Discontinuous variational time integrators for complex multibody collisions

G. Johnson¹,‡, S. Leyendecker²,*,# and M. Ortiz¹

¹Computational Mechanics Group, California Institute of Technology, Mail Code 105-50, Pasadena, CA 91125, USA
²Chair of Applied Dynamics, University of Erlangen-Nuremberg, Haberstrasse 1, D-91058 Erlangen, Germany

SUMMARY

The objective of the present work is to formulate a new class of discontinuous variational time integrators that allow the system to adopt two possibly different configurations at each sampling time $t_k$, representing predictor and corrector configurations of the system. The resulting sequence of configuration pairs then represents a discontinuous—or non-classical—trajectory. Continuous or classical trajectories are recovered simply by enforcing a continuity constraint at all times. In particular, in systems subject to one-sided contact constraints simulated via discontinuous variational time integrators, the predictor configuration is not required to satisfy the one-sided constraints, whereas the corrector configuration is obtained by a closest-point projection (CPP) onto the admissible set. The resulting trajectories are generally discontinuous, or non-classical, but are expected to converge to classical or continuous solutions for decreasing time steps. We account for dissipation, including friction, by means of a discrete Lagrange–d’Alembert principle, and make extensive use of the spacetime formalism in order to ensure exact energy conservation in conservative systems, and the right rate of energy decay in dissipative systems. The structure, range and scope of the discontinuous variational time integrators, and their accuracy characteristics are illustrated by means of examples of application concerned with rigid multibody dynamics. Copyright © 2014 John Wiley & Sons, Ltd.

Received 27 August 2013; Revised 27 June 2014; Accepted 18 July 2014

KEY WORDS: contact; dynamical systems; Lagrangian; time integration; implicit; variational methods

1. INTRODUCTION

Discrete Lagrangian mechanics (cf., e.g., [1, 2] and references therein) supplies a self-contained theory of dynamics of mechanical systems in which time is a discrete variable ab initio. Because of the discrete nature of time, discrete Lagrangian mechanics provides a basis for the formulation of time-integration algorithms, a connection that has been extensively investigated in the past, including specifically systems with one-sided contact constraints [3–5]. Because of its Lagrangian structure, discrete Lagrangian mechanics results in variational time integrators with exact conservation properties. However, the enforcement of contact constraints is cumbersome in situations involving many bodies undergoing complex collision sequences, as each individual collision needs to be resolved in time. On an entirely parallel track, incremental energy-dissipation functionals [6] combining inertia, potential energy, and dissipation provide an alternative characterization of discrete trajectories by means of a sequence of minimum problems. The connection between these two paradigms has been investigated in [7]. Within the incremental energy minimization framework, one-sided constraints

*Correspondence to: S. Leyendecker, Chair of Applied Dynamics, University of Erlangen-Nuremberg, Haberstrasse 1, D-91058 Erlangen, Germany.
#E-mail: sigrid.leyendecker@ltd.uni-erlangen.de
‡Presently Research Fellow, University of Texas at Austin/NASA Johnson Spaceflight Center.

Copyright © 2014 John Wiley & Sons, Ltd.
can be accounted for by means of nonsmooth calculus [8] and closest-point projections [9–11]. The resulting time integrators can deal efficiently with large many-body systems undergoing complex collision sequences, as multiple collisions can be accounted for within a single time step. However, in general, incremental minimization does not possess a Lagrangian structure, and the resulting algorithms lack exact conservation properties.

The objective of the present work is to formulate a new class of *discontinuous variational (DV)* time integrators that combine the best attributes of variational and incremental minimization algorithms. We particularly wish to account for collisions involving multibody contact between many—possibly nonsmooth—bodies. One aspect of incremental minimization and projection methods [6, 9–11] that renders them especially well suited to such problems is their predictor–corrector structure. Thus, in those approaches, the updated configuration is reached in two steps, namely, an unconstrained predictor step followed by a corrector step in the form of a projection to the admissible set. In order to build this structure into variational integrators, we allow the system to adopt two possibly different configurations \( (q^+_k, q^-_k) \) at each sampling time \( t_k \), representing predictor and corrector configurations of the system. The sequence \( (q^+_k) \) of configuration pairs then represents a discontinuous—or non-classical—trajectory of the system. Within this extended framework, continuous or classical trajectories are recovered simply by enforcing the continuity constraint \( q^-_k = q^+_k \) at all times.

The full benefit of the DV time integrators is reaped for systems subject to one-sided contact constraints requiring the system to remain within a certain admissible set \( C \). For these systems, we relax the continuity constraint and require, instead, that \( q^-_k \in C \) and \( q^+_k = P_C(q^+_k) \), where \( P_C \) is the closest-point projection onto the admissible set \( C \). In this setting, we regard \( q^+_k \) as a trial or predictor configuration, not required to satisfy the one-sided constraints, and \( q^-_k \) as the corresponding corrector configuration, required to satisfy the one-sided constraints. We note that, if the predictor configuration is admissible, that is, if \( q^-_k \in C \), then \( q^-_k = q^+_k = P_C(q^+_k) \) and the classical continuity constraint is recovered. By contrast, when the predictor configuration is inadmissible, that is, if \( q^-_k \notin C \), the trajectory becomes discontinuous, or non-classical.

In systems with dissipation, the trajectories of the system are characterized by means of a discrete Lagrange–d’Alembert principle (e.g., [2]). Following [12], we make extensive use of the *spacetime formalism* in order to ensure exact energy conservation over discontinuities in conservative systems, and the right rate of energy decay in dissipative systems. In the spacetime formulation, time is regarded as an additional generalized coordinate. The spacetime formulation of dissipative dynamics is particularly revealing as it effectively provides a unified framework that encompasses both the equations of motion and the energy equation. The spacetime formalism not only affords considerable economy of notation, by placing the space and time coordinates on an equal footing, but also, more importantly, becomes essential in the presence of collisions, where energy conservation plays a fundamental role, and in systems with dissipation, where the energy equation is no longer redundant. Remarkably, as in the conservative case [12], the spacetime formalism supplies a precise statement of discrete energy balance, as well as the means to determine discrete times ensuring that energy is dissipated at exactly the right rate. It also bears emphasis that, in the discrete spacetime formulation, the discretizations of the equations of motion and the energy equation are not arbitrary but follow jointly as spatial and temporal sections of a single spacetime discretization.

The DV integrators, thus defined, allow for discontinuous trajectories in the presence of one-sided contact constraints, thus formalizing and placing predictor–corrector collision algorithms within a Lagrangian framework. In particular, the discontinuous or non-classical trajectories allow the predictor configurations to violate the one-sided constraints. However, the extent of this violation, that is, the extent of penetration of the predictors into the inadmissible region, may be expected to decrease with decreasing time step and to vanish entirely in the limit. In this manner, discontinuous or non-classical trajectories are expected to converge to classical trajectories of the time-continuous system. The structure, range and scope of the DV time integrators, and their accuracy and convergence characteristics are illustrated by way of a number of examples of application concerned with the dynamics of systems of rigid bodies undergoing complex collisions in the sense that the bodies are three-dimensional possibly nonconvex with possibly nonsmooth boundaries (having edges and corners) chained together by joints in multibody systems. Then, (elastic or inelastic) collisions do
affect the rotational as well as translational motion, and furthermore, frictional effects can be taken into account. In developing these examples, we make extensive use of the constrained formulation of rigid body dynamics [13], further developed in [14, 15], and the supporting separating hyperplane (SSH) representation [16] of the corresponding one-sided contact constraints. The versatility and good conservation and accuracy properties of the DV time integrators are amply demonstrated by these examples.

2. TIME-CONTINUOUS FORMULATION

In this section, we summarize the formulation of constrained Lagrangian mechanics that provides the basis for subsequent developments (cf., e.g., [3–5, 9–11, 17] for additional background). We particularly wish to account for collisions involving multibody contact between possibly nonsmooth bodies. For these systems, situations such as corner-to-corner contact are likely to occur, and smooth analysis requiring on differentiable domain boundaries with uniquely defined normals fails to apply in general. By contrast, nonsmooth analysis (cf., e.g., [8, 10] and Appendix A) provides an efficient analytical tool capable of handling nonsmooth collisions that couples well with the variational methods of Lagrangian mechanics.

2.1. Unconstrained systems

We begin by considering an unconstrained system characterized by configurations described by points on a smooth manifold \( Q \), and a Lagrangian \( L : TQ \to \mathbb{R} \). The system executes motions \( q : \mathbb{R} \to Q \) in some suitable space of trajectories. Let \( \mathcal{E} \) denote the collection of all open bounded intervals of \( \mathbb{R} \). Then, the unconstrained action functional \( S : X \times \mathcal{E} \to \mathbb{R} \equiv \mathbb{R} \cup \{-\infty, +\infty\} \) is defined as

\[
S(q, A) = \int_A L(q(t), \dot{q}(t)) \, dt.
\]

Furthermore, if \( S(\cdot, A) \) is differentiable, its first variation is given by

\[
\delta S(q, \varphi, A) = \int_A \left( \frac{\partial L}{\partial q}(q(t), \dot{q}(t)) \cdot \varphi(t) + \frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t)) \cdot \varphi(t) \right) \, dt,
\]

where \( (q, \varphi) : \mathbb{R} \to TQ \) is an admissible variation. The physical trajectories obey Hamilton’s principle of stationary action

\[
\delta S(q, \varphi, A) = 0
\]

for all admissible variations \( \varphi \) and open sets \( A \in \mathcal{E} \). Smooth trajectories, if they exist, satisfy the Euler–Lagrange equations

\[
- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q}(q(t), \dot{q}(t)) = 0,
\]

which may be regarded as a balance of inertial and energetic forces.

---

\( ^1 \)For general background on manifolds, confirm for example, [18]

\( ^2 \)Following conventional notation, we use the same symbol \( q \) to variously denote a point of \( Q \), representing a configuration of the system, and a trajectory \( q : \mathbb{R} \to Q \), the precise meaning of the symbol to be disambiguated by the context.

\( ^3 \)Confirm for example, [5, 17, 19] for functional analytical frameworks for Lagrangian mechanics.

\( ^* \)Here and subsequently, we proceed formally and assume that invoked mathematical operations are well defined.

---

DOI: 10.1002/nme
2.2. The embedded case

In the applications considered here, and often in practice, the configuration manifold \( Q \) is embedded in \( \mathbb{R}^n \), that is, \( Q \) is a smooth submanifold of \( \mathbb{R}^n \) of dimension \( \dim(Q) < n \). We recall that the conormal bundle \( N^*Q \) consists of pairs \((q, f_Q) \in Q \times \mathbb{R}^n\) such that \( f_Q \) is orthogonal to \( T_qQ \) in \( \mathbb{R}^n \). If \( q \) is a stationary point of the action, then the embedded trajectory (not renamed) in \( \mathbb{R}^n \) satisfies the constrained Euler–Lagrange equations

\[
- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + f_Q(t) = 0, \tag{5}
\]

where \((q, f_Q) : \mathbb{R} \to N^*Q\) is determined so that \( q(t) \in Q \) at all times. In particular, owing to the normality of the confining forces \( f_Q \), we have

\[
f_Q(t) \cdot \dot{q}(t) = 0 \tag{6}
\]

at all times, that is, the confining forces do no work. Equation (5) expresses a balance between inertial forces, energetic forces, and normal reactions that keep the trajectory within the embedded configurational manifold.

In the applications that follow, concerned with rigid body dynamics, and often in practice, the embedded configurational manifold \( Q \) is defined implicitly by \( N_Q \equiv \mathbb{R}^n / \text{dim}(Q) = \text{codim}(Q) \) holonomic, scleronomous constraints

\[
g_Q(q) = 0, \tag{7}
\]

expressed in terms of a smooth constraint function \( g_Q : \mathbb{R}^n \to \mathbb{R}^{N_Q} \). In particular, the Jacobian \( G_Q(q) = Dg_Q(q) \) has full rank \( N_Q \) everywhere in \( Q \) and, therefore, every normal reaction \( f_Q \in N^*_qQ \) at \( q \in Q \), can be expressed as

\[
f_Q = G_Q^T(q)\lambda_Q, \tag{8}
\]

where \( \lambda_Q : \mathbb{R} \to \mathbb{R}^{N_Q} \) are Lagrange multipliers \([2, 4]\). Using this representation, the constrained Euler–Lagrange Equation (5) can be expressed in the form

\[
- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + G_Q^T(q(t))\lambda_Q(t) = 0. \tag{9}
\]

These equations of motion, in conjunction with the constraint Equation (7), define a system of equations to be solved jointly for the embedded trajectory \( q \) and the Lagrange multipliers \( \lambda_Q \). Let the \((n \times (n - N_Q))\)-matrix \( P(q) \) be a so called null space matrix whose columns form a basis for the tangent space of the constraint manifold, that is, \( \text{range}(P(q)) = \text{null}(G_Q(q)) \) holds. Premultiplication of (9) by \( P^T \) eliminates the normal reactions from the system. Correspondingly, the conjugate momentum \( p = \frac{\partial L}{\partial q} \) obtained via the Legendre transform can be reduced to

\[
p_P = P^T(q)p, \tag{10}
\]

and using the reduced mass matrix \( M_{red}(q) = P^T(q)M(q)P(q) \) with \( M \) being the mass matrix corresponding to the redundant state, the kinetic energy \( \frac{1}{2} \dot{q}^TM\dot{q} \) can be written as \( \frac{1}{2} p^T M_{red}^{-1} p \) along solutions of (7) and (9).
2.3. Systems with one-sided contact constraints

Suppose that the motion of the system is further constrained by non-interpenetration or contact constraints restricting the configurations of the system to a subset \( C \subset Q \). Begin by supposing further that the trajectory is differentiable over an open interval of time \( \mathcal{E} \). Then, the equations of motion take the form

\[
- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + f_Q(t) + f_C(t) = 0, \tag{11}
\]

where the contact forces \((q, f_C) : A \rightarrow NC \subset T^*Q\) are determined so that \( q \in C \) at all times. Here, \( N_qC \) denotes the relative normal cone \( N_qC \subset T_q^*Q \) to \( C \) at \( q \) in the sense of nonsmooth analysis (cf. [8], Appendix A). We recall that the relative normal cone \( N_qC \) is related to the relative tangent cone \( T_qC \) by polarity, and that \( T_qC \) is the set of all possible outgoing velocities from \( q \in C \) into \( T_qQ \). This normality and tangentiality conditions can be expressed jointly in Kuhn–Tucker form as

\[
\dot{q} (t) \in T_{q(t)}C, \quad f_C(t) \in N_{q(t)}C, \quad f_C(t) \cdot \dot{q}(t) = 0, \tag{12}
\]

which, in particular, state that the contact forces do no work. Furthermore, if \( q(t) \) is in the interior of \( C \), then \( f_C(t) = 0 \), that is, the contact forces can only be nonzero if \( q(t) \) is on the boundary of \( C \).

