SIAM J. Applied Dynamical Systems
Vol. 2, No. 3, pp. 381–416

Nonsmooth Lagrangian Mechanics and Variational Collision Integrators*

R. C. Fetecau†, J. E. Marsden‡, M. Ortiz§, and M. West¶

Abstract. Variational techniques are used to analyze the problem of rigid-body dynamics with impacts. The theory of smooth Lagrangian mechanics is extended to a nonsmooth context appropriate for collisions, and it is shown in what sense the system is symplectic and satisfies a Noether-style momentum conservation theorem.

Discretizations of this nonsmooth mechanics are developed by using the methodology of variational discrete mechanics. This leads to variational integrators which are symplectic-momentum preserving and are consistent with the jump conditions given in the continuous theory. Specific examples of these methods are tested numerically, and the long-time stable energy behavior typical of variational methods is demonstrated.

Key words. discrete mechanics, variational integrators, collisions

AMS subject classifications. 37M15, 70F35, 58E30

DOI. 10.1137/S1111111102406038

1. Introduction.

In this paper, we investigate nonsmooth Lagrangian mechanics and its discretization by means of variational, numerical, and geometric methods. In particular, we are interested in the problem of rigid-body collisions, for which the velocity, acceleration, and forces are all nonsmooth or even discontinuous.

We shall begin with a survey of some history and literature to put our own work into context. The literature and history is of course quite complex with many points of view, so we focus on selected highlights only.

History and literature: Theory. The problem of collisions has been extensively treated in the literature since the early days of mechanics. More recently, much work has been done on the rigorous mathematical foundation of impact problems, in particular, by generalizing Newton's law to include forces which are measure-valued and hence can include impulses at the point of impact. The contact dynamics is thus governed by a *measure differential inclusion*, a general

*Received by the editors April 23, 2002; accepted for publication (in revised form) by M. Golubitsky May 1, 2003; published electronically August 23, 2003.

http://www.siam.org/journals/siads/2-3/40603.html

†Applied Mathematics 217-50, California Institute of Technology, Pasadena, CA 91125 (van@acm.caltech.edu). The research of this author was partially supported by the California Institute of Technology and the National Science Foundation.

‡Control and Dynamical Systems 107-81, California Institute of Technology, Pasadena, CA 91125 (marsden@cds.caltech.edu). The research of this author was partially supported by the California Institute of Technology and the National Science Foundation.

§Graduate Aeronautical Laboratories 105-50, California Institute of Technology, Pasadena, CA 91125 (ortiz@aero.caltech.edu).

¶Department of Mathematics, University of California Davis, One Shields Avenue, Davis, CA 95616 (mwest@math.ucdavis.edu).

formulation that can directly incorporate impulsive forces and nonsmooth solutions. In this context, a measure differential inclusion has the form

$$\frac{dv}{dt} \in F(t,x), \ \frac{dx}{dt} = g(t,x,v),$$

where $v(t)$ and $x(t)$ denote the velocity and the position, F is a set-valued function, and $v(\cdot)$ is required only to have bounded variation.

The extension of the concept of a differential equation to that of a *differential inclusion* was first considered in [12, 13, 14]. These works provide a deep study of ODEs with a discontinuous right-hand side, but the fact that solutions are required to be continuous in the phase space makes the theory inapplicable to collisions. Measure differential inclusions can be found in different contexts in [51, 52], and the use of this concept in rigid-body dynamics was further developed in [38, 39], where the (unilateral) contact between rigid bodies received a formulation (called by the author a *sweeping process*) that combines differential inclusions with convex analysis. Since then, an extensive literature has been devoted to the theoretical and numerical study of nonsmooth dynamics within the mathematical framework of measure differential inclusions.

Substantial progress has been made in the last two decades on the existence and uniqueness theory for the generalized solutions of rigid-body dynamics. The first rigorous results in this area were produced in [33] for the case of an inelastic collision with a single convex constraint. Further results generalized the existence theory to more general contacts in [42, 43], to more general (nonconvex, but of class C^1) constraints in [34], or even to a less regular constraint for an arbitrary frictionless impact in [32]. The recent works [55, 56] consider the impact dynamics with friction and give a rigorous mathematical solution to the famous problem of Painlevé.

In the same elegant framework of differential inclusions, but oriented toward the control and stability of nonsmooth dynamical systems, we mention the works of Brogliato [6, 7] and Brogliato, Niculescu, and Orhant [5].

History and Literature: Computations. The measure differential inclusion has also been proved to be an excellent mathematical foundation for the study of numerical methods for discontinuous ODEs. It is not within our scope to give a complete account of these methods, but we refer the reader to the excellent overviews of numerical methods for differential inclusions in [11] and [30]. In particular, such numerical approaches have been pursued to develop efficient numerical methods for rigid-body dynamics in the sweeping process formalism in [38, 40, 44, 56].

Various other numerical methods for rigid-body systems have been studied extensively in the engineering and mathematics literature. We refer to the excellent book [47] for a comprehensive account of some of these methods. We particularly note the approach that reduces the contact to a *complementarity problem*, a concept frequently used in constrained optimization, to decide at each step which constraints are active.

However, most existing practical codes are based on smoothing techniques, a class of methods which use a penalty formulation to regularize the problem. This approach relies on the definition of a proper gap function as a means to detect and penalize the interpenetration; see, for example, [54, 8, 60, 57]. An obvious weakness of the penalty methods is that they

cannot handle collisions of irregularly shaped bodies (bodies with corners), where neither normals nor gap functions can be defined. An elegant solution to this problem is offered by the *nonsmooth analysis* approach from [25], where new robust contact algorithms are derived using the powerful tools of nonsmooth calculus (see [10]).

An important issue in contact dynamics is how to formulate physically correct friction models, and an extensive body of literature has addressed this problem. Frictional effects are generally accounted for by introducing a friction law (Coulomb's law is an example) which relates the sliding velocity to the contact forces. An alternative approach uses the *maximum dissipation principle*, where the friction force c_f is required to maximize the rate of energy dissipation $-c_f^T v_{rel}$, where v_{rel} is the relative velocity at the contact, out of all possible friction forces allowed by a given contact force c_n. However, the correct modeling of friction still has many open questions which generate controversy in various engineering and mathematical communities. All the various numerical methods for contact that we mentioned above have introduced friction in the dynamics, and we refer to [2] and [39] for measure differential inclusion methods, [23, 28, 45, 48, 58, 59, 46, 3] for the complementarity and gap function formulation, and [41] for the nonsmooth analysis approach.

Variational methodology. Our approach, in contrast, is based on a variational methodology that goes back to [61] which allows the direct handling of the nonsmooth nature of contact problems. We also use a variational approach to develop numerical integrators for nonsmooth rigid-body dynamics. The procedure is based on a discrete Lagrangian principle and automatically generates a symplectic-momentum preserving integrator. Near impact, we introduce a collision point and a collision time and solve for them using a variational method.

Variational integrators are known to have remarkable near-energy preserving properties, and we will recover this excellent energy behavior even in the nonsmooth case. We want to emphasize that the variational point of view is not confined to conservative systems but also applies to forced and dissipative systems, as demonstrated in [26]. In future works, we will investigate how forces and friction can be added to our collision algorithm and also how to incorporate other dissipative effects (inelastic collisions).

Issues addressed in this paper. We first show that, by introducing a space of configuration trajectories extended by introducing curve parameterizations as variables, the traditional approach to the calculus of variations can be applied. Moreover, the formulation in the extended setting enables us to address and give a rigorous interpretation of the sense in which the flow map of a mechanical system subjected to dissipationless impact dynamics is symplectic. The nonautonomous variational approach also leads to Weierstrass–Erdmann-type conditions for impact, in terms of energy and momentum conservation at the contact point (see [19] and [61]).

On the discrete side, the variational formalism leads to symplectic-momentum preserving integrators that are consistent with the jump conditions and the continuous theory.

The theory of geometric integration (see, for example, [50] and [18]) is typically concerned with smooth Hamiltonian or Lagrangian systems posed on smooth spaces. These techniques do not immediately apply to nonsmooth settings, and naive applications can result in extremely bad behavior, as demonstrated in [56].

Our methods answer an important question posed by [56]: *How can geometric integrators be formulated and implemented for collision problems?* In fact, the algorithms developed in

the present paper show how a symplectic method can be constructed for nonsmooth systems so that it retains the good behavior normally associated with symplectic methods.

Some work on extensions of geometric integration to collision problems exists. In particular, [4] has constructed time-symmetric methods for contact, and [21] (see also [20]) has developed methods for impacts of hard spheres. To date there have been no symplectic methods for collisions presented, in part due to difficulties with understanding symplecticity in a nonsmooth setting. However, the variational formulation of continuous time nonsmooth systems that we develop here is a key which allows us to understand the geometric structure of the problem, both before and after discretization. Our methods can be considered extensions of the large body of work on geometric integration of ODEs (see, for example, [16, 17, 18, 22, 29, 37]).

We caution that the algorithm presented in this paper is implicit and very expensive and thus may not be appropriate for use with large collision systems. Nonetheless, it is the first geometric integrator for collision problems and thus serves as a basis for the construction of more efficient methods in the future. In fact, the methods of this paper have already led to the development of more computationally feasible collision integrators (see [9]).

We also discuss how nonsmooth analysis techniques [25] can be incorporated into the variational procedure such that the integrator can cope with nonsmooth contact geometries (such as corner-to-corner collisions). As we mentioned before, this is the case which most existing algorithms cannot handle (the standard penalty methods simply fail since no proper gap function can be defined for such geometries).

Organization of the paper. In section 2, we first consider the time-continuous situation and extend the conventional setting of geometric Lagrangian mechanics (see, for example, [36]) to include nonsmooth but still continuous trajectories. This allows us to recover the standard jump conditions at impact and to prove that the flow map of the system is symplectic in the extended sense.

To apply the standard geometric mechanical tools in nonsmooth situations, it is necessary to formulate the problem so that the space of admissible trajectories of the system has a smooth manifold structure. To do this, we work in the extended framework, where both configuration variables and time are considered as functions of a fixed parameter space. This is the same approach as that used in multisymplectic mechanics (see [15, 35]), where it was introduced to allow the consideration of right, or horizontal, transformations of the system.

Next, in section 3, we discretize the variational structure of the system, based on the concept of discrete mechanics (see [37] for an overview and history), to obtain variational integrators for collision problems. By discretizing the variational structure, rather than some generalized equations of motion, we are able to show that our methods have various geometric properties, including the preservation of momentum maps and symplectic structures.

Finally, in section 4, we consider particular examples of our variational integrators for collision problems and investigate their behavior on a number of sample problems of rigid-body collisions. In the appendix, we lay the foundations of some future work by briefly discussing possible uses of the nonsmooth calculus approach (see [25]) in the context of variational collision integrators.

1.1. Overview of the continuous model. Before we begin with the body of the paper, we will first give a brief overview of the main ideas and techniques used. This is not a rigorous treatment, but everything stated here will be precisely defined and proven later.

Consider the system defined by the Lagrangian $L(q,\dot{q}) = \frac{1}{2}\dot{q}^T M\dot{q} - V(q)$, where M is a mass matrix and V is a potential function. Here $q = (q^1,\dots,q^n)$ is a vector of configuration variables which lives in the configuration manifold $q \in Q$. We now consider a subset $C \subset Q$, which we call the admissible set and which represents those configurations for which no contact is occurring. The boundary ∂C of the admissible set is all of those points at which contact has just occurred but for which no interpenetration is taking place.

We now consider a trajectory $q(t)$ which maps $q : [0,T] \to Q$ such that $q(t) \in C$, except at a particular time t_i for which $q(t_i) \in \partial C$. The time t_i is thus the time at which contact occurs, and we allow the trajectory $q(t)$ to be nonsmooth but still continuous at this time.

Proceeding in the standard way for Lagrangian mechanics, we construct the action of the trajectory by integrating the Lagrangian along $q(t)$. We then compute variations of the action with respect to variations in both the curve $q(t)$ and the impact time t_i, holding the endpoints of the curve fixed, to give

$$\begin{aligned}
&\delta \int_0^T L(q(t),\dot{q}(t))\,dt \\
&\quad = \int_0^{t_i} \left[\frac{\partial L}{\partial q}\cdot\delta q + \frac{\partial L}{\partial \dot{q}}\cdot\delta\dot{q}\right] dt + \int_{t_i}^{T} \left[\frac{\partial L}{\partial q}\cdot\delta q + \frac{\partial L}{\partial \dot{q}}\cdot\delta\dot{q}\right] dt \\
&\qquad - \Big[L(q,\dot{q})\cdot\delta t_i\Big]_{t_i^-}^{t_i^+} \\
&\quad = \int_0^{t_i} \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right]\cdot\delta q\;dt + \int_{t_i}^{T} \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right]\cdot\delta q\;dt \\
&\qquad - \left[\frac{\partial L}{\partial \dot{q}}\cdot\delta q + L\right]_{t_i^-}^{t_i^+},
\end{aligned}$$

where we have used integration by parts and the condition $\delta q(T) = \delta q(0) = 0$. Requiring that the variations of the action be zero for all δq implies that on the intervals away from t_i the integrand must be zero, giving the well-known Euler–Lagrange equations

$$\frac{\partial L}{\partial q}(q,\dot{q}) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q,\dot{q})\right) = 0.$$

For the particular form of the Lagrangian chosen above, this is simply

$$M\ddot{q} = -\nabla V(q),$$

which is Newton's equation of mass times acceleration equals force, and this equation describes the motion of the system away from impact.

Not only must the two integrals in the variation equation be zero, but the jump term at t_i must also be zero. Here it is necessary to recall that the curve at time t_i must lie in the boundary ∂C of the admissible set, and differentiating this relationship $q(t_i) \in \partial C$ gives the

condition

$$\delta q(t_i) + \dot{q}(t_i) \cdot \delta t_i \in T\partial C,$$

which states that the combined variation on the left-hand side must be in the tangent plane to ∂C at the impact point. The space of allowable $\delta q(t_i)$ and δt_i is spanned by the set of $\delta q(t_i) \in T\partial C$ with $\delta t_i = 0$, together with the additional variation $\delta q(t_i) = -\dot{q}(t_i)$ with $\delta t_i = 1$. Substituting each of these into the jump term in the variation equation gives the two relations

$$\left[\frac{\partial L}{\partial \dot{q}}\bigg|_{t_i^+} - \frac{\partial L}{\partial \dot{q}}\bigg|_{t_i^-} \right] \cdot \delta q(t_i) = 0 \text{ for all } \delta q(t_i) \in T\partial C,$$
$$\left[\frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L \right]_{t_i^+} - \left[\frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L \right]_{t_i^-} = 0.$$

When the Lagrangian is of the form kinetic minus potential, as above, these can be written as

$$\dot{q}(t_i^+) - \dot{q}(t_i^-) \in N_C(q_i(t)), \tag{1a}$$
$$E_L(t_i^+) - E_L(t_i^-) = 0, \tag{1b}$$

where the energy is $E_L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} + V(q)$ and $N_C(q)$ is the normal cone to ∂C at q. The first of these two equations states that the jump in velocity at the impact point must be orthogonal to the boundary ∂C, while the second equation states that energy must be conserved during the impact. Together these two equations constitute a system of n equations which describe the evolution of the system during the collision.

