = \text{diag}(0.0295723,\ 0.029669, 0.030068)^T$. The shape of the pin is specified in `http://www.bowlingball.com/BowlVersity/bowling-pin-specification`. The pin is originally symmetric about the $z$-axis with its center of mass at the origin. To engage in contact, it rotates about the $y$-axis through $\tan^{-1}(3.15435)$ and then translates by $(0.068436, 0, 0.0737351)^T$. The ball has radius 0.108458. Its three holes are each created from subtracting a cylinder parallel to the $y$-axis. Holes 1 and 2 each has radius 0.0127, while hole 3 has radius 0.0142875. The bottom faces of the three cylinders are centered respectively at $(0.03302, 0.067845, -0.126238)^T$, $(0.03302, 0.067845, -0.090678)^T$, and $(-0.0381, 0.0676971, -0.108458)^T$.

For each impact, we obtain the "true value" of $\mathbf{I}_r$ using numerical integration with very small step size until closed-form evaluation becomes possible. They are given in (15) for the icosahedron-tetrahedron impact and as $(0.7330654, -0.0216048, 2.585018)^T$ for the pin-ball impact. Then we compare three different solutions: numerical integration (NI) with fixed step size and no use of closed-form evaluation even when it becomes possible, as conducted in previous works, Algorithm 1 using fixed step size of the normal impulse in integration (FS), and Algorithm 1 using adjustable integration step size (36) (VS) with $\varepsilon = 0.5$ and $h_1 = h_2 = 0.01$. The error of a method is measured as the norm of the difference between its obtained value of $\mathbf{I}_r$ and the "true value".

The code is written in Mathematica 10.3 and executed on a Lenovo Thinkpad X230i with Intel Core i3-3120M Processors and 4 GB system RAM. The results are shown in Table 1. On the icosahedron-tetrahedron impact, NI, FS, VS generate outcomes with comparable errors. VS is 74 times faster than NI and 50% faster than FS. On the pin-ball impact, VS achieves 25 times the accuracy of NI in 1/662 amount of time. It is also more than one time faster than FS with a slightly higher accuracy.

We have also compared the integration scheme VS with Mathematica's numerical solver `NDSolve` for ODEs. `NDSolve` is executed to solve the hodograph governing equation (31) for the same icosahedron-tetrahedron instance. Solution of this ODE is the most time consuming part of impact computation. Three different schemes of `NDSolve` are carried out separately: `Default` adaptive in both step size and integration method, `ExplicitRungeKutta` with adaptive step size, and `ExplicitRungeKutta` with a fixed step size. These three schemes, each executed in 1000 runs, respectively take average times of 0.00252722, 0.00368162, and 0.00316682 seconds. VS, which spends less than 0.0004134 second on integration as indicated in Table 1, is five to eight times faster than `NDSolve`, which also produces no better accuracies with its three schemes.

Use of a general ODE solver for impact solution must be able to track the contact velocity $\mathbf{v}_\perp$ during integration, and then switch to closed-form evaluation as soon as it reaches zero or assumes a constant direction. The above, if possible, would make the solver even slower. Hence, there would be no improvement to combine Algorithm 1 with a general ODE solver.

| | Icosa-Tetra | | | | Pin-Ball | | | |
|---|---|---|---|---|---|---|---|---|
| | "Truth" | NI | FS | VS | "Truth" | NI | FS | VS |
| #Step size | 1E-6 | 0.00095 | 0.001 | — | 1E-6 | 0.001 | 0.002 | — |
| #Iter. | 68212 | 107 | 69 | 29 | 165073 | 2586 | 83 | 29 |
| Time (s) | 18.9697 | 0.0312002 | 0.000624 | 0.0004134 | 46.9563 | 0.2652 | 0.000936 | 0.0004006 |
| Error | — | 0.000877 | 0.00108669 | 0.00111229 | —- | 0.00143 | 6.0469E-5 | 5.70441E-5 |

Table 1. Impact computation comparison from executing Algorithm 1 respectively with pure numerical integration (NI), integration with fixed step size coupled with closed-form evaluation (FS), and integration with adaptive step size coupled with close-form evaluation (VS).

## 7 Discussion

This paper describes a complete algorithm for solving the 3D frictional impact problem with energy-based restitution. The presented formal analysis takes an extra step from the existing inquiries. We have shown the positive definiteness of the inverse inertia matrix $W$ and the governing matrix $B$ of the hodograph evolution, and based on these properties, proved the existence and uniqueness of the direction for resumed contact sliding. A lower bound for the number of invariant sliding directions is formally established. Such directions were only argued to exist before.

Closed forms exist for the growth of the hodograph from the point where it assumes an invariant direction or reaches the origin. With a linear relationship (26) to the sliding velocity, the impulse accumulation also begins to assume a closed form, which makes possible a case-by-case analysis as that applicable to the two-dimensional impact.

Numerical integration is generally unavoidable for generating the hodograph from the beginning of the impact. By exploiting the hodograph's differential geometry, we have described an integration scheme to achieve some significant speed-up.

One challenge is to prove the existence (and uniqueness) of the impact solution under the contemporary scheme which combines Coulomb's friction model with either Poisson's impulse-based or Stronge's energy-based restitution. Such existence has only been assumed so far. A proof would ultimately verify the consistency between these hypotheses about two different phenomenons: friction and impact.

Another challenge, from the practical side, is how to predict whether the sliding velocity will reach zero or converge to an invariant direction, and if so, determine (or estimate) the normal impulse value at which this happens. Knowing this value holds the key to avoid numerical integration. Resolution of the issue would reduce the simulation time from tenths of a millisecond (e.g., 0.4 ms for the two impact instances in Table 1) to microseconds. This would make it possible to simulate collisions in large quantities at high speeds, affecting a number of applications areas such as high energy physics, chemistry, robotics, and movie and gaming industries.

Extension to fast computation of multi-body impacts also awaits investigation. Such an impact can be either sequenced into two-body impacts (Chatterjee and Ruina, 1998) or treated as simultaneous impacts interacting with each other (Liu et al. 2008; Jia et al. 2013). With the first approach, Algorithm 1 is directly applicable to solving individual two-body impacts. In the second approach, normal impulses at different contacts are related to each other differentially, while the normal impulse and the tangential impulse at each contact are also related differentially. It solves a system of ordinary differential equations driven successively by the normal impulses at some contacts. While the introduced fast integration scheme in Section 6.1 may be generalized to multi-body impacts, it is unlikely that closed forms of impulse accumulation will exist over any period.


### Acknowledgements

Support of this research was provided by the National Science Foundation through the grant IIS-1421034. Thanks to undergraduate student Jacob Stimes for the use of his code to evaluate the angular inertia matrices of the objects in the impact instances. We would also like to thank the anonymous reviewers for their careful reading and valuable comments. Any opinions, findings and thoughts discussed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.


