TYPE II HAMILTONIAN LIE GROUP VARIATIONAL INTEGRATORS WITH APPLICATIONS TO GEOMETRIC ADJOINT SENSITIVITY ANALYSIS

BRIAN K. TRAN[1] AND MELVIN LEOK[2]

Abstract. Variational integrators for Euler–Lagrange equations and Hamilton's equations are a class of structure-preserving numerical methods that respect the conservative properties of such systems. Lie group variational integrators are a particular class of these integrators that apply to systems which evolve over the tangent bundle and cotangent bundle of Lie groups. Traditionally, these are constructed from a variational principle which assumes fixed position endpoints. In this paper, we instead construct Lie group variational integrators with a novel Type II variational principle on the cotangent bundle of a Lie group which allows for Type II boundary conditions, i.e., fixed initial position and final momenta; these boundary conditions are particularly important for adjoint sensitivity analysis, which is the motivating application in our paper. In general, such Type II variational principles are only globally defined on vector spaces or locally defined on general manifolds; however, by left translation, we are able to define this variational principle globally on cotangent bundles of Lie groups. By developing the continuous and discrete Type II variational principles over Lie groups, we construct a structure-preserving Lie group variational integrator that is both symplectic and momentum-preserving. Subsequently, we introduce adjoint systems on Lie groups, and show how these adjoint systems can be used to perform geometric adjoint sensitivity analysis for optimization problems on Lie groups. Finally, we conclude with two numerical examples to show how adjoint sensitivity analysis can be used to solve initial-value optimization problems and optimal control problems on Lie groups.

Contents

[1]Los Alamos National Laboratory, Theoretical Division, Los Alamos, NM 87545, USA.
[2]Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0112, USA.
E-mail address: btran@lanl.gov, mleok@ucsd.edu.

1. Introduction

In this paper, we aim to develop Lie group variational integrators from a Type II variational principle with the motivating application of performing intrinsic geometric adjoint sensitivity analysis on Lie groups. Lie Group variational integrators are a class of geometric structure-preserving integrators for integrating Lagrangian or Hamiltonian systems evolving over tangent and cotangent bundles of Lie groups (see [5; 23; 26–28; 34; 37]). Such methods generally have good conservation properties, such as respecting the symplecticity and momentum-preservation of these systems. Adjoint systems provide an efficient method for performing dynamically-constrained optimization and sensitivity analysis. The geometry of these systems has gained interest as it can be described by a Hamiltonian structure. Particularly, the Hamiltonian structure of adjoint systems encode a quadratic conservation law which is the key to adjoint sensitivity analysis [44]. We aim to synthesize these two areas of research, by developing Lie group variational integrators which are applicable to the maximally degenerate Hamiltonian structures found in adjoint systems and hence, develop geometric integrators which respect the quadratic conservation law enjoyed by adjoint systems, making them particularly useful for adjoint sensitivity analysis on Lie groups. We begin with a brief introduction and review of these topics.

1.1. **Lagrangian and Hamiltonian Variational Integrators.**

Geometric numerical integration aims to preserve geometric conservation laws under discretization, and this field is surveyed in the monograph by Hairer et al. [18]. Discrete variational mechanics [30; 36] provides a systematic method of constructing symplectic integrators. It is typically approached from a Lagrangian perspective by introducing the *discrete Lagrangian*, $L_d : Q \times Q \to \mathbb{R}$, which is a Type I generating function of a symplectic map and approximates the *exact discrete Lagrangian*, which is constructed from the Lagrangian $L : TQ \to \mathbb{R}$ as

$$L_d^E(q_0, q_1; h) = \operatorname{ext}_{\substack{q \in C^2([0,h],Q) \\ q(0)=q_0, q(h)=q_1}} \int_0^h L(q(t), \dot{q}(t))dt, \tag{1.1}$$

which is equivalent to Jacobi's solution of the Hamilton–Jacobi equation. The exact discrete Lagrangian generates the exact discrete-time flow map of a Lagrangian system, but, in general, it cannot be computed explicitly. Instead, this can be approximated by replacing the integral with a quadrature formula, and replacing the space of C^2 curves with a finite-dimensional function space. Given a finite-dimensional function space $\mathbb{M}^n([0, h]) \subset C^2([0, h], Q)$ and a quadrature formula $\mathcal{G} : C^2([0, h], Q) \to \mathbb{R}$, $\mathcal{G}(f) = h \sum_{j=1}^m b_j f(c_j h) \approx \int_0^h f(t)dt$, the *Galerkin discrete Lagrangian* is

$$L_d(q_0, q_1) = \operatorname{ext}_{\substack{q \in \mathbb{M}^n([0,h]) \\ q(0)=q_0, q(h)=q_1}} \mathcal{G}(L(q, \dot{q})) = \operatorname{ext}_{\substack{q \in \mathbb{M}^n([0,h]) \\ q(0)=q_0, q(h)=q_1}} h \sum_{j=1}^m b_j L(q(c_j h), \dot{q}(c_j h)).$$