It is well known that the system described by the Euler–Lagrange equations has many special properties. In particular, the flow on state space is symplectic, meaning that it conserves a particular two-form, and if there are symmetry actions on phase space, then there are corresponding conserved quantities of the flow, known as momentum maps. All of these geometric properties can be proven directly from the variational principle used above, and so they also hold for nonsmooth systems. Later we will see how this can be precisely formulated.

1.2. Overview of the discrete model. Discrete variational mechanics is based on replacing the position q and velocity $\dot{q}$ with two nearby positions q_0 and q_1 and a timestep h. These positions should be thought of as being two points on a curve at time h apart so that $q_0 \approx q(0)$ and $q_1 \approx q(h)$ for some short curve segment $q(t)$.

We next consider a discrete Lagrangian $L_d(q_0, q_1, h)$, which we think of as approximating the action integral along the curve segment between q_0 and q_1. For concreteness, consider the very simple approximation given by

$$L_d(q_0, q_1, h) = h \left[\left(\frac{q_1 - q_0}{h} \right)^T M \left(\frac{q_1 - q_0}{h} \right) - V(q_0) \right].$$

This is simply the rectangle rule applied to approximate the action integral, with the velocity being approximated by the difference operator.

Now consider a discrete curve of points $\{q_k\}_{k=0}^N$ in C and corresponding times $t_k = kh$, together with a special impact point $\tilde{q} \in \partial C$ and an impact time $\tilde{t} = \alpha t_{i-1} + (1-\alpha)t_i$. Here $\alpha \in [0,1]$ is a parameter which interpolates $\tilde{t}$ with the interval $[t_{i-1}, t_i]$. Given such a discrete trajectory

$$(q_0, t_0), \dots, (q_{i-1}, t_{i-1}), (\tilde{q}, \tilde{t}), (q_i, t_i), \dots, (q_N, t_N),$$

we calculate the discrete action along this sequence by summing the discrete Lagrangian on each adjacent pair, with the timestep being the difference between the pair of times. Following the continuous derivation above, we compute variations of this action sum with respect to variations in the q_k as well as $\tilde{q}$ and α (and hence $\tilde{t}$), with the boundary points q_0 and q_N held fixed. This gives

$$\begin{aligned}
\delta \Bigg[\sum_{k=0}^{i-2} & L_d(q_k, q_{k+1}, h) + L_d(q_{i-1}, \tilde{q}, \alpha h) \\
& + L_d(\tilde{q}, q_i, (1-\alpha)h) + \sum_{k=i}^{N-1} L_d(q_k, q_{k+1}, h) \Bigg] \\
&= \sum_{k=0}^{N-1} \Big[D_1 L_d(q_k, q_{k+1}, h) \cdot \delta q_k + D_2 L_d(q_k, q_{k+1}, h) \cdot \delta q_{k+1} \Big] \\
&= \sum_{k=1}^{i-2} \Big[D_2 L_d(q_{k-1}, q_k, h) + D_1 L_d(q_k, q_{k+1}, h) \Big] \cdot \delta q_k \\
&\qquad + \sum_{k=i+1}^{N-1} \Big[D_2 L_d(q_{k-1}, q_k, h) + D_1 L_d(q_k, q_{k+1}, h) \Big] \cdot \delta q_k \\
&\qquad + \Big[D_2 L_d(q_{i-2}, q_{i-1}, h) + D_1 L_d(q_{i-1}, \tilde{q}, \alpha h) \Big] \cdot \delta q_{i-1} \\
&\qquad + \Big[D_2 L_d(q_{i-1}, \tilde{q}, \alpha h) + D_1 L_d(\tilde{q}, q_i, (1-\alpha)h) \Big] \cdot \delta \tilde{q} \\
&\qquad + \Big[D_2 L_d(\tilde{q}, q_i, (1-\alpha)h) + D_1 L_d(q_i, q_{i+1}, h) \Big] \cdot \delta q_i \\
&\qquad + \Big[D_3 L_d(q_{i-1}, \tilde{q}, \alpha h) - D_3 L_d(\tilde{q}, q_i, (1-\alpha)h) \Big] \cdot h \delta \alpha,
\end{aligned}$$

where we have rearranged the summation and we have used the fact that $\delta q_0 = \delta q_N = 0$. This calculation is illustrated graphically in Figure 1.

If we now require that the variations of the action be zero for any choice of δq_k, then we obtain the discrete Euler–Lagrange equations

$$D_2 L_d(q_{k-1}, q_k, h) + D_1 L_d(q_k, q_{k+1}, h) = 0,$$

which must hold for each k away from the impact time. For the particular L_d chosen above, we compute

$$\begin{aligned}
D_2 L_d(q_{k-1}, q_k, h) &= M\left(\frac{q_k - q_{k-1}}{h}\right), \\
D_1 L_d(q_k, q_{k+1}, h) &= -\left[M\left(\frac{q_{k+1} - q_k}{h}\right) + h \nabla V(q_k) \right],
\end{aligned}$$

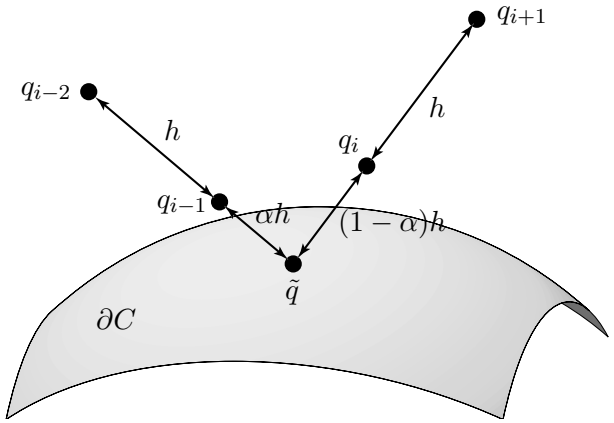

Figure 1. *The discrete variational principle for collisions.*

and so the discrete Euler–Lagrange equations are

$$M\left(\frac{q_{k+1}-2q_k+q_{k-1}}{h^2}\right) = -\nabla V(q_k).$$

This is clearly a discretization of Newton's equations, using a simple finite difference rule for the derivative.

If we take initial conditions (q_0, q_1), then the discrete Euler–Lagrange equations define a recursive rule for calculating the sequence $\{q_k\}_{k=0}^N$. Regarded in this way, they define a map $F_{L_d} : (q_k, q_{k+1}) \mapsto (q_{k+1}, q_{k+2})$, which we can think of as a one-step integrator for the system defined by the continuous Euler–Lagrange equations, away from impact.

Near impact, we must consider the other equations which are implied by the discrete variation equation being zero. Assume that we have used the discrete Euler–Lagrange equations to compute the trajectory up until the pair (q_{i-2}, q_{i-1}), just before impact. Now we have the equation

$$D_2L_d(q_{i-2}, q_{i-1}, h) + D_1L_d(q_{i-1}, \tilde{q}, \alpha h) = 0,$$

which becomes

$$M\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right) - M\left(\frac{q_{i-1}-q_{i-2}}{h}\right) = -\alpha h \nabla V(q_{i-1}).$$

Combining this with the condition that $\tilde{q} \in \partial C$ we obtain $n+1$ equations to be solved for the $n+1$ unknowns $\tilde{q}$ and α. We thus now know the point and time of contact. Next, we recall that $\tilde{q} \in \partial C$, and so its variations must lie in the tangent space. This means that we have the pair of equations

$$\begin{aligned}\Big[D_2L_d(q_{i-1}, \tilde{q}, \alpha h) + D_1L_d(\tilde{q}, q_i, (1-\alpha)h)\Big] \cdot \delta\tilde{q} = 0 \text{ for all } \delta\tilde{q} \in T\partial C,\\ D_3L_d(q_{i-1}, \tilde{q}, \alpha h) - D_3L_d(\tilde{q}, q_i, (1-\alpha)h) = 0,\end{aligned}$$

which become

$$M\left(\frac{q_i-\tilde{q}}{(1-\alpha)h}\right)-M\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)+(1-\alpha)h\nabla V(\tilde{q})\in N_C(\tilde{q}),$$

$$\left[\frac{1}{2}\left(\frac{q_i-\tilde{q}}{(1-\alpha)h}\right)^T M\left(\frac{q_i-\tilde{q}}{(1-\alpha)h}\right)+V(\tilde{q})\right]$$

$$-\left[\frac{1}{2}\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)^T M\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)+V(q_{i-1})\right]=0.$$

These are discrete versions of (1), and they give n equations to be solved for q_i. Finally, we use the equation

$$D_2L_d(\tilde{q},q_i,(1-\alpha)h)+D_1L_d(q_i,q_{i+1},h)=0,$$

which is

$$M\left(\frac{q_{i+1}-q_i}{h}\right)-M\left(\frac{q_i-\tilde{q}}{(1-\alpha)h}\right)=-\nabla V(q_i),$$

to solve for q_{i+1}, and then we can revert to using the standard discrete Euler–Lagrange equations to continue away from the impact.

The power of the variational approach becomes apparent when we consider the geometric properties of the discrete system. Just as in the continuous case, we can derive conservation laws of the discrete system directly from the variational principle. In particular, we will see that there is a conserved discrete symplectic form, and conserved discrete momentum maps arise from symmetries. In addition, in section 4, we will investigate the numerical behavior of the discrete system, regarded as an integrator for the continuous problem, and we will see that it also has excellent long-time stable energy behavior.

To understand the geometry and properties of both the continuous and discrete nonsmooth mechanics, however, we now need to return to the beginning and develop a more rigorous treatment of the variational procedure.

2. Continuous model. As noted in the introduction, the basic methodology used here is that of variational mechanics and variational discretizations. Clearly, a generalization to the nonsmooth setting of the autonomous, smooth variational mechanics cannot be done in a straightforward way. One of the major obstacles is that the lack of smoothness for the mappings prevents us from using the differential calculus on the manifold of mappings, as one essentially does in the smooth case (see [36]).

The main issue addressed in this section is how to overcome this difficulty and how to derive the conservation of quantities such as energy, momentum maps, and the symplectic form using a variational approach. The approach we use is to extend the problem to the nonautonomous case so that both configuration variables and time are functions of a separate parameter τ. This allows the impact to be fixed in τ space while remaining variable in both configuration and time spaces, and it means that the relevant space of configurations will indeed be a smooth manifold, as we shall prove.

To make our variational procedure clear, we initially consider only the frictionless, purely elastic impact problem. In the last subsection, however, we show how the results can be extended to deal with friction and nonelastic impacts.

2.1. Lagrangian mechanics in a nonsmooth setting. Consider a configuration manifold Q and a submanifold with boundary $C \subset Q$ which represent the subset of admissible configurations. Let ∂C be called the contact set, and let $L : TQ \to \mathbb{R}$ be a regular Lagrangian.

Remark. Similar results are obtained if we considered the configuration Q a manifold with boundary and the contact set to be ∂Q.

Let us now consider the *path space* defined by

$$\mathcal{M} = \mathcal{T} \times \mathcal{Q}([0,1], \tau_i, \partial C, Q),$$

where

$$\begin{aligned} \mathcal{T} &= \{c_t \in C^\infty([0,1], \mathbb{R}) \mid c_t' > 0 \text{ in } [0,1]\}, \\ \mathcal{Q}([0,1], \tau_i, \partial C, Q) &= \{c_q : [0,1] \to Q \mid c_q \text{ is a } C^0, \text{ piecewise } C^2 \text{ curve}, \\ &\qquad c_q(\tau) \text{ has only one singularity at } \tau_i, c_q(\tau_i) \in \partial C\}. \end{aligned}$$

A path $c \in \mathcal{M}$ is thus a pair $c = (c_t, c_q)$. Given a path, we can form the *associated curve* $q : [c_t(0), c_t(1)] \to Q$ by

$$q(t) = c_q(c_t^{-1}(t)),$$

and we denote by $\mathcal{C}$ the space of all these paths $q(t) \in Q$.

The theory we will develop applies to rigid-body impact problems, such as a particle bouncing on a rigid wall or two rigid bodies colliding, where the submanifold ∂C is obtained from the condition that interpenetration of matter cannot occur. The moment of impact τ_i is fixed in the τ space but is allowed to vary in the t space according to $t_i = c_t(\tau_i)$; thus the setting we suggest is not restrictive in this sense.

We use a nonautonomous formulation of an autonomous mechanical system in order to achieve smoothness of the manifold of mappings, as one can see from the following lemmas.

Lemma 2.1. *$\mathcal{T}$ is a smooth manifold.*

Proof. $\mathcal{T}$ is an open set in $C^\infty([0,1], \mathbb{R})$, which is a smooth manifold (see [36]). Then $\mathcal{T}$ is a submanifold of $C^\infty([0,1], \mathbb{R})$ and thus a manifold. ∎

Lemma 2.2. *$\mathcal{Q}([0,1], \tau_i, \partial C, Q)$ is a smooth manifold.*

Proof. Fix a chart U in Q such that $U \cap \partial C \neq \emptyset$ and $U \cap \partial C$ is a chart in ∂C. Consider the set

$$\mathcal{Q}_U = \mathcal{Q}([0, \tau_i], U) \times \mathcal{Q}([\tau_i, 1], U) \times (U \cap \partial C),$$

where

$$\begin{aligned} \mathcal{Q}([0, \tau_i], U) &= \{q : [0, \tau_i] \to Q \mid q \text{ is a } C^\infty \text{ curve}, q(\tau_i) \in U\}, \\ \mathcal{Q}([\tau_i, 1], U) &= \{q : [\tau_i, 1] \to Q \mid q \text{ is a } C^\infty \text{ curve}, q(\tau_i) \in U\}. \end{aligned}$$

An element $c \in \mathcal{Q}([0,1], \tau_i, \partial C, Q)$ is the inverse image of the origin for some map $g_U : \mathcal{Q}_U \to \mathbb{R}^{2n}$ given by

$$g_U(q_1(\tau), q_2(\tau), q_i) = \begin{pmatrix} q_1(\tau_i) - q_i \\ q_2(\tau_i) - q_i \end{pmatrix},$$

where we denote the dimension of Q by n. One can prove that 0 is a regular value of g_U, and then the set $g_U^{-1}(0)$ is a submanifold of $\mathcal{Q}_U$, and thus it has a manifold structure.