### Appendix A: Invariant Directions

In this section, we show that generally there exist up to four unit vectors that will either preserve or reverse their directions under an affine mapping defined by a diagonal, positive definite matrix. The result is used in the proof of Theorem 5 in Section 4.

**Theorem 8.** *Let* $\Sigma = \mathrm{diag}(\lambda_1, \lambda_2)$ *with* $\lambda_1, \lambda_2 > 0$*, and* $\mathbf{p} = (p_x, p_y)^T$ *be a vector. Consider the equation*

$$\hat{\mathbf{u}} \times (\Sigma \hat{\mathbf{u}} + \mathbf{p}) = \mathbf{0} \tag{48}$$

*in the unit vector* $\hat{\mathbf{u}}$*. The following hold:*

*(i) When* $\lambda_1 = \lambda_2$*, (48) is satisfied by any unit vector* $\hat{\mathbf{u}}$ *if* $\mathbf{p} = \mathbf{0}$*, and by* $\hat{\mathbf{u}} = \pm \mathbf{p}/\|\mathbf{p}\|$ *if* $\mathbf{p} \neq \mathbf{0}$*.*

*(ii) When* $\lambda_1 \neq \lambda_2$ *and* $p_x = 0$*, (48) is satisfied by the two vectors* $(0, \pm 1)$*, and if* $p_y^2 < (\lambda_1 - \lambda_2)^2$*, also by the fol-*

*lowing two vectors:*

$$\left(\pm\sqrt{1-\left(\frac{p_y}{\lambda_1-\lambda_2}\right)^2},\frac{p_y}{\lambda_1-\lambda_2}\right)^T. \qquad (49)$$

*(iii) When $\lambda_1 \neq \lambda_2$ and $p_y = 0$, (48) is satisfied by the two vectors $(\pm 1, 0)$, and if $p_x^2 < (\lambda_1-\lambda_2)^2$, also by the following two vectors:*

$$\left(\frac{p_x}{\lambda_1-\lambda_2},\pm\sqrt{1-\left(\frac{p_x}{\lambda_1-\lambda_2}\right)^2}\right)^T.$$

*(iv) When $\lambda_1 \neq \lambda_2$, $p_x \neq 0$, and $p_y \neq 0$, (48) is satisfied by two to four vectors.*

*(v) If $\|\Sigma^{-1}\mathbf{p}\| \leq 1$, every $\hat{\mathbf{u}}$ satisfying (48) must be in the direction of $\Sigma\hat{\mathbf{u}}+\mathbf{p}$ unless the latter is zero.*

*(vi) If $\|\Sigma^{-1}\mathbf{p}\| > 1$, at least one $\hat{\mathbf{u}}$ satisfying (48) is in the direction of $\Sigma\hat{\mathbf{u}}+\mathbf{p}$, and exactly one such $\hat{\mathbf{u}}$ in the opposite direction of $\Sigma\hat{\mathbf{u}}+\mathbf{p}$.*

*Proof.* Part (i) is trivial to establish. We begin with proving part (ii). Then, part (iii) will follow by symmetry. Suppose $p_x = 0$. Substitute $\hat{\mathbf{u}} = (\cos\theta, \sin\theta)^T$ into (48):

$$\cos\theta\Big((\lambda_2-\lambda_1)\sin\theta + p_y\Big) = 0. \qquad (50)$$

The above equation holds if either $\cos\theta = 0$ or, otherwise, $(\lambda_2-\lambda_1)\sin\theta + p_y = 0$. The former condition implies $\hat{\mathbf{u}} = (0,\pm 1)^T$. The latter condition (plus $\cos\theta \neq 0$) implies that no other $\hat{\mathbf{u}}$ satisfies (50) if $p_y^2 > (\lambda_1-\lambda_2)^2$, and the two unit vectors given in (49) also satisfy it if $p_y^2 < (\lambda_1-\lambda_2)^2$. (Note $p_y^2 \neq (\lambda_1-\lambda_2)^2$ in order to satisfy $(\lambda_2-\lambda_1)\sin\theta + p_y = 0$ under $\cos\theta \neq 0$, i.e., $\sin\theta \neq \pm 1$.)

Let us now establish (iv). It is easy to show $\mathbf{u} \neq (-1,0)^T$ to satisfy (48) under $p_y \neq 0$. Substitute the bijective mapping $\hat{\mathbf{u}} = ((1-\eta^2)/(1+\eta^2), 2\eta/(1+\eta^2))$ over $(-\infty,\infty)$ into (48), and multiply away the denominators in the resulting equation. We end up with a quartic polynomial equation in $\eta$:

$$-p_y\eta^4 + 2(\lambda_1-\lambda_2-p_x)\eta^3 + 2(\lambda_2-\lambda_1-p_x)\eta + p_y = 0.$$

The above equation has 0, 2, or 4 real roots with multiplicities counted. Each distinct root corresponds to a unique unit vector $\hat{\mathbf{u}}$ satisfying (48). Below we show that there exist at least two such unit vectors.

Note that the mapping $\phi : \hat{\mathbf{u}} \mapsto \Sigma\hat{\mathbf{u}} + \mathbf{p}$ maps the unit circle $\mathcal{C}$ to an ellipse $\mathcal{E}$ centered at $\mathbf{p}$. The origin lies inside, on, or outside the ellipse if and only if it does the same with the image of the unit circle under the mapping $\hat{\mathbf{u}} \mapsto \hat{\mathbf{u}} + \Sigma^{-1}\mathbf{p}$. It is easy to show that the latter happens if and only if $\|\Sigma^{-1}\mathbf{p}\|$ is less than, equal to, or greater than 1, respectively. The three case are considered separately below.