Given a discrete Lagrangian L_d, the *discrete Hamilton–Pontryagin principle* imposes the discrete second-order condition $q_k^1 = q_{k+1}^0$ using Lagrange multipliers p_{k+1}, which yields a variational principle on $(Q \times Q) \times_Q T^*Q$,

$$\delta \left[\sum\nolimits_{k=0}^{n-1} L_d(q_k^0, q_k^1) + \sum\nolimits_{k=0}^{n-2} p_{k+1}(q_{k+1}^0 - q_k^1)\right] = 0.$$

This in turn yields the *implicit discrete Euler–Lagrange equations*,

$$q_k^1 = q_{k+1}^0, \qquad p_{k+1} = D_2 L_d(q_k^0, q_k^1), \qquad p_k = -D_1 L_d(q_k^0, q_k^1), \tag{1.2}$$

where D_i denotes the partial derivative with respect to the i-th argument. Making the identification $q_k = q_k^0 = q_{k-1}^1$, we obtain the *discrete Lagrangian map* and *discrete Hamiltonian map* which are $F_{L_d} : (q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$ and $\tilde{F}_{L_d} : (q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$, respectively. The last two equations of (1.2) define the *discrete fiber derivatives*, $\mathbb{F}L_d^{\pm} : Q \times Q \to T^*Q$,

$$\mathbb{F}L_d^+(q_k, q_{k+1}) = (q_{k+1}, D_2 L_d(q_k, q_{k+1})),$$
$$\mathbb{F}L_d^-(q_k, q_{k+1}) = (q_k, -D_1 L_d(q_k, q_{k+1})).$$

These two discrete fiber derivatives induce a single unique *discrete symplectic form* $\Omega_{L_d} = (\mathbb{F}L_d^{\pm})^*\Omega$, where Ω is the canonical symplectic form on T^*Q, and the discrete Lagrangian and Hamiltonian maps preserve Ω_{L_d} and Ω, respectively. The discrete Lagrangian and Hamiltonian maps can be expressed as $F_{L_d} = (\mathbb{F}L_d^-)^{-1} \circ \mathbb{F}L_d^+$ and $\tilde{F}_{L_d} = \mathbb{F}L_d^+ \circ (\mathbb{F}L_d^-)^{-1}$, respectively. This characterization allows one to relate the approximation error of the discrete flow maps to the approximation error of the discrete Lagrangian.

The variational integrator approach simplifies the numerical analysis of symplectic integrators. The task of establishing the geometric conservation properties and order of accuracy of the discrete Lagrangian map F_{L_d} and discrete Hamiltonian map $\tilde{F}_{L_d}$ reduces to the simpler task of verifying certain properties of the discrete Lagrangian L_d instead.

Theorem 1.1 (Discrete Noether's theorem (Theorem 1.3.3 of [36]))**.** *If a discrete Lagrangian L_d is invariant under the diagonal action of G on $Q \times Q$, then the single unique discrete momentum map, $\mathbf{J}_{L_d} = (\mathbb{F}L_d^{\pm})^*\mathbf{J}$, is invariant under the discrete Lagrangian map F_{L_d}, i.e., $F_{L_d}^*\mathbf{J}_{L_d} = \mathbf{J}_{L_d}$.*

Theorem 1.2 (Variational error analysis (Theorem 2.3.1 of [36]))**.** *If a discrete Lagrangian L_d approximates the exact discrete Lagrangian L_d^E to order p, i.e., $L_d(q_0, q_1; h) = L_d^E(q_0, q_1; h) + \mathcal{O}(h^{p+1})$, then the discrete Hamiltonian map $\tilde{F}_{L_d}$ is an order p accurate one-step method.*

The bounded energy error of variational integrators can be understood by performing backward error analysis, which then shows that the discrete flow map is approximated with exponential accuracy by the exact flow map of the Hamiltonian vector field of a modified Hamiltonian [3; 47].

Given a degenerate Hamiltonian, where the Legendre transform $\mathbb{F}H : T^*Q \to TQ$, $(q, p) \mapsto (q, \frac{\partial H}{\partial p})$, is noninvertible, there is no equivalent Lagrangian formulation. Thus, a characterization of variational integrators directly in terms of the continuous Hamiltonian is desirable. This is achieved by considering the Type II analogue of Jacobi's solution, given by

$$H_d^{+,E}(q_k, p_{k+1}) = \operatorname*{ext}_{\substack{(q,p)\in C^2([t_k,t_{k+1}],T^*Q) \\ q(t_k)=q_k, p(t_{k+1})=p_{k+1}}} \left[p(t_{k+1})q(t_{k+1}) - \int_{t_k}^{t_{k+1}} [p\dot{q} - H(q,p)]\, dt\right].$$

A computable Galerkin discrete Hamiltonian H_d^+ is obtained by choosing a finite-dimensional function space and a quadrature formula,

$$H_d^+(q_0, p_1) = \operatorname*{ext}_{\substack{q\in\mathbb{M}^n([0,h]) \\ q(0)=q_0 \\ (q(c_jh),p(c_jh))\in T^*Q}} \left[p_1 q(t_1) - h\sum\nolimits_{j=1}^m b_j[p(c_jh)\dot{q}(c_jh) - H(q(c_jh), p(c_jh))]\right].$$

Interestingly, the Galerkin discrete Hamiltonian does not require a choice of a finite-dimensional function space for the curves in the momentum, as the quadrature approximation of the action integral only depend on the momentum values $p(c_j h)$ at the quadrature points, which are determined by the extremization principle. In essence, this is because the action integral does not depend on the time derivative of the momentum $\dot{p}$. As such, both the Galerkin discrete Lagrangian and the Galerkin discrete Hamiltonian depend only on the choice of a finite-dimensional function space for curves in the position, and a quadrature rule. It was shown in Proposition 4.1 of [31] that when the Hamiltonian is hyperregular, and for the same choice of function space and quadrature rule, they induce equivalent numerical methods.