Now $\{g_U^{-1}(0)\}_U$ represents a covering of $\mathcal{Q}([0,1],\tau_i,\partial C,Q)$, where each element of the covering is a manifold. The elements of the covering satisfy the compatibility conditions necessary to ensure that $\mathcal{Q}([0,1],\tau_i,\partial C,Q)$ itself is a manifold (see [1]). ■

Corollary 2.3. *$\mathcal{M}$ is a smooth manifold.*

Remark. The theory can be easily extended to a problem involving more than one impact by simply taking multiple points τ_i at which the trajectory is nonsmooth.

Note that the tangent space at $q \in \mathcal{Q}$ can be written as

$$T_q\mathcal{Q} = \{v : [0,1] \to TQ \mid v \text{ is a } C^0 \text{ piecewise } C^2 \text{ map }, v(\tau_i) \in T_{q(\tau_i)}\partial C\},$$

which will be a convenient form below when we consider variations of trajectories. The tangent space to the path space $\mathcal{M}$ is then given by $T\mathcal{M} = T\mathcal{T} \times T\mathcal{Q}$.

Remark. As we have noted above, fixing the impact point τ_i in τ space allows us to rigorously define what we mean by a variation of the impact point in t space. This is similar to the introduction of a parameterized spacetime in [35] and [31].

The *action map* $\mathfrak{G} : \mathcal{M} \to \mathbb{R}$ is given by

$$\mathfrak{G}(c_t, c_q) = \int_0^1 L\left(c_q(\tau), \frac{c_q'(\tau)}{c_t'(\tau)}\right) c_t'(\tau)\, d\tau, \tag{2}$$

where c' denotes the derivative with respect to τ.

Remark. $c_q'(\tau)$ does not exist at τ_i, but the definition makes sense nonetheless.

If q is the associated curve for $c \in \mathcal{M}$, by the change of coordinates $s = c_t(\tau)$ we can also write $\mathfrak{G}$ as

$$\mathfrak{G}(q) = \int_{c_t(0)}^{c_t(1)} L(q(s), \dot{q}(s))ds, \tag{3}$$

where $\dot{q}$ denotes the derivative with respect to t.

Define the *extended configuration manifold* to be $Q_e = \mathbb{R} \times Q$ and the *second order submanifold* of $T(TQ_e)$ to be

$$\ddot{Q}_e = \left\{ \frac{d^2c}{d\tau^2}(0) \in T(TQ_e) \mid c : [0,1] \to Q_e \text{ is a } C^2 \text{ curve} \right\}. \tag{4}$$

Now we can derive the equations of motion and the jump conditions in a purely variational way, by taking variations of the actions with respect to the path. This leads to the following fundamental theorem.

Theorem 2.4. *Given a C^k Lagrangian L, $k \geq 2$, there exist a unique C^{k-2} mapping $EL : \ddot{Q} \to T^*Q_e$ and a unique C^{k-1} one-form Θ_L on TQ_e such that for all variations $\delta c \in T_c\mathcal{M}$ of c we have*

$$d\mathfrak{G}(c) \cdot \delta c = \int_0^{\tau_i} EL(c'') \cdot \delta c\, d\tau + \int_{\tau_i}^1 EL(c'') \cdot \delta c\, d\tau + \Theta_L(c') \cdot \hat{\delta} c|_0^{\tau_i^-} + \Theta_L(c') \cdot \hat{\delta} c|_{\tau_i^+}^1, \tag{5}$$

where

$$\hat{\delta}c(\tau) = \left(\left(c(\tau), \frac{\partial c}{\partial \tau}(\tau) \right), \left(\delta c(\tau), \frac{\partial \delta c}{\partial \tau}(\tau) \right) \right).$$

The mapping EL is called the Euler–Lagrange derivative *and the one-form* Θ_L *is called the* Lagrangian one-form. *In coordinates they have the expression*

$$EL(c'') = \left[\frac{\partial L}{\partial q} c_t' - \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{q}} \right) \right] dc_q + \left[\frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{q}} \frac{c_q'}{c_t'} - L \right) \right] dc_t, \tag{6}$$

$$\Theta_L(c') = \left[\frac{\partial L}{\partial \dot{q}} \right] dc_q - \left[\frac{\partial L}{\partial \dot{q}} \frac{c_q'}{c_t'} - L \right] dc_t. \tag{7}$$

Proof. Consider $\delta c \in T_c\mathcal{M}$. We calculate $d\mathfrak{G}(c) \cdot \delta c$ using the definition (see [36])

$$d\mathfrak{G}(c) \cdot \delta c = \frac{d}{d\lambda} \mathfrak{G}(c^\lambda) \Big|_{\lambda=0}, \tag{8}$$

where c^λ is a curve in $\mathcal{M}$ with $c^0 = c$ and $\frac{dc^\lambda}{d\lambda}|_{\lambda=0} = \delta c$. Splitting c^λ into components $c^\lambda = (c_t^\lambda, c_q^\lambda)$, we then have $(\frac{d}{d\lambda} c_t^\lambda|_{\lambda=0}, \frac{d}{d\lambda} c_q^\lambda|_{\lambda=0}) = (\delta c_t, \delta c_q)$, and we can calculate

$$d\mathfrak{G} \cdot \delta c = \int_0^1 \left[\frac{\partial L}{\partial q} \delta c_q + \frac{\partial L}{\partial \dot{q}} \left(\frac{\delta c_q'}{c_t'} - \frac{c_q' \delta c_t'}{(c_t')^2} \right) \right] c_t' \, d\tau + \int_0^1 L \delta c_t' \, d\tau.$$

Now we split the integral $\int_0^1$ into $\int_0^{\tau_i} + \int_{\tau_i}^1$ in order to integrate the $\delta c_q'$ and $\delta c_t'$ terms by parts. Some straightforward algebra then leads to (5). ∎

2.2. Hamilton's principle of critical action. Hamilton's principle of critical action tells us that we should consider critical points of the action function. Therefore, let us define the *space of solutions* $\mathcal{M}_L \subset \mathcal{M}$ to be the set of all paths $c \in \mathcal{M}$ which satisfy $d\mathfrak{G}(c) \cdot \delta c = 0$ for all variations $\delta c \in T_c\mathcal{M}$ which are zero at the boundary points 0 and 1.

Using (5), we can see that c is a solution if it satisfies

$$\int_0^{\tau_i} EL(c'') \cdot \delta c \, d\tau + \int_{\tau_i}^1 EL(c'') \cdot \delta c \, d\tau + \Theta_L(c')|_{\tau_i^-}^{\tau_i^+} \cdot \hat{\delta} c(\tau_i) = 0 \tag{9}$$

for all variations $\delta c \in T_c\mathcal{M}$.

From (9) it is clear that c is a solution iff the Euler–Lagrange derivative is zero on smooth portions and the Lagrangian one-form has a zero jump at τ_i. Splitting $EL(c'')$ into the two components, we obtain

$$\frac{\partial L}{\partial q} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) = 0 \quad \text{in} \quad [t_0, t_i) \cup (t_i, t_1], \tag{10}$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \dot{q} - L \right) = 0 \quad \text{in} \quad [t_0, t_i) \cup (t_i, t_1], \tag{11}$$

where $t_0 = c_t(0)$, $t_1 = c_t(1)$, and $t_i = c_t(\tau_i)$.

In fact, (11) is redundant, as it is a consequence of (10). Indeed, if c is a path satisfying (10) for all $t \in (t_0, t_i) \cup (t_i, t_1)$, then the second component (11) of the Euler–Lagrange equations is identically satisfied. To see this, we may calculate

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\dot{q} - L\right) &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\dot{q} + \frac{\partial L}{\partial \dot{q}}\ddot{q} - \frac{dL}{dt} \\ &= \left[\frac{\partial L}{\partial q}\dot{q} + \frac{\partial L}{\partial \dot{q}}\ddot{q}\right] - \frac{dL}{dt} \\ &= 0,\end{aligned}$$

where we used (10) to pass from the first to the second line.

The second part (11) of the Euler–Lagrange equations represents the conservation of energy for an autonomous system, provided the motion is smooth. The *energy* $E : TQ \to \mathbb{R}$ is defined to be

$$E(q, \dot{q}) = \frac{\partial L}{\partial \dot{q}}(q, \dot{q}) \cdot \dot{q} - L(q, \dot{q}).$$

It is not surprising that the second part of the Euler–Lagrange equations (11) is redundant, since the first part (10) already has the energy evolution built into it.

The previous definition of the energy function allows us to write the Lagrangian one-form in the compact notation

$$\Theta_L = \frac{\partial L}{\partial \dot{q}} dq - E dt, \tag{12}$$

where we use (q, t) to refer to the two components of c. The conservation of the Lagrangian one-form at the impact time reads

$$\Theta_L|_{\tau_i^-} = \Theta_L|_{\tau_i^+} \quad \text{on} \quad TQ_e|(\mathbb{R} \times \partial C). \tag{13}$$

Splitting this into the two components gives

$$\left.\frac{\partial L}{\partial \dot{q}}\right|_{t=t_i^-} \cdot \delta q = \left.\frac{\partial L}{\partial \dot{q}}\right|_{t=t_i^+} \cdot \delta q \tag{14}$$

for any $\delta q \in T_{q(t_i)}\partial C$ and

$$E(q(t_i^-), \dot{q}(t_i^-)) = E(q(t_i^+), \dot{q}(t_i^+)). \tag{15}$$

These equations are the Weierstrass–Erdmann-type conditions for impact. That is, (14) states that the linear momentum must be conserved in the tangent direction to ∂C, while (15) states that the energy must be conserved during an elastic impact.

The system of (14) and (15) must be solved for $\dot{q}(t_i^+)$. An obvious solution is $\dot{q}(t_i^+) = \dot{q}(t_i^-)$, but this is ruled out since the resulting trajectory would no longer lie in the admissible set. That is, it would violate the physical noninterpenetration condition.

Remark. Of course, existence and uniqueness for nonsmooth systems are very deep questions. Here, we will simply remark that for a codimension-one smooth boundary ∂C and quadratic kinetic energy, solutions to the system (14), (15) exist and are unique locally. The questions of global existence and uniqueness of solutions for more general Lagrangians is left for future works.

2.3. Lagrangian flows and conservation of the symplectic form. As we have already seen, a path $c \in \mathcal{M}$ is a solution of the variational principle if its associated curve $q(t)$ satisfies the Euler–Lagrange equations (10) and the jump conditions (14) and (15). It is a well-known fact that, in the smooth case, such a trajectory is uniquely determined by an initial condition in TQ. Since we work in a nonsmooth context, we must assume uniqueness of the physical trajectory at the impact point; we have already discussed in the previous subsection some conditions under which this actually occurs.

Under this hypothesis, the space $\mathcal{C}_L$, defined to be the space of curves $q(t)$ that satisfy (10), (14), and (15), may be identified with the space of initial conditions $(t_0, q_0, \dot{q}_0)$ on $\mathbb{R} \times TQ$.

Based on these remarks, we can define a flow $F_t : \mathbb{R} \times TQ \to \mathbb{R} \times TQ$ as

$$F_t(t_0, q_0, \dot{q}_0) = (t_0 + t, q(t_0 + t), \dot{q}(t_0 + t)), \tag{16}$$

where $q(t)$ is the unique trajectory in $\mathcal{C}_L$ corresponding to $(t_0, q_0, \dot{q}_0) \in \mathbb{R} \times TQ$. The mapping F_t is called the *Lagrangian flow*. In the nonsmooth setting, F_t will not necessarily be a smooth map on the whole of its domain. Later, we will restrict our attention to the parts of the domain on which F_t is smooth in order to use the derivatives of F_t with respect to the initial conditions and to time.

Remark. Even though we have worked within an extended configuration manifold formulation up until this point, here we have defined a flow on TQ rather than taking a flow on TQ_e with initial conditions in TQ_e. The reason for doing this is that the derivative t_0' has no physical meaning, and no mechanical problem has the derivative of time with respect to some parameter as an initial condition.

Next, we will show in which sense the Lagrangian flow F_t is symplectic. We begin by relating the previous approach to the one used in the rest of the section.

As we noted above, to any initial condition $(t_0, q_0, \dot{q}_0)$ in $\mathbb{R} \times TQ$ there corresponds a unique trajectory $q(t) \in \mathcal{C}_L$ such that $(q(t_0), \dot{q}(t_0)) = (q_0, \dot{q}_0)$. Trajectories in $\mathcal{C}_L$ are unique up to reparameterization in τ. Accordingly, we can define an equivalence relation in $\mathcal{M}_L$ by

$$c^0 \sim c^1 \text{ iff } c_q^0 \circ (c_t^0)^{-1} = c_q^1 \circ (c_t^1)^{-1}, \tag{17}$$

where $c^0, c^1 \in \mathcal{M}_L, c^0 = (c_q^0, c_t^0), c^1 = (c_q^1, c_t^1)$. That is, two paths are equivalent if they have the same associated curve, and so to a given trajectory $q(t)$ in $\mathcal{C}_L$ there corresponds an equivalence class $\hat{c}$ of curves in the extended space.

In a similar manner, we can define an equivalence relation on TQ_e by

$$(t_0, q_0, t_0', q_0') \sim (t_1, q_1, t_1', q_1') \text{ iff } t_0 = t_1,\ q_0 = q_1, \text{ and } \frac{q_0'}{t_0'} = \frac{q_1'}{t_1'}, \tag{18}$$

which is a pointwise version of the previous equivalence relation (17).

Now, the quotient space $TQ_e/\sim$ may be identified with the product $\mathbb{R}\times TQ$, and the flow F_t may be regarded not as a flow on TQ_e (which would not be desirable, as explained in the previous remark) but as a flow on the equivalence classes of TQ_e.

To prove symplecticity for the flow F_t in a precise sense, we must reinterpret Theorem 2.4 by slightly modifying the definition of the Lagrangian one-form Θ_L.