Suppose, in addition, that the admissible set \( C \) is defined by a set of one-sided constraints

\[
g_C(q) \leq 0, \tag{13}
\]

where, \( g_C : Q \rightarrow \mathbb{R}^{NC} \) is a smooth function with Jacobian \( G_C : TQ \rightarrow \mathbb{R}^{NC} \) and the notation \( \mathbb{R}^N \ni a \leq b \in \mathbb{R}^N \) is used to denote the set of component-wise inequalities \( a_1 \leq b_1, \ldots, a_N \leq b_N \). When the contact constraints admit representation (13), the normal cone is given by [8]

\[
N_qC = \left\{ G_C^T(q) \lambda_C \cdot g_C(q) \leq 0, \lambda_C \geq 0, g_C i \lambda_C i = 0, i = 1, \ldots, NC \right\}, \tag{14}
\]

and \( \lambda_C \in \mathbb{R}^{NC} \) are one-sided Lagrange multipliers. Using this representation, the embedded Euler–Lagrange equations with one-sided constraints can be expressed in the form

\[
- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + G_C^T(q(t)) \lambda_C(t) + G^T_C(q(t)) \lambda_C(t) = 0. \tag{15}
\]

These equations of motion, in conjunction with the constraint Equation (7), and the constraint inequalities (13), define a system of equations to be solved jointly for the trajectory \( q(t) \in C \) and the Lagrange multipliers \( \lambda_C(t) \) and \( \lambda_C(t) \).

In the presence of percussive, that is, isolated collisions, the trajectories may not be differentiable and the contact forces and the equation of motion (11) must be interpreted in the sense of distributions in general. Suppose, for instance, that the solution is piecewise smooth and that the velocity jumps discontinuously at \( t_0 \). Then, a distributional treatment of (11) gives the jump condition

\[
\begin{bmatrix}
\frac{\partial L}{\partial \dot{q}} (q(t_0), \dot{q}(t_0)) \\
\lambda_C(t_0)
\end{bmatrix} + I_C(t_0) = 0, \tag{16}
\]

for some impulse \( I_C(t_0) \in N_{q(t_0)}C \). Alternatively, the trajectories may be characterized directly from the constrained action

\[
S_C(q, A) = \begin{cases} \int_A L (q(t), \dot{q}(t)) \, dt, & \text{if } q(t) \in C, \text{ a. e. in } A, \\ +\infty, & \text{otherwise}. \end{cases} \tag{17}
\]
This constrained action can also be expressed as

\[
S_C(q, A) = \int_A L(q(t), \dot{q}(t)) \, dt + I_{A \times C}(q),
\]

where \(I_{A \times C}\) is the indicator function of \(A \times C\), namely,

\[
I_{A \times C}(q) = \begin{cases} 
0, & \text{if } q(t) \in C, \text{ a.e. in } A, \\
+\infty, & \text{otherwise.}
\end{cases}
\]

The corresponding spacetime Euler–Lagrange equations are then

\[
-\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + \partial I_C(q)(t) \equiv 0.
\]

We recall that the admissible set \(C\) need be neither smooth nor convex, and, consequently, the variation of \(S_C\) and the derivatives of \(I_C\), which define the contact force system, are generalized derivatives in the sense of nonsmooth calculus (cf. Appendix A) [8]. In particular, \(\partial I_C\) is set-valued in general, which renders the corresponding discrete Euler–Lagrange equation a discrete differential inclusion.

Yet another device for facilitating the consideration of collisions and nondifferentiable trajectories is to regularize the contact constraint by means of a contact potential. For instance, a regularization of the Yoshida type consists of considering the sequence of actions

\[
S_{C,\epsilon}(q, A) = \int_A \left( L(q(t), \dot{q}(t)) + \frac{1}{\epsilon} \text{dist}^2(q(t), C) \right) dt,
\]

where \(\text{dist}(q, C)\) is the geodesic distance in \(Q\) between \(q\) and \(C\). The actions \(S_{C,\epsilon}(q, A)\) determine sequences \((q_\epsilon)\) of regularized trajectories and, as discussed, for example, in [5, 20], as \(\epsilon \to 0, S_{C,\epsilon}(q, A) \to S_C(q, A)\) pointwise and interpenetration is increasingly penalized.

2.4. Conservation of momentum maps

The theory of symmetry of smooth Lagrangian systems is well known [18] but may stand a brief summary, specifically as it bears on the present work. Let \(G\) be a Lie group with Lie algebra \(\mathfrak{g} = T_e G\). A left action of \(G\) on \(Q\) is a mapping \(\Phi : G \times Q \to Q\) such that (1) \(\Phi(e, \cdot) = \text{id}\) and (2) \(\Phi(g, \Phi(h, \cdot)) = \Phi(gh, \cdot) \quad \forall g, h \in G\). Let \(\xi \in \mathfrak{g}\). Then, the infinitesimal generator of \(\Phi\) corresponding to \(\xi\) is the vector field \(\xi_Q \in TQ\) given by

\[
\xi_Q(q) = \frac{d}{d\epsilon} [\Phi(\exp(\epsilon \xi), q)]_{\epsilon=0}.
\]

The momentum map \(J : TQ \to \mathfrak{g}^*\) defined by the action \(\Phi\) then follows from the identity

\[
\langle J(q, \dot{q}), \xi \rangle = \left\langle \partial_q L(q, \dot{q}), \xi_Q(q) \right\rangle, \quad \forall \xi \in \mathfrak{g}.
\]

We say that the Lagrangian \(L\) is invariant under the action \(\Phi\) if

\[
L(\Phi_g(q), T\Phi_g(q)\dot{q}) = L(q, \dot{q}), \quad \forall g \in G, (q, \dot{q}) \in TQ,
\]

where we write \(\Phi_g(\cdot) = \Phi(g, \cdot)\). Under these conditions, we additionally say that \(G\) is a symmetry group of the system and that \(\Phi\) expresses a symmetry of the system. The classical theorem of Noether states that if \(L\) is invariant under the action \(\Phi\), then the corresponding momentum map \(J\) is a constant of the motion, that is, it remains constant along trajectories.
Classical examples include the following. (1) Conservation of linear momentum, in which case $Q = E(n)^N, G = E(n), \Phi(u, q) = \{q_1 + u, \ldots, q_N + u\}$ represents a rigid translation of the system by $u \in E(n)$ and the corresponding momentum map is the total linear momentum of the system, $J = p_1 + \cdots + p_N$. (2) Conservation of angular momentum, in this case $Q = E(n)^N, G = SO(n), \Phi(R, q) = \{Rq_1, \ldots, Rq_N\}$ represents a rigid rotation of the system by $R \in SO(n)$ and the corresponding momentum map is the total angular momentum of the system, $J = q_1 \times p_1 + \cdots + q_N \times p_N$.

2.5. Extension to dissipative systems

The preceding conservative framework can be extended to systems with dissipation, including friction, by recourse to the Lagrange–d’Alembert principle. The Lagrange–d’Alembert principle is of particular value for systems that, while dissipative, have a strong conservative component that sets the framework for the dynamics of the system. In such cases, it may be beneficial to treat the conservative component of the dynamics variationally, and the Lagrange–d’Alembert principle supplies an effective device to that end. In particular, the Lagrangian variational framework developed in the foregoing is recovered identically when the dissipative forces vanish.

Recall that, according to the Lagrange–d’Alembert principle, the trajectories of the system satisfy the stationarity condition

$$\delta S(q, \dot{q}, A) + \int_A F(t, q(t), \dot{q}(t)) \cdot \dot{q}(t) \, dt = 0,$$

where $S(q, A)$ is a differentiable action, Equation (1), $\delta S(q, \dot{q}, A)$ is its variation, Equation (2), $(q, \dot{q}) : \mathbb{R} \to TQ$ is an admissible variation and the force field $F : \mathbb{R} \times TQ \to T^*Q$ describes the—possibly time dependent—dissipative processes operative in the system. If the trajectory is differentiable over an open interval of time $A \in \mathcal{E}$, then it satisfies the equations of motion

$$- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + \mathbf{f}_D(t) = 0$$

with

$$\mathbf{f}_D(t) = F(t, q(t), \dot{q}(t)),$$

which may be regarded as a balance of inertial, energetic, and dissipative forces. If the configuration manifold is embedded, the system is subject to one-sided constraints, and the trajectories are differentiable, then the equations of motion are of the form

$$- \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} (q(t), \dot{q}(t)) \right) + \frac{\partial L}{\partial q} (q(t), \dot{q}(t)) + \mathbf{f}_Q(t) + \mathbf{f}_C(t) + \mathbf{f}_D(t) = 0.$$

where, as before, the contact forces $(q, \mathbf{f}_C) : A \to N_qC \subset T_q^*Q$ are determined so that $q \in C$ at all times.

An important consequence of dissipation is that, in the presence of dissipative forces, the total energy, Equation (43), is no longer a constant of the motion. Instead, along trajectories, we have

$$\dot{E} = \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L \right) = \left( \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} \right) \cdot \dot{q} = \mathbf{f}_D \cdot \dot{q}.$$
We expect, on physical grounds, the dissipation inequality

$$f_D \cdot \dot{q} \leq 0$$

(30)
to hold and, therefore, the total energy of the system to decrease monotonically in time. In a number of cases of interest, the dissipative forces derive jointly from a scalar dissipative potential $\psi(q, \dot{q})$, that is,

$$F(t, q, \dot{q}) = -\frac{\partial \psi}{\partial \dot{q}}(t, q, \dot{q}).$$

(31)

For these systems, the dissipation inequality (30) is automatically satisfied if the dissipation potential is convex and is minimized at $\dot{q} = 0$.

In the presence of percussive collisions, the dissipative force system and the equations of motion must be interpreted in a distributional sense, especially if the dissipation is rate independent and, therefore, allows for discontinuous velocities. Suppose, for instance, that the solution is piecewise smooth and that the velocity jumps discontinuously at $t_0$. Then, a distributional treatment of (28) gives the jump condition

$$\left[ \frac{\partial L}{\partial \dot{q}} (q(t_0), \dot{q}(t_0)) \right] + I_C(t_0) + I_D(t_0) = 0,$$

(32)
in terms of a normal impulse $I_C(t_0) \in N_q(t_0)C$ and a dissipative impulse $I_D(t_0) \in T_q(t_0)C$ tending to decelerate the tangential component of the motion.

**Example 2.1 (Linear damping)**

For a system with $Q = \mathbb{R}^N$, Lagrangian

$$L(q, \dot{q}) = K(\dot{q}) - V(q)$$

(33)
potential energy $V(q)$ and kinetic energy

$$K(\dot{q}) = \frac{1}{2} \dot{q}^T M \dot{q}$$

(34)
linear damping is characterized by the quadratic dissipation potential

$$\psi(\dot{q}) = \frac{1}{2} \dot{q}^T C \dot{q},$$

(35)
where $C$ is the damping matrix. We note that $\psi(\dot{q})$ is minimized at $\dot{q} = 0$ and is convex if $C$ is positive definite, in which case the dissipation inequality (30) is automatically satisfied.

**Example 2.2 (Coulomb friction)**

In the applications pursued subsequently, the dissipative forces of interest arise as a result of frictional contact during collisions in multibody systems. In order to formulate the corresponding dissipative force field, suppose that, as in Section 2.3, the contact constraints restrict the configurations of the system to a subset $C \subset Q$. Suppose, in addition, that the frictional forces obey Coulomb friction and that the trajectory is differentiable over an open interval of time $A \in E$. Then, the dissipative potential is of the form (cf. [11])

$$\psi(q, \dot{q}) = \mu |f_C \times \dot{q}|.$$

(36)
where $\mu$ is Coulomb’s coefficient of friction, $f_C$ are the normal contact forces required to keep the trajectory $q$ in $C$ over the time interval $A$, and $\times$ denotes the vector product. The dissipation potential is proportional to the magnitude of the contact forces, in accordance with Coulomb’s law of friction. If, in particular, $q$ is in the interior $\text{Int}\ C$ of $C$, then it follows that $f_C = 0$, that is, $\psi$ vanishes identically in the absence of contact, as required. The dissipation potential (36) is a convex—though not differentiable—function of $\dot{q}$ that is minimized at the origin, which ensures the satisfaction of the dissipation inequality (30). We also note that

$$ |f_C \times \dot{q}| = |f_C| \, |\dot{q}_\parallel|, \quad (37) $$

where $\dot{q}_\parallel$ is the component of $\dot{q}$ normal to $f_C$, or sliding component of the velocity field. Thus, only the sliding component of the velocity field gives rise to frictional dissipation, as expected. However, if the contact forces are nonzero over an interval of time, $f_C(t) \neq 0$, $t \in A \in \mathcal{E}$, or persistent contact, then necessarily, $q(t)$ must lie on the boundary $\partial C$ of $C$—and $\dot{q}(t)$ on the tangent cone $T_{\dot{q}(t)}C$—over the same interval of time, which corresponds to a relative sliding motion over the region of contact. Thus, under these conditions of persistent contact, we have $\dot{q}_\parallel = \dot{q}$, and the distinction between the velocity field and its sliding component becomes superfluous and inconsequential.

Inserting (36) into (31), we obtain

$$ F(t, q, \dot{q}) = \mu f_C \times \partial |f_C \times \dot{q}|, \quad (38) $$

where

$$ \partial |v| = \begin{cases} \{ f = v/|v|, \text{ if } v \neq 0 \}, & \text{if } |v| \leq 1, \\ \{ |f| \leq 1 \}, & \text{if } v = 0. \end{cases} \quad (39) $$

where we assume that $f_C \times \dot{q} \neq 0$. Under conditions of persistent contact, $|f_C \times \dot{q}| = |f_C| \, |\dot{q}|$, thus

$$ F(t, q, \dot{q}) = -\mu |f_C| \, \frac{\dot{q}}{|\dot{q}|} = -\mu |f_C| \, \partial |\dot{q}|, \quad (40) $$

and the frictional force opposes the sliding motion in proportion to the contact force, as expected.