The *Type II discrete Hamilton's phase space variational principle* states that

$$\delta \left\{ p_N q_N - \sum_{k=0}^{N-1} \left[p_{k+1} q_{k+1} - H_d^+(q_k, p_{k+1}) \right] \right\} = 0,$$

for discrete curves in T^*Q with fixed (q_0, p_N) boundary conditions. This yields the *discrete Hamilton's equations*, which are given by

$$q_{k+1} = D_2 H_d^+(q_k, p_{k+1}), \qquad p_k = D_1 H_d^+(q_k, p_{k+1}). \tag{1.3}$$

Given a discrete Hamiltonian H_d^+, we introduce the *discrete fiber derivatives* (or discrete Legendre transforms), $\mathbb{F}^+ H_d^+$,

$$\begin{aligned} \mathbb{F}^+ H_d^+ &: (q_0, p_1) \mapsto (D_2 H_d^+(q_0, p_1), p_1), \\ \mathbb{F}^- H_d^+ &: (q_0, p_1) \mapsto (q_0, D_1 H_d^+(q_0, p_1)). \end{aligned}$$

The discrete Hamiltonian map can be expressed in terms of the discrete fiber derivatives,

$$\tilde{F}_{H_d^+}(q_0, p_0) = \mathbb{F}^+ H_d^+ \circ (\mathbb{F}^- H_d^+)^{-1}(q_0, p_0) = (q_1, p_1),$$

Similar to the Lagrangian case, we have a discrete Noether's theorem and variational error analysis result for Hamiltonian variational integrators.

Theorem 1.3 (Discrete Noether's theorem (Theorem 5.3 of [31]))**.** *Given the action Φ on the configuration manifold Q, let Φ^{T^*Q} be the cotangent lifted action on T^*Q. If the generalized discrete Lagrangian $R_d(q_0, q_1, p_1) = p_1 q_1 - H_d^+(q_0, p_1)$ is invariant under the cotangent lifted action Φ^{T^*Q}, then the discrete Hamiltonian map $\tilde{F}_{H_d^+}$ preserves the momentum map, i.e., $\tilde{F}_{H_d^+}^* \mathbf{J} = \mathbf{J}$.*

Theorem 1.4 (Variational error analysis (Theorem 2.2 of [45]))**.** *If a discrete Hamiltonian H_d^+ approximates the exact discrete Hamiltonian $H_d^{+,E}$ to order p, i.e., $H_d^+(q_0, p_1; h) = H_d^{+,E}(q_0, p_1; h) + \mathcal{O}(h^{p+1})$, then the discrete Hamiltonian map $\tilde{F}_{H_d^+}$ is an order p accurate one-step method.*

It should be noted that there is an analogous theory of discrete Hamiltonian variational integrators based on Type III generating functions $H_d^-(p_0, q_1)$.

Remark 1.1. *It should be noted that the current construction of Hamiltonian variational integrators is only valid on vector spaces and local coordinate charts as it involves Type II/Type III generating functions $H_d^+(q_k . p_{k+1})$, $H_d^-(p_k, q_{k+1})$, which depend on the position at one boundary point, and the momentum at the other boundary point. However, this does not make intrinsic sense on a manifold, since one needs the base point in order to specify the corresponding cotangent space. One possible approach to constructing an intrinsic formulation of Hamiltonian variational integrators on general cotangent bundles is to start with discrete Dirac mechanics [30], and consider a generating function $E_d^+(q_k, q_{k+1}, p_{k+1})$, $E_d^-(q_k, p_k, q_{k+1})$, that depends on the position at both boundary points* *and the momentum at one of the boundary points. This approach can be viewed as a discretization of the generalized energy $E(q,v,p) = \langle p,v\rangle - L(q,v)$, in contrast to the Hamiltonian $H(q,p) = \text{ext}_v \langle p,v\rangle - L(q,v) = \langle p,v\rangle - L(q,v)\big|_{p=\frac{\partial L}{\partial v}}$.*

As mentioned in the previous remark, an issue with Type II Hamiltonian variational integrators is that they are only valid on vector spaces or on local charts, due to the Type II boundary conditions $q(t_k) = q_k, p(t_{k+1}) = p_{k+1}$, which requires a local trivialization of T^*Q. Furthermore, these methods cannot in general be extended to arbitrary parallelizable manifolds M, i.e., $T^*M \cong M\times V$ for some vector space V, since the isomorphism $T^*M \cong M \times V$ may be neither explicit nor computable. However, for a Lie group G, the trivialization $T^*G \cong G \times \mathfrak{g}^*$ is given simply by left or right translation. Using this trivialization, we will extend the construction of Type II Hamiltonian variational integrators to the setting of Hamiltonian systems on the cotangent bundle of a Lie group.