That is, Theorem 2.4 stands with the same statement and fundamental relation (5) if we replace Θ_L with the one-form $\bar\Theta_L$ on $\mathbb{R}\times TQ\cong TQ_e/\sim$, where $\bar\Theta_L$ is given by the same coordinate expression as Θ_L, i.e., relation (7). More precisely, (5) becomes

$$d\mathfrak{G}(c)\cdot\delta c=\int_0^{\tau_i} EL(c'')\cdot\delta c\,d\tau+\int_{\tau_i}^1 EL(c'')\cdot\delta c\,d\tau+\bar\Theta_L(\tilde c)\cdot\delta\tilde c\big|_0^{\tau_i^-}+\bar\Theta_L(\tilde c)\cdot\delta\tilde c\big|_{\tau_i^+}^1, \tag{19}$$

where

$$\begin{aligned}\tilde c(\tau)&=\left(c_t(\tau),c_q(\tau),\frac{c_q'(\tau)}{c_t'(\tau)}\right),\\ \delta\tilde c(\tau)&=\left(\left(c_t(\tau),c_q(\tau),\frac{c_q'(\tau)}{c_t'(\tau)}\right),\left(\delta c_t(\tau),\delta c_q(\tau),\left(\frac{\delta c_q'}{c_t'}-\frac{c_q'\delta c_t'}{(c_t')^2}\right)(\tau)\right)\right).\end{aligned}$$

It is exactly this one-form $\bar\Theta_L$ on $\mathbb{R}\times TQ$ which is preserved by the flow F_t, as we will now show.

To any fixed $(t_0,q_0,\dot q_0)\in\mathbb{R}\times TQ$ we associate the integral curve $s\mapsto F_s(t_0,q_0,\dot q_0)$ for $s\in[0,t]$; the value of $\mathfrak{G}$ on that curve is denoted by $\mathfrak{G}_t$ and again called the *action*. Thus we define the map $\mathfrak{G}_t:\mathbb{R}\times TQ\to\mathbb{R}$ by

$$\mathfrak{G}_t(t_0,q_0,\dot q_0)=\int_{t_0}^{t_0+t} L\left(q(s),\dot q(s)\right)ds, \tag{20}$$

where $q(t)\in\mathcal{C}_L$ is the solution corresponding to $(t_0,q_0,\dot q_0)$.

If $c=(c_t,c_q)$ is any representative in the equivalence class $\hat c$ corresponding to q, we can write

$$F_t(t_0,q_0,\dot q_0)=\left(c_t(\tau),c_q(\tau),\frac{c_q'(\tau)}{c_t'(\tau)}\right), \tag{21}$$

where $\tau=c_t^{-1}(t_0+t)$.

Consider now an arbitrary curve $\lambda\mapsto(t_0^\lambda,q_0^\lambda,\dot q_0^\lambda)$ in $\mathbb{R}\times TQ$ which passes through $(t_0,q_0,\dot q_0)$ at $\lambda=0$. Denote by $q^\lambda(t)$ the unique trajectories in $\mathcal{C}_L$ corresponding to $(t_0^\lambda,q_0^\lambda,\dot q_0^\lambda)$ and by $\hat c^\lambda$ their equivalence classes in $\mathcal{M}_L$. (At $\lambda=0$ they reduce to $q(t)$ and $\hat c$, respectively.) We pick representatives $(c_t^\lambda,c_q^\lambda)$ in $\hat c^\lambda$ such that for any $\lambda\ge0$ we have

$$(c_t^\lambda)^{-1}(t_0^\lambda+t)=\text{const} \tag{22}$$

for some $t>0$; we denote this common value by τ.

Then, using (21) and (22), the fundamental (19) becomes

$$d\mathfrak{G}_t((t_0,q_0,\dot{q}_0))\cdot(\delta t_0,\delta q_0,\delta\dot{q}_0)=\bar{\Theta}_L(F_t(t_0,q_0,\dot{q}_0))\cdot\frac{d}{d\lambda}F_t(t_0^\lambda,q_0^\lambda,\dot{q}_0^\lambda)\Big|_{\lambda=0}$$
$$-\bar{\Theta}_L(t_0,q_0,\dot{q}_0)\cdot\frac{d}{d\lambda}(t_0^\lambda,q_0^\lambda,\dot{q}_0^\lambda)\Big|_{\lambda=0}, \tag{23}$$

where $(\delta t_0,\delta q_0,\delta\dot{q}_0)=\frac{d}{d\lambda}|_{\lambda=0}(t_0^\lambda,q_0^\lambda,\dot{q}_0^\lambda)$.

Taking the exterior derivative of (23), we derive

$$0=dd\mathfrak{G}_t=F_t^*(d\bar{\Theta}_L)-d\bar{\Theta}_L. \tag{24}$$

Defining the *Lagrangian symplectic form* by $\Omega_L=-d\bar{\Theta}_L$, we now see that relation (24) gives the symplecticity of the flow in the extended sense

$$F_t^*\Omega_L=\Omega_L. \tag{25}$$

Thus we derived conservation of the canonical symplectic structure in the extended sense (see [24]), namely,

$$\Omega_L=\omega_L+dE\wedge dt, \tag{26}$$

where $\omega_L=-d\theta_L$ is the canonical symplectic form. Here, θ_L represents the component of the Lagrangian one-form given by (12)

$$\theta_L=\frac{\partial L}{\partial\dot{q}}dq. \tag{27}$$

It is the term $dE\wedge dt$ that distinguishes the nonautonomous structure used here from the autonomous approach, for which the symplectic structure is given only by the canonical symplectic form ω_L.

2.4. Noether's theorem. Suppose that a Lie group G, with Lie algebra $\mathfrak{g}$, acts on Q by the (left or right) action $\Phi: G\times Q\to Q$. Consider the tangent lift of this action to $T\Phi: G\times TQ\to TQ$ given by $(T\Phi)_g(v_q)=T(\Phi_g)\cdot v_q$, and for $\xi\in\mathfrak{g}$ define the *infinitesimal generators* $\xi_Q: Q\to TQ$ and $\xi_{TQ}: TQ\to T(TQ)$ by

$$\xi_Q(q)=\frac{d}{dt}_{|t=0}\exp(t\xi)\cdot q,$$
$$\xi_{TQ}(v_q)=\frac{d}{dt}_{|t=0}T_q\phi_t(v_q),$$

where ϕ_t is the flow of the vector field ξ_Q.

In this subsection, we will not use the extended configuration manifold setting with variations in both time and configuration variables, as is done in the rest of the paper. This means that we are restricted to symmetries of the configuration variables, which do not involve altering the time variable. This allows us to deal with most of the interesting physical problems,

while still keeping the theory relatively simple. For a full account of the conservation of momentum maps in the extended setting, see [37].

For a fixed initial time $t_0 \in \mathbb{R}$, define the flow map $\tilde{F}_t : TQ \to TQ$ by

$$\tilde{F}_t(q_0, \dot{q}_0) = (q(t_0 + t), \dot{q}(t_0 + t)), \tag{28}$$

where $q(t)$ is the unique trajectory in $\mathcal{C}_L$ corresponding to $(q_0, \dot{q}_0) \in TQ$, as initial condition at t_0.

In the autonomous setting, the Lagrangian one-form $\bar{\Theta}_L$ reduces to the configuration component θ_L given by (27), and the action $\mathfrak{G}_t$ from (20) becomes the map $\bar{\mathfrak{G}}_t : TQ \to \mathbb{R}$ defined by

$$\bar{\mathfrak{G}}_t(q_0, \dot{q}_0) = \int_{t_0}^{t_0+t} L\left(q(s), \dot{q}(s)\right) ds. \tag{29}$$

Define the *Lagrangian momentum map* $J_L : TQ \to \mathfrak{g}^*$ to be

$$J_L(v_q) \cdot \xi = \theta_L \cdot \xi_{TQ}(v_q).$$

We will now show that when the group action is a symmetry of both the Lagrangian and the submanifold ∂C, then the momentum maps are conserved quantities of the flow.

A Lagrangian $L : TQ \to \mathbb{R}$ is said to be *infinitesimally invariant* under the lift of the group action $\Phi : G \times Q \to Q$ if $dL \cdot \xi_{TQ} = 0$ for all $\xi \in \mathfrak{g}$, and in this case the group action is said to be a *symmetry* of the Lagrangian.

In proving the following theorem, we will essentially use the assumption that the group action Φ leaves the boundary ∂C of the collision set invariant (locally). An example where this assumption is valid is the case of two or more irregular bodies (for example, binary astroids) moving in space under gravitational forces. In this case, the collision set is invariant to translations and rotations ($G = SE(3)$).

Theorem 2.5 (Noether's theorem). *Consider a Lagrangian system $L : TQ \to \mathbb{R}$ which is infinitesimally invariant under the lift of the (left or right) group action $\Phi : G \times Q \to Q$. Under the assumption that the group action leaves ∂C invariant (locally), the corresponding Lagrangian momentum map $J_L : TQ \to \mathfrak{g}^*$ is a conserved quantity of the flow so that $J_L \circ \tilde{F}_t = J_L$ for all times t.*

Proof. The group action of G on Q induces a group action of G on the space $\mathcal{C}$ of paths $q(t)$ in Q by pointwise action so that $\Phi_g(q)(t) = \Phi_g(q(t))$. The tangent lift of Φ acting on $\mathcal{C}$ will thus be the pointwise group action of the tangent lift of Φ group action on Q. From this we derive

$$d\mathfrak{G}(q) \cdot \xi_{\mathcal{C}}(q) = \int_{t_0}^{t_1} dL \cdot \xi_{TQ}\, dt,$$

and so, symmetries of the Lagrangian induce symmetries of the action. This implies that Φ_g leaves the space of solutions $\mathcal{C}_L$ of the Euler–Lagrange equations invariant, and so we may restrict Φ_g to $\mathcal{C}_L$.

Furthermore, the flow map $\tilde{F}_t : TQ \to TQ$ commutes with the tangent lift of Φ on $\mathcal{C}$: $\tilde{F}_t \circ T\Phi_g = T\Phi_g \circ \tilde{F}_t$. Differentiating this with respect to g in the direction ξ gives

$$T(\tilde{F}_t) \cdot \xi_{TQ} = \xi_{TQ} \circ \tilde{F}_t.$$

We now follow the same idea used to prove symplecticity of the flow map F_t and identify the space of solutions $\mathcal{C}_L$ with the space of initial conditions TQ. For an initial condition $v_q \in TQ$ and corresponding solution curve $q \in \mathcal{C}_L$, we thus have

$$\begin{aligned} d\mathfrak{G}(q) \cdot \xi_{\mathcal{C}}(q) &= d\bar{\mathfrak{G}}_t(v_q) \cdot \xi_{TQ}(v_q) \\ &= ((\tilde{F}_t)^*(\theta_L) - \theta_L)(v_q) \cdot \xi_{TQ}(v_q) \end{aligned} \tag{30}$$

from (23).

To derive (30), one uses the assumption that the group action Φ leaves ∂C invariant (locally). More precisely, it is essential that the path curves $q^\lambda \in \mathcal{C}$ corresponding to $v_q^\lambda = \eta_\lambda(v_q)$ (by η_λ we denote the flow of ξ_{TQ} on TQ) have exactly the same impact time t_i as the curve q. We conclude this from the relation

$$q^\lambda(t_i) = \Phi_{\exp(\lambda\xi)}(q)(t_i) = \exp(\lambda\xi) \cdot q(t_i)$$

as well as the assumption on the group action and the condition that $q(t_i) \in \partial C$.

As the left-hand side of (30) is always zero, the previous identity gives

$$(\theta_L \cdot \xi_{TQ}) \circ \tilde{F}_t = \theta_L \cdot \xi_{TQ},$$

which is the definition of conservation of the momentum map. We complete the proof by noting that the argument above is valid for all times $t \in \mathbb{R}$. ∎

2.5. Forcing and friction. In this subsection, we extend the theory developed so far to include forcing and friction. To do this in the variational framework, we turn from using Hamilton's principle to the Lagrange–d'Alembert extension of it.

The usual force field description of impact dynamics contains a given external force, a normal contact force field over the area in contact, and a friction force field required to be self-equilibrated and tangential to the surfaces in contact.

Following [36], we define the *exterior force field* as a fiber-preserving map $F : TQ_e \to T^*Q_e$ over the identity, which we write in coordinates as

$$F : (c, c') \mapsto (c, F(c, c')). \tag{31}$$

We use a unified treatment of contact forces (the normal and the frictional forces) by defining the *contact force field* to be a map $f^{\text{con}} : TQ_e|(\partial C \times \mathbb{R}) \to T^*(\partial C \times \mathbb{R})$.

Given a Lagrangian L and the exterior and contact force fields defined as above, the *integral Lagrange–d'Alembert principle* for a curve $c \in \mathcal{M}$ states that

$$\begin{aligned} \delta \int_0^1 L\left(c_q(\tau), \frac{c_q'(\tau)}{c_t'(\tau)}\right) c_t'(\tau)\, d\tau + \int_0^1 F(c(\tau), c'(\tau)) \cdot \delta c(\tau)\, d\tau \\ + f^{\text{con}}(c(\tau_i), c'(\tau_i)) \cdot \delta c(\tau_i) = 0 \end{aligned} \tag{32}$$

for all admissible variations δc vanishing at the endpoints.

Using integration by parts and notation from section 2.1, one can show that (32) is equivalent to

$$\int_0^{\tau_i} \left[EL(c'') + F(c')\right] \cdot \delta c \, d\tau + \int_{\tau_i}^1 (EL(c'') + F(c')) \cdot \delta c \, d\tau + \Theta_L(c')\big|_{\tau_i^-}^{\tau_i^+} \cdot \hat{\delta} c(\tau_i) + f^{\mathrm{con}}(c(\tau_i), c'(\tau_i)) \cdot \delta c(\tau_i) = 0. \tag{33}$$

From (33) we obtain the *extended forced Euler–Lagrange equations*, which have coordinate expressions

$$\frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} c_t' = F_q \text{ in } [0, \tau_i) \cup (\tau_i, 1], \tag{34}$$

$$-\frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{q}}\dot{q} - L\right) = F_t \text{ in } [0, \tau_i) \cup (\tau_i, 1], \tag{35}$$

where (F_t, F_q) denote the corresponding components of F.

However, the first part (34) of the extended forced Euler–Lagrange equations has the energy evolution built into it, as can be seen from

$$\begin{aligned} \frac{dE}{dt} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\dot{q} - L\right) \\ &= \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q}\right)\dot{q} \\ &= \frac{F_q}{c_t'}\dot{q}, \end{aligned} \tag{36}$$

where we used (34) to pass from the first to the second line.