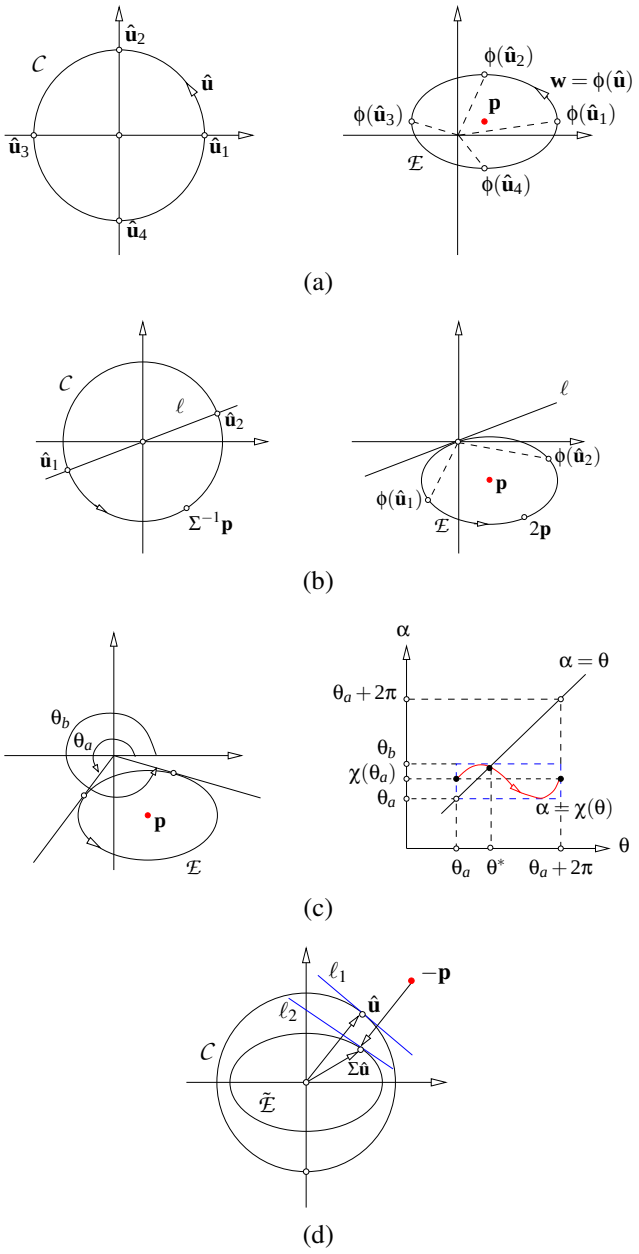



Fig. 12. Proofs of part (iv) of Theorem 8: (a) $\|\Sigma^{-1}\mathbf{p}\| < 1$, (b) $\|\Sigma^{-1}\mathbf{p}\| = 1$, (c) $\|\Sigma^{-1}\mathbf{p}\| > 1$, and (d) of part (v) of the theorem.

**Case 1** ($\|\Sigma^{-1}\mathbf{p}\| < 1$) Without loss of generality, we suppose that the center $\mathbf{p}$ of the image ellipse $\mathcal{E}$ lies in the first quadrant, as shown in Fig. 12(a). On the unit circle $\mathcal{C}$, let $\hat{\mathbf{u}}_i$, $1 \leq i \leq 4$, be its intersections of $\mathcal{C}$ with the coordinate axes. Picture a unit vector $\hat{\mathbf{u}}$ rotating counterclockwise from $\hat{\mathbf{u}}_1$ to $\hat{\mathbf{u}}_2$. Its polar angle varies across the interval $[0, \frac{\pi}{2}]$. Simultaneously, the image point $\mathbf{w} = \phi(\hat{\mathbf{u}})$ moves across an arc on $\mathcal{E}$ that subtends a polar angle interval properly contained in $[0, \frac{\pi}{2}]$. Because the movements of $\hat{\mathbf{u}}$ and $\mathbf{w}$ start and end simultaneously, $\hat{\mathbf{u}}$ and $\mathbf{w}$ must achieve the same polar angle at some point $\hat{\mathbf{u}}^*$, subsequently, $\hat{\mathbf{u}}^* \times \phi(\hat{\mathbf{u}}^*) = \mathbf{0}$.
Meanwhile, the polar angle range $[\pi, \frac{3}{2}\pi]$ of the third quarter of $\mathcal{C}$ between $\hat{\mathbf{u}}_3$ and $\hat{\mathbf{u}}_4$ is a proper subset of

that subtended by its image arc on $\mathcal{E}$ between $\phi(\hat{\mathbf{u}}_3)$ and $\phi(\hat{\mathbf{u}}_4)$. A similar reasoning establishes the existence of another $\hat{\mathbf{u}}$ on the third quarter of $\mathcal{C}$ to satisfy (48).

**Case 2** ($\|\Sigma^{-1}\mathbf{p}\| = 1$) See Fig. 12(b). There exists a $\mathbf{u}$ which maps to the origin, and trivially satisfies (48). Below we show that a second $\mathbf{u}$ exists. Let $\ell$ be the line through the origin and tangent to the image ellipse $\mathcal{E}$. This line intersects the unit circle at $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$, which are labeled in the way such that a counterclockwise traversal from $\hat{\mathbf{u}}_1$ to $\hat{\mathbf{u}}_2$ will pass through $\Sigma^{-1}\mathbf{p}$, which is mapped to the point $2\mathbf{p}$ on $\mathcal{E}$. As a point $\hat{\mathbf{u}}$ moves from $\hat{\mathbf{u}}_1$ to $\Sigma^{-1}\mathbf{p}$ and to $\hat{\mathbf{u}}_2$, its image point $\phi(\hat{\mathbf{u}})$ moves on $\mathcal{E}$ from $\phi(\hat{\mathbf{u}}_1)$ to $2\mathbf{p}$ to $\phi(\hat{\mathbf{u}}_2)$.[9] Since the traversed half circle (between $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$) and $\mathcal{E}$ are on the same side of $\ell$, the polar angle range subtended by the half circle must contain that subtended by the elliptic arc between $\phi(\hat{\mathbf{u}}_1)$ and $\phi(\hat{\mathbf{u}}_2)$. As both traversals start simultaneous and end simultaneously, there exists some $\hat{\mathbf{u}}^*$ during the traversal that is in the same direction as $\phi(\hat{\mathbf{u}}^*)$. Thus, $\hat{\mathbf{u}}^*$ satisfies (48).

**Case 3** ($\|\Sigma^{-1}\mathbf{p}\| > 1$) The origin lies outside the image ellipse $\mathcal{E}$. This is illustrated in Fig. 12(c). Let $\theta_a$ and $\theta_b$ be the polar angles of the two points of tangency from the origin to $\mathcal{E}$ such that $0 \le \theta_a < 2\pi$ and $\theta_a < \theta_b < \theta_a + \pi$. Consider a point $\hat{\mathbf{u}}$ starting at $(\cos\theta_a, \sin\theta_a)^T$ on the unit circle $\mathcal{C}$ to complete a full counterclockwise rotation. As its polar angle $\theta$ increases from $\theta_a$ to $\theta_a + 2\pi$, the polar angle $\chi(\theta)$ of the image point varies in the range $[\theta_a, \theta_b]$. In the $\theta$-$\alpha$ plane shown in Fig. 12(c), the curve $\alpha = \chi(\theta)$ starts at $(\theta_a, \chi(\theta_a))$, where $\chi(\theta_a) \ge \theta_a$, and ends at $(\theta_a + 2\pi, \chi(\theta_a))$, where

$$\theta_a + 2\pi > \theta_a + \pi > \theta_b \ge \chi(\theta_a).$$

Since its two endpoints are on the opposite sides of the line $\alpha = \theta$, the curve must cross the line at some point $(\theta^*, \theta^*)$, implying $\chi(\theta^*) = \theta^*$. Then $\hat{\mathbf{u}}^* = (\cos\theta^*, \sin\theta^*)^T$ is in the same direction as $\phi(\hat{\mathbf{u}}^*)$, and thus, satisfies (48).