1.2. **Lie Group Variational Integrators.** Lie group variational integrators preserve the Lie group structure of the configuration space without the use of local charts, reprojection, or constraints. Instead, the discrete solution is updated using the exponential of a Lie algebra element that satisfies a discrete variational principle. These yield highly efficient geometric integration schemes for rigid body dynamics that automatically remain on the rotation group. We avoid coordinate singularities associated with local charts, such as Euler angles, by representing the attitude globally as a rotation matrix, which is important for accurately simulating chaotic orbital motion.

These ideas were introduced in [26], and applied to a system of extended rigid bodies interacting under their mutual gravitational potential in [27; 28], wherein symmetry reduction to a relative frame is also addressed. The superior computational efficiency of Lie group variational integrators for the full body simulation of systems of extended rigid bodies in the context of astrodynamics was demonstrated in [15].

Lie group variational integrators can be seen as the synthesis of Lie group methods (see, for example, [22]) and variational integrators that serves as the basis for constructing efficient geometric structure-preserving integrators for the dynamics of mechanical systems which evolve on Lie groups.

The basic idea of a Lie group method is to express the solution in terms of an element of the Lie algebra,

$$g(t) = g_0 \exp(\xi(t)),$$

as opposed to a group element, and to use the exponential map and group composition to ensure that the solution remains on the group. The problem reduces to finding an appropriate Lie algebra element $\xi \in \mathfrak{g}$, which is desirable, as the Lie algebra is always linear, even when the Lie group is nonlinear, and interpolants can be easily obtained. We construct an interpolant on the Lie group by using polynomial interpolation at the level of the Lie algebra.

The exponential (or an approximation thereof, such as the Cayley transform, or more generally, the diagonal Padé approximants, for quadratic matrix Lie groups [9]) allows one to approximate a curve on G by a discrete time curve on the Lie algebra. One can combine this with an approximation space for the fibers of $TG \cong G\times\mathfrak{g}$ to obtain a discrete Lagrangian. Enforcing a discrete variational principle then results in a Lie group variational integrator. For more details, see [5; 23; 26–28; 34; 37].

1.3. **Adjoint Systems and their Geometry.** The solution of many nonlinear problems involves successive linearization, and as such variational equations and their adjoints play a critical role in a variety of applications. Adjoint equations are of particular interest when the parameter space has significantly higher dimension than that of the output or objective. In particular, the simulation of adjoint equations arise in sensitivity analysis [7; 8], adaptive mesh refinement [33], uncertainty quantification [51], automatic differentiation [17], superconvergent functional recovery [41], optimal control [42], optimal design [16], optimal estimation [40], and deep learning viewed as an optimal control problem [4].

The study of geometric aspects of adjoint systems arose from the observation that the combination of any system of differential equations and its adjoint equations are described by a formal Lagrangian [20; 21]. This naturally leads to the question of when the formation of adjoints and discretization commutes [46], and prior work on this include the Ross–Fahroo lemma [43], and the observation by Sanz-Serna [44] that the adjoints and discretization commute if and only if the discretization is symplectic.

We will briefly review the geometry of adjoint and variational systems. Let M be a manifold; throughout this paper, we will assume that all manifolds are smooth and finite-dimensional and all maps between them are smooth, unless otherwise stated. Let $\dot{q} = F(q)$ be an ODE on M, specified by a vector field F on M. Then, the adjoint system associated to F is the ODE on T^*M with the coordinate expression

$$\begin{aligned} \dot{q} &= F(q), \\ \dot{p} &= -[DF(q)]^* p, \end{aligned}$$

where DF is the linearization of F. The adjoint system can be viewed as the ODE on T^*M corresponding to the vector field given by the cotangent lift of F. Intrinsically, the adjoint system can be understood as a Hamiltonian system on T^*M relative to the canonical symplectic form on T^*M, with Hamiltonian given by

$$H(q, p) = \langle p, F(q) \rangle.$$

Furthermore, we associate to F the variational system, which is the ODE on TM with coordinate expression

$$\begin{aligned} \dot{q} &= F(q), \\ \dot{v} &= DF(q)v. \end{aligned}$$

The variational system can be viewed as the ODE on TM corresponding to the vector field given by the tangent lift of F.

The importance of the adjoint and variational systems is that they satisfy an adjoint-variational quadratic conservation law: for any solution curves (q, p) of the adjoint system and (q, v) of the variational system, covering the same base curve q, one has

$$\frac{d}{dt} \langle p, v \rangle = 0.$$

This quadratic conservation law is the key to adjoint sensitivity analysis [44]. The interest in studying the geometry of these adjoint and variational systems arises from the fact that this quadratic conservation law can also be interpreted as symplecticity of the Hamiltonian flow of the adjoint system [49].