Suppose, contrariwise, that the solution is only piecewise smooth and that the velocity jumps discontinuously at $t_0$, corresponding to a percussive collision. Then, formally integrating (28) across the jump gives the jump condition and using (40) gives the frictional impulse

$$ I_D(t_0) = -\mu |L_C(t_0)| \left( \partial | \dot{q} |^+ + (\partial | \dot{q} |)^- \right) / 2, \quad (41) $$

where $L_C(t_0)$ is the normal impulse reaction. If $\partial C$ is smooth at $q(t_0)$ then $N_{q(t_0)}C$ reduces to the normal outward half ray and the normal components of $(\partial | \dot{q} |)^+$ and $(\partial | \dot{q} |)^-$ cancel out, whereupon it follows that the frictional impulse $I_D(t_0)$ opposes the parallel or sliding motion in proportion to the normal impulse reaction, as expected.

### 2.6. The spacetime formalism

Conservation of energy can be fit into the Lagrangian framework by means of a spacetime formulation in which time is regarded as a generalized coordinate, for example, $q_0$. The corresponding spacetime Lagrangian is

$$ L(q, q') = L \left( q_0, q \cdot q' / q_0' \right) q_0', \quad (42) $$
where $L(t,q,\dot{q})$ is a general time-dependent Lagrangian, $q = (q_0, q)$ is a point in the space–time configuration manifold $Q = \mathbb{R} \times Q$ and the trajectories $q(\tau)$ are parameterized in terms of an arbitrary parameter $\tau$, and $q'$ denotes the derivative with respect to that parameter. Let $G = \mathbb{R}$ and $\Phi(s,q) = (q_0 + s, q)$ be a time-shift and suppose that $L(q, q')$ is invariant under $\Phi$, that is, $L(q, q')$ is autonomous or time independent. Then, the corresponding momentum map is

$$ J = L - \dot{q} L \cdot \dot{q} = -E, $$

(43)

and the total energy $E$ of the system is a constant of the motion. We also note that Euler–Lagrange Equation (4) and the energy conservation equation follow jointly as the Euler–Lagrange equations of the spacetime Lagrangian, namely,

$$ -\frac{d}{d\tau} \frac{\partial L}{\partial q'} + \frac{\partial L}{\partial q} = 0. $$

(44)

Indeed, decomposing this equation into temporal and spatial components yields

$$ -\frac{d}{d\tau} \frac{\partial L}{\partial q_0'} + \frac{\partial L}{\partial q_0} = \left( \frac{dE}{dt} + \frac{\partial L}{\partial t} \right) q_0' = 0, $$

(45a)

$$ -\frac{d}{d\tau} \frac{\partial L}{\partial q'} + \frac{\partial L}{\partial q} = \left( -\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} + \frac{\partial L}{\partial q} \right) q_0' = 0. $$

(45b)

This compact spacetime framework, and extensions to dissipating systems thereof, will be found useful as a basis for formulating time-discrete variational integrators, especially in the presence of collisions and dissipation.

The spacetime formulation of dissipative dynamics is particularly revealing as it effectively provides a unified framework that encompasses both the equations of motion and the energy equation. We begin by introducing the spacetime dissipative force field

$$ F(q, q') = \left( -F \left( q_0, q \cdot q'/q_0' \right) \cdot q' \right), $$

(46)

which combines the rate of dissipation with the primitive dissipative force field. By virtue of this definition, the equations of motion (4) and energy equation (29) can be expressed jointly as

$$ -\frac{d}{d\tau} \frac{\partial L}{\partial q'} + \frac{\partial L}{\partial q} + F = 0, $$

(47)

which extends (44) to systems with dissipation. Decomposing into temporal and spatial components, we find

$$ \left( \frac{dE}{dt} + \frac{\partial L}{\partial t} - F \cdot \dot{q} \right) q_0' = 0, $$

(48a)

$$ \left( -\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} + \frac{\partial L}{\partial q} + F \right) q_0' = 0. $$

(48b)

which also generalize (45) to the dissipative case. For systems characterized by a dissipative potential, we may additionally introduce the spacetime dissipative potential

$$ \Psi(q, q') = \psi \left( q_0, q \cdot q'/q_0' \right), $$

(49)

with the property that

$$ F = -\frac{\partial \Psi}{\partial q'} q_0'^2. $$

(50)
Again, separating temporal and spatial components, we find

\[- \frac{\partial \Psi}{\partial q_0} q_0^2 = \frac{\partial \Psi}{\partial \dot{q}} \cdot \dot{q} = -F \cdot \dot{q} \cdot q_0, \quad (51a)\]

\[- \frac{\partial \Psi}{\partial q} q_0^2 = - \frac{\partial \Psi}{\partial \dot{q}} q_0^2 = F \cdot q_0, \quad (51b)\]

as required. This remarkable joint potential structure of the equations of motion and energy equation of dissipative systems, which does not appear to have been noted heretofore in the literature, will play an important role in the development of the time-discrete formulation.

3. CLASSICAL VARIATIONAL INTEGRATORS

Next, we change paradigms and proceed to regard time as a discrete variable. In this new framework, the objective is to characterize discrete trajectories \( q_0, \ldots, q_k, q_{k+1}, \ldots, q_N \), sampled at corresponding discrete times \( t_0, \ldots, t_k, t_{k+1}, \ldots, t_N \). In this section, we briefly review some aspects of variational time integrators that lay the foundation for subsequent developments. Detailed accounts of variational time integrators may be found in [1, 2] and references therein. Background on incremental minimization and projection methods may be found in [6, 9–11].

3.1. The unconstrained and the embedded case

We begin by considering a time-dependent discrete Lagrangian \( L_d : (\mathbb{R} \times Q) \times (\mathbb{R} \times Q) \rightarrow \mathbb{R} \) characterizing the dynamics of the system over a time interval. For purposes of convergence, the discrete Lagrangian should approximate the continuous Lagrangian in the sense that

\[ L_d (t_k, q_k, t_{k+1}, q_{k+1}) \approx S (q, (t_k, t_{k+1})). \quad (52) \]

By formally taking variations, we additionally have

\[ D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_k + D_4 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_{k+1} \approx \delta S (q, \delta q, (t_k, t_{k+1})), \quad (53) \]

which establishes the consistency between the time-discrete and time-continuous Euler–Lagrange equations. Next, consider a discrete trajectory of points \( \{q_k\}_{k=0}^N \) and postulate the discrete action

\[ S_d \left(\{q_k\}_{k=0}^N\right) = \sum_{k=0}^{N-1} L_d (t_k, q_k, t_{k+1}, q_{k+1}). \quad (54) \]

Assuming differentiability, the corresponding variation is

\[ \delta S_d \left(\{q_k\}_{k=0}^N\right) = \sum_{k=1}^{N-1} \left[ D_4 L_d (t_{k-1}, q_{k-1}, t_k, q_k) + D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \right] \cdot \delta q_k, \quad (55) \]

where we have used the identities \( \delta q_0 = \delta q_N = 0 \). Requiring the variations of the action to be zero for all \( \delta q_k \), we obtain the discrete Euler–Lagrange equations

\[ D_4 L_d (t_{k-1}, q_{k-1}, t_k, q_k) + D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) = 0, \quad k = 1, \ldots, N - 1. \quad (56) \]

We note that the discrete Euler–Lagrange equation at time \( t_k \) can be solved for \( q_{k+1} \) given \( (q_{k-1}, q_k) \), thus defining a variational time integrator.
We can rewrite the discrete Euler–Lagrange equations in position-momentum form by introducing the discrete conjugate momenta

\begin{align}
  p_k^+ &= D_4L_d (t_{k-1}, q_{k-1}, t_k, q_k), \\
  p_k^- &= -D_2L_d (t_k, q_k, t_{k+1}, q_{k+1}),
\end{align}

(57a, b)

whereupon (56) takes the simple form

\[ p_k^+ - p_k^- = 0. \] (58)

which establishes the continuity of linear momenta. In calculations, the initial conditions are most commonly specified in terms of initial position and momenta \((q_0, p_0)\). More generally, we may know \((q_k, p_k)\) at time \(t_k\). These data and (58) then determine \(q_{k+1}\), whereupon (57b) delivers \(p_{k+1}\), which defines a recursive relation for the discrete trajectory referred to as the \((q, p)\) scheme, as opposed to the original \((q, q)\) scheme (56).

Corresponding to (10) in the embedded case, reduced discrete Legendre transforms are given by

\[ \begin{align*}
  p_q^+ &= P^T(q_k) D_4L_d (t_{k-1}, q_{k-1}, t_k, q_k), \\
  p_q^- &= -P^T(q_k) D_2L_d (t_k, q_k, t_{k+1}, q_{k+1}),
\end{align*} \] (59)

in the embedded case. Using the local reparametrization, \(q_{k+1} = \bar{F}_d (u_{k+1}, q_k)\) which ensures that \(q_{k+1} \in Q\), the time stepping scheme

\[ p_q^+ - p_q^- = 0 \] (60)

can be solved for \(u_{k+1} \in \mathbb{R}^{n-NQ}\). A full explanation of this topic including remarks on the initialization of the simulation and the recovery of the Lagrange multipliers when solving (60) may be found in [21].

**Example 3.1 (Trapezoidal rule in the unconstrained case)**

A simple discrete Lagrangian is obtained by restricting the incremental action over \((t_k, t_{k+1})\) to trajectories of the form

\[ q(t) = \frac{t_{k+1} - t}{t_{k+1} - t_k} q_k + \frac{t - t_k}{t_{k+1} - t_k} q_{k+1}, \] (61)

and further approximating the resulting action integral by means of the trapezoidal rule, with the result

\[ L_d (t_k, q_k, t_{k+1}, q_{k+1}) = \frac{t_{k+1} - t_k}{2} \left\{ L \left( t_k, q_k, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) + L \left( t_{k+1}, q_{k+1}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \right\}. \] (62)

For a system characterized by the Lagrangian (33), the trapezoidal-rule discrete Lagrangian (62) specifically evaluates to

\[ L_d (t_k, q_k, t_{k+1}, q_{k+1}) = \left\{ K \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) - \frac{V(q_k) + V(q_{k+1})}{2} \right\} (t_{k+1} - t_k). \] (63)
and the corresponding discrete Euler–Lagrange equations follow as

\[
M \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} - \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) + \frac{t_{k+1} - t_{k-1}}{2} DV(q_k) = 0.
\] (64)

Evidently, these discrete Euler–Lagrange equations coincide with the result of a central-difference discretization of the time-continuous Euler–Lagrange equations. An analysis of the trapezoidal-rule variational integrator may be found in [1].

3.2. Exact conservation properties

An important property of variational integrators is that symmetries of the system result in the exact conservation of discrete momentum maps, a result known as discrete Noether’s theorem [2]. Let \( G \) be a Lie group and \( \Phi : G \times Q \rightarrow Q \) a left action of \( G \) on \( Q \). Let \( \xi \in \mathfrak{g} = T_e G \) and \( \xi_Q \in TQ \) its corresponding infinitesimal generator. We say that the discrete Lagrangian \( L_d \) is invariant under the action \( \Phi \) if

\[
L_d \left( t_k, \Phi_g(q_k), t_{k+1}, \Phi_g(q_{k+1}) \right) = L_d \left( t_k, q_k, t_{k+1}, q_{k+1} \right), \quad \forall \xi \in \mathfrak{g}, \quad (q_k, q_{k+1}) \in Q \times Q.
\] (65)

In particular, for all \( \xi \in \mathfrak{g} \), the sequence of actions

\[
S_d \left( \{ \Phi_{\exp(\epsilon \xi)}q_k \}_{k=0}^N \right) = \sum_{k=0}^{N-1} L_d \left( t_k, \Phi(\exp(\epsilon \xi), q_k), t_{k+1}, \Phi(\exp(\epsilon \xi), q_{k+1}) \right)
\] (66)

is constant and independent of \( \epsilon \). Therefore, differentiating with respect to \( \epsilon \) at \( \epsilon = 0 \) and using the discrete Euler–Lagrange equations, we obtain the discrete Noether’s theorem

\[
J_d \left( t_{N-1}, q_{N-1}, t_N, q_N \right) = J_d \left( t_0, q_0, t_1, q_1 \right),
\] (67)

where the discrete momentum map \( J_d : \mathbb{R} \times Q \times \mathbb{R} \times Q \rightarrow \mathfrak{g}^* \) follows from the identity

\[
J_d \left( t_k, q_k, t_{k+1}, q_{k+1} \right) \cdot \xi = D_d L_d \left( t_k, q_k, t_{k+1}, q_{k+1} \right) \cdot \xi_Q(q_{k+1}).
\] (68)

Discrete conservation of energy can be established by treating the discrete times \( t_k \) as additional generalized variables and recalling that energy is the momentum conjugate to time [12]. The discrete action thus becomes

\[
S_d \left( \{(t_k, q_k)\}_{k=0}^N \right) = \sum_{k=0}^{N-1} L_d \left( t_k, q_k, t_{k+1}, q_{k+1} \right),
\] (69)

which, in contrast to (54), is now to be regarded as a function of spatial and temporal coordinates. The Euler–Lagrange equations corresponding to the generalized time variables are

\[
D_3 L_d \left( t_{k-1}, q_{k-1}, t_k, q_k \right) + D_4 L_d \left( t_k, q_k, t_{k+1}, q_{k+1} \right) = 0, \quad k = 1, \ldots, N - 1,
\] (70)

which supply additional equations for the determination of the sequence of discrete times \( \{t_k\}_{k=0}^N \). As before, we can rewrite the discrete Euler–Lagrange equations (70) in conservation form by introducing the discrete energies
whereupon (70) becomes

\[ E_k^+ - E_k^- = 0, \]  

which may be regarded as an expression of discrete conservation of energy.

**Example 3.2 (Trapezoidal rule and energy conservation)**

For the simple trapezoidal-rule discrete Lagrangian (62), the discrete energy balance Equation (70) evaluates to

\[ K \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) + \frac{V(q_k) + V(q_{k+1})}{2} = K \left( \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) + \frac{V(q_{k-1}) + V(q_k)}{2}, \]  

which indeed represents a statement of discrete energy balance with discrete energies

\[ E_k^- = E_k^+ = K \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) + \frac{V(q_k) + V(q_{k+1})}{2}. \]  

Evidently, these discrete energies supply an approximation of the time-continuous energy, with the distinguishing property—relative to any other arbitrary discrete approximation of the energy—of being exactly conserved along discrete trajectories determined by (58) and (72), or by (60) and (72) in the embedded case.

**Remark 3.1**

It is well known that in praxis, (70) is not always solvable, see, for example, [22]. Thus, away from any collision, we simply solve (58) or (60) on a given equidistant time grid. However, when treating collisions in the framework of discontinuous discrete Lagrange mechanics, the discrete energy conservation is no longer redundant but can be used to compute the post-collision momentum as described in Section 5.3.