Since $\|\Sigma^{-1}(-\mathbf{p})\| > 1$, it follows from Theorem 12 that there also exists a unique $\hat{\mathbf{u}}$ that is in the direction of $-\Sigma\hat{\mathbf{u}} - \mathbf{p}$, and therefore also satisfies (48).

Moving on, we now prove part (v). Suppose $\|\Sigma^{-1}\mathbf{p}\| \le 1$. The origin must be inside or on the image ellipse $\mathcal{E}$. Also, suppose that $\hat{\mathbf{u}}$ satisfies (48) with $\Sigma\hat{\mathbf{u}} + \mathbf{p} \ne \mathbf{0}$, and without loss of generality, is in the first quadrant. Fig. 12(d) displays $\hat{\mathbf{u}}$ on the unit circle $\mathcal{C}$. Let $\tilde{\mathcal{E}}$ be the image of $\mathcal{C}$ under the mapping $\hat{\mathbf{u}} \mapsto \Sigma\hat{\mathbf{u}}$. Since $\Sigma$ is diagonal and positive definite, $\Sigma\hat{\mathbf{u}}$ must be in the first quadrant on $\tilde{\mathcal{E}}$. Therefore, the polar angles of $\hat{\mathbf{u}}$ and $\Sigma\hat{\mathbf{u}}$ has a difference in the range $(-\frac{\pi}{2}, \frac{\pi}{2})$. The lines $\ell_1$ and $\ell_2$ are tangent to $\mathcal{C}$ and $\tilde{\mathcal{E}}$ at $\hat{\mathbf{u}}$ and $\Sigma\hat{\mathbf{u}}$, respectively. They can be aligned with each other under a rotation in the range $(-\frac{\pi}{2}, \frac{\pi}{2})$ as well.

Since $\hat{\mathbf{u}} \times (\Sigma\hat{\mathbf{u}} + \mathbf{p}) = \mathbf{0}$, the non-zero vector $\Sigma\hat{\mathbf{u}} + \mathbf{p}$ is either in the direction of $\hat{\mathbf{u}}$ or in its opposite direction. Suppose that $\hat{\mathbf{u}}$ and $\Sigma\hat{\mathbf{u}} + \mathbf{p}$, the vector from $-\mathbf{p}$ to $\Sigma\hat{\mathbf{u}}$, are in opposite directions. We will arrive at a contradiction in two steps below.

1. The points $-\mathbf{p}$ and $\Sigma\hat{\mathbf{u}}$ must be on different sides of $\ell_2$. Suppose that instead $-\mathbf{p}$ is either on the same side of $\ell_2$ with $\tilde{\mathcal{E}}$ or on $\ell_2$. Then, $\Sigma\hat{\mathbf{u}} + \mathbf{p}$ can be rotated through an angle in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$ to be aligned with the line $\ell_2$. Meanwhile, $\hat{\mathbf{u}}$ can be rotated through either $-\frac{\pi}{2}$ or $\frac{\pi}{2}$ to be aligned with the line $\ell_1$. These two lines differ by an rotation in $(-\frac{\pi}{2}, \frac{\pi}{2})$. The difference between the polar angles of $\hat{\mathbf{u}}$ and $\Sigma\hat{\mathbf{u}} + \mathbf{p}$ must be in the range $(-\pi, \pi)$. Because the range is an open interval, these two vectors cannot be opposite to each other. Hence a contradiction. So $-\mathbf{p}$ and $\Sigma\hat{\mathbf{u}}$ are on different sides of $\ell_2$.
2. Consequently, $-\mathbf{p}$ must be outside the ellipse $\tilde{\mathcal{E}}$. Equivalently, the origin (translation of the point $-\mathbf{p}$ by $\mathbf{p}$) must be outside the image ellipse $\mathcal{E}$ (translation of $\tilde{\mathcal{E}}$ by $\mathbf{p}$). But this is impossible under $\|\Sigma^{-1}\mathbf{p}\| \le 1$ as argued earlier. Another contradiction. The vectors $\hat{\mathbf{u}}$ and $\Sigma\hat{\mathbf{u}} + \mathbf{p}$ cannot oppose each other.

Finally, we prove part (vi). The existence of a unit vector $\hat{\mathbf{u}}$ in the direction of $\Sigma\mathbf{u} + \mathbf{p}$ follows from (i), (ii) (either $(0,1)$ or $(0,-1)$), (iii) (either $(1,0)$ or $(-1,0)$), and Case 3 of (iv). Under $\|\Sigma^{-1}(-\mathbf{p})\| > 1$, Theorem 12, to be established in Appendix B, will imply that there exists a unique $\hat{\mathbf{u}}$ in the direction of $-\Sigma\hat{\mathbf{u}} - \mathbf{p}$.

## Appendix B: Centrifugal Invariant Direction

Let $\Sigma$ be a diagonal $2 \times 2$ matrix with positive diagonal entries, and $\mathbf{p}$ be a vector such that $\|\Sigma^{-1}\mathbf{p}\| > 1$. In this appendix, we will prove that there exists a unique unit vector $\hat{\mathbf{u}} \in \mathbb{R}^2$ such that $\hat{\mathbf{u}} \sim -\Sigma\hat{\mathbf{u}} + \mathbf{p}$ (that is, both vectors are non-zero and in the same direction).

Let us introduce two mappings over $\mathbb{R}^2$:

$$\alpha : \mathbf{u} \mapsto \mathbf{w} = -\Sigma\mathbf{u} \quad \text{and} \quad \beta : \mathbf{w} \mapsto \mathbf{s} = \mathbf{w} + \mathbf{p}.$$

Clearly, $(\beta \circ \alpha)(\hat{\mathbf{u}}) = -\Sigma\hat{\mathbf{u}} + \mathbf{p}$. As shown in Fig. 13, due to the negative definiteness of $-\Sigma$, $\alpha$ maps the unit circle $\mathcal{C}$ to an ellipse $\mathcal{E}_1$ centered at $\mathbf{p}$, symmetric about the $x$- and $y$-axes, and with semimajor and semiminor axes equal to the diagonal elements of $\Sigma$. A point on $\mathcal{C}$ is mapped to one on the ellipse in the opposite quadrant. Under $\beta$, the ellipse $\mathcal{E}_1$ is translated by $\mathbf{p}$ to become the ellipse $\mathcal{E}_2$ centered at $\mathbf{p}$. We reason that $\mathcal{E}_2$ does not contain the origin as follows. This ellipse contains the origin if and only if the image ellipse $\mathcal{E}_3$ of $\mathcal{C}$ under the mapping $\mathbf{u} \mapsto -\mathbf{u} + \Sigma^{-1}\mathbf{p}$ does. But $\mathcal{E}_3$ does not because $\|\Sigma^{-1}\mathbf{p}\| > 1$.