In [49], we developed Type II variational integrators for adjoint systems for ODEs and DAEs on vector spaces by utilizing their respective symplectic and presymplectic structures. One of the goals of this paper is to extend this construction to the nonlinear setting and in particular, adjoint systems over Lie groups.

1.4. **Main Contributions.** In this paper, we develop a continuous and discrete theory for Type II variational principles on cotangent bundles of Lie groups, which gives an intrinsic meaning to Hamiltonian systems with fixed initial position g_0 and fixed terminal momenta p_1 boundary conditions, in contrast to traditional variational principles for Lie group variational integrators which assume fixed initial and final positions. The motivation for developing Type II variational principles is that the corresponding Type II boundary conditions arise in adjoint sensitivity analysis, which is the motivating application of this paper. Traditionally, such Type II variational principles are only globally defined on vector spaces, or locally defined on charts on a general manifold; however, for Lie groups, left-trivialization allows us to define such a Type II variational principle globally on the cotangent bundle of a Lie group. Specifically, in Section 2, we develop a novel Type II variational principle for Hamiltonian systems on cotangent bundles of Lie groups by introducing a d'Alembert variational principle. This is a novel variational principle since typically, variational principles are given by a stationarity condition for the action corresponding to fixed initial and terminal positions, g_0 and g_1; however, in our setting, since the final position g_1 is not fixed, virtual work can be deposited into the system by varying g_1. This is accounted for in our variational principle by demanding that the action is stationary only modulo this virtual work term arising from g_1 boundary variations. Subsequently, we discretize the variational principle to develop structure-preserving numerical methods for such systems. We prove that such methods are symplectic and also momentum-preserving. We also develop a discrete reduction theory for left-invariant systems, and show that the discrete reduction theory can be interpreted as momentum preservation associated to left-invariance.

In Section 3, we apply the continuous and discrete theory developed in Section 2 to the particular case of adjoint Hamiltonian systems on Lie groups. In the continuous setting, we introduce the adjoint and variational equations associated to an ODE on a Lie group, and prove global existence and uniqueness results for these equations. In the discrete setting, we show how our variational integrators can be used to perform intrinsic structure-preserving adjoint sensitivity analysis on Lie groups. In particular, we show how initial condition sensitivities and parameter sensitivities of cost functions can be computed exactly within this framework. Finally, we conclude with two numerical examples, which utilizes this geometric adjoint sensitivity analysis to solve an initial condition optimization problem and an optimal control problem on $SO(3)$.

2. Hamiltonian Variational Integrators on Cotangent Bundles of Lie Groups

In this paper, we aim to construct and analyze geometric integration methods for Hamiltonian dynamics on the cotangent bundle T^*G of a Lie group G, subject to Type II boundary conditions $g(0) = g_0, p(T) = p_1$. Throughout, our our motivating class of examples is adjoint systems.

Example 2.1 (Adjoint Systems on Lie Groups)**.** *Consider an ODE on a Lie group G given by $\dot{g} = f(g)$, specified by a vector field F on G. We associate to F the adjoint Hamiltonian $H : T^*G \to \mathbb{R}$, given by*

$$H(g, p) = \langle p, F(q)\rangle.$$

We refer to the Hamiltonian system $i_{X_H}\Omega = dH$, relative to the canonical symplectic form Ω on T^*G, as the adjoint system associated to the ODE $\dot{g} = F(g)$. The motivation for considering Type II boundary conditions arises from the fact that, viewing the ODE on G as flowing forward in time, the momenta p can be interpreted as flowing backward in time, which backpropagates sensitivity information back to the initial time. We will describe this in more detail in Section 3.

We also provide as another motivating example the class of mechanical systems on T^*G. Although we will not be particularly concerned with this class of examples in this paper, it is worthwhile pointing out the distinctions between these two classes of examples (see Remark 2.1).

Example 2.2 (Mechanical Systems on TG)**.** *We consider a mechanical system on a Lie group G described by a Lagrangian $L : TG \to \mathbb{R}$. By left-trivialization of the tangent bundle,*

$$TG \ni (g, \dot{g}) \mapsto (g, \xi) = (g, g^{-1}\dot{g}) \in G \times \mathfrak{g},$$

the Euler–Lagrange equations for $l(g, \xi) \equiv L(g, g\xi)$ can be expressed as

$$\begin{aligned} \frac{d}{dt}\frac{\delta l}{\delta \xi} &= ad_\xi^* \frac{\delta l}{\delta \xi} + d_g l, \\ \frac{d}{dt} g &= g\xi. \end{aligned}$$

Assuming that the Lagrangian is hyperregular, i.e., $\mathbb{F}L : TG \to T^*G$ is a diffeomorphism, the system can be equivalently described as a Hamiltonian system on T^*G; or by left-trivialization, it can be equivalently described as a Hamiltonian system on $G \times \mathfrak{g}^* \ni (g, \mu)$ given by

$$\begin{aligned} \frac{d}{dt}\mu &= ad_\xi^* \mu + g\frac{\delta l}{\delta g}, \\ \frac{d}{dt} g &= \xi, \\ \mu &= \frac{\delta l}{\delta \xi}. \end{aligned}$$

For more details on this class of examples, see [5; 6; 23].