### 3.3. Incremental minimum principles

Radovitzky and Ortiz [6] noted that the classical time integrators of dynamics can be recast as a sequence of minimum problems in terms of a suitably defined incremental energy that combines inertia, potential energy, and dissipation. Lew [7] investigated the correspondence between the incremental minimization and discrete Lagrangian approaches and showed that the symmetry of the cross derivatives,

\[ D_4 D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \rightarrow \text{symmetric}, \]  

provides a sufficient condition for the discrete Euler–Lagrange equations (56) to derive from an incremental minimum principle. Indeed, if (75) holds, then we can define an incremental energy \( F_d (t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, q_{k+1}) \) such that

\[ D_4 F_d (t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, q_{k+1}) = D_4 L_d (t_{k-1}, q_{k-1}, t_k, q_k) + D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}), \]  

whereupon the incremental minimum problem

\[ q_{k+1} \in \text{argmin} \ F_d (t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, \cdot) \]  

is equivalent to (56).
The time-adaptive equations (56) and (70) likewise derive from a joint incremental minimum principle if the symmetric condition

\[
\begin{pmatrix}
    D_4 D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \\
    D_4 D_1 L_d (t_k, q_k, t_{k+1}, q_{k+1})
\end{pmatrix}
\begin{pmatrix}
    D_3 D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \\
    D_3 D_1 L_d (t_k, q_k, t_{k+1}, q_{k+1})
\end{pmatrix} \rightarrow \text{symmetric,}
\]

(78)
is satisfied by the discrete Lagrangian. Then, introducing a function \( F_d (t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, q_{k+1}) \) such that

\[
\begin{align*}
    D_4 F_d (t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, q_{k+1}) &= D_4 L_d (t_{k-1}, q_{k-1}, t_k, q_k) \\
    &\quad + D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) , \\
    D_3 F_d (t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, q_{k+1}) &= D_3 L_d (t_{k-1}, q_{k-1}, t_k, q_k) \\
    &\quad + D_1 L_d (t_k, q_k, t_{k+1}, q_{k+1}) .
\end{align*}
\]

(79a)

(79b)

the incremental minimum problem

\[
(t_{k+1}, q_{k+1}) \in \text{argmin} \, F_d (t_{k-1}, q_{k-1}, t_k, q_k, \cdot, \cdot)
\]

(80)
is equivalent to (56) and (70).

3.4. The spacetime formalism

Time-adaptive discrete Lagrangian dynamics such as described in the foregoing can be rendered in a more compact form by recourse to spacetime notation. Thus, letting \( q = (t, q) \in \mathbb{Q} = \mathbb{R} \times \mathbb{Q} , p = (-E, p) \in T^*Q \), and \( L_d (q_k, q_{k+1}) = L_d (t_k, q_k, t_{k+1}, q_{k+1}) \), the action (69) becomes

\[
S_d (q_k)_{k=0}^{N-1} = \sum_{k=0}^{N-1} L_d (q_k, q_{k+1}) .
\]

(81)

and the Euler–Lagrange Equations (56) and (70) can be jointly expressed as

\[
D_1 L_d (q_{k-1}, q_k) + D_2 L_d (q_k, q_{k+1}) = 0 , \quad k = 1, \ldots, N-1 .
\]

(82)

The spacetime discrete Lagrangian \( L_d (q_k, q_{k+1}) \) may also be regarded as an approximation of the spacetime Lagrangian (42) in the sense that

\[
L_d (q_k, q_{k+1}) \approx S (q_k, q_{k+1}) ,
\]

(83)

with the time parametrization normalized so that \( \tau_k = k \in \mathbb{Z} \). By formally taking variations, we then have

\[
\begin{align*}
    D_1 L_d (q_k, q_{k+1}) \cdot \delta q_k + D_2 L_d (q_k, q_{k+1}) \cdot \delta q_{k+1} \\
    = D_1 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \delta t_k + D_2 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_k \\
    + D_3 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \delta t_{k+1} + D_4 L_d (t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_{k+1} \\
    \approx \delta S (q, q_k, (t_k, \tau_{k+1})) ,
\end{align*}
\]

(84)

which jointly establishes the consistency between the time-discrete and time-continuous space Euler–Lagrange and energy conservation equations.
Finally, writing $F_d(q_{k-1}, q_k, q_{k+1}) = F_d(t_{k-1}, q_{k-1}, t_k, q_k, t_{k+1}, q_{k+1})$, the incremental minimum principle (80) can be compactly recast in the equivalent form

$$q_{k+1} \in \arg\min F_d(q_{k-1}, q_k, \cdot).$$

(85)

**Example 3.3 (Spacetime trapezoidal rule)**

The discrete spacetime formalism may be illustrated by means of the simple trapezoidal-rule approximation (cf. Examples 3.1 and 3.2). We begin by considering incremental spacetime trajectories of the form

$$q_0(\tau) = (1 - \tau)t_k + \tau t_{k+1},$$

(86a)

$$q(\tau) = (1 - \tau)q_k + \tau q_{k+1},$$

(86b)

along which the incremental action becomes

$$S(q,(\tau_k, \tau_{k+1})) = \int_{\tau_k}^{\tau_{k+1}} L\left(q_0(\tau), q(\tau), \frac{q_{k+1} - q_k}{t_{k+1} - t_k}, (t_{k+1} - t_k) d\tau\right).$$

(87)

An additional application of the trapezoidal rule to the integral returns (62). The spatial and temporal sections of the spacetime discrete Euler–Lagrange Equation (82) then return the discrete equations of motion (64) and the discrete energy conservation Equation (73), respectively.

As will become apparent in the developments that follow, the spacetime formalism not only affords considerable economy of notation, by placing the space and time coordinates on an equal footing, but, more importantly, becomes essential in the presence of collisions, where energy conservation plays a fundamental role, and in systems with dissipation, where the energy equation is no longer redundant.

### 3.5. Extension to dissipative systems

The time-discrete counterpart to the time-continuous Lagrange–d’Alembert principle (25) takes the general form (e.g., [2])

$$\sum_{k=1}^{N-1} \left[ D_4 L_d(t_{k-1}, q_{k-1}, t_k, q_k) + D_2 L_d(t_k, q_k, t_{k+1}, q_{k+1}) \right] \cdot \delta q_k$$

$$+ \sum_{k=1}^{N-1} F_d^+(t_{k-1}, q_{k-1}, t_k, q_k) + F_d^-(t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_k = 0,$$

(88)

where $F_d^+$ and $F_d^-$ are right and left discrete dissipative forces, respectively. For purposes of convergence, these discrete dissipative forces should approximate the time-continuous dissipative forces in the sense

$$F_d^-(t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_k + F_d^+(t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_{k+1} \approx \int_{t_k}^{t_{k+1}} F(t, q(t), \dot{q}(t)) \cdot \delta q(t) dt.$$  

(89)

The corresponding discrete Euler–Lagrange equations then become

$$D_4 L_d(t_{k-1}, q_{k-1}, t_k, q_k) + D_2 L_d(t_k, q_k, t_{k+1}, q_{k+1})$$

$$+ F_d^+(t_{k-1}, q_{k-1}, t_k, q_k) + F_d^-(t_k, q_k, t_{k+1}, q_{k+1}) = 0, \quad k = 1, \ldots, N - 1.$$  

(90)

These equations may be regarded as expressing a balance of linear momentum of the form (58) with extended momenta

$$p_k^+ = D_4 L_d(t_{k-1}, q_{k-1}, t_k, q_k) + F_d^+(t_{k-1}, q_{k-1}, t_k, q_k),$$

(91a)

$$p_k^- = -D_2 L_d(t_k, q_k, t_{k+1}, q_{k+1}) - F_d^-(t_k, q_k, t_{k+1}, q_{k+1}).$$

(91b)

Copyright © 2014 John Wiley & Sons, Ltd. 

DOI: 10.1002/nme
Lagrange–d’Alembert principle now takes the form and the corresponding discrete spacetime Euler–Lagrange equations follow as

\[ F_d^+(q_k, q_{k+1}) \cdot \delta q_k + F_d^-(q_k, q_{k+1}) \cdot \delta q_{k+1} \approx \int_{t_k}^{t_{k+1}} F(q(\tau), q'(\tau)) \cdot \delta q(\tau) q'_0(\tau) \, d\tau, \]  

(92)

where \( F(q, q') \) is the time-continuous spacetime dissipative force field (46). The discrete spacetime Lagrange–d’Alembert principle now takes the form

\[ \sum_{k=1}^{N-1} [D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1})] \cdot \delta q_k + \sum_{k=1}^{N-1} [F_d^+(q_{k-1}, q_k) + F_d^-(q_k, q_{k+1})] \cdot \delta q_k = 0, \]

(93)

and the corresponding discrete spacetime Euler–Lagrange equations follow as

\[ D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) + F_d^+(q_{k-1}, q_k) + F_d^-(q_k, q_{k+1}) = 0, \quad k = 1, \ldots, N-1. \]  

(94)

A spatial section of these equations returns the discrete equations of motion (90). In addition, a temporal section supplies the sought discrete energy equation. Indeed, inserting definition (46) into (92) and (94) and extracting the temporal component yields the discrete energy equation

\[ D_3 L_d(t_{k-1}, q_{k-1}, t_k, q_k) + D_1 L_d(t_k, q_k, t_{k+1}, q_{k+1}) + G_d^+(t_{k-1}, q_{k-1}, t_k, q_k) + G_d^-(t_k, q_k, t_{k+1}, q_{k+1}) = 0, \quad k = 1, \ldots, N-1, \]  

(95)

where

\[ G_d^-(q_k, q_{k+1}) \delta t_k + G_d^+(q_k, q_{k+1}) \delta t_{k+1} \approx - \int_{t_k}^{t_{k+1}} F(q_0(\tau), q(\tau), q'(\tau)/q'_0(\tau)) \cdot q'(\tau) \, dt \]  

(96)

defines a discrete dissipation rate. Remarkably, as in the conservative case, the spacetime formalism supplies a precise statement of discrete energy balance, as well as the means to determine discrete times ensuring that energy is dissipated at exactly the right rate. It also bears emphasis that, in the discrete spacetime formulation, the discretizations of the equations of motion and the energy equation are not arbitrary, but follow jointly as spatial and temporal sections of a single spacetime discretization.

**Example 3.4 (Damped spacetime trapezoidal rule)**

Consider, as in Example 3.3, incremental spacetime trajectories of the form (86). Inserting these trajectories into (89) and approximating the resulting integrals by the trapezoidal rule give

\[ F_d^+(t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_k + F_d^-(t_k, q_k, t_{k+1}, q_{k+1}) \cdot \delta q_{k+1} \]

\[ = \left\{ F(t_k, q_k, q_{k+1} - q_k/t_{k+1} - t_k) \cdot \delta q_k + F(t_{k+1}, q_{k+1}, q_{k+1} - q_k/t_{k+1} - t_k) \cdot \delta q_{k+1} \right\} \frac{t_{k+1} - t_k}{2}. \]

(97)
Identifying terms, we obtain

\[ F_d^-(q_k, q_{k+1}) = F \left( t_k, q_k, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \frac{t_{k+1} - t_k}{2}, \]  

(98a)

\[ F_d^+(q_k, q_{k+1}) = F \left( t_{k+1}, q_{k+1}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \frac{t_{k+1} - t_k}{2}. \]  

(98b)

Inserting now (86) into (96) and approximating the resulting integrals by the trapezoidal rule give

\[ G_d(q_k, q_{k+1}) \delta t_k + G_d^+(q_k, q_{k+1}) \delta t_{k+1} \]

\[ = - \left\{ F \left( t_k, q_k, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \delta t_k + F \left( t_{k+1}, q_{k+1}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \delta t_{k+1} \right\} \frac{q_{k+1} - q_k}{2}. \]

(99)

Identifying terms, we obtain

\[ G_d^-(q_k, q_{k+1}) = -F \left( t_k, q_k, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \frac{q_{k+1} - q_k}{2}, \]  

(100a)

\[ G_d^+(q_k, q_{k+1}) = F \left( t_{k+1}, q_{k+1}, \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) \frac{q_{k+1} - q_k}{2}. \]  

(100b)

For the system of Example 2.1, the damped discrete equations of motion follow as

\[ M \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} - \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) + \frac{t_{k+1} - t_k - t_k - t_{k-1}}{2} DV(q_k) + C \frac{q_{k+1} - q_{k-1}}{2} = 0, \]  

(101)

and the corresponding energy equation as

\[ K \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) + \frac{V(q_k) + V(q_{k+1})}{2} - K \left( \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) - \frac{V(q_{k-1}) + V(q_k)}{2} \]

\[ = -(t_{k+1} - t_k) \psi \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) - (t_k - t_{k-1}) \psi \left( \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right), \]  

(102)

which determines the rate of decay of the discrete energy. Given \((t_{k-1}, q_{k-1})\) and \((t_k, q_k)\), Equations (101) and (102) can be jointly solved for \((t_{k+1}, q_{k+1})\).

### 3.6. One-sided contact constraints

We conclude this section by briefly referring to two existing approaches to dynamic contact: discrete variational integrators [3] and predictor-projection methods [9–11]. We specifically emphasize the advantages and disadvantages of these approaches by way of segue to the subsequent development of time-discontinuous integrators. We recall that a variational theory of collisions, including moving obstacles, can be obtained by restricting the spacetime configurational space to a subset \(C \subset Q\), representing the admissible configurations of the system satisfying non-penetration constraints. Here and henceforth, we adopt spacetime notation for simplicity of notation. We also note that the admissible set is allowed to be time dependent, for example, in order to represent moving obstacles.
3.6.1. Variational time integrators. The admissibility of the trajectories can be enforced by means of
\[
S_d (\{q_k\}_{k=0}^N) = \sum_{k=0}^{N-1} L_d(q_k, q_{k+1}) + \sum_{k=1}^{N-1} I_C(q_k),
\]
(103)
where \(I_C\) is the indicator function of \(C\). The corresponding spacetime Euler–Lagrange equations are
\[
D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) + f_C(\tau_k) = 0, \quad k = 1, \ldots, N - 1,
\]
(104)
where
\[
f_C(\tau_k) \in \partial I_C(q_k)
\]
(105)
is the spacetime contact force at \(\tau_k\) and \(\partial I_C(q_k)\) is the set of possible spacetime contact forces. The precise discrete contact force history \(f_C\) in \(I_C(q)\) follows from the condition that the discrete trajectory \(q\) be in \(C\) at all times. We note that if \(C\) is independent of time, then the temporal component of \(f_C\) vanishes and the discrete energy is a constant of the motion. We also note that dissipation can again be accounted for by recourse to the discrete Lagrange–d’Alembert principle discussed in Section 3.5.