Therefore, from (35), the time component F_t of the exterior force field must necessarily be of the form

$$F_t = -F_q \cdot \dot{q}. \tag{37}$$

This compatibility condition is a consequence of the fact that the mechanical system is autonomous and the equations must depend only on the associated curve $q(t)$. The nonautonomous approach is relevant only in the context of nonsmooth mechanics, and it is not surprising that there is no particular gain from this approach wherever the motion is smooth.

Now we turn to (33) and write the remaining terms on the left-hand side in components to obtain

$$\frac{\partial L}{\partial \dot{q}}\bigg|_{t_i^-}^{t_i^+} \cdot \delta q + f_q^{\mathrm{con}} \cdot \delta q = 0 \tag{38}$$

for any $\delta q \in T_{q(t_i)}\partial C$, and

$$E(q(t_i^+), \dot{q}(t_i^+)) - E(q(t_i^-), \dot{q}(t_i^-)) - f_t^{\mathrm{con}} = 0. \tag{39}$$

Equations (38) and (39) represent the standard jump conditions for an inelastic impact with friction. Equation (38) gives the jump in the tangential component of the linear momentum due to the frictional forces acting on the tangent plane of the contact submanifold ∂C. The energy dissipation, given by (39), is due to the tangential frictional forces as well as to the normal reaction force exerted by the constraint. For frictionless collisions, f_t^{con} plays the same role as the coefficient of restitution from the measure differential inclusion formulation of contact dynamics [27, 56].

3. Discrete model. We now turn to considering discrete models of contact problems, in which the continuous time variable is replaced with a discrete time index. The equations of motion are thus algebraic rather than differential equations, and they can be regarded as an integrator for the continuous system.

The approach we use is based on discrete variational mechanics (see [37]), in which the variational principle is discretized and the discrete equations and their conservation properties are derived as in the continuous case. This has the advantage of automatically capturing much of the geometric structure of the true problem even in the approximate discrete setting.

3.1. Discrete configurations and equations of motion. Disregard for the moment the continuous formulation of the previous section, and introduce a fixed *timestep* $h \in \mathbb{R}$. Consider a *discrete Lagrangian* $L_d : Q \times Q \to \mathbb{R}$, which is a function of two configuration points and the timestep, so that $L_d = L_d(q_0, q_1, h)$. The discrete Lagrangian will be chosen to approximate the continuous action integral over an interval of length h so that

$$L_d(q_0, q_1, h) \approx \int_0^h L(q, \dot{q}) dt,$$

where $q : [0, h] \to \mathbb{R}$ is an exact solution of the Euler–Lagrange equations, for L satisfies the boundary conditions $q(0) = q_0$ and $q(h) = q_1$.

We now consider an increasing sequence of times

$$t_k = kh \text{ for } k = 0, \dots, N,$$

we also fix $\tilde{\alpha} \in [0, 1]$, and we let $\tilde{\tau} = t_{i-1} + \tilde{\alpha} h$ denote the fixed impact time (corresponding to τ_i from the continuous model) and $\tilde{t} = t_{i-1} + \alpha h$ denote the actual impact time (corresponding to t_i). We take $\alpha = t_d(\tilde{\alpha})$, where t_d is some strictly increasing function which maps $[0, 1]$ onto $[0, 1]$. Thus we assumed only that the step at which the impact occurs is known and not the impact time $\tilde{t}$, which is allowed to vary according to variations in α.

The *discrete path space* is defined by

$$\mathcal{M}_d = \mathcal{T}_d \times \mathcal{Q}_d(\tilde{\alpha}, \partial C, Q), \tag{40}$$

where

$$\mathcal{T}_d = \{t_d(\tilde{\alpha}) \mid t_d \in C^\infty([0, 1], [0, 1]), t_d \text{ onto}, t_d' > 0 \text{ in } [0, 1]\}, \tag{41}$$

$$\mathcal{Q}_d(\tilde{\alpha}, \partial C, Q) = \{q_d : \{t_0, \dots, t_{i-1}, \tilde{\tau}, t_i, \dots, t_N\} \to Q, q_d(\tilde{\tau}) \in \partial C\}. \tag{42}$$

Remark. The set $\mathcal{T}_d$ is actually the real interval $[0,1]$, but we used (41) to define it in order to emphasize the analogy with the continuous case.

We identify the discrete trajectory with its image

$$(\alpha, q_d) = (\alpha, \{q_0, \ldots, q_{i-1}, \tilde{q}, q_i, \ldots, q_N\}),$$

where $q_k = q_d(t_k)$ for $k \in \{0, \ldots, N\}$, $\tilde{q} = q_d(\tilde{\tau})$, and $\alpha = t_d(\tilde{\alpha})$. Thus a discrete trajectory can be regarded as a sequence of points in Q, one of which must be in ∂C, and a single real number $\alpha \in [0,1]$.

The *discrete action map* $\mathfrak{G}_d : \mathcal{M}_d \to \mathbb{R}$ is defined by

$$\mathfrak{G}_d(\alpha, q_d) = \sum_{k=0}^{i-2} L_d(q_k, q_{k+1}, h) + \sum_{k=i}^{N-1} L_d(q_k, q_{k+1}, h) + L_d(q_{i-1}, \tilde{q}, \alpha h) + L_d(\tilde{q}, q_i, (1-\alpha)h). \tag{43}$$

As the discrete path space $\mathcal{M}_d$ is isomorphic to $[0,1] \times Q \times \cdots \times \partial C \times \cdots \times Q$ (N copies of Q), it can be given a smooth manifold structure.

For $q_d \in \mathcal{Q}_d(\tilde{\alpha}, \partial C, Q)$, the tangent space $T_{q_d}\mathcal{Q}_d(\tilde{\alpha}, \partial C, Q)$ is the set of all maps $v_{q_d} : \{t_0, \ldots, t_{i-1}, \tilde{\tau}, t_i, \ldots, t_N\} \to TQ$ such that $\pi_Q \circ v_{q_d} = q_d$ and $v_{q_d}(\tilde{\tau}) \in T_{\tilde{q}}\partial C$. For simplicity we will identify v_{q_d} with its image in TQ.

The tangent space to the full discrete path space is now $T\mathcal{M}_d = T\mathcal{T}_d \times T\mathcal{Q}_d$. At a given point $(\alpha, q_d) \in \mathcal{M}_d$ we will write a tangent vector in $T_{(\alpha,q_d)}\mathcal{M}_d$ as

$$(\delta\alpha, \delta q_d) = (\delta\alpha, \{\delta q_0, \ldots, \delta q_{i-1}, \delta\tilde{q}, \delta q_i, \ldots, \delta q_N\}).$$

Define the *discrete second order manifold* to be

$$\ddot{Q}_d = Q \times Q \times Q,$$

which has the same information content as the continuous second order manifold $\ddot{Q}$.

We now proceed, as in the continuous case, to derive the discrete equations of motion and the conservation laws from Hamilton's principle of critical action. We take variations of the discrete action sum with respect to the discrete path and to the parameter α, as stated in the following theorem.

Theorem 3.1. *Given a C^k discrete Lagrangian $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$, $k \geq 1$, there exist a unique C^{k-1} mapping $EL_d : \ddot{Q}_d \to T^*Q$ and unique C^{k-1} one-forms $\Theta^-_{L_d}$ and $\Theta^+_{L_d}$ on the discrete Lagrangian phase space $Q \times Q$ such that, for all variations $(\delta\alpha, \delta q_d) \in T_{(\alpha,q_d)}\mathcal{M}_d$ of (α, q_d), we have*

$$\begin{aligned}
&d\mathfrak{G}_d(\alpha, q_d) \cdot (\delta\alpha, \delta q_d) \\
&= \sum_{k=1}^{i-2} EL_d(q_{k-1}, q_k, q_{k+1}) \cdot \delta q_k + \sum_{k=i+1}^{N-1} EL_d(q_{k-1}, q_k, q_{k+1}) \cdot \delta q_k \\
&\quad + \Theta^+_{L_d}(q_{N-1}, q_N) \cdot (\delta q_{N-1}, \delta q_N) - \Theta^-_{L_d}(q_0, q_1) \cdot (\delta q_0, \delta q_1) \\
&\quad + [D_2 L_d(q_{i-2}, q_{i-1}, h) + D_1 L_d(q_{i-1}, \tilde{q}, \alpha h)] \cdot \delta q_{i-1} \\
&\quad + h\,[D_3 L_d(q_{i-1}, \tilde{q}, \alpha h) - D_3 L_d(\tilde{q}, q_i, (1-\alpha)h)] \cdot \delta\alpha \\
&\quad + i^*(D_2 L_d(q_{i-1}, \tilde{q}, \alpha h) + D_1 L_d(\tilde{q}, q_i, (1-\alpha)h)) \cdot \delta\tilde{q} \\
&\quad + [D_2 L_d(\tilde{q}, q_i, (1-\alpha)h) + D_1 L_d(q_i, q_{i+1}, h)] \cdot \delta q_i,
\end{aligned} \tag{44}$$

where $i^* : T^*Q \to T^*\partial C$ is the cotangent lift of the embedding $i : \partial C \to Q$.

The map EL_d is called the discrete Euler–Lagrange derivative *and the one-forms $\Theta^+_{L_d}$ and $\Theta^-_{L_d}$ are the* discrete Lagrangian one-forms. *In coordinates these have the expressions*

$$EL_d(q_{k-1}, q_k, q_{k+1}) = [D_2L_d(q_{k-1}, q_k, h) + D_1L_d(q_k, q_{k+1}, h)]\, dq_k \tag{45}$$

for $k \in \{1, \dots, i-2, i, \dots, N-1\}$ and

$$\Theta^+_{L_d}(q_k, q_{k+1}) = D_2L_d(q_k, q_{k+1}, h)\, dq_{k+1},$$
$$\Theta^-_{L_d}(q_k, q_{k+1}) = -D_1L_d(q_k, q_{k+1}, h)\, dq_k.$$

Proof. The formula is derived by straightforward algebra, by computing the derivative of the discrete action map, and by some rearrangement of the summation. This rearrangement corresponds to a discrete version of integration by parts, resulting in two boundary terms which are interpreted as the discrete Lagrangian one-forms. ■

By using the discrete version of Hamilton's principle, we consider the paths (α, q_d), which are critical points of the discrete action. Therefore, we define the *discrete space of solutions* to be the set of all paths which satisfy $d\mathfrak{G}_d(\alpha, q) \cdot (\delta\alpha, \delta q) = 0$ for all variations $(\delta\alpha, \delta q_d) \in T_{(\alpha,q_d)}\mathcal{M}_d$ which are zero at the boundary points 0 and N.

From (44) we conclude that (α, q_d) is a solution iff the discrete Euler–Lagrange derivative is zero at all k other than $\{0, i-1, i, N\}$. This statement at an arbitrary k reads

$$D_2L_d(q_{k-1}, q_k, h) + D_1L_d(q_k, q_{k+1}, h) = 0 \tag{46}$$

and is known as *discrete Euler–Lagrange equations*. These describe the motion of the system away from the impact point by implicitly defining a map $(q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$.

Near the point of impact, the discrete Hamilton's principle gives three additional sets of equations, namely,

$$D_2L_d(q_{i-2}, q_{i-1}, h) + D_1L_d(q_{i-1}, \tilde{q}, \alpha h) = 0, \tag{47a}$$
$$\tilde{q} \in \partial C, \tag{47b}$$

which is a system of $n+1$ equations to be solved for $\tilde{q}$ and α, and

$$D_3L_d(q_{i-1}, \tilde{q}, \alpha h) - D_3L_d(\tilde{q}, q_i, (1-\alpha)h) = 0, \tag{48a}$$
$$i^*(D_2L_d(q_{i-1}, \tilde{q}, \alpha h) + D_1L_d(\tilde{q}, q_i, (1-\alpha)h)) = 0, \tag{48b}$$

which is a system of n equations for the unknown q_i. Finally, we also have

$$D_2L_d(\tilde{q}, q_i, (1-\alpha)h) + D_1L_d(q_i, q_{i+1}, h) = 0, \tag{49}$$

which gives n equations to be solved for q_{i+1}.

A discrete trajectory can thus be formed by starting from an initial condition (q_0, q_1), using (46) to solve successively for the q_k until the impact time is reached, and then solving the systems (47), (48), and (49) in turn to obtain $\tilde{q}$, α and then q_i and q_{i+1}, before once again continuing with (46) to complete the trajectory.

Remark. The discrete energy conservation through the collision depends critically on exactly resolving the collision time with the parameter α. This is also the key feature of the recent improvements of [41] to the nonsmooth collision methods developed by [25].

3.2. Relationship between discrete and continuous models. Having established the basic discrete variational mechanics, we now consider how the discrete model can be regarded as an approximation to the continuous model.

At first glance it appears that the discrete Euler–Lagrange equations are defined only in terms of pairs of configuration positions. We will now see, however, that they can also be interpreted as defining a mapping on the cotangent bundle T^*Q. Define the *discrete Legendre transforms* or *discrete fiber derivatives* $\mathbb{F}^+L_d, \mathbb{F}^-L_d : Q \times Q \to T^*Q$ as given by

$$\begin{aligned}\mathbb{F}^+L_d(q_0, q_1) \cdot \delta q_1 &= D_2L_d(q_0, q_1, h) \cdot \delta q_1,\\ \mathbb{F}^-L_d(q_0, q_1) \cdot \delta q_0 &= -D_1L_d(q_0, q_1, h) \cdot \delta q_0,\end{aligned}$$

where h is the timestep in between q_0 and q_1. We note the implicit dependence on the timestep of the definition above. This dependence is completely neglected in the constant timestep discrete variational mechanics or rigorously treated in the nonautonomous setting (using adaptive timesteps), and we refer to [37] for a complete account of these ideas.

These also can be written

$$\mathbb{F}^+L_d : (q_0, q_1) \mapsto (q_1, p_1) = (q_1, D_2L_d(q_0, q_1, h)), \tag{50a}$$

$$\mathbb{F}^-L_d : (q_0, q_1) \mapsto (q_0, p_0) = (q_0, -D_1L_d(q_0, q_1, h)). \tag{50b}$$

If both discrete fiber derivatives are locally isomorphisms, then we say that L_d is *regular*. We will generally assume that we are working with regular discrete Lagrangians.

We introduce the notation

$$p^+_{k,k+1} = p^+(q_k, q_{k+1}, h) = \mathbb{F}^+L_d(q_k, q_{k+1}), \tag{51a}$$

$$p^-_{k,k+1} = p^-(q_k, q_{k+1}, h) = \mathbb{F}^-L_d(q_k, q_{k+1}) \tag{51b}$$

for the momentum at the two endpoints of each interval $[k, k + 1]$.