Remark 2.1. *It is interesting to note that these two classes of examples, Example 2.1 and Example 2.2, are at the opposing ends of the spectrum of regularity and degeneracy for Hamiltonian systems.*

Recall that a Hamiltonian is said to be regular if the Hessian of the Hamiltonian $D_p^2 H(q, p)$ is invertible for all $(q, p) \in T^*G$ and degenerate otherwise. If the Hamiltonian is regular, then the (inverse) Legendre transform $\mathbb{F}H : T^*G \to TG$ is a local diffeomorphism.

On the one hand, systems of the form described in Example 2.2 are regular and furthermore, they are maximally regular (or hyperregular) in the sense that $\mathbb{F}H : T^*G \to TG$ is a global diffeomorphism.

On the other hand, adjoint systems of the form described in Example 2.1 are maximally degenerate, in the sense that the Hessian $D_p^2 H(q, p)$ is the zero matrix. While this appears to be a deficiency of such systems, we will see that this is a key property of these systems, arising from the fact that these systems are lifts of differential equations on the base space G.

Because adjoint systems are degenerate, they do not admit an equivalent Lagrangian description. As such, we aim to construct integrators for Hamiltonian systems on T^*G without assuming that they arise from a Lagrangian system.

2.1. **A Type II Variational Principle for Hamiltonian Systems on Cotangent Bundles of Lie Groups.** A common approach to constructing geometric integrators for Lagrangian and Hamiltonian systems is to restrict the variational principle, from which these systems arise, to some appropriate finite-dimensional space of possible trajectories, and solve the approximate problem on this restricted space. We thus aim to construct integrators for Hamiltonian systems on T^*G by first formulating a variational principle for these systems in the continuous setting and subsequently restricting to a discrete variational principle.

To develop a variational principle for Hamiltonian systems on T^*G, we first consider the boundary conditions that we wish to place on the system. Note that fixed endpoint conditions on the basespace $g(0) = g_0, g(T) = g_1$ are generally incompatible with systems of the form Example 2.1, since adjoint systems on T^*G cover first-order ODEs on G and thus, one cannot freely specify both $g(0)$ and $g(T)$. As such, we instead consider Type II boundary conditions of the form $g(0) = g_0, p(T) = p_1$. For general Hamiltonian systems on the cotangent bundle of a manifold, the issue with these boundary conditions is that one cannot intrinsically specify a covector $p(T) = p_1$ without specifying the basepoint $q(T) = q_1$. This is not an issue for adjoint systems in particular, since the time-T flow of the underlying ODE on G determines the basepoint where p_1 is specified. However, since we would like our theory to apply to general Hamiltonian systems on T^*G, we do not want to restrict to adjoint systems in particular. Fortunately, we can make sense of Type II boundary conditions, since T^*G is trivializable by left-translation.

Let $\mathfrak{g} = T_eG$ denote the Lie algebra of G and $\mathfrak{g}^* = T_e^*G$ be its dual. We will denote the duality pairing between $v \in T_gG$ and $p \in T_g^*G$ as $\langle p, v\rangle$, where the base point is understood in context. Let $L_g : G \to G$ denote left-translation by g, $L_g(x) = gx$. Left-translation induces maps on the tangent bundle and cotangent bundle of G by pushforward and pullback, respectively, which we denote as

$$\begin{aligned}
T_xL_g &: T_xG \to T_{gx}G, \\
T_x^*L_g &: T_{gx}^*G \to T_x^*G.
\end{aligned}$$

For $v_g \in T_gG, p_g \in T_g^*G$, we will denote their left-translations to their respective fibers over the identity as simply

$$\begin{aligned}
g^{-1}v_g &\equiv T_gL_{g^{-1}}(v_g) \in \mathfrak{g}, \\
g^*p_g &\equiv T_e^*L_g(p_g) \in \mathfrak{g}^*.
\end{aligned}$$

This notation is suggestive, since in the case that G is a matrix Lie group, the left-translation of a tangent vector to the fiber over the identity acts by matrix multiplication by the inverse of g and the left-translation of a covector to the fiber over the identity acts by matrix multiplication by the adjoint of g.

A useful fact is that the pairing $\langle p_g, v_g\rangle$ is preserved under left-translation,

$$\begin{aligned}
\langle g^* \cdot p_g, g^{-1} \cdot v_g\rangle &= \langle T_e^*L_g(p_g), T_gL_{g^{-1}}(v_g)\rangle \\
&= \langle p_g, T_eL_g \circ T_gL_{g^{-1}}(v_g)\rangle \\
&= \langle p_g, T_g(L_g \circ L_{g^{-1}})(v_g)\rangle \\
&= \langle p_g, v_g\rangle.
\end{aligned}$$

By left-translation on the cotangent bundle, we get the left-trivialization $T^*G \cong G \times \mathfrak{g}^*$. With this trivialization, we can make sense of Type II boundary conditions $g(0) = g_0 \in G, \mu(T) = \mu_1 \in \mathfrak{g}^*$, with coordinates (g, μ) on $G \times \mathfrak{g}^*$.