As shown in Fig. 14(a), from the origin $\mathbf{o}$ there are two rays tangent to $\mathcal{E}_2$ at $\mathbf{s}_l$ and $\mathbf{s}_r$, respectively. The rays form the left and right edges (viewed from $\mathbf{o}$) of a cone in which $\mathcal{E}_2$ is inscribed. Clearly, the apex angle of the cone is less than $\pi$ because the ellipse is convex and the origin $\mathbf{o}$ lies outside of it. The points $\mathbf{s}_l$ and $\mathbf{s}_r$ split the ellipse into two arcs: $\mathcal{E}_2^+$ inside the triangle $\triangle\mathbf{o}\mathbf{s}_l\mathbf{s}_r$ (including $\mathbf{s}_l$ and $\mathbf{s}_r$), and $\mathcal{E}_2^-$ outside it.

[9] Note that $\phi(\hat{\mathbf{u}}_1)$, $\mathbf{p}$, and $\phi(\hat{\mathbf{u}}_2)$ are collinear because $\hat{\mathbf{u}}_1$, the origin $\mathbf{o}$, and $\hat{\mathbf{u}}_2$ are, given that an affine mapping preserves collinearity.

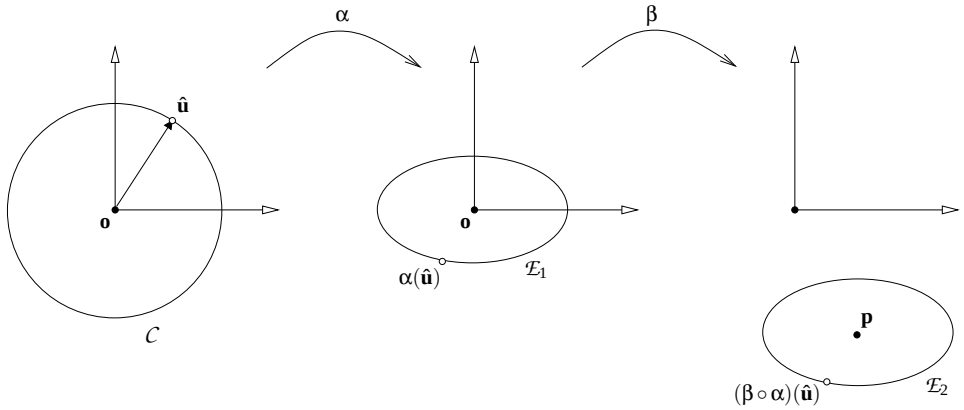
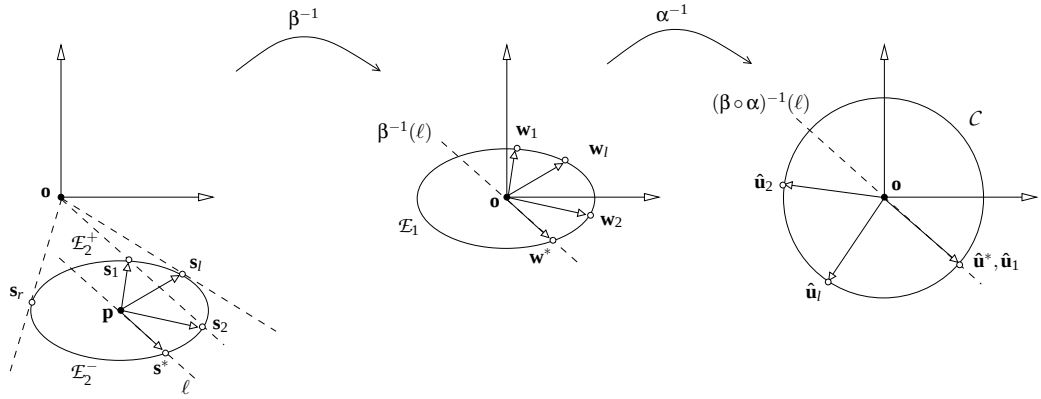



Fig. 13. Unit circle $\mathcal{C}$ maps to an ellipse $\mathcal{E}_2$ centered at $\mathbf{p}$ under $\beta \circ \alpha$.




(a) (b) (c)

Fig. 14. Inverse mapping sequence $\alpha^{-1} \circ \beta^{-1}$.

**Lemma 9.** *If a unit vector* $\hat{\mathbf{u}}^*$ *satisfies* $\hat{\mathbf{u}}^* \sim (\beta \circ \alpha)(\hat{\mathbf{u}}^*)$, *then* $(\beta \circ \alpha)(\hat{\mathbf{u}}^*)$ *must lie on the elliptic arc* $\mathcal{E}_2^+$.

*Proof.* Suppose $\hat{\mathbf{u}}^* \sim (\beta \circ \alpha)(\hat{\mathbf{u}}^*)$. The line through the origin $\mathbf{o}$ in the direction of $\hat{\mathbf{u}}^*$ intersects the ellipse $\mathcal{E}_2$ at up to two points $\mathbf{s}_1$ and $\mathbf{s}_2$. See Fig. 14(a). If $\mathbf{s}_1 = \mathbf{s}_2$, then both must coincide with $(\beta \circ \alpha)(\hat{\mathbf{u}}^*)$, which must be either $\mathbf{s}_l$ or $\mathbf{s}_r$, and thus lies on $\mathcal{E}_2^+$. The statement is thus true.

Assume $\mathbf{s}_1 \neq \mathbf{s}_2$. Permute the indices such that $\mathbf{s}_1$ is closer to $\mathbf{o}$ than $\mathbf{s}_2$. Either $\mathbf{s}_1$ or $\mathbf{s}_2$ must be $(\beta \circ \alpha)(\hat{\mathbf{u}}^*)$. Let $\mathbf{s}^*$ be the point on $\mathcal{E}_2$ such that $\mathbf{s}^* - \mathbf{p} \sim \hat{\mathbf{u}}^*$, and $\ell$ the line through $\mathbf{p}$ and $\mathbf{s}^*$ (see Fig. 14(a)). The points $\mathbf{s}_1$ and $\mathbf{s}_2$ lie either on the side of $\ell$ containing $\mathbf{s}_l$ or on the other side of the line containing $\mathbf{s}_r$. Without loss of generality, we assume that $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_l$ are on the same side of $\ell$. A clockwise traversal of the half ellipse on this side visits $\mathbf{s}_1$, $\mathbf{s}_l$, $\mathbf{s}_2$, and $\mathbf{s}^*$ sequentially.

The inverse mapping $\beta^{-1}$ takes $\mathcal{E}_2$ back to $\mathcal{E}_1$, as shown in Fig. 14(b), with $\mathbf{p}$ translated to the origin $\mathbf{o}$, and $\mathbf{s}_1$, $\mathbf{s}_l$, $\mathbf{s}_2$, and $\mathbf{s}^*$ to $\mathbf{w}_1$, $\mathbf{w}_l$, $\mathbf{w}_2$, and $\mathbf{w}^*$, respectively. It follows from $\mathbf{s}^* - \mathbf{p} \sim \hat{\mathbf{u}}^*$ that $\mathbf{w}^* \sim \hat{\mathbf{u}}^*$. The order of traversal of the four points is preserved under the translation.