We recall that the admissible set \(C\) need be neither smooth nor convex and, consequently, the derivatives \(\partial I_C\), which define the contact force system, are generalized derivatives in the sense of nonsmooth calculus (cf. Appendix A) [8]. In particular, \(\partial I_C\), representing the set of possible spacetime contact forces, is set-valued in general, which renders the corresponding discrete Euler–Lagrange equation a discrete differential inclusion.

Example 3.5 (Free mass particle on half-plane)
We consider a free point mass constrained to move on the half-plane \(C = \{q \equiv (x, y) \in \mathbb{R}^2, \; y \geq b\}\), corresponding to an impenetrable wall at \(y = b\). The particle adopts three consecutive configurations \(q_1, q_2,\) and \(q_3\) at times \(t_1, t_2,\) and \(t_3\). We characterize the discrete dynamics of the system by means of the trapezoidal-rule discrete Lagrangian (62). Under these conditions, the action (103) specializes to
\[
S_d (\{q_2\}) = \frac{m}{2} \frac{|q_2 - q_1|^2}{t_2 - t_1} + \frac{m}{2} \frac{|q_3 - q_2|^2}{t_3 - t_2} + I_C(q_2).
\]
(106)
The corresponding Euler–Lagrange equations are
\[
m \frac{x_2 - x_1}{t_2 - t_1} - m \frac{x_3 - x_2}{t_3 - t_2} = 0,
\]
(107a)
\[
m \frac{y_2 - y_1}{t_2 - t_1} - m \frac{y_3 - y_2}{t_3 - t_2} = \frac{t_3 - t_1}{2} f_C,
\]
(107b)
\[
m \frac{\left(\frac{x_2 - x_1}{t_2 - t_1}\right)^2 + \frac{m}{2} \left(\frac{y_2 - y_1}{t_2 - t_1}\right)^2}{2} = \frac{m}{2} \left(\frac{x_3 - x_2}{t_3 - t_2}\right)^2 + \frac{m}{2} \left(\frac{y_3 - y_2}{t_3 - t_2}\right)^2,
\]
(107c)
where
\[
f_C = 0, \quad \text{if } y > b,
\]
(108a)
\[
f_C \geq 0, \quad \text{if } y = b.
\]
(108b)
The discrete Euler–Lagrange Equations (107) enforce the exact conservation of discrete momentum and energy. We see from (107a) that the horizontal motion parallel to the wall is uniform and is undisturbed by the wall. In addition, the contact reaction $\partial l_C(q_2)$ vanishes if $q$ lies in the interior of $C$ and is vertical, but otherwise arbitrary, if it lies on its boundary. Suppose that $y_1 > b$ and $y_2 = b$, corresponding to a collision with the obstacle at time $t_2$. Finally, from (107c), which ensures conservation of energy, we obtain

$$\frac{y_3 - b}{t_3 - t_2} = -\frac{b - y_1}{t_2 - t_1}. \quad (109)$$

This relation represents a reversal of the velocity of the particle upon impact, gives its final location $y_3$ as a function of the final $t_3$, and ensures conservation of linear momentum and energy through the collision. In particular, if we choose $t_3 - t_2 = t_2 - t_1 = \Delta t$, we have

$$y_3 - b = y_1 - b, \quad (110)$$

corresponding to a symmetric reflection of the particle about $b$. Finally, Equation (107b) gives the contact reaction.

**Example 3.6 (Trapezoidal rule applied to Coulomb friction)**

In order to formulate discrete equations of motion and the discrete energy equation in the presence of Coulomb friction, we must supply an appropriate discrete frictional force system. To this end, we consider, as in Example 3.3, incremental spacetime trajectories of the form (86). Inserting these trajectories into (89), with $F(t, \dot{q}, \ddot{q})$ given by (38), and approximating the resulting integrals by the trapezoidal rule gives

$$F_{\dot{q}}(t_k, q_k, t_{k+1}, q_{k+1}) \delta q_k + F_{\ddot{q}}(t_k, q_k, t_{k+1}, q_{k+1}) \delta \ddot{q}_k = \frac{\mu}{2} (t_{k+1} - t_k) f_C(t_k) \times \delta \bigg| f_C(t_k) \times \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \bigg| \cdot \delta q_k$$

$$= \frac{\mu}{2} (t_{k+1} - t_k) f_C(t_k) \times \delta \bigg| f_C(t_k) \times \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \bigg| \cdot \delta q_k$$

$$+ \frac{\mu}{2} (t_{k+1} - t_k) f_C(t_{k+1}) \times \delta \bigg| f_C(t_{k+1}) \times \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \bigg| \cdot \delta q_{k+1} \quad (111)$$

$$\approx \int_{t_k}^{t_{k+1}} \mu f_C(t) \times \delta \bigg| f_C(t) \times \dot{q}(t) \bigg| \cdot \delta q(t) \, dt.$$
and, identifying terms, we obtain

\[ G^-_d(q_k, q_{k+1}) = -\frac{\mu}{2} \left( t_{k+1} - t_k \right) (q_{k+1} - q_k) \cdot f_C(t_k) \times \frac{q_{k+1} - q_k}{t_{k+1} - t_k}, \]  
\[ G^+_d(q_k, q_{k+1}) = \frac{\mu}{2} \left( t_{k+1} - t_k \right) (q_{k+1} - q_k) \cdot f_C(t_{k+1}) \times \frac{q_{k+1} - q_k}{t_{k+1} - t_k}. \]  

(114a)

(114b)

Example 3.7 (Free mass particle on half-plane with Coulomb friction)

We consider the free point mass system of Example 3.5 modified to account for Coulomb friction between the particle and the wall. For the free particle in a half-plane under consideration, the trapezoidal-rule Lagrange–d’Alembert equations reduce to

\[ m \frac{x_2 - x_1}{t_2 - t_1} - m \frac{x_3 - x_2}{t_3 - t_2} - \frac{t_3 - t_1}{2} - \mu|f_C| = 0, \]  
\[ m \frac{y_2 - y_1}{t_2 - t_1} - m \frac{y_3 - y_2}{t_3 - t_2} + \frac{t_3 - t_1}{2} f_C = 0, \]  
\[ \frac{m}{2} \left( \frac{x_2 - x_1}{t_2 - t_1} \right)^2 + \frac{m}{2} \left( \frac{y_2 - y_1}{t_2 - t_1} \right)^2 - \frac{m}{2} \left( \frac{x_3 - x_2}{t_3 - t_2} \right)^2 - \frac{m}{2} \left( \frac{y_3 - y_2}{t_3 - t_2} \right)^2 = \mu|f_C| \frac{|x_2 - x_1| + |x_3 - x_2|}{2}. \]  

(115a)

(115b)

(115c)

We see from (115a) that the horizontal motion parallel to the wall is decelerated as a result of friction. In addition, we see from (115c) that an amount of energy is dissipated in proportion to the distance traveled, as expected. In general, Equations (115) can be solved, for example, for \( f_C, x_3, \) and \( y_3, \) for known values of the remaining variables.

3.6.2. Incremental minimization. As discussed in Section 3.2, the variational structure of the time integrator (104) ensures exact conservation properties, including energy. However, this exact conservation comes at the expense of resolving all collision times exactly. In large many-body systems undergoing complex collisions, that expense can be prohibitive, rendering the approach impractical. An alternative approach results from the constrained incremental minimum principle

\[ q_{k+1} \in \text{argmin} \ F_d(q_{k-1}, q_k, \cdot) + I_C(\cdot). \]  

(116)

The corresponding discrete dynamic-equilibrium equations are

\[ D_2L_d(q_{k-1}, q_k) + D_1L_d(q_k, q_{k+1}) + f_C(t_{k+1}) = 0, \quad k = 1, \ldots, N - 1. \]  

(117)

where

\[ f_C(t_{k+1}) \in \partial I_C(q_{k+1}) \]  

(118)

are the contact forces at time \( t_{k+1}. \) We note that the discrete equations of motion (117) resulting from incremental energy minimization differ from the discrete equations of motion (104) resulting from the discrete Hamilton’s principle in that the constraints are enforced on the updated configuration in the former and on the current configuration in the latter.
Example 3.8 (Trapezoidal rule)
For purposes of illustration, suppose that the system is characterized by the Lagrangian (33) and the corresponding trapezoidal-rule discrete Lagrangian (62), whereupon the discrete dynamic-equilibrium equations (117) specialize to

\[
M \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} - \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) + \frac{t_{k+1} - t_{k-1}}{2} DV(q_k) + \partial I_C(q_{k+1}) \equiv 0. \tag{119a}
\]
\[
K \left( \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \right) + V(q_k) + \frac{V(q_{k+1})}{2} = K \left( \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) + \frac{V(q_{k-1}) + V(q_k)}{2}. \tag{119b}
\]

Define now an unconstrained predictor \( q_{k+1}^{\text{pre}} \) as the solution of the unconstrained equations

\[
M \left( \frac{q_{k+1}^{\text{pre}} - q_k}{t_{k+1} - t_k} - \frac{q_k - q_{k-1}}{t_k - t_{k-1}} \right) + \frac{t_{k+1} - t_{k-1}}{2} DV(q_k) = 0. \tag{120}
\]

Then (119a) can be rewritten in the form

\[
M (q_{k+1} - q_{k+1}^{\text{pre}}) + \partial I_C(q_{k+1}) \equiv 0. \tag{121}
\]

We recall from nonsmooth analysis that the solution of this equation is

\[
q_{k+1} = P_C (q_{k+1}^{\text{pre}}), \tag{122}
\]

where \( P_C \) is the closest-point projection to \( C \) in the mass norm. The updated solution \((t_{k+1}, q_{k+1})\) follows by solving (122) and (119b), thus ensuring conservation of energy.

The structure of the discrete dynamic-equilibrium Equations (117) is now apparent from the preceding example. Thus, the updated configuration \( q_{k+1} \) is attained in two steps. The first predictor step neglects the contact constraints and results in an unconstrained predictor \( q_{k+1}^{\text{pre}} \). The second corrector step projects the predictor back to the admissible set. The strength of this approach resides in its ability to handle complex collision sequences without the need to resolve each individual collision. For large multibody systems, this attribute can result in considerable computational efficiencies. However, the discrete dynamic-equilibrium Equation (117) do not have a Lagrangian structure, and the corresponding time integrator lacks exact conservation and symplecticity in general.

4. DISCONTINUOUS VARIATIONAL INTEGRATORS

Next, we endeavor to formulate time integrators that combine the best attributes of discrete Lagrangian mechanics, such as symplecticity and exact conservation properties, and incremental minimization, such as the convenient predictor–corrector structure of the integrators and the lack of need to resolve all individual collision times. We accomplish these objectives by allowing the discrete trajectories of the discrete Lagrangian system to be discontinuous and by penalizing discontinuities weakly. In this manner, the resulting discrete dynamics of the system is Lagrangian and, in particular, obeys a discrete Hamilton or Lagrange–d’Alembert principle.

4.1. Discontinuous discrete Lagrangian mechanics

One aspect of incremental minimization and projection methods that renders them especially well suited to contact problems is their predictor–corrector structure. Thus, in those approaches, the solution, which is now denoted by \( q_{k+1}^{+} \), is reached in two steps, namely, an unconstrained predictor step yields \( q_{k+1}^{+} \), and is followed by a corrector step in the form of a projection to the admissible set. In order to formalize this concept within the framework of variational integrators, we may
allow the system to adopt two possibly different configurations \((q_k^+, q_k^-)\) at each sampling time \(t_k\), representing predictor and corrector configurations, respectively. The discrete sampling times can be treated likewise. The sequence \((q_k^-, q_{k+1}^+)\) of spacetime configuration pairs then represents a discontinuous trajectory of the system. In this representation, the discrete action takes the form

\[
S_d \left( \{ q_k^\pm \}_{k=0}^N \right) = \sum_{k=0}^{N-1} L_d \left( q_k^-, q_{k+1}^+ \right),
\]

(123)

and the continuous, or classical, trajectories of the system follow from an application of Hamilton’s principle of stationary action subject to the continuity constraints

\[
q_-^k - q_+^k = 0, \quad k = 1, \ldots, N - 1.
\]

(124)

Enforcement of these constraints by means of Lagrange multipliers yields the extended action

\[
S_d \left( \{ q_k^\pm, \lambda_k \}_{k=0}^N \right) = \sum_{k=0}^{N-1} L_d \left( q_k^-, q_{k+1}^+ \right) + \sum_{k=1}^{N-1} \lambda_k \cdot (q_-^k - q_+^k),
\]

(125)

and the extended discrete Euler–Lagrange equations

\[
D_1 L_d \left( q_k^-, q_{k+1}^+ \right) + \lambda_k = 0, \quad k = 1, \ldots, N - 1,
\]

(126a)

\[
D_2 L_d \left( q_{k-1}^-, q_k^+ \right) - \lambda_k = 0, \quad k = 1, \ldots, N - 1.
\]

(126b)

which, together with the continuity constraint (124), determine the extended trajectory \(\{ q_k^\pm, \lambda_k \}_{k=0}^N\). Evidently, adding up (126a) and (126b) and using the constraint (124) yield (82), which establishes the consistency between the discontinuous and classical approaches.

### 4.2. Discontinuous discrete Lagrangian mechanics with one-sided constraints

We wish to extend the preceding discontinuous framework so as to allow for trial or predictor configurations in violation of the contact constraints that are subsequently projected back, or corrected, to the admissible domain. To this end, we relax the continuity constraint (124) to the following constraints

\[
q_-^k \in C, \quad k = 1, \ldots, N - 1,
\]

(127a)

\[
q_-^k = P_C(q_k^+), \quad k = 1, \ldots, N - 1,
\]

(127b)

where \(P_C\) is the closest-point projection onto \(C\), in the sense of the geodesic distance of \(\mathbb{R} \times Q\). In writing the constraints (127), we regard \(q_k^+\) as a trial or predictor configuration, not required to satisfy the one-sided constraints, and \(q_k^-\) as the corresponding corrector configuration, required to satisfy the one-sided constraints. We recall that, if \(q \in C\), then \(P_C(q) = q\). Thus, if the predictor configuration is admissible, that is, if \(q_k^+ \in C\), then (127b) reduces to (124) and the classical trajectory is recovered. By contrast, when the predictor configuration is inadmissible, that is, \(q_k^+ \notin C\), the trajectory becomes discontinuous, or non-classical. We note that, if the configurational manifold \(Q\) is Euclidean, then (127b) is equivalent to the differential inclusion

\[
q_-^k - q_+^k \in \partial I_C(q_k^+),
\]

(128)

(cf. Appendix A), which supplies an additional means of enforcing the corrector constraint. Finally, we note that dissipation can again be accounted for by recourse to the discrete Lagrange–d’Alembert principle discussed in Section 3.5.
It bears emphasis that the DV integrators just defined allow for discontinuous trajectories in the presence of one-sided contact constraints, thus formalizing and placing the predictor–corrector collision algorithms within a Lagrangian framework. In particular, the discontinuous or non-classical trajectories allow the predictor configurations to violate the one-sided constraints. However, the extent of this violation, that is, the extent of penetration of the predictors into the inadmissible region, may be expected to decrease with decreasing time step and to vanish entirely in the limit. In this manner, discontinuous or non-classical trajectories are expected to converge to classical trajectories of the time-continuous system.