What remains is to construct a variational principle. Recall that the action for a Hamiltonian system on T^*G is given by

$$S[g, p] = \int_0^T \Big(\langle p, \dot{g}\rangle - H(g, p)\Big)dt,$$

where $H : T^*G \to \mathbb{R}$. By left-translation, with $\mu = g^* \cdot p$, we define the left-trivialized Hamiltonian $h : G \times \mathfrak{g}^* \to \mathbb{R}$ as

$$h(g, \mu) \equiv H(g, g^{*-1} \cdot \mu) = H(g, p).$$

The action can then be expressed as

$$\begin{aligned} S[g,p] &= \int_0^T \Big(\langle p, \dot g\rangle - H(g,p)\Big) dt \\ &= \int_0^T \Big(\langle g^*\cdot p, g^{-1}\cdot \dot g\rangle - H(g,p)\Big) dt \\ &= \int_0^T \Big(\langle \mu, g^{-1}\cdot \dot g\rangle - h(g,\mu)\Big) dt =: s[g,\mu]; \end{aligned}$$

we refer to $s[g,\mu]$ as the left-trivialized action.

Now, we prescribe boundary conditions $g(0) = g_0 \in G$, $\mu(T) = \mu_1 \in \mathfrak{g}^*$. Given a curve $(g(t), p(t))$ on T^*G, by left-translation, the terminal momenta condition $\mu(T) = \mu_1$ on $\mathfrak{g}^*$ corresponds to $p(T) = g(T)^{*-1}\cdot \mu_1 \in T^*_{g(T)}G$. To state a variational principle, we observe that by left-translation, we can prescribe a boundary condition on $p(T)$ (equivalently, on $\mu(T)$) but we cannot fix the terminal point $g(T)$. As such, a variation δg can always introduce virtual work on the system by varying the terminal point $g(T)$; the virtual work done by varying the terminal point is given by $\langle p(T), \delta g(T)\rangle$, or equivalently, $\langle \mu(T), \eta(T)\rangle$ where we defined the left-trivialization of the variation $\eta = g^{-1}\cdot\delta g$. Thus, we cannot demand the the action S (equivalently, s) is stationary since one can always introduce virtual work as described above; however, we can demand that it is stationary modulo the virtual work that is introduced into the system by varying the terminal point $g(T)$. Thus, we impose the variational principle

$$\delta S[g,p] = \langle p(T), \delta g(T)\rangle,$$

or equivalently, by left translation

$$\delta s[g,\mu] = \langle \mu(T), \eta(T)\rangle,$$

where the variations fix $g(0)$ and $p(T)$ (equivalently, $\mu(T)$). We refer to this variational principle as the Type II d'Alembert variational principle, due to its similarity to the d'Alembert variational principle which utilizes virtual work to derive forced Lagrangian or Hamiltonian systems [36].

Theorem 2.1 (Type II d'Alembert Variational Principle)**.** *The following are equivalent*

(i) *The Type II d'Alembert variational principle*

$$\delta S[g,p] = \langle p(T), \delta g(T)\rangle,$$

on T^*G is satisfied, where the variations of the action $\delta g, \delta p$ satisfy $\delta g(0) = 0, \delta p(T) = 0$, corresponding to boundary conditions $g(0) = g_0, p(T) = g(T)^{*-1}\cdot\mu_1$.

(ii) *Hamilton's equations hold in canonical coordinates on T^*G, with the above Type II boundary conditions,*

$$\dot g = D_p H(g,p), \tag{2.1a}$$

$$\dot p = -D_g H(g,p), \tag{2.1b}$$

$$g(0) = g_0, \tag{2.1c}$$

$$p(T) = g(T)^{*-1}\cdot\mu_1. \tag{2.1d}$$

(iii) *The Type II d'Alembert variational principle*

$$\delta s[g,\mu] = \langle \mu(T), \eta(T)\rangle$$

on $G\times\mathfrak{g}^*$ is satisfied, where the variation δg is left-trivialized as $\eta = g^{-1}\cdot\delta g$ and the variation δp is left-trivialized as $\delta\mu = g^*\cdot\delta p$, with $\delta\eta(0) = 0, \delta\mu(T) = 0$, corresponding to boundary conditions $g(0) = g_0, \mu(T) = \mu_1$.

(iv) *The Lie–Poisson equations hold on $G \times \mathfrak{g}^*$, with the above Type II boundary conditions,*

$$\dot{g} = g \cdot D_\mu h(g,\mu), \tag{2.2a}$$

$$\dot{\mu} = -g^* \cdot D_g h(g,\mu) + \mathrm{ad}^*_{D_\mu h(g,\mu)}\mu, \tag{2.2b}$$

$$g(0) = g_0, \tag{2.2c}$$

$$\mu(T) = \mu_1. \tag{2.2d}$$

Remark 2.2. *Above, we denote by $D_gH, D_pH, D_gh, D_\mu h$ the functional derivatives satisfying*

$$dH(g,p)\cdot(\delta g,\delta p) = \langle D_gH(g,p),\delta g\rangle + \langle \delta p, D_pH(g,p)\rangle,$$
$$dh(g,\mu)\cdot(\delta g,\delta\mu) = \langle D_gh(g,\mu),\delta g\rangle + \langle \delta \mu, D_\mu h(g,\mu)\rangle.$$