The image line $\beta^{-1}(\ell)$ splits the ellipse $\mathcal{E}_1$ into two halves. The half containing $\mathbf{w}_1$, $\mathbf{w}_l$, $\mathbf{w}_2$, and $\mathbf{w}^*$ is mapped under $\alpha^{-1}$ to a half unit circle shown in Fig. 14(c). Because $\alpha$ is defined by the negative definite matrix $-\Sigma$, a clockwise traversal along the half circle starting at $\hat{\mathbf{u}}^*$ will visit $\hat{\mathbf{u}}_1 = \alpha^{-1}(\mathbf{w}_1)$, $\hat{\mathbf{u}}_l = \alpha^{-1}(\mathbf{w}_l)$, and $\hat{\mathbf{u}}_2 = \alpha^{-1}(\mathbf{w}_2)$, sequentially. Since $\hat{\mathbf{u}}_1 \neq \hat{\mathbf{u}}_2$ and both on the same half circle starting at $\hat{\mathbf{u}}^*$ (coinciding with either $(\beta \circ \alpha)^{-1}(\mathbf{s}_1) = \hat{\mathbf{u}}_1$ or $(\beta \circ \alpha)^{-1}(\mathbf{s}_2) = \hat{\mathbf{u}}_2$), the only possibility is $\hat{\mathbf{u}}^* = \hat{\mathbf{u}}_1$, namely, $(\beta \circ \alpha)(\hat{\mathbf{u}}^*) = \mathbf{s}_1$, which lies on the elliptic arc $\mathcal{E}_1^+$.

In Fig. 15(a), picture a point $\mathbf{s}$ moving on the arc $\mathcal{E}_2^+$ from $\mathbf{s}_l$ to $\mathbf{s}_r$. The unit vector $\hat{\mathbf{s}} = \mathbf{s}/\|\mathbf{s}\|$ traces out an arc $\mathcal{A}$ on the unit circle $\mathcal{C}$ from $\hat{\mathbf{s}}_l = \mathbf{s}_l/\|\mathbf{s}_l\|$ to $\hat{\mathbf{s}}_r = \mathbf{s}_r/\|\mathbf{s}_r\|$ (see Fig. 15(b)). This arc is less than half of the circle. Its image $\mathcal{F} = (\beta \circ \alpha)(\mathcal{A})$ is accordingly less than half of $\mathcal{E}_2$. Let $\mathbf{q}_l$ and $\mathbf{q}_r$ be the endpoints of $\mathcal{F}$ such that $\mathbf{q}_l = (\beta \circ \alpha)(\hat{\mathbf{s}}_l)$ and $\mathbf{q}_r = (\beta \circ \alpha)(\hat{\mathbf{s}}_r)$.

**Lemma 10.** *There exists a unique unit vector* $\hat{\mathbf{u}}$ *such that* $\hat{\mathbf{u}} \sim -\Sigma\hat{\mathbf{u}} + \mathbf{p}$ *if and only if* $\mathcal{E}_2^+ \cap \mathcal{F} \neq \emptyset$.

*Proof.* ($\Rightarrow$) Suppose there exists a unique $\hat{\mathbf{u}}$ with $\hat{\mathbf{u}} \sim -\Sigma\hat{\mathbf{u}} + \mathbf{p}$. Since all the points on the ellipse $\mathcal{E}_2$, including $-\Sigma\hat{\mathbf{u}} + \mathbf{p}$, must lie inside the cone positively spanned by the vectors $\overrightarrow{\mathbf{os}_l}$ and $\overrightarrow{\mathbf{os}_r}$, $\hat{\mathbf{u}}$ must be on the circular arc $\mathcal{A}$ shown in Fig. 15(b). Let $\mathbf{s} = -\Sigma\hat{\mathbf{u}} + \mathbf{p}$. Apparently, $\mathbf{s}$ is on $\mathcal{F} = (\beta \circ \alpha)(\mathcal{A})$ in Fig. 15(c). Meanwhile, it follows from Lemma 9 that $\mathbf{s}$ is also on $\mathcal{E}_2^+$. Thus, $\mathcal{E}_2^+$ and $\mathcal{F}$ share at least one point $\mathbf{s}$.

($\Leftarrow$) Suppose $\mathcal{E}_2^+ \cap \mathcal{F} \neq \emptyset$. Picture a point $\mathbf{s}$ moving counterclockwise from $\mathbf{s}_l$ to $\mathbf{s}_r$ on $\mathcal{E}_2^+$ (cf. Fig. 15(a)). Simultaneously, the point $\mathbf{q} = (\beta \circ \alpha)(\hat{\mathbf{s}})$, where $\hat{\mathbf{s}} = \mathbf{s}/\|\mathbf{s}\|$, is

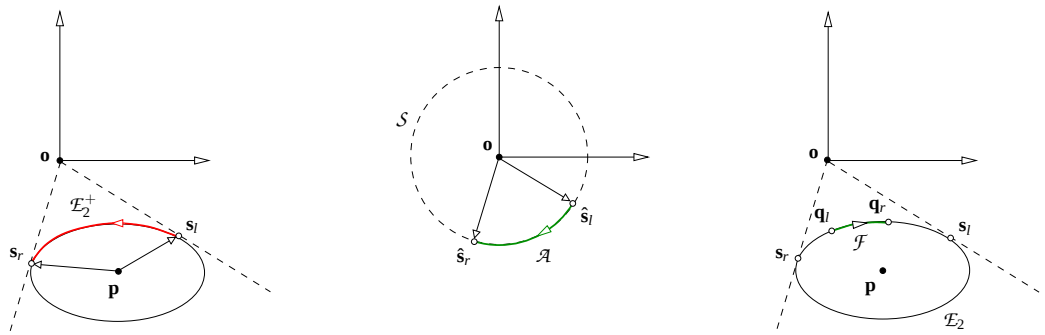



Fig. 15. The unit vectors $\hat{\mathbf{s}}_l$ and $\hat{\mathbf{s}}_r$ in the directions from $\mathbf{o}$ to the points of tangency $\mathbf{s}_l$ and $\mathbf{s}_r$ on $\mathcal{E}_2$ delimits an arc $\mathcal{A}$ on the unit circle that contains the point $\hat{\mathbf{u}}$ with $\hat{\mathbf{u}} \sim (\beta \circ \alpha)\hat{\mathbf{u}}$, if it exists.