The structure of the DV integrators is illustrated by the following examples. The performance and accuracy properties of DV integrators are ascertained in Section 6 by way of numerical testing.

Example 4.1 (Free mass particle on half-plane)

We consider a free point mass constrained to move on the half-plane \( C = \{ q \equiv (x, y) \in \mathbb{R}^2, \ y \geq b \} \), corresponding to an impenetrable wall at \( y = b \in \mathbb{R} \) and three consecutive configurations as in Example 3.5. In contrast to that example, we proceed to consider a discontinuous trajectory determined by the extended action (125), which under present conditions specializes to

\[
S_d (\{ q_2^\pm \}) = \frac{m}{2} \frac{|q_2^+ - q_1|}{t_2 - t_1}^2 + \frac{m}{2} \frac{|q_3 - q_2^-|}{t_3 - t_2}^2.
\]  

(129)

Suppose that \( y_1 > b \) and \( y_2^+ < b \), that is, the predictor \( y_2^+ \) overshoots the wall. Then, the constraints (127) specialize to

\[
t_2^- = t_2^+, \\
x_2^- = x_2^+, \\
y_2^- \geq b, \\
y_2^- - y_2^+ \geq 0.
\]

(130)

We note that (130a) is a direct consequence of the time-independence of the admissible set \( C \). Using (130a) and (130b), and enforcing the remaining constraints by means of Lagrange multipliers \( \lambda_2, v_2 \) leads to the extended action (125) reading

\[
S_d (\{ q_2^\pm \}, \lambda_2, v_2) = \frac{m}{2} \frac{|q_2^+ - q_1|}{t_2 - t_1}^2 + \frac{m}{2} \frac{|q_3 - q_2^-|}{t_3 - t_2}^2 + \lambda_2 (y_2^- - b) + v_2 (y_2^- - y_2^+),
\]

(131)

and the corresponding discrete Euler–Lagrange equations follow as

\[
m \frac{x_2 - x_1}{t_2 - t_1} - m \frac{x_3 - x_2}{t_3 - t_2} = 0, \\
m \frac{y_3 - y_2^+}{t_3 - t_2} + \lambda_2 + v_2 = 0, \\
m \frac{y_2^+ - y_1}{t_2 - t_1} - v_2 = 0, \\
\frac{m}{2} \left( \frac{x_2 - x_1}{t_2 - t_1} \right)^2 + \frac{m}{2} \left( \frac{y_2^+ - y_1}{t_2 - t_1} \right)^2 = \frac{m}{2} \left( \frac{x_3 - x_2}{t_3 - t_2} \right)^2 + \frac{m}{2} \left( \frac{y_3 - y_2^-}{t_3 - t_2} \right)^2.
\]

(132)
subject to the Kuhn–Tucker conditions

\[ \begin{align*}
y_2^- - b & \geq 0, & \lambda_2 & \geq 0, & \lambda_2 (y_2^- - b) & = 0, \\
y_2^- - y_2^+ & \geq 0, & v_2 & \geq 0, & v_2 (y_2^- - y_2^+) & = 0. 
\end{align*} \] (133a) (133b)

Eliminating \( v_2 \) from these equations by the addition of (132b) and (132c) further gives

\[ \begin{align*}
mx_2 - x_1 - mx - x_2 & = 0, \\
m \left( \frac{y_2^- - y_1}{t_2 - t_1} \right)^2 + m \left( \frac{y_1^+ - y_1}{t_2 - t_1} \right)^2 & = m \left( \frac{x_3 - x_2}{t_3 - t_2} \right)^2 + m \left( \frac{y_2^- - y_2^+}{t_3 - t_2} \right)^2, \\
m \left( \frac{x_2 - x_1}{t_2 - t_1} \right)^2 & = m \left( \frac{x_3 - x_2}{t_3 - t_2} \right)^2. 
\end{align*} \] (134a) (134b) (134c)

where we have made explicit use of the constraints to simplify the equations and we have redefined \( \lambda_2 \) in terms of a contact force \( f_C \). As in the classical discrete trajectories, the motion parallel to the obstacle is uniform and remains undisturbed through the collision. In addition, from (134c), we obtain

\[ \frac{y_3 - b}{t_3 - t_2} = \frac{y_2^+ - y_1}{t_2 - t_1}, \] (135)

which reverses the incoming velocity, gives \( y_3 \) as a function of \( t_3 \) and ensures conservation of linear momentum and energy through the collision. If we choose \( t_3 - t_2 = t_2 - t_1 = \Delta t \), then we obtain

\[ y_3 - b = y_1 - y_2^+, \] (136)

corresponding to a projection of the unconstrained predictor to the wall followed by a reflection of the trajectory. Finally, Equation (134c) yields the contact reaction \( f_C \). This discontinuous trajectory is shown in Figure 1(b), with a typical classical trajectory shown in Figure 1(a) by way of reference.

**Example 4.2 (Free mass particle on half-plane with Coulomb friction)**

We extend the preceding example by allowing frictional interaction between the free mass particle and the wall. The effect of friction is to introduce an additional dissipative force system into the discrete equations of motion. Assuming Coulomb friction (cf. Example 2.2) and the trapezoidal-rule discrete frictional force system (112) leads to the discrete equations of motion

\[ \begin{align*}
& \begin{array}{c}
q_1 \bullet \\
q_2 \bullet \\
q_3 \bullet \\
q_4 \bullet \\
q_5 \bullet \\
q_6 \bullet \\
q_7 \bullet \\
q_8 \bullet \\
\end{array}, \\
& \begin{array}{c}
\text{wall} \ \\
\text{wall} \\
\text{wall} \ \\
\text{wall} \ \\
\text{wall} \ \\
\text{wall} \ \\
\text{wall} \ \\
\text{wall} \ \\
\end{array}, \\
& \begin{array}{c}
q_9 \bullet \\
q_10 \bullet \\
q_11 \bullet \\
q_12 \bullet \\
q_13 \bullet \\
q_14 \bullet \\
q_15 \bullet \\
q_16 \bullet \\
\end{array}
\end{align*} \]

Figure 1. Free point mass in half-plane. (a) Classical variational trajectory and (b) discontinuous variational trajectory.
Evidently, these discrete equations of motion differ from (134) by the presence of additional frictional forces identical to those in (115).

We see from these examples that, as in the continuous time integrators, the discontinuous time integrators, owing to their variational structure, enjoy exact conservation properties through collisions. However, in contrast to the classical variational integrators and in analogy to incremental minimization approaches, the discontinuous time integrators do not require the determination of trajectory intersections with the admissible domain boundary, and have a convenient and computational efficient predictor-corrector structure, with the corrector taking the form of a closest-point projection to the admissible set. We may thus conclude that, when applied to collisions, the discontinuous time integrators do indeed combine the best properties of variational time integrators and incremental minimization approaches.

5. IMPLEMENTATION DETAILS

In Section 6, we collect a number of examples of application and numerical tests concerned with rigid multibody dynamics that illustrate the versatility and good conservation and accuracy properties of the DV time integrators. In all the examples, we make extensive use of the constrained formulation of rigid body dynamics [13], further developed in [14, 15], and the SSH representation [16] of the corresponding one-sided contact constraints. We begin this section with brief accounts of these numerical tools.

5.1. Supporting separating hyperplane algorithm

To detect contact between convex bodies (nonconvex bodies are considered as being composed by convex subbodies), a subdifferentiable global contact detection algorithm, the so called SSH algorithm, developed in [16] is used, see Figure 2. The efficiency and convenience of integration schemes based on an implicit definition (13) of the admissible set also hinges critically on the identification of a suitable set of constraint functions \( g \). Recently, Johnson et al. [16] have formulated a class of contact constraints, termed SSH constraints, based on separation theorems of convex analysis. For the positive integer \( n \), let \( \alpha \notin 0 \in \mathbb{R}^n \) and \( a \in \mathbb{R} \), then the affine and convex set

\[
H_{\alpha,a} = \{ x \in \mathbb{R}^n | (\alpha, x) - a = 0 \}
\]

is called a hyperplane in \( \mathbb{R}^n \). Here, \( \alpha \) is the normal vector to the plane, while \( a = (\alpha, a) \) for some point \( a \in \mathbb{R}^n \) belonging to the plane. The signed distance between a point \( y \in \mathbb{R}^n \) and a hyperplane can be computed as \( (\alpha, y) - a \). In particular, Johnson et al. [16] have shown that, for every pair \((K_1, K_2)\) of convex bodies, the function \( g_C(K_1, K_2) \) determined via the quadratically constrained linear programming problem

\[
\frac{x_2 - x_1}{t_2 - t_1} - \frac{x_3 - x_2}{t_3 - t_2} - \frac{t_3 - t_1}{2} \mu |f_C| = 0,
\]

\[
m\frac{y_2^+ - y_1}{t_2 - t_1} - m\frac{y_3 - y_2}{t_3 - t_2} + \frac{t_3 - t_1}{2} f_C = 0,
\]

\[
m \left( \frac{x_2 - x_1}{t_2 - t_1} \right)^2 + m \left( \frac{y_2^+ - y_1}{t_2 - t_1} \right)^2 - m \left( \frac{x_3 - x_2}{t_3 - t_2} \right)^2 - m \left( \frac{y_3 - y_2}{t_3 - t_2} \right)^2 = \mu |f_C| \frac{|x_2 - x_1| + |x_3 - x_2|}{2}.
\]
Figure 2. Schematic of the supporting separating hyperplane algorithm. Figure 2(a)–2(c) shows the domain of potentially separating supporting hyperplanes for two separable polyhedral sets. Figure 2(d)–2(f) shows the same domain for sets that are not separable.

\[ g_C(K_1, K_2) = - \max_{a_1, a_2 \in \mathbb{R}^n} (a_1 - a_2), \]

subject to \( \langle \alpha, x \rangle - a_1 \geq 0, \forall x \in \text{ext } K_1, \)
\( \langle \alpha, x \rangle - a_2 \leq 0, \forall x \in \text{ext } K_2, \)
\( \langle \alpha, \alpha \rangle = 1 \) \tag{138}

is positive if the sets interpenetrate and negative otherwise and, therefore, supplies a contact constraint function. In (138), \( \text{ext } K \) denotes the set of extremal points of the a convex set \( K \). If \( K \) is a polytope, then \( \text{ext } K \) coincides with its vertex set. If \( K_1 \) and \( K_2 \) are polytopes, then the evaluation of \( g_C(K_1, K_2) \) and its derivatives can be conveniently reduced to a linearly constrained linear programming problem (SSH LP)

\[ g_C(K_1, K_2) = - \max_{a_1, a_2 \in \mathbb{R}^n} (a_1 - a_2), \]

subject to \( \langle \alpha, x \rangle - a_1 \geq 0, \forall x \in \text{ext } K_1, \)
\( \langle \alpha, x \rangle - a_2 \leq 0, \forall x \in \text{ext } K_2, \)
\( \langle \beta, \alpha \rangle = 1 \) \tag{139}

see [16] and Remark 5.1 for remarks on the choice of \( \beta \) in praxis. The admissible set of an arbitrary number of convex bodies can be represented by means of pairwise contact constraint functions, one per pair of bodies. Contact constraints for nonconvex bodies can also be formulated using the SSH approach by decomposition into convex subbodies. For instance, general contact constraints for finite-element models can be formulated by applying the SSH LP algorithm to every pair of elements.
5.1.1. Subgradients for the supporting separating hyperplane linear program. As shown in [16] and mentioned in Section 2.3, the generalized derivative of the SSH LP lies in the normal cone for configurations in which \( g_C(q) = 0 \). Furthermore, if the set of optimal primal solutions to the SSH LP for the variables \( \alpha \) is denoted \( A^*_p \), and the set of optimal dual solutions is denoted \( \Pi^* \subseteq \mathbb{R}^{\text{ext } K_1 + \text{ext } K_2} \), then a direction \( \nabla_{x_i} g_C \in \partial g_C \) is given by

\[
\nabla_{x_i} g_C = -\pi_i \alpha, \quad \forall x_i \in \text{ext } K_1,
\]

\[
\nabla_{x_i} g_C = \pi_i \alpha, \quad \forall x_i \in \text{ext } K_2,
\]

(140)

for \( \alpha \in A^*_p \) and the components \( \pi_i \) of a dual variable vector \( \pi \in \Pi^* \). To simplify notation, introduce the index sets \( I_1 \) and \( I_2 \), to indicate the vertex sets of \( K_1 \) and \( K_2 \), respectively. Whence, the dual feasibility conditions at the optimum of the SSH LP are given by

\[
\pi_i \geq 0, \quad \sum_{i \in I_1} \pi_i = 1, \quad \sum_{i \in I_2} \pi_i = 1,
\]

\[
\sum_{i \in I_1} x_i \pi_i - \sum_{i \in I_2} x_i \pi_i = h,
\]

(141)

with \( h \) being the separation vector, that is, \( h = g_C \beta \).

This implies that \( \sum_{i \in I_1} x_i \pi_i \) and \( \sum_{i \in I_2} x_i \pi_i \) are points on the convex hull of their respective sets, and are the same point when there is no overlap, that is, when \( g_C = 0 \). Thus, a specific contact point on each body can be recovered directly from the solution to the SSH LP.