We can now use definitions (50a) and (51) of the discrete fiber derivatives and of the discrete momenta to see that the discrete Euler–Lagrange equations (46) can be written as

$$\mathbb{F}^+L_d(q_{k-1}, q_k) = \mathbb{F}^-L_d(q_k, q_{k+1}) \tag{52}$$

or simply

$$p^+_{k-1,k} = p^-_{k,k+1}. \tag{53}$$

That is, the discrete Euler–Lagrange equations enforce the condition that the momentum at time k should be the same when evaluated from the lower interval $[k - 1, k]$ or the upper interval $[k, k + 1]$.

In this interpretation, (48b) represents conservation of the projection of momentum (by i^*, on $T^*\partial C$) at the moment of impact

$$i^*p^+(q_{i-1}, \tilde{q}, \alpha h) = i^*p^-(\tilde{q}, q_i, (1 - \alpha)h), \tag{54}$$

which is a discrete version of the jump condition (14) from the continuous case.

To give an interpretation of the discrete equations around the impact time, we define the *discrete energy* to be

$$E_d(q_k, q_{k+1}, h) = -D_3 L_d(q_k, q_{k+1}, h). \tag{55}$$

Using this, we can write (48a) as

$$E_d(q_{i-1}, \tilde{q}, \alpha h) = E_d(\tilde{q}, q_i, (1-\alpha)h), \tag{56}$$

so this equation simply represents conservation of discrete energy at the impact time, a discrete analogue of (15).

Remark. The discrete energy defined in this way is used in [24] and can be motivated in several ways; first, for Lagrangians of the form of kinetic minus potential energy, and with the choice of discrete Lagrangians given by

$$L_d(q_0, q_1, h) = L\left(\gamma q_0 + (1-\gamma)q_1, \frac{q_1 - q_0}{h}\right), \tag{57}$$

where $\gamma \in [0, 1]$ is an interpolation parameter, the discrete energy gets the usual expression

$$E_d(q_0, q_1, h) = \frac{1}{2}\left(\frac{q_1 - q_0}{h}\right)^T M \left(\frac{q_1 - q_0}{h}\right) + V(\gamma q_0 + (1-\gamma)q_1). \tag{58}$$

A second motivation is the fact that the discrete energy becomes exactly the Hamiltonian when one uses the exact discrete Lagrangian L_d^E—that is, the discrete Lagrangian is equal to the action integral taken along exact solutions of the Euler–Lagrange equations.

3.3. Symplecticity of the flow. Define the *discrete Lagrangian map* $F_{L_d} : Q \times Q \to Q \times Q$ by

$$(q_0, q_1) \mapsto (q_1, q_2), \tag{59}$$

where q_2 is obtained by using the algorithm from section 3.1. A solution $(\alpha, q_d) \in \mathcal{M}_d$ is formed by iteration of the map F_{L_d}, and it is uniquely determined by the initial condition $(q_0, q_1) \in Q \times Q$ and the choice of timestep h. Hence we parameterize the discrete solutions of the variational principle by the initial conditions (q_0, q_1), and we consider the restriction of $\mathfrak{G}_d$ to that solution space.

The discrete fiber derivatives enable us to push the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ forward to T^*Q. We define the *discrete Hamiltonian map* $\tilde{F}_{L_d} : T^*Q \to T^*Q$ by

$$\tilde{F}_{L_d} = \mathbb{F}^+ L_d \circ F_{L_d} \circ (\mathbb{F}^+ L_d)^{-1}, \tag{60}$$

with the coordinate expression

$$\tilde{F}_{L_d} : (q_0, p_0) \mapsto (q_1, p_1). \tag{61}$$

We note that the discrete Hamiltonian map can be equivalently defined using the other discrete Legendre transform

$$\tilde{F}_{L_d} = \mathbb{F}^- L_d \circ F_{L_d} \circ (\mathbb{F}^- L_d)^{-1}. \tag{62}$$

Define the *restricted discrete action map* $\hat{\mathfrak{G}}_d : Q \times Q \to \mathbb{R}$ to be $\hat{\mathfrak{G}}_d(q_0, q_1) = \mathfrak{G}_d(\alpha, q_d)$, where (α, q_d) is the corresponding solution in $\mathcal{M}_d$ such that $(q_d(t_0), q_d(t_1)) = (q_0, q_1)$. Then (44) becomes

$$d\hat{\mathfrak{G}}_d = (F^N_{L_d})^* \Theta^+_{L_d} - \Theta^-_{L_d}. \tag{63}$$

Taking a further derivative of this expression and using the fact that $d^2\hat{\mathfrak{G}}_d = 0$, we obtain

$$(F^N_{L_d})^*(\Omega_{L_d}) = \Omega_{L_d}, \tag{64}$$

where $\Omega_{L_d} = d\Theta^+_{L_d} = d\Theta^-_{L_d}$ is the unique *discrete Lagrangian symplectic form*, with coordinate expression

$$\Omega_{L_d}(q_0, q_1) = \frac{\partial^2 L_d}{\partial q_0^i \partial q_1^j} dq_0^i \wedge dq_1^j. \tag{65}$$

We have thus proven that the discrete evolution map exactly preserves a discrete symplectic structure, so, regarding F_d as an integrator for the continuous system, we see that it is automatically a symplectic method.

Note that the discrete Lagrangian symplectic form is the pullback under either discrete Legendre transform of the canonical symplectic form on T^*Q. The discrete Hamiltonian map $\tilde{F}_{L_d} : T^*Q \to T^*Q$ thus preserves the canonical symplectic form and the canonical momentum maps on T^*Q.

3.4. Discrete Noether theorem. Consider the (left or right) group action $\Phi : G \times Q \to Q$ of a Lie group G on Q, with infinitesimal generator as defined in section 2.4. This action can be lifted to $Q \times Q$ by the product $\Phi_g(q_0, q_1) = (\Phi_g(q_0), \Phi_g(q_1))$, which has the *infinitesimal generator* $\xi_{Q\times Q} : Q \times Q \to T(Q \times Q)$ given by

$$\xi_{Q\times Q}(q_0, q_1) = (\xi_Q(q_0), \xi_Q(q_1)). \tag{66}$$

The two *discrete Lagrangian momentum maps* $J^+_{L_d}, J^-_{L_d} : Q \times Q \to g^*$ are

$$J^+_{L_d}(q_0, q_1) \cdot \xi = \Theta^+_{L_d} \cdot \xi_{Q\times Q}(q_0, q_1), \tag{67a}$$

$$J^-_{L_d}(q_0, q_1) \cdot \xi = \Theta^-_{L_d} \cdot \xi_{Q\times Q}(q_0, q_1). \tag{67b}$$

As in the continuous approach to Noether's theorem from section 2.4, we are restricted to symmetries of the configuration variables only. We consider symmetries which do not involve altering the time variable and thus consider the timestep h to be a fixed constant.

If a discrete Lagrangian $L_d : Q \times Q \to \mathbb{R}$ is such that $dL_d \cdot \xi = 0$, then L_d is said to be *infinitesimally invariant* under the group action, and Φ is said to be a *symmetry* of the discrete Lagrangian. Note that

$$dL_d \cdot \xi = (\Theta^+_{L_d} - \Theta^-_{L_d}) \cdot \xi_{Q\times Q},$$

and so, when L_d is infinitesimally invariant under the group action Φ, the two discrete momentum maps are equal. In such cases, we will use the notation $J_{L_d} : Q \times Q \to \mathfrak{g}^*$ for the unique single *discrete Lagrangian momentum map*.

Theorem 3.2 (discrete Noether's theorem). *Consider a discrete Lagrangian system $L_d : Q \times Q \times \mathbb{R} \to \mathbb{R}$ which is infinitesimally invariant under the lift of the (left or right) action $\Phi : G \times Q \to Q$. If we assume that the action leaves ∂C invariant (locally), then the corresponding discrete Lagrangian momentum map $J_{L_d} : Q \times Q \to \mathfrak{g}^*$ is a conserved quantity of the discrete Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ so that $J_{L_d} \circ F_{L_d} = J_{L_d}$.*

Proof. We introduce an action of G on the discrete path space $\mathcal{M}_d$ by pointwise action on the configuration components so that $\Phi_g : \mathcal{M}_d \to \mathcal{M}_d$ is given by $\Phi_g(\alpha, q_d) = (\alpha, \Phi_g(q_d))$. Then the infinitesimal generator $\xi_{\mathcal{M}_d} : \mathcal{M}_d \to T\mathcal{M}_d$ is given by

$$\xi_{\mathcal{M}_d}(\alpha, q_d) = (0, \xi_Q(q_0), \dots, \xi_Q(q_{i-1}), \xi_Q(\tilde{q}), \xi_Q(q_i), \dots, \xi_Q(q_N)).$$

From (43) we derive

$$d\mathfrak{G}_d(\alpha, q_d) \cdot \xi_{\mathcal{M}_d}(\alpha, q_d) = \sum_{k=0}^{N-1} dL_d \cdot \xi, \tag{68}$$

and so the space of solutions of the discrete Euler–Lagrange equations is invariant under the action of G, and the Lagrangian map $F_{L_d} : Q \times Q \to Q \times Q$ commutes with the lifted action $\Phi_g : Q \times Q \to Q \times Q$.

Identifying the space of solutions with the space of initial conditions $Q \times Q$ and using (63), we obtain

$$\begin{aligned} d\mathfrak{G}_d(\alpha, q_d) \cdot \xi_{\mathcal{M}_d}(\alpha, q_d) &= d\hat{\mathfrak{G}}_d(q_0, q_1) \cdot \xi_{Q\times Q}(q_0, q_1) \\ &= ((F_{L_d}^N)^*(\Theta_{L_d}^+) - \Theta_{L_d}^-)(q_0, q_1) \cdot \xi_{Q\times Q}(q_0, q_1). \end{aligned}$$

From (68) and the invariance of the discrete Lagrangian, the left-hand side of the previous equation is zero, and so we have

$$(\Theta_{L_d}^+ \cdot \xi_{Q\times Q}) \circ F_{L_d}^N = \Theta_{L_d}^- \cdot \xi_{Q\times Q}. \tag{69}$$

The last relation is simply the statement of preservation of the discrete momentum map, given that for symmetry actions there is only a single unique discrete momentum map and that the above argument holds for all subintervals, including a single timestep. ∎

Observe that J_{L_d} is the pullback under $\mathbb{F}^{\pm}L_d$ of the canonical momentum map J_H on T^*Q and that J_H is thus preserved by $\tilde{F}_{L_d}$.

4. Numerical examples.

In this section, we will choose a particular discrete Lagrangian and illustrate the performance of the algorithm from the previous section on two simple conservative systems. Here we are particularly interested in the extent to which the variational integrator preserves the energy for very long time simulations.

The examples that we present very much simplify the issues regarding grazing impacts and multiple nearby solutions, such as one would encounter in complex collisions (simulation studies of powder flows, for example). Our algorithm, as presented in this paper, is limited to

relatively simple situations, when one can readily identify and resolve the impacts. However, considerable progress has already been made in extending these methods to more practical schemes which are demonstrated in examples involving very complicated collision sequences (see [9]).

4.1. The discrete algorithm. For systems of the form

$$L(q,\dot{q}) = \frac{1}{2}\dot{q}^T M \dot{q} - V(q), \tag{70}$$

where M is a mass matrix and V is a potential function, the Euler–Lagrange equations are given by

$$M\ddot{q} = -\nabla V(q),$$

which is simply Newton's equation of mass times acceleration equals force. We consider the second order discrete Lagrangian

$$L_d(q_0,q_1,h) = \frac{h}{2}\left(\frac{q_1-q_0}{h}\right)^T M\left(\frac{q_1-q_0}{h}\right) - h\left(\frac{V(q_0)+V(q_1)}{2}\right), \tag{71}$$

which is clearly an approximation to the action integral over an interval of length h. The discrete energy function for this choice of discrete Lagrangian is

$$E_d(q_0,q_1,h) = \frac{1}{2}\left(\frac{q_1-q_0}{h}\right)^T M\left(\frac{q_1-q_0}{h}\right) + \left(\frac{V(q_0)+V(q_1)}{2}\right), \tag{72}$$

and the discrete Euler–Lagrange equations are

$$M\frac{q_{k+2}-2q_{k+1}+q_k}{h^2} = -\nabla V(q_{k+1}). \tag{73}$$

Using the discrete Legendre transform (50a), we can push this algorithm forward on T^*Q and obtain a map

$$(q_k,p_k) \mapsto (q_{k+1},p_{k+1})$$

given by

$$\begin{aligned} q_{k+1} &= q_k + hM^{-1}p_k - \frac{h^2}{2}M^{-1}\nabla V(q_k), \\ p_{k+1} &= p_k - \frac{h}{2}\left(\nabla V(q_k) + \nabla V(q_{k+1})\right). \end{aligned}$$

The integrator defined by the previous set of equations is called the leap-frog/Verlet integrator and is one of the most popular integration schemes in molecular dynamics. It is a second order accurate integrator, as one can also infer from the fact that the discrete Lagrangian is second order (see [37] for details about this theory).

This equation describes the motion of the discrete system away from the point of impact. Given a point (q_{i-1}, p_{i-1}) just before impact, we must then solve (47) for $\tilde{q}$ and α, which are

$$M\frac{\tilde{q}-q_{i-1}}{\alpha h} - M\frac{q_{i-1}-q_{i-2}}{h} + (1+\alpha)\frac{h}{2}\nabla V(q_{i-1}) = 0, \tag{74a}$$

$$\tilde{q} \in \partial C. \tag{74b}$$

Next we solve (48) for q_i, which reads

$$\frac{1}{2}\left(\frac{q_i-\tilde{q}}{(1-\alpha)h}\right)^T M\left(\frac{q_i-\tilde{q}}{(1-\alpha)h}\right) - \frac{1}{2}\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)^T M\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)$$

$$+\frac{1}{2}(V(q_i)-V(q_{i-1})) = 0, \tag{75a}$$

$$i^*\left(M\frac{q_i-\tilde{q}}{(1-\alpha)h} - M\frac{\tilde{q}-q_{i-1}}{\alpha h} + \frac{h}{2}\nabla V(\tilde{q})\right) = 0. \tag{75b}$$

To implement the system (75), we write (75b) in a form using Lagrange multipliers. More precisely, we consider ∂C to have a local representation $\partial C = \phi^{-1}(0) \subset Q$, where 0 is a regular point of the constraint function $\phi : Q \to \mathbb{R}$. Then we solve (75a) together with the system

$$M\frac{q_i-\tilde{q}}{(1-\alpha)h} - M\frac{\tilde{q}-q_{i-1}}{\alpha h} + \frac{h}{2}\nabla V(\tilde{q}) + \lambda\nabla\phi(\tilde{q}) = 0 \tag{76}$$

for the unknowns $q_i \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

Finally, we solve for q_{i+1} by (49), which is

$$M\frac{q_{i+1}-q_i}{h} - M\frac{q_i-\tilde{q}}{(1-\alpha)h} + (2-\alpha)\frac{h}{2}\nabla V(q_i) = 0, \tag{77}$$

and we then continue integrating with (73) above.