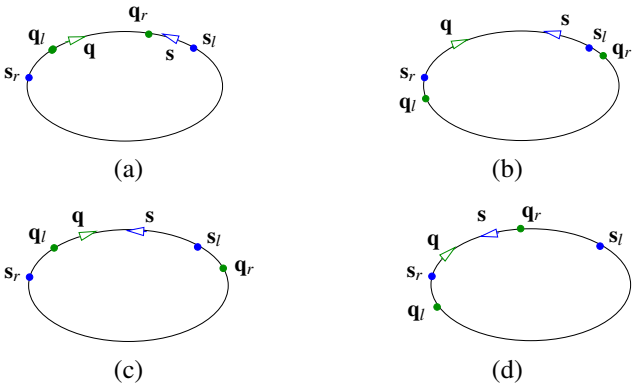



Fig. 16. Four possible cases (a)–(d) of overlapping between the elliptic segments $\mathcal{E}_2^+$ between $\mathbf{s}_l$ and $\mathbf{s}_r$, and $\mathcal{A}$ between $\mathbf{q}_l$ and $\mathbf{q}_r$.

moving clockwise from $\mathbf{q}_l$ to $\mathbf{q}_r$ due to the negative definiteness of $-\Sigma$. The two moving points will reach $\mathbf{s}_r$ and $\mathbf{q}_r$ at the same time. Both arcs $\mathcal{E}_2^+$ and $\mathcal{F}$ traversed by $\mathbf{s}$ and $\mathbf{q}$, respectively, are less than half of the ellipse. If $\mathbf{s}$ meets $\mathbf{q}$ during the movement, then $\mathbf{s} = \mathbf{q} = (\beta \circ \alpha)(\hat{\mathbf{s}})$, which implies $\hat{\mathbf{s}} \sim (\beta \circ \alpha)\hat{\mathbf{s}}$. We consider four cases of $\mathcal{E}_2^+ \cap \mathcal{F} \neq \emptyset$:

**Case 1** $\mathcal{F} \subseteq \mathcal{E}_2^+$ (Fig. 16(a)) The counterclockwise moving $\mathbf{s}$ needs to cross $\mathbf{q}_r$ and then $\mathbf{q}_l$ before reaching $\mathbf{s}_r$. At the time of $\mathbf{s}$ crossing $\mathbf{q}_r$, $\mathbf{q}$ must have not reached $\mathbf{q}_r$ yet. Therefore, $\mathbf{s}$ will meet the clockwise moving $\mathbf{q}$ on the arc $\mathcal{F}$ between $\mathbf{q}_l$ and $\mathbf{q}_r$.

**Case 2** $\mathcal{F} \supset \mathcal{E}_2^+$ (Fig. 16(b)) Similarly, the clockwise moving point $\mathbf{q}$ needs to cross $\mathbf{s}_r$ and $\mathbf{s}_l$ sequentially before reaching $\mathbf{q}_r$. Therefore, it will meet the counterclockwise moving $\mathbf{s}$ on the arc $\mathcal{E}_2^+$ between $\mathbf{s}_l$ and $\mathbf{s}_r$.

**Case 3** $\mathbf{q}_l \in \mathcal{E}_2^+$ but $\mathbf{q}_r \notin \mathcal{E}_2^+$ (Fig. 16(c)) The moving points $\mathbf{s}$ and $\mathbf{q}$ need to cross $\mathbf{q}_l$ and $\mathbf{s}_l$ to reach $\mathbf{s}_r$ and $\mathbf{q}_r$, respectively. They will be moving toward each other on the elliptic arc between $\mathbf{q}_l$ and $\mathbf{s}_l$, and thus will eventually meet.

**Case 4** $\mathbf{q}_l \notin \mathcal{E}_2^+$ but $\mathbf{q}_r \in \mathcal{E}_2^+$ (Fig. 16(d)) Similarly, $\mathbf{s}$ and $\mathbf{q}$ will be moving toward each other on the elliptic arc between $\mathbf{s}_r$ and $\mathbf{q}_r$. They will cross $\mathbf{q}_r$ and $\mathbf{s}_r$, respectively, to reach $\mathbf{s}_r$ and $\mathbf{q}_r$. Hence they will meet on the elliptic arc between $\mathbf{q}_r$ and $\mathbf{s}_r$ as well.

In all the four cases above, after $\mathbf{s}$ and $\mathbf{q}$ meet, they will continue moving in opposite directions to traverse less than half of the ellipse. Hence they will not meet again. The uniqueness of $\hat{\mathbf{u}}$ with $\hat{\mathbf{u}} \sim -\Sigma\hat{\mathbf{u}} + \mathbf{p}$ thus follows.

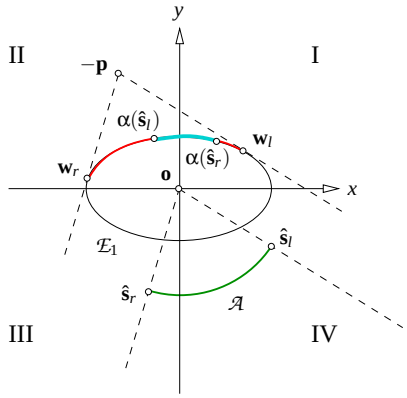
**Lemma 11.** *The arcs $\mathcal{E}_2^+$ and $\mathcal{F}$ intersect.*

*Proof.* Apply the inverse mapping $\beta^{-1}$ to the plane containing $\mathcal{E}_2$. The origin and the tangency points $\mathbf{s}_l$, $\mathbf{s}_r$ are translated to $-\mathbf{p}$, $\mathbf{w}_l = \mathbf{s}_l - \mathbf{p}$ and $\mathbf{w}_r = \mathbf{s}_r - \mathbf{p}$, respectively. The ellipse $\mathcal{E}_2$ is translated back to $\mathcal{E}_1$, the image of the unit circle $\mathcal{C}$ under $\alpha$, its arc $\mathcal{E}_2^+$ translated to $\mathcal{E}_1^+$, the arc between $\mathbf{w}_l$ and $\mathbf{w}_r$, and $\mathcal{F}$ to $\alpha(\mathcal{A})$. We need only establish $\mathcal{E}_1^+ \cap \alpha(\mathcal{A}) \neq \emptyset$.

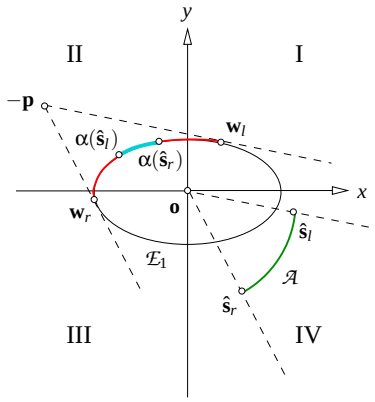
The arc $\mathcal{A}$ is Less than half of the unit circle $\mathcal{C}$ centered at the origin. It is contained in either two adjacent quadrants, or in just one quadrant.