5.2. Constrained formulation of rigid body dynamics

The efficiency and convenience of integration schemes based on an implicit definition (7) of the configuration manifold hinges critically on the identification of a suitable set of constraint functions \( g_Q \). In the examples presented subsequently, we make use of the constrained formulation of rigid body dynamics [13] and further developed in [14, 15], where it is shown that the constrained equations of motion are consistent with the classical Newton–Euler equations for the rigid body dynamics. In this formulation, the motion of a rigid body is described by the mapping (cf. Figure 3)

\[
x(X, t) = \varphi(t) + X_I d_I(t),
\]

(142)

where \( X_I \in \mathbb{R} \) and \( I = 1, 2, 3 \) are material coordinates in the body-fixed director triad \( \{d_I(t)\} \), and \( \varphi(t) \) describes the motion of the origin of the triad. Thus, in this representation, the generalized coordinates \( q(t) \) of the rigid body consist of \( \varphi(t) \) and \( \{d_I(t)\} \), and the corresponding conjugate momentum \( p(t) \) comprises \( p_\varphi(t) \) and \( \{p_I(t)\} \). In the absence of external forcing, the unconstrained Lagrangian reduces to
which represents the kinetic energy of the body. In order to represent the rigidity of the body, the directors are constrained to stay orthonormal during the motion via the constraint functions $g_Q(q) = 0$. For the rigid body formulation in use, angular momentum can be computed as

$$L(q, p) = \p \times p_\psi + \sum_{I=1}^{3} d_I \times p_I.$$  \hfill (144)

The constrained description of rigid bodies in terms of directors is also well suited to multibody dynamics. In particular, the use of directors facilitates the coupling of bodies by means of joints. Suitable constraint functions describing different types of joints may be found in [14, 15].

### 5.2.1. Subgradients for the constrained rigid body formulation.

It can be seen in (142) that in this formulation, each point in the rigid body, and in particular each vertex $x_i$, has an associated vector $X_i \in \mathbb{R}^3$. For concreteness, let the points in $I_1$ be associated with a rigid body in configuration $q^1$, and likewise, let the points in $I_2$ be associated with a rigid body configuration $q^2$. By the application of the chain rule, the relationship

$$\nabla q^j g_C = \sum_{i \in I_i} \left( \frac{\partial x_i}{\partial q^j} \right)^T \nabla x_i g_C = -A^j \alpha,$$

holds with the matrices

$$A^j = \begin{bmatrix} I \\ \sum_{i \in I_1} \pi_j X_{i1} I \\ \sum_{i \in I_2} \pi_j X_{i2} I \\ \sum_{i \in I_3} \pi_j X_{i3} I \end{bmatrix}$$  \hfill (146)

for $j = 1, 2$ and for the separation vector, $\nabla q^j h$ is given by $\pm (A^j)^T$, where the (+) corresponds to $j = 1$ and the (−) to $j = 2$.

**Remark 5.1**

Note that the vector $\beta$ in the SSH LP can be taken as the direction from the center of mass of $K_2$ to the center of mass of $K_1$, thus, for rigid bodies, $\beta = \varphi^1 - \varphi^2$.

### 5.3. Decomposition-based contact dynamics

The implementation of the theory developed in Section 4.2 requires solving the closest-point projection together with a reflection of the trajectory at the boundary of the admissible set. This section gives an overview on the implementation of these steps when the algorithm is used to simulate contact problems of rigid multibody system dynamics. The accuracy of these approximations, for example, as regards strict conservation properties, is subsequently assessed by means of selected numerical examples in Section 6.

The collision response is decomposed into a closest-point projection operation $P_C : Q \rightarrow Q$, mapping the inadmissible configuration $q^+_k$ to an admissible configuration $q^-_k = P_C(q^+_k)$ on the boundary of the admissible set via the closest-point projection $P_C$, and the use of a momentum

$$L(q, \dot{q}) = \int_B \frac{\rho_0}{2} |\dot{x}(X, t)|^2 \, dV,$$  \hfill (143)

Figure 4. The closest-point projection maps the inadmissible configuration to the closest admissible one, that is, $q_k^- = P_C(q_k^+)$. Correspondingly, $P_C$ maps the momentum $p_k^-$ at the inadmissible configuration into the new cotangent space. For the point mass system shown here, $P_C$ is the identity. The momentum decomposition reflects the part of the momentum being normal to the contact surface resulting in the post-collision momentum $p_k^\perp$.

update according to the jump condition. The latter involves the map $P_C : T^*_{q_k^+} Q \rightarrow T^*_{q_k^-} Q$ between the cotangent spaces at the inadmissible and admissible configurations and a momentum decomposition. Details can be observed from Figure 4 for a point mass. Note, first of all, that, in the presence of a configuration dependent potential, the total energy is not conserved in the closest-point projection, and secondly, that for a point mass, $P_C$ is the identity. The lifted pre-collision momentum $P_C(p_k^\perp)$ is decomposed into one part $P_C(p_k^\perp)\parallel$ being tangential and another part $P_C(p_k^\perp)\perp$ being normal to the contact surface. Then the normal part is reflected to obtain the post-collision momentum $p_k^\perp$ in accordance with the jump conditions (16).

The procedure for rigid body dynamics is illustrated in Figure 5. Because of the presence of rotational DOFs, the closest-point projection of the configuration involves translation and rotation of the square, and $P_C$ rotates the momentum vector correspondingly. Here, the momentum decomposition reflects the part of the momentum being normal to the constraint manifold in accordance with the constraints on momentum level.

The determination of an admissible configuration from the inadmissible configuration via a closest-point projection operation is one key approximation in the algorithm. Using the nodal reparametrization $\tilde{F}_d$ introduced in Section 3.1 and setting $q_k^- = \tilde{F}_d(u_k^-, q_k^+)$, the closest-point projection can be formulated in terms of the mass norm as follows:

$$u_k^- = \arg \min_{u_k^-} \frac{1}{2h^2} \left[ \tilde{F}_d(u_k^-, q_k^+) - q_k^+ \right]^T M \left[ \tilde{F}_d(u_k^-, q_k^+) - q_k^+ \right],$$

subject to $I_{A\times C}(\tilde{F}_d(u_k^-, q_k^+)) \leq 0$.

(147)

where the use of the mass norm has proved to minimize the artificial change in the conservation quantities because of the CPP in praxis. The indicator function $I_{A\times C}$ of the admissible set defined in (19) is approximated by a penalty function involving a parameter $\epsilon$ as in (21). The numerical examples give evidence that independent of the choice of $\epsilon$, the solution of (147) lies exactly on the boundary of the admissible set. Furthermore, the regularized indicator function is a single nonlinear locally convex and continuously differentiable inequality constraint for the minimization. From a
Figure 5. The closest-point projection of the configuration of a rigid body involves translation and rotation, and $\tilde{P}_C$ rotates the momentum accordingly. The redundant momentum vectors representing angular momentum about the $z$-axis are shown in bold off-center arrows on the body. The effect of the collision is evident at configuration $q_{k+1}$ because the direction of rotation has changed.

computational perspective, this significantly simplifies the algorithm compared to the use of many nonlinear nonsmooth constraints.

As noted by [9], momentum decompositions are a particularly easy way to solve the impact equations if a contact configuration is known. The essence of these decompositions is to explicitly separate the normal (and sliding, if friction is considered) components of the momentum at the collision point and then modify these components according to the type of collision being considered. For example, for perfectly elastic collisions, the normal component of the momentum is reflected about the collision surface. If only translational DOFs are present in the system, this method is especially simple. However, if rotational DOFs and constraints are present, then the momenta are configuration dependent, and care must be taken to ensure that the momentum is decomposed and updated in the appropriate subspace.

In the absence of frictional forces, the tangential components of the momentum remain fixed; that is, only the components of the momentum in the direction normal to the contact surface are affected by the collision. After the pre-collision momentum is mapped into the new cotangent space, the algorithm proceeds by identifying the normal and tangential components of the momentum to the contact surface using the normality property of the gradient of the interpenetration constraint $\nabla g_C (q_k)$ at the contact point. The component of the momentum being constraint consistent and normal to the contact surface can be found explicitly via an orthogonal projection as

$$P_{\perp} = P^T \nabla g_C^T \left[ \nabla g_C P M^{-1} P^T \nabla g_C^T \right]^{-1} \nabla g_C P M^{-1} P^T \tilde{P}_C \left( P_k^+ \right).$$

Note that as already stated earlier in this section, the (sub-)gradients $\nabla g_C (q_k)$ are introduced in Section 5.1.1. Inelastic frictionless collisions can be modeled using a coefficient of restitution.
\[ e \in [0,1], \text{ with } e = 1 \text{ being perfectly elastic, and } e = 0 \text{ perfectly plastic. The post-collision momentum can be computed explicitly according to} \]

\[
P_k^- = P^T \tilde{F}_C (p_k^+) - (1 + e)P_{\perp},
\]

(149)

and the time stepping (60) is continued.

Next, we develop an impulse-based approach to model Coulombic frictional contact between two bodies. To begin, note that the tangential component of the momentum \( p_k \) can be further decomposed into fixed \( p_{\parallel fix} \) and sliding \( p_{\parallel slide} \) components. The momentum associated with (nonzero) fixed components does not cause any relative motion between the points of contact in the bodies; thus, it instantaneously keeps the separation vector defined in (141) zero. The sliding component represents relative motion along the tangential direction of a contact surface, which will be related to frictional forces. Analogously to (148) using orthogonal projections, we obtain

\[
P_{\perp} + p_{\parallel slide} = P^T \nabla h^T \left[ \nabla h \frac{f_{\text{red}}}{M_{\text{red}}^{-1}} P^T \nabla h^T \right]^{-1} \nabla h \frac{f_{\text{red}}}{M_{\text{red}}^{-1}} P^T \tilde{F}_C \left( p_k^+ \right).
\]

(150)

Again, the (sub-)gradients \( \nabla h \in \partial h \) are introduced in Section 5.1.1. What remains is to distinguish the sliding component of the momentum, which amounts to subtracting the normal momentum components determined in (148) from the results of (150).

A frictional impulse that captures slip-stick frictional behavior of Coulomb friction can be stated as follows. The maximum frictional impulse is given by the negative sliding component, which corresponds to perfect stick. In the Coulomb model with friction coefficient \( \mu \in \mathbb{R}^+ \), the tangential forces are applied according to the ratio between normal and tangential forces. In the extension of the key result of [9] to the present case of constrained rigid body dynamics, the sliding impulse can be defined as

\[
I_{\text{slide}} = \begin{cases} 
-p_{\parallel slide}, & \text{if } \frac{p_{\parallel slide}M_{\text{red}}^{-1}p_{\perp}}{p_{\perp}M_{\text{red}}^{-1}p_{\perp}} \leq \mu, \\
-\mu \frac{pp_{\parallel slide}M_{\text{red}}^{-1}pp_{\perp}}{p_{\perp}M_{\text{red}}^{-1}pp_{\perp}}, & \text{otherwise}.
\end{cases}
\]

(151)

Thus, in the presence of Coulombic friction, the explicit momentum update reads

\[
P_k^- = P^T \tilde{F}_C \left( p_k^+ \right) - (1 + e)P_{\perp} + I_{\text{slide}}.
\]

(152)

**Multiple collisions** Because of the assumption that collisions are simultaneous, accounting for multiple collisions amounts to using the gradient of \( \sum_{i \in C} g_{C_i}(q) \) (the sum of the components of the vector \( g_C \)) to resolve the normal momenta for each collision group, and solve one system for each group. Frictional forces, which depend on local relative motions, can be included by the sum of local impulses in the group.

6. NUMERICAL Examples

Away from collisions, the (DV) algorithm exactly conserves momentum and shows the good energy behavior, which is typical for structure preserving integrators. When collisions happen, the nonphysical changes in these quantities are exactly due to the closest-point projection operation—or to intentional dissipation in the momentum updates due to inelastic or frictional contacts. The momentum decomposition does not change the kinetic energy (for perfectly elastic frictionless collisions) nor does it change the angular momentum. In summary, in the event that the closest-point projection is the identity, both the total energy and the total angular momentum of the system are exactly preserved over a collision.
6.1. Newton’s cradle

The dynamics of the well-known Newton’s cradle toy, modified so that the balls are replaced by cubes and the strings are not modeled so there is only one series of collisions, is investigated for perfectly elastic collisions. Two cases of initial conditions are considered. In case 1, four cubes of side length $l = 0.2\sqrt{2}$ m and uniform density $\rho = 0.27\ \text{kg/m}^3$ are aligned at rest, being separated by $l=10$. One cube of the same dimension starts with its face a distance of $l=2$ from the face of the end cube and moves toward the bunch with initial velocity $0.105\ \text{m/s}$. As can be seen in the images in Figure 6, the impulse of the initial impact causes the cube on the opposite end of the bunch from the original moving cube to depart from the cluster. The fixed time step is $h = 0.01\ \text{s}$. The evolution of the energy and the deviation in total angular momentum from the initial value for this case are shown in Figure 7. Because of the presence of rotational symmetry in the Lagrangian of this problem, the total angular momentum is an exactly conserved quantity. The variational integrator is able to represent this before any collision happens. However, comparing the upper and lower plots in Figure 7, note that the CPP correction of approximately $10^{-4}\ \text{m}$ of each overlap induces a small change in the total angular momentum of the system (upper plot). Removing this yields a value that is zero to numerical accuracy (lower plot).

In case 2, three of the described cubes are aligned at rest being separated by $l=100$. Two cubes start with their faces a distance of $l/2$ from the face of the end cubes and move toward the bunch with initial velocities $0.105\ \text{m/s}$, respectively. As can be seen in the images in Figure 8, the impulse of the initial impact causes both of the end cubes moving away from the cluster. The evolution of the energy and the change in angular momentum for this case for $h = 0.01\ \text{s}$ are shown in Figure 9. In this example, it happens that the overlap of approximately $10^{-10}\ \text{m}$ corrected by the CPP is so small that the CPP induces no noticeable change in the angular momentum of the system.

In both of these examples, no relative rotation between the bodies is induced by the collision, which is the expected result for perfectly flat face–face contact.
Before proceeding to more involved examples, two cases of perfectly plastic collisions between two cubes are treated. In the first perfectly plastic contact example, one cube moves toward a stationary cube with a speed of 0.5 m/s. In the second perfectly plastic contact example, both cubes move toward each other with a speed of 0.5 m/s, so that the relative speed is 1.0 m/s. Several time lapse images of the motion as well as the evolution of the energy are shown in Figures 10 and 11, as well as Figure 12, respectively. They show the results expected from a simple algebraic calculation.

### 6.2. Metronome

A simple kinematic chain consisting of two kite-shaped prisms of uniform density $\rho = 0.27 \text{ kg/m}^3$ with side length $0.2\sqrt{3}$ m connected by a spherical joint at the top of the first body and the bottom of the second body is introduced. In the initial configuration, the third directors of both bodies are aligned, that is, $d_1^3 = d_2^3$, and the remaining directors of the second body are rotated by $\frac{\pi}{4}$ rad about
Figure 11. Time lapse images of a perfectly plastic collision between two moving cubes. After the collision, the cubes remain stationary. (a) $t = 0\, \text{s}$, (b) $t = 0.2\, \text{s}$, and (c) $t = 0.4\, \text{s}$.