We can also handle multiple impacts within a single timestep by dividing the impact step into as many substeps as we need and solving (74) and (75) sequentially for any constraint involved in the impact. We will explicitly derive the equations for the case of two impacts solved within the timestep (t_{i-1}, t_i); generalization to an arbitrary number would be immediate.

If multiple impacts are realized in the timestep (t_{i-1}, t_i), then the system (75) will return a solution q_i which is not admissible ($q_i \notin C$). Let us assume that there is only one additional impact in the subinterval $(\tilde{t}, t_i)$ which occurs at the contact point $\tilde{q}' \in \partial C$ and time $\tilde{t}' = \tilde{t} + \beta h$, with $0 < \beta \leq 1-\alpha$. Then the conservation of the discrete energy and the momentum at the impact point $\tilde{q}$, in addition to the condition that $\tilde{q}'$ must lie on ∂C, give the system of $n+1$ equations

$$\frac{1}{2}\left(\frac{\tilde{q}'-\tilde{q}}{\beta h}\right)^T M\left(\frac{\tilde{q}'-\tilde{q}}{\beta h}\right) - \frac{1}{2}\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)^T M\left(\frac{\tilde{q}-q_{i-1}}{\alpha h}\right)$$

$$+\frac{1}{2}(V(\tilde{q}')-V(q_{i-1})) = 0, \tag{78a}$$

$$i^*\left(M\frac{\tilde{q}'-\tilde{q}}{\beta h} - M\frac{\tilde{q}-q_{i-1}}{\alpha h} + (\alpha+\beta)\frac{h}{2}\nabla V(\tilde{q})\right) = 0, \tag{78b}$$

$$\tilde{q}' \in \partial C, \tag{78c}$$

to be solved for $\tilde{q}'$ and β.

Next, the analogous versions of (78a) and (78b) for the second impact point $\tilde{q}'$ give n equations for q_i:

$$\frac{1}{2}\left(\frac{q_i-\tilde{q}'}{(1-\alpha-\beta)h}\right)^T M\left(\frac{q_i-\tilde{q}'}{(1-\alpha-\beta)h}\right)-\frac{1}{2}\left(\frac{\tilde{q}'-\tilde{q}}{\beta h}\right)^T M\left(\frac{\tilde{q}'-\tilde{q}}{\beta h}\right)$$

$$+\frac{1}{2}(V(q_i)-V(\tilde{q}))=0, \tag{79a}$$

$$i^*\left(M\frac{q_i-\tilde{q}'}{(1-\alpha-\beta)h}-M\frac{\tilde{q}'-\tilde{q}}{\beta h}+(1-\alpha)\frac{h}{2}\nabla V(\tilde{q}')\right)=0. \tag{79b}$$

The case of an arbitrary number of impacts is treated in a similar manner, by dividing the timestep (t_{i-1}, t_i) into as many substeps as needed and sequentially solving systems of type (78) to find all the contact points and times. Finally, we solve a system of type (79) for q_i, and then we revert to the standard discrete Euler–Lagrange equations to continue away from the impact.

In the numerical examples, we solve the implicit sets of (74) and (75) with nested Newton loops.

4.2. Particle colliding with a rigid surface. The first example we consider consists of a particle with unit mass moving under gravity in the (x, y)-plane and successively colliding and bouncing on a horizontal rigid floor located at $y = 0$. This simple system has two degrees of freedom (the coordinates of the particle) $q = (x, y)$, the configuration manifold is $Q = \mathbb{R}^2$, and the contact submanifold ∂C is the line $y = 0$. The particle moves with trajectory $q(t) \in \mathbb{R}^2$ in the admissible set $y \geq 0$.

The Lagrangian describing this problem is in the form (70), where M is the diagonal 2×2 mass matrix with diagonal elements (m, m) (m denotes the mass of the particle) and V is the gravitational potential given by

$$V(q) = mgy. \tag{80}$$

Here, g denotes the gravitational acceleration.

The discretization we use is (71), the one for which the variational collision integrator was explicitly derived in the last subsection. The integrator is run with a step size of $h = 0.01$; the initial conditions we used in the simulation are $q_0 = (0, 1)$ and $\dot{q}_0 = (-2, 0)$. We considered a unitary mass particle ($m = 1$).

The energy behavior in this case is shown in Figure 2 for a relatively large number of impacts (1000 impacts). The same pattern is observed if the simulation is carried out for essentially arbitrarily long times. This fluctuating energy behavior is typical of symplectic methods. A detailed account on how the variational symplectic methods perform on smooth conservative systems can be found in [26].

4.3. Rotating nonconvex rigid body colliding with a rigid surface. Now consider a sequence of collisions and bounces on a horizontal rigid floor for a three-degree-of-freedom system, namely, a rotating four-point star-shaped rigid body (see Figure 3) moving in a plane. The convex hull of the star-shaped body is a square with sides of length L. The rigid

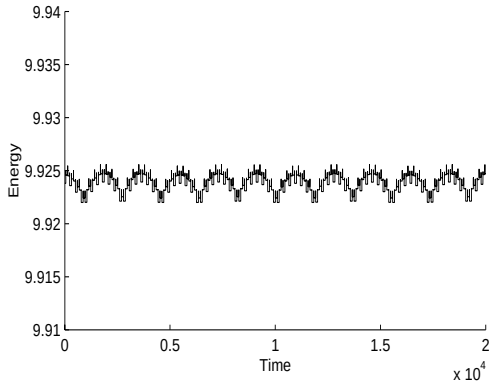

Figure 2. *The long-time energy behavior for a particle bouncing on a rigid floor.*

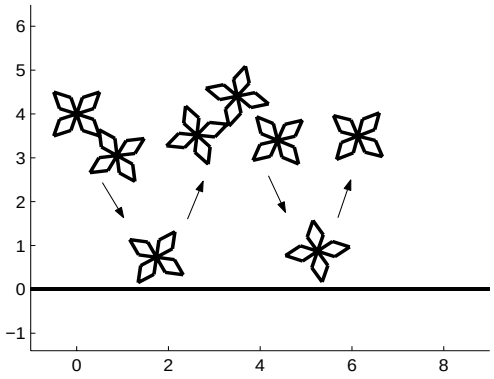

Figure 3. *A rotating four point, star shaped rigid body colliding and bouncing on a horizontal rigid floor.*

body moves under the gravitational force field in the vertical (x, y) plane. The configuration manifold Q is $SE(2)$ with local coordinates $q = (x, y, \theta)$, where $(x, y) \in \mathbb{R}^2$ stand for the coordinates of the center of mass and $\theta \in [0, 2\pi]$ stands for the oriented angle that a line moving rigidly with the body makes with the horizontal axis. The contact set ∂C given by the nonpenetration condition is given explicitly by

$$y = \frac{L}{2}\left(|\sin\theta| + |\cos\theta|\right). \tag{81}$$

The subset of points where $y \geq \frac{L}{2}\left(|\sin\theta| + |\cos\theta|\right)$ represents the admissible set $C \subset Q$, and contact occurs whenever the relation becomes an equality.

The Lagrangian describing this problem has the expression (70), where V is the gravitational potential (80) and M is the diagonal 3×3 mass matrix with diagonal elements (m, m, I), where m is the mass of the body and I is the moment of inertia of the star-shaped body with respect to the z-axis through its center of symmetry. In terms of m and L, I is given by $I = \frac{29}{192}mL^2$.

We use again the discretization given by (71) and run the variational collision integrator from section 4.1 with a timestep $h = 0.005$ and initial conditions $q_0 = (0, 3.5, 0)$ and $\dot{q}_0 =$

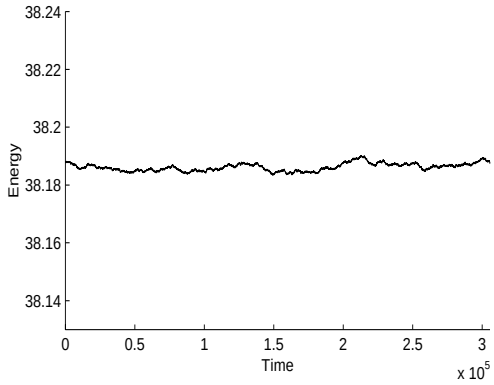

Figure 4. *The long-time energy behavior for a star-shaped rigid body bouncing on a rigid floor. Note the fluctuating energy behavior typical of symplectic methods.*

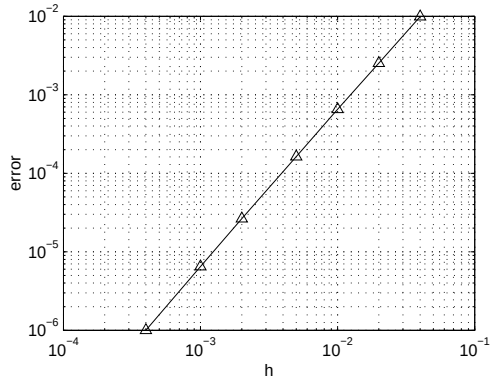

Figure 5. *Log-log error diagram for the method in the star bounce example, after one collision. The second order accuracy of the integrator from the smooth setting is preserved through collision. In this case a nonlinear gravity was used to avoid the degeneracy of the exactly integrable linear gravitational potential.*

$(-2, 0, 5)$. We considered the body to have unitary mass $m = 1$ and a square convex hull of size $L = 1$. A long-time (1,500 impacts) energy plot is shown in Figure 4. The long-time energy behavior appears to be reasonably stable. It is not clear from these numerical experiments whether this is an indication of a nearby conserved energy, as exists for variational integrators applied to smooth systems, or simply a fairly stable random walk. More numerical investigations and analytical work are needed to resolve this question.

We numerically checked the order of accuracy of the algorithm for the star bounce example. A log-log error diagram after one collision is presented in Figure 5. The numerical results show that the integrator is second order accurate; i.e., the order of the method is the same as the order of the discrete Lagrangian L_d. This is in fact a fundamental property of the variational integrators developed in smooth settings (see [37]), and we believe that it extends to the nonsmooth setting as well. In our future work on the subject, we intend to formulate and prove such results for the variational collisional algorithms presented in this paper.

5. Appendix: Nonsmooth analysis approach. The purpose of this appendix is to discuss collisions with multibody nonsmooth contact geometries when the contact set ∂C has a large number of singularities. For example, in granular flows or fragmentation of brittle solids, there are a large number of fragments undergoing complex collision sequences. For these collisions, situations like corner-to-corner contact are very likely to occur, and the variational algorithm from section 3.1 cannot cope with contact in singular points of the contact set ∂C.

However, the nonsmooth analysis (see [10]) provides an efficient analytical tool to formulate and treat algorithmically complex contact situations, as shown in [25]. The goal of this appendix is to combine discrete Lagrangian mechanics with nonsmooth calculus to derive a variational formulation of the nonsmooth contact (in the sense of nonsmooth admissible configuration sets). The symplectic nature of such an algorithm is poorly understood, but one can conjecture that future theory on that would depend on approaches like this one.

If C is the admissible set (possibly nonsmooth and nonconvex) of the system, we must have $q(t) \in C$ for all times or, in the discrete case, $q_k \in C$ for all k. These constraints may be enforced by adding to the Lagrangian the indicator function I_C of C defined by

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases} \tag{82}$$

In the discrete context, this translates into defining a constrained discrete Lagrangian $\tilde{L}_d$ by adding contributions from the indicator function. One particular way to do this is

$$\tilde{L}_d(q_k, q_{k+1}, h) = L_d(q_k, q_{k+1}, h) - \frac{1}{2}\left[I_C(q_k) + I_C(q_{k+1})\right]. \tag{83}$$

We use, as before, the variational principle of Hamilton to derive the discrete equations of motions. Thus the discrete Euler–Lagrange equations become

$$D_2L_d(q_{k-1}, q_k, h) + D_1L_d(q_k, q_{k+1}, h) - \partial I_C(q_k) \ni 0, \tag{84}$$

where ∂I_C denotes the generalized gradient of the indicator function.

For points q in the interior of C, $\partial I_C(q) = \{0\}$, while for points on the boundary of C, $\partial I_C(q) = N_C(q)$, where $N_C(q)$ represents the normal cone to C at q defined in the nonsmooth analysis framework (see [10] for a complete account of the nonsmooth calculus used here). However, if q is a convex point, $N_C(q)$ reduces to the normal cone in the usual convex analysis sense.

The constrained discrete equations (84) are thus the usual discrete Euler–Lagrange equations (46) away from the impact. The generalized gradient ∂I_C is not trivial only for $\tilde{q} \in \partial C$. If we specialize (84) for points $q_{i-1}, \tilde{q}$, and q_{i+1}, then we obtain

$$D_2L_d(q_{i-1}, \tilde{q}, \alpha h) + D_1L_d(\tilde{q}, q_{i+1}, (1-\alpha)h) - \partial I_C(\tilde{q}) \ni 0, \tag{85}$$

which is a natural generalization of (48b) in the case when $\tilde{q}$ is a singular point of ∂C.

Alternatively, using the previous notation for discrete momenta (51), (85) can be written as

$$p^+(q_{i-1}, \tilde{q}, \alpha h) - p^-(\tilde{q}, q_{i+1}, (1-\alpha)h) \in N_C(\tilde{q}), \tag{86}$$

where we used $\partial I_C(\tilde{q}) = N_C(\tilde{q})$ for $\tilde{q} \in \partial C$. Therefore, (86) generalizes (54) in the case when $\tilde{q}$ is a corner of ∂C and we cannot define a tangent plane at that point.