**Case 1** The arc $\mathcal{A}$ is contained in two adjacent quadrants. Then, it must contain a half coordinate axis, say, without loss of generality, the negative $y$-axis as shown in Fig. 17(a). Its left endpoint $\hat{\mathbf{s}}_l$ and the right endpoint $\hat{\mathbf{s}}_r$ (viewed from the origin) must be in the quadrants IV and III, respectively. Their image points $\alpha(\hat{\mathbf{s}}_l) = -\Sigma\hat{\mathbf{s}}_l$ and $\alpha(\hat{\mathbf{s}}_r) = -\Sigma\hat{\mathbf{s}}_r$, given the negative definiteness of the diagonal matrix $-\Sigma$, must lie inside quadrants II and I opposite to quadrants IV and III, respectively. Hence the arc $\alpha(\mathcal{A})$ contains the point of $\mathcal{E}_1$ on the positive $y$-axis. Meanwhile, the tangent line through $-\mathbf{p}$ to $\mathcal{E}_1$ in the direction of $\hat{\mathbf{s}}_l$ (which is in quadrant IV) must be incident at a point on the ellipse in quadrant I. This is the right endpoint $\mathbf{w}_l$ of $\mathcal{E}_1^+$. Similarly, we infer that the left endpoint $\mathbf{w}_r$ must be in quadrant II. Thus, the arc $\mathcal{E}_1^+$ also contains the point of $\mathcal{E}_1$ on the positive $y$-axis, establishing $\mathcal{E}_1^+ \cap \alpha(\mathcal{A}) \neq \emptyset$.

**Case 2** The arc $\mathcal{A}$ is contained in one quadrant, say, quadrant IV, as shown in Fig. 17(b). Its image $\alpha(\mathcal{A})$ must be inside quadrant II. Similarly, based on the geometry of an ellipse we infer that the points of tangency $\mathbf{w}_l$ and $\mathbf{w}_r$ must lie inside quadrants I and III, respectively. Therefore, the elliptic arc $\mathcal{E}_1^+$ between $\mathbf{w}_l$ and $\mathbf{w}_r$ contains the quarter of the ellipse inside quadrant II, which contains $\alpha(\mathcal{A})$. We have $\mathcal{E}_1^+ \cap \alpha(\mathcal{A}) = \alpha(A) \neq \emptyset$.




(a)




(b)

Fig. 17. The arc $\mathcal{A}$ is contained in either (a) two adjacent quadrants or (b) just one quadrant.

Lemmas 10 and 11 lead to the following theorem.

**Theorem 12.** *Let $\Sigma$ be a diagonal $2\times 2$ matrix with positive diagonal elements, and $\mathbf{p}\in\mathbb{R}^2$ such that $\|\Sigma^{-1}\mathbf{p}\|>1$. Then there exists a unique unit vector $\hat{\mathbf{u}}\in\mathbb{R}^2$ such that $\hat{\mathbf{u}}\sim -\Sigma\hat{\mathbf{u}}+\mathbf{p}$.*

## References


[1] Newton, I., 1686. *Philosophiae Naturalis Principia Mathematica*. Royal Society Press, London.

[2] Poisson, S. D., 1827. "Note sur i'extension des fils et des plaques élastiques". *Annales de Chimie et de Physique,* **36**, pp. 384–387.

[3] Stronge, W. J., 1990. "Rigid body collisions with friction". *Proceedings of the Royal Society of London A,* **431**(1881), pp. 168–181.

[4] Stronge, W. J., 2000. *Impact Mechanics*. Cambridge University Press, Cambridge, UK.

[5] Cross, R., 2010. "Impact of a ball on a surface with tangential compliance". *American Journal of Physics,* **78**(7), pp. 716–720.

[6] Jia, Y.-B., 2013. "Three-dimensional impact: energy-based modeling of tangential compliance". *International Journal of Robotics Research,* **32**(1), pp. 56–83.

[7] Brach, R. M., 1989. "Rigid body collisions". *ASME Journal of Applied Mechanics,* **56**(1), pp. 133–137.

[8] Smith, C. E., 1991. "Predicting rebounds using rigid-body dynamics". *Transactions of the ASME,* **58**(3), pp. 754–758.

[9] Glocker, G., and Pfeiffer, F., 1995. "Multiple impacts with friction in rigid multibody systems". *Nonlinear Dynamics,* **7**(4), pp. 471–497.

[10] Stewart, D. E., 2000. "Rigid-body dynamics with friction and impact". *SIAM Review,* **42**(1), pp. 3–39.

[11] Chatterjee, A., and Ruina, A., 1998. "A new algebraic rigid-body collision law based on impulse space considerations". *ASME Journal of Applied Mechanics,* **65**(4), pp. 939–951.

[12] Routh, E. J., 1905. *Dynamics of a System of Rigid Bodies*. MacMillan and Co., London.

[13] Wang, Y., and Mason, M. T., 1992. "Two-dimensional rigid-body collisions with friction". *ASME Journal of Applied Mechanics,* **59**(3), pp. 635–642.

[14] Darboux, G., 1880. "Etude géométrique sur les percussions et le choc des corps". *Bulletin des Sciences Mathématiques et Astronomiques, deuxième série, tome* **4**(1), pp. 126–160.

[15] Keller, J. B., 1986. "Impact with friction". *ASME Journal of Applied Mechanics,* **53**(1), pp. 1–4.

[16] Zhao, Z., and Liu, C., 2007. "The analysis and simulation for three-dimensional impact with friction". *Multibody System Dynamics,* **18**(4), pp. 511–530.

[17] Zhang, Y., and Sharf, I., 2007. "Rigid body impact modeling using integral formulation". *ASME Journal of Computational and Nonlinear Dynamics,* **2**(1), pp. 98–102.

[18] Bhatt, V., and Koechling, J., 1995. "Three-dimensional frictional rigid-body impact". *ASME Journal of Applied Mechanics,* **62**(4), pp. 893–898.

[19] Batlle, J. A., 1996. "The sliding velocity flow of rough collisions in multibody systems". *ASME Journal of Applied Mechanics,* **63**(3), pp. 804–809.

[20] Elkaranshawy, H. A., 2007. "Rough collision in three-dimensional rigid multi-body systems". *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics,* **221**(4), pp. 541–550.

[21] Kane, T. R., and Levinson, D. A., 1985. *Dynamics: Theory and Applications*. McGraw-Hall, New York.

[22] Jia, Y.-B., Gardner, M., and Mu, X., 2016. Batting flying objects to targets in two dimensions. Manuscript under preparation.

[23] Pressley, A., 2001. *Elementary Differential Geometry*. Springer-Verlag, London.

[24] Liu, C., Zhao, Z., and Brogliato, B., 2008. "Frictionless multiple impacts in multibody systems. I. Theoretical framework". *Proceedings of the Royal Society of London A*, **464**(2100): 3193–3211.

[25] Jia Y.-B., Mason, M., and Erdmann, M., 2013. "Multiple impacts: A state transition diagram approach". *International Journal of Robotics Research,* **32**(1), pp. 84–114.