Figure 12. Energy evolution for perfectly plastic Newton’s cradle examples for $h = 0.01\, \text{s}$. (a) Energy, $v_{rel} = 0.5\%$ (one moving cube) and (b) energy, $v_{rel} = 1.0\%$ (two moving cubes).

Figure 13. Time lapse images for a prismatic metronome connected by a spherical joint. (a) $t = 0\, \text{s}$, (b) $t = 1.8\, \text{s}$, (c) $t = 3.6\, \text{s}$, (d) $t = 5.4\, \text{s}$, (e) $t = 7.2\, \text{s}$, and (f) $t = 9.0\, \text{s}$.

d\frac{1}{4} relative to the first body. The initial angular velocities are $\omega_1 = [0\, 1\, 0]$ and $\omega_2 = [0\, -1\, 0]$, so that the bodies rotate about the joint and swing toward each other, collide when the edge of the second body strikes the face of the first body, and so on. Several time lapse images of this motion are shown in Figure 13. The evolution of the energies and the change in angular momentum...
Figure 14. Change in angular momentum and energy evolution for a prismatic metronome connected by a spherical joint solved for $h = 0.015$ s. (a) Angular momentum and (b) energy.

for the metronome example are shown in Figure 14. Comparing the upper and lower plots, note that the CPP correction causes small changes in otherwise conserved components of the angular momentum. In the example shown, the fixed time step is $h = 0.015$ s, and (clearly) only pairwise collisions take place.

As in the Newton’s cradle examples, no additional relative rotation is induced by the collisions, which is the expected result for this particular edge–face contact configuration. As shown in Figure 14, removing the changes due to the CPP operation in the angular momentum leads to conservation up to numerical accuracy.

6.3. Falling polygons

The next example considers seven polygons starting in a flower-like cluster with centers of mass placed at 0.8 m above the floor. All polygons have a small initial velocity of absolute value 0.01 m/s in random directions, and no initial angular velocity. For the purposes of this example, the floor is also modeled as a polygon occupying the $(x, y)$-plane with vertices at $(x, y) = (\pm 10, \pm 10)$ cm, and normal $e_z$. In the evaluation of the SSH LP to determine overlap and the necessary (sub-)gradients for body-floor interactions, $\beta = e_z$ is always chosen. Three different shapes are used: a cube with side length 0.2 m, a right square pyramid with base dimension 0.2 m and height 0.4 m, and a regular kite-shaped prism with side length $0.2\sqrt{3}$ m, all with a uniform density of $\rho = 0.27$ kg/m$^3$.

Figure 15. Time lapse images for falling polygons using the discontinuous variational algorithm. (a) $t = 0.0$ s, (b) $t = 0.2$ s, (c) $t = 0.4$ s, (d) $t = 0.6$ s, (e) $t = 0.8$ s, and (f) $t = 1.0$ s.
The polygons collide with each other and with the floor under the influence of a gravitational potential with $g = 9.81 \, \text{m/s}^2$. This example differs from previous examples due to the presence of a potential; now, the CPP operation effects the overall energy of the system and not just the momentum. A series of time lapse images of the system are shown in Figure 15, and the evolution of the total energy and the change in angular momentum are shown in Figure 16. Note that after changes due to the CPP operation are removed, $L_3$ is exactly preserved. The small incremental increases in the total energy are also due to the CPP operation, which has the effect of instantaneously increasing the gravitational potential without green affecting the kinetic energy (because the mass matrix in the employed rigid body formulation is constant), thus causing a small increase in the overall energy. During the simulation, several incidents of multiple collisions happen, which are dealt with seamlessly by the algorithm.

Before moving on, we return to the simple system of the cube hitting the floor, but now model the collisions as inelastic ($e = 0.6$) and frictional ($\mu = 0.8$). Because energy is dissipated from the system at every collision, a static solution with the cube resting on the floor is eventually reached, as shown in Figure 17. Even though the algorithm is intended as a collision integrator, the algorithm does a reasonable job dealing with an essentially static problem, with the kinetic energy tending to zero and the gravitational energy attaining a minimum. Note that the presence of friction prevents the conservation of angular momentum in the initial series of collisions, but the angular momentum oscillates about zero, which is the expected result. The evolution of the energy and momentum less changes due to the CPP operation are shown in Figure 18.

Figure 16. Change in angular momentum and energy evolution for falling polygons with $h = 0.001 \, \text{s}$. Small changes are due to the CPP operation. (a) Angular momentum, (b) angular momentum, and (c) energy.
6.4. Hexagons

This example illustrates an instance in which the (DV) algorithm overcomes the limitations of the discrete variational exact collision integrator in [3, 4]. In addition, by the introduction of dissipation into the system via friction or restitution, the formerly ‘potentially’ clumping rigid right hexagonal prisms successfully assemble into a coherent structure. The hexagons have height 0.4 m and a side length of 0.2 m. As in previous examples, a series of alternating positive and negative Coulombic charges are centered behind each side face (Figure 19), with the constant $K = 5e - 4$ in the potential.
This type of system has several complicating factors for the exact integrator, which are overcome here. For the exact integrator, aside from the obvious issues of multiple contacts being compounded by the attractive potentials, the potential depends on the relative configuration of each pair of bodies, so that the discrete collision equations must be solved for the full system, so that the impact equations could not be reduced to local problems. In the present (DV) algorithm, these issues do not pose a problem because the basis of the algorithm is the assumption of simultaneous collisions at the end of a times step. That being said, as in the example with a gravitational potential, the CPP operation causes an increase in the overall energy of the system if no dissipation is present, as can be seen in Figure 20, which also shows the evolution of the change in angular momentum with changes due the CPP operation removed. Time lapse images of the motion for a time step of $h = 1e^{-3}$ s are shown in Figure 21.

Figure 20. Change in angular momentum and energy evolution for rigid hexagons with $h = 0.001$ s and no dissipation. The incremental increase in energy is due to the CPP operation, which increases the Coulombic potential energy. (a) Angular momentum and (b) energy.

Figure 21. Time lapse images for rigid hexagons with the motion driven by internal Coulombic charges, as shown in Figure 19. No dissipation is present in this system. (a) $t = 0$ s, (b) $t = 0.4$ s, (c) $t = 0.8$ s, (d) $t = 1.2$ s, (e) $t = 1.6$ s, and (f) $t = 2.0$ s.
Figure 22. Time lapse images for rigid hexagons with the motion driven by internal Coulombic charges, as shown in Figure 19. The collisions are modeled as inelastic and frictional with $\mu = 0.8$ and $e = 0.8$. The steady state solution at a minimal potential energy is successfully captured. (a) $t = 0$ s, (b) $t = 0.4$ s, (c) $t = 0.8$ s, (d) $t = 1.2$ s, (e) $t = 1.6$ s, and (f) $t = 2.0$ s.

Figure 23. Energy evolution for rigid hexagons with $h = 0.001$ s with $\mu = 0.8$ and $e = 0.8$. The solution approaches a steady state with only small oscillations about the minimum.

By making all collisions frictional with $\mu = 0.8$ and inelastic with $e = 0.8$, the present algorithm successfully captures the assembly of the hexagons into a coherent flower-like structure with only small oscillations about the minimum energy, as shown in Figures 22 and 23.

7. SUMMARY AND CONCLUSIONS

We have formulated a new class of DV time integrators that allow the system to adopt two possibly different configurations at each sampling time $t_k$, representing predictor and corrector configurations of the system. The resulting sequence of configuration pairs then represents a discontinuous—or non-classical—trajectory of the system. Continuous or classical trajectories are recovered simply by enforcing a continuity constraint at all times. In particular, when simulating systems subject to one-
sided contact constraints with DV time integrators, the predictor configuration is a trial configuration not required to satisfy the one-sided constraints, whereas the corrector configuration is obtained by a closest-point projection of the predictor onto the admissible set. The resulting trajectories are generally discontinuous, or non-classical, but are expected to converge to classical or continuous solutions in the limit of a vanishingly small time step. We account for dissipation, including friction, by means of a discrete Lagrange–d’Alembert principle, and we make extensive use of the spacetime formalism in order to ensure exact energy conservation in conservative systems, and the right rate of energy decay in dissipative systems. The structure, range and scope of the DV time integrators, and their accuracy and convergence characteristics are illustrated by means of a number of examples of application concerned with rigid multibody dynamics.

In closing, we point out some of the limitations of the present work and possible directions for further development of DV time integrators. Whereas the numerical examples presented are suggestive of strong conservation and accuracy properties, a mathematically rigorous analysis of convergence of DV time integrators would be highly desirable. Because, in the presence of collisions, trajectories are necessarily discontinuous and defined in a weak sense, robust variational notions of convergence, such as $\Gamma$ convergence, suggest themselves. Examples of application of $\Gamma$ convergence to the analysis of time-integration algorithms may be found in [5, 19].

APPENDIX A: FINITE-DIMENSIONAL NONSMOOTH ANALYSIS

In this appendix, we review basic concepts of nonsmooth analysis used tacitly or explicitly in the preceding derivations. A complete account on nonsmooth analysis may be found in the monograph [8]. In essence, nonsmooth analysis deals with physical objects and functions for which strict differentiability may not be postulated. In particular, one of the goals of nonsmooth analysis is to develop a collection of tools enabling the study of differential properties of nondifferentiable functions.

Let $X = \mathbb{R}^n$ be equipped with the Euclidean norm $\| \cdot \|$. We identify the dual space $X^*$ with $\mathbb{R}^n$ itself, and denote the duality pairing $\langle \cdot, \cdot \rangle$. Physically, $X$ is to be regarded as the space of positions and velocities, whereas $X^*$ is the space of forces. The duality pairing $\langle f, v \rangle$, $v \in X$, $f \in X^*$, is then the power developed as a point moves with velocity $v$ under the action of force $f$. Throughout this section, unless otherwise stated, a function $f$ is understood to be defined over $X$ and to take values over the real line $\mathbb{R}$.

Let $f$ be Lipschitz near a point $x$ and let $v$ be a vector in $X$. The generalized directional derivative of $f$ at $x$ in the direction $v$ is

$$f^\circ(x; v) = \limsup_{y \to x, t \to 0^+} \frac{f(y + tv) - f(y)}{t}, \tag{A.1}$$

where $y \in X$ and $t$ is a positive scalar. This definition generalizes the conventional concept of directional derivative in that it does not presuppose the existence of any limits, as it involves the upper limit only, and the base point $y$ is allowed to vary. The generalized differential (aka gradient) of a Lipschitz function $f$ at $x$ is the subset of $X^*$ given by

$$\partial f(x) = \{ f \in X^* \mid f^\circ(x; v) \geq \langle f, v \rangle, \text{ for all } v \in X \}. \tag{A.2}$$

The generalized directional derivative is recovered from the generalized gradient as

$$f^\circ(x; v) = \max\{ \langle f, v \rangle \mid f \in \partial f(x) \}. \tag{A.3}$$

The generalized gradient reduces to the ordinary derivative at points where the function is continuously differentiable. It also reduces to the subdifferential in the case of convex functions. Generalized directional derivatives and gradients provide a means of characterizing local minima of nonsmooth functions. Thus, if $f$ is Lipschitz and if $x$ is a local minimum of $f$, then
There are a number of connections between nonsmooth analysis and geometry, which will be proven useful in contact applications. We begin by considering a nonempty subset $C \subset X$ and defining the distance function as

$$d_C(x) = \inf\{\|x - y\| \mid y \in C\}. \quad (A.5)$$

This function is globally Lipschitz of rank 1 but not differentiable in the classical sense. The distance function $d_C(x)$ provides a useful device for characterizing certain aspects of the geometry of $C$. For instance, the tangent cone $T_C(x)$ to $C$ at $x$ is the set

$$T_C(x) = \{v \in X \mid d_C^2(x; v) = 0\}. \quad (A.6)$$

The normal cone $N_C(x)$ to $C$ at $x$ is defined by polarity as

$$N_C(x) = \{f \in X^* \mid \langle f, v \rangle \leq 0, \text{ for all } v \in T_C(x)\}. \quad (A.7)$$

Evidently, if $x$ is in the interior of $C$ then $T_C(x) = X$ and $N_C(x) = \{0\}$. Furthermore, if $C$ is closed and $x$ is a regular (i.e., smooth) point of $\partial C$, then $T_C(x)$ is the interior tangent halfspace to $C$ at $x$ and $N_C(x)$ is the outward normal ray.

Clarke [8] has shown how generalized gradients and the related concepts outlined earlier can be extended to functions that are not locally Lipschitz and take values in the extended real line $\mathbb{R} \cup \{\pm \infty\}$. In particular, we recall that the indicator of a set $C$ in $X$ is the extended-valued function:

$$I_C(x) = \begin{cases} 0, & \text{if } x \in C, \\ \infty, & \text{otherwise}. \end{cases} \quad (A.8)$$

Let $x \in C$. Then it follows that

$$\partial I_C(x) = N_C(x), \quad x \in C, \quad (A.9)$$

that is, the generalized differential of the indicator function at $x \in C$ coincides with the normal cone at $x$.

The set identity (A.9) establishes an important connection between generalized gradients and geometry. Another such connection may be effected by introducing the concept of closest-point projection. Thus, let $C$ be a set in $X$ and $y$ be a point in $X$. The closest-point projection of $y$ onto $C$ is the set

$$P_C(y) = \{x \in \text{cl } C \mid d_C(y) = \|x - y\|\}. \quad (A.10)$$

It follows from the definition that $P_C^2 = P_C$ and, therefore, $P_C$ does indeed define a nonlinear projection from $X$ onto the closure of $C$. Evidently, if $C$ is closed, then $P_C(y)$ reduces to the singleton $\{y\}$ iff $y \in C$. For a convex set $C$, the corresponding closest-point projection $P_C$ is uniquely defined and maps $X$ onto $C$. However, $P_C$ is generally set-valued for arbitrary nonconvex sets. Let $C$ be a nonempty subset of $X$. Then, it follows that $x = P_C(y)$ if and only if

$$y - x \in \partial I_C(x) = N_C(x), \quad (A.11)$$

which establishes the sought relation between the closest-point projection, the generalized differential and the normal cone.

ACKNOWLEDGEMENTS

The support of the W. M. Keck Foundation through Caltech’s Keck Institute for Space Studies is gratefully acknowledged.
REFERENCES