For Lagrangians consisting of only kinetic energy and for the particular discretization (57), the momentum conservation (85) leads to a very interesting geometrical interpretation. Indeed, for a unitary mass matrix, (85) becomes

$$\frac{\tilde{q} - q_{i-1}}{\alpha h} - \frac{q_i - \tilde{q}}{(1-\alpha)h} - \partial I_C(\tilde{q}) \ni 0. \tag{87}$$

The inclusion (87) can be rewritten in the form

$$(1-\alpha)q_{i-1} + \alpha q_i \in (I - \partial I_C)(\tilde{q}). \tag{88}$$

Now we will connect (88) with the concept of the *closest point projection*. First recall the definition of the *resolvent* of the set-valued operator ∂I_C as

$$R = (I + \partial I_C)^{-1}. \tag{89}$$

It is a well-known fact that the resolvent of the subgradient of the indicator function of a convex set is the closest-point projection onto that set (see [49]). Under the assumption that $\bar{C}$, the complement of C, is a convex set (see Figure 6), the inclusion (88) can be written as

$$\tilde{q} = P_{\bar{C}}(q_{i-\alpha}), \tag{90}$$

where $P_{\bar{C}}$ represents the closest-point projection operator onto $\bar{C}$ and $q_{i-\alpha}$ is the convex combination of the points q_{i-1} and q_i

$$q_{i-\alpha} = (1-\alpha)q_{i-1} + \alpha q_i.$$

We will conclude this appendix by the following two remarks which re-emphasize the particular benefit of the nonsmooth calculus approach.

Remark. Besides its theoretical attractiveness, the nonsmooth analysis approach has a great advantage over the standard penalty formulation methods in dealing with complex nonsmooth contact geometries (see [25]) where neither normals nor gap functions may be defined. Indeed, for such problems penalty methods simply fail.

Remark. The nonsmooth approach also gives the natural framework for constructing time-adaptive variational integrators for collisions (see [24] and [37]), but we will leave the development and illustration of such contact algorithms for future work.

6. Future directions.

Order of accuracy. For systems without collisions, the order of accuracy of the discrete Lagrangian L_d and the discrete Hamiltonian map $\tilde{F}_d$ are the same. In principle, this will also be true for contact algorithms as developed in this paper (see the numerical results presented in Figure 5), but precise proofs remain to be formulated.

Elastic bodies. Although the numerical simulations presented in section 4 were all for rigid-body collisions, the discrete variational formalism applies for arbitrary potential energies, such as those for hyperelastic materials. We have not yet tested these methods for such systems, however. In this context, it will also be very interesting to use the techniques of section 2.5 to include external forces and dissipative effects.

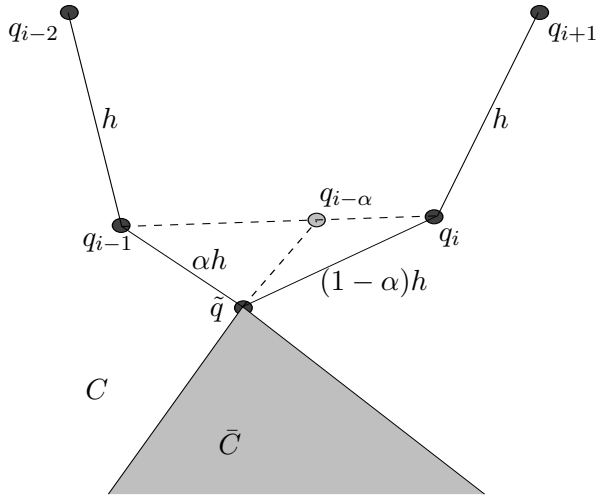

Figure 6. *Collision at a singular point of the contact set ∂C.*

Multisymplectic extensions to PDEs . While PDE contact systems can be first discretized in space and then treated as a system of contact ODEs in time with the algorithms developed in this paper, much greater understanding can be gained by a fully space-time variational formulation of both the continuous and discrete problems. The framework of multisymplectic mechanics [15] and multisymplectic discretizations [35] is particularly appropriate for this, and we will treat this subject in a forthcoming paper.

REFERENCES

[1] R. Abraham, J. E. Marsden, and T. S. Ratiu, *Manifolds, Tensor Analysis and Applications*, 2nd ed., Appl. Math. Sci. 75, Springer-Verlag, New York, 1988.
[2] M. Anitescu, F. A. Potra, and D. E. Stewart, *Time-stepping for three-dimensional rigid body dynamics*, Comput. Methods Appl. Mech. Engrg., 177 (1999), pp. 183–197.
[3] F. Armero and E. Petöcz, *A new dissipative time-stepping algorithm for frictional contact problems: Formulation and analysis*, Comput. Methods Appl. Mech. Engrg., 179 (1999), pp. 151–178.
[4] E. Barth, B. Leimkuhler, and S. Reich, *A time-reversible variable-stepsize integrator for constrained dynamics*, SIAM J. Sci. Comput., 21 (1999), pp. 1027–1044.
[5] B. Brogliato, S. Niculescu, and P. Orhant, *On the control of finite dimensional mechanical systems with unilateral constraints*, IEEE Trans. Automat. Control, 42 (1997), pp. 200–215.
[6] B. Brogliato, *Nonsmooth Impact Mechanics: Models, Dynamics and Control*, Lecture Notes in Control and Inform. Sci. 220, Springer-Verlag, New York, 1996.
[7] B. Brogliato, *On the control of non-smooth complementarity dynamical systems*, R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci., 359 (2001), pp. 2369–2383.
[8] N. J. Carpenter, R. L. Taylor, and M. G. Katona, *Lagrange constraints for transient finite-element surface-contact*, Internat. J. Numer. Methods Engrg., 32 (1991), pp. 103–128.
[9] F. Cirak and M. West, *Scalable Explicit Collision Response Using Momentum Decomposition*, manuscript.
[10] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[11] A. Dontchev and F. Lempio, *Difference methods for differential inclusions: A survey*, SIAM Rev., 34 (1992), pp. 263–294.

[12] A. F. Filippov, *Differential equations with discontinuous right-hand side*, American Mathematical Society Translations, 42 (1964), pp. 199–231.
[13] A. F. Filippov, *Classical solutions of differential equations with multivalued right-hand side*, SIAM J. Control, 5 (1967), pp. 609–621.
[14] A. F. Filippov, *Differential Equations with Discontinuous Right-Hand Side*, Kluwer Academic, Norwell, MA, 1988.
[15] M. Gotay, J. Isenberg, and J. E. Marsden, *Momentum Maps and the Hamiltonian Structure of Classical Relativistic Field Theories* I, http://www.cds.caltech.edu/~marsden/ (1997).
[16] E. Hairer and C. Lubich, *Long-time energy conservation of numerical methods for oscillatory differential equations*, SIAM J. Numer. Anal., 38 (2000), pp. 414–441.
[17] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations* I: *Nonstiff problems*, 2nd ed., Springer Ser. Comput. Math. 8, Springer-Verlag, New York, 1993.
[18] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations* II: *Stiff and Differential-Algebraic Problems*, 2nd ed., Springer Ser. Comput. Math. 14, Springer-Verlag, New York, 1996.
[19] M. R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley & Sons, New York, 1966.
[20] Y. A. Houndonougbo and B. B. Laird, *Constant-temperature molecular dynamics algorithms for mixed hard-core/continuous potentials*, J. Chem. Phys., 117 (2002), pp. 1001–1009.
[21] Y. A. Houndonougbo, B. B. Laird, and B. J. Leimkuhler, *A molecular dynamics algorithm for mixed hard-core/continuous potentials*, Mol. Phys., 98 (2000), pp. 309–316.
[22] A. Iserles, H. Munthe-Kaas, and A. Zanna, *Lie group methods*, in Acta Numerica, Acta Numer. 9, Cambridge University Press, Cambridge, UK, 2000, pp. 215–365.
[23] M. Jean, *Unilateral contact and dry friction: Time and space discrete variables formulation*, Arch. Mech. (Arch. Mech. Stos.), 40 (1988), pp. 677–691.
[24] C. Kane, J. E. Marsden, and M. Ortiz, *Symplectic-energy-momentum preserving variational integrators*, J. Math. Phys., 40 (1999), pp. 3353–3371.
[25] C. Kane, E. A. Repetto, M. Ortiz, and J. E. Marsden, *Finite element analysis of nonsmooth contact*, Comput. Methods Appl. Mech. Engrg., 180 (1999), pp. 1–26.
[26] C. Kane, J. E. Marsden, M. Ortiz, and M. West, *Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems*, Internat. J. Numer. Methods Engrg., 49 (2000), pp. 1295–1325.
[27] M. Kunze and D. P. Monteiro Marques, *An introduction to Moreau's sweeping process*, in Impacts in Mechanical Systems, B. Brogliato, ed., Lecture Notes in Phys. 551, Springer-Verlag, Berlin, 2000, pp. 1–61.
[28] C. Y. Lee and J. T. Oden, *Theory and approximation of quasi-static frictional contact problems*, Comput. Methods Appl. Mech. Engrg., 106 (1993), pp. 407–429.
[29] B. Leimkuhler and S. Reich, *A reversible averaging integrator for multiple time-scale dynamics*, J. Comput. Phys., 171 (2001), pp. 95–114.
[30] F. Lempio and V. Veliov, *Discrete approximations of differential inclusions*, Bayreuth. Math. Schr., 54 (1998), pp. 149–232.
[31] A. Lew, J. E. Marsden, M. Ortiz, and M. West, *Asynchronous variational integrators*, Arch. Ration. Mech. Anal., 167 (2003), pp. 85–146.
[32] M. Mabrouk, *A unified variational model for the dynamics of perfect unilateral constraints*, Eur. J. Mech. A Solids, 17 (1998), pp. 819–842.
[33] M. D. P. Monteiro Marques, *Chocs inélastiques standards: Un rèsultat d'existence*, Sèm. Anal. Convexe, 15 (1985), pp. 1–32.
[34] D. P. Monteiro Marques, *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*, Progress. in Nonlinear Differential Equations and Their Applications, 9, Birkhäuser Verlag, Basel, 1993.
[35] J. E. Marsden, G. W. Patrick, and S. Shkoller, *Multisymplectic geometry, variational integrators and nonlinear PDEs*, Comm. Math. Phys., 199 (1998), pp. 351–395.
[36] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Texts Appl. Math. 17, Springer-Verlag, New York, 1994.

[37] J. E. Marsden and M. West, *Discrete mechanics and variational integrators*, in Acta Numerica, Acta Numer. 10, Cambridge University Press, Cambridge, UK, 2001, pp. 357–514.
[38] J.-J. Moreau, *Une formulation du contact à frottement sec; application au calcul numérique*, C. R. Acad. Sci. Paris Sér. II, 302 (1986), pp. 799–801.
[39] J.-J. Moreau, *Unilateral contact and dry friction in finite freedom dynamics*, in Nonsmooth Mechanics and Applications, J.-J. Moreau and P. D. Panagiotopoulos, eds., CISM Courses and Lectures 302, Springer-Verlag, Vienna, 1988, pp. 1–82.
[40] J.-J. Moreau, *Numerical aspects of the sweeping process*, Comput. Methods Appl. Mech. Engrg., 177 (1999), pp. 329–349.
[41] A. Pandolfi, C. Kane, J. E. Marsden, and M. Ortiz, *Time-discretized variational formulation of nonsmooth frictional contact*, Internat. J. Numer. Methods Engrg., 53 (2002), pp. 1801–1829.
[42] L. Paoli and M. Schatzman, *Schéma numérique pour un modèle de vibrations avec constraints unilatérale et perte d'énergie aux impacts, en dimension finie*, C. R. Acad. Sci. Paris Sér. I Math., 317 (1993), pp. 211–215.
[43] L. Paoli and M. Schatzman, *Mouvement à un nombre fini de degrés de liberté avec contraintes unilatérales: Cas avec perte d'énergie*, RAIRO Modél. Math. Anal. Numér., 27 (1993), pp. 673–717.
[44] L. Paoli and M. Schatzman, *Approximation et existence en vibro-impact*, C. R. Acad. Sci. Paris Sér. I Math., 329 (1999), pp. 1103–1107.
[45] D. Peric and D. R. J. Owen, *Computational model for* 3*-d contact problems with friction based on the penalty method*, Internat. J. Numer. Methods Engrg., 35 (1992), pp. 1289–1309.
[46] F. Pfeiffer, *Unilateral problems of dynamics*, Archive of Applied Mechanics, 69 (1999), pp. 503–527.
[47] F. Pfeiffer and C. Glocker, *Dynamics of Rigid Body Systems with Unilateral Constraints*, Wiley, New York, 1996.
[48] E. B. Pires and J. T. Oden, *Analysis of contact problems with friction under oscillating loads*, Comput. Methods Appl. Mech. Engrg., 39 (1983), pp. 337–362.
[49] R. T. Rockafellar, *Convex Analysis*, Princeton Math. Ser. 28, Princeton University Press, Princeton, NJ, 1970.
[50] J. M. Sanz-Serna and M. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall, London, 1994.
[51] M. Schatzman, *Sur une classe de problèmes hyperboliques non linéaires*, C. R. Acad. Sci. Paris Sér. A-B, 277 (1973), pp. A671–A674.
[52] M. Schatzman, *A class of nonlinear differential equations of second order in time*, Nonlinear Anal., 2 (1978), pp. 355–373.
[53] M. Schatzman, C.-H. Lamarque, and J. Bastien, *An ill-posed mechanical problem with friction*, Eur. J. Mech. A Solids, 18 (1999), pp. 415–420.
[54] J. C. Simo, P. Wriggers, and R. L. Taylor, *A perturbed Lagrangian formulation for the finite-element solution of contact problems*, Comput. Methods Appl. Mech. Engrg., 50 (1985), pp. 163–180.
[55] D. E. Stewart, *Convergence of a time-stepping scheme for rigid-body dynamics and resolution of Painleve's problem*, Arch. Ration. Mech. Anal., 145 (1998), pp. 215–260.
[56] D. E. Stewart, *Rigid-body dynamics with friction and impact*, SIAM Rev., 42 (2000), pp. 3–39.
[57] R. L. Taylor and P. Papadopoulos, *On a finite-element method for dynamic contact impact problems*, Internat. J. Numer. Methods Engrg., 36 (1993), pp. 2123–2140.
[58] L. White and J. T. Oden, *Dynamics and control of viscoelastic solids with contact and friction effects*, Nonlinear Anal., 13 (1989), pp. 459–474.
[59] P. Wriggers, T. V. Van, and E. Stein, *Finite-element formulation of large deformation impact-contact problems with friction*, Comput. & Structures, 37 (1990), pp. 319–331.
[60] P. Wriggers and G. Zavarise, *Application of augmented Lagrangian techniques for nonlinear constitutive laws in contact interfaces*, Comm. Numer. Methods Engrg., 9 (1993), pp. 815–824.
[61] L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory*, W. B. Saunders Company, Philadelphia, 1969, corrected printing, Chelsea, UK, 1980.