Proof. To see that (i) and (ii) are equivalent, compute the variation of S,

$$\begin{aligned}\delta S[g,p] &= \int_0^T \Big(\langle \delta p,\dot{g}\rangle + \Big\langle p, \frac{d}{dt}\delta g\Big\rangle - \langle D_gH,\delta g\rangle - \langle \delta p, D_pH\rangle\Big)dt \\ &= \int_0^T \Big(\langle \delta p,\dot{g} - D_pH\rangle + \langle -\dot{p} - D_gH,\delta g\rangle\Big)dt + \langle p,\delta g\rangle\Big|_0^T.\end{aligned}$$

If (ii) holds, the integrand above vanishes by the equations of motion; furthermore, $\delta g(0) = 0$. Thus, the above expression reduces to

$$\delta S[g,p] = \langle p(T),\delta g(T)\rangle,$$

i.e., (i) holds. Conversely, if (i) holds, we have

$$0 = \delta S[g,p] - \langle p(T),\delta g(T)\rangle = \int_0^T \Big(\langle \delta p,\dot{g} - D_pH\rangle + \langle -\dot{p} - D_gH,\delta g\rangle\Big)dt.$$

Then, by the fundamental lemma of the calculus of variations, (ii) holds.

To see that (iii) is equivalent to (iv), compute the variation of s. For simplicity, we denote the left-translation of $\dot{g}$ by $\xi = g^{-1}\cdot\dot{g}$ and similarly $\eta = g^{-1}\cdot\delta g$.

$$\begin{aligned}\delta s[g,\mu] &= \int_0^T \Big(\langle \delta\mu, g^{-1}\cdot\dot{g}\rangle + \Big\langle \mu, g^{-1}\frac{d}{dt}\delta g - g^{-1}\cdot\delta g g^{-1}\cdot\dot{g}\Big\rangle - \langle\delta\mu, D_\mu h\rangle - \langle D_gh,\delta g\rangle\Big)dt \\ &= \int_0^T \Big(\langle \delta\mu, g^{-1}\cdot\dot{g} - D_\mu h\rangle + \langle \mu, \dot{\eta} + \mathrm{ad}_\xi\eta\rangle - \langle g^*\cdot D_gh,\eta\rangle\Big)dt \\ &= \int_0^T \Big(\langle \delta\mu, g^{-1}\cdot\dot{g} - D_\mu h\rangle + \langle -\dot{\mu} + \mathrm{ad}^*_\xi\mu - g^*\cdot D_gh,\eta\rangle\Big)dt + \langle\mu,\eta\rangle\Big|_0^T.\end{aligned}$$

If (iv) holds, the integrand above vanishes by the equations of motion, noting that $\xi = g^{-1}\cdot\dot{g} = D_\mu h$; furthermore, $\eta(0) = 0$. Thus, the above expression reduces to

$$\delta s[g,\mu] = \langle\mu(T),\eta(T)\rangle,$$

i.e., (iii) holds. Conversely, if (iii) holds, we have

$$0 = \delta s[g,\mu] - \langle\mu(T),\eta(T)\rangle = \int_0^T \Big(\langle \delta\mu, g^{-1}\cdot\dot{g} - D_\mu h\rangle + \langle -\dot{\mu} + \mathrm{ad}^*_\xi\mu - g^*\cdot D_gh,\eta\rangle\Big)dt.$$

Then, by the fundamental lemma of the calculus of variations, (iv) holds.

Finally, (i) and (iii) are equivalent by left-translation, since $S[g,p] = s[g,\mu]$ and $\langle p(T),\delta g(T)\rangle = \langle\mu(T),\eta(T)\rangle$. $\square$

Remark 2.3. *Note that one can also modify the above variational principle to include external forces by adding the virtual work done by the external force. Given a left-trivialized external force $f : [0,T] \to \mathfrak{g}^*$, one can modify the above variational principle to*

$$\delta s[g,\mu] = \langle \mu(T), \eta(T)\rangle + \int_0^T \langle f, \eta\rangle dt,$$

or equivalently,

$$\delta S[g,p] = \langle p(T), \delta g(T)\rangle + \int_0^T \langle g^{*-1}\cdot f, \delta g\rangle dt.$$

This modifies the momenta equations (2.2b) *on $G \times \mathfrak{g}$ to include the external force,*

$$\dot{\mu} = -g^* \cdot D_g h(g,\mu) + \mathrm{ad}^*_{D_\mu h(g,\mu)}\mu + f,$$

or equivalently, modifies the momenta equation (2.1b) *on T^*G to be*

$$\dot{p} = -D_g H(g,p) + g^{*-1}\cdot f.$$

2.2. **Discrete Hamiltonian Variational Integrators on Cotangent Bundles of Lie Groups.** In this section, we develop a discrete counterpart to the continuous Type II variational principle on cotangent bundles of Lie groups developed in the previous section.

Consider the action

$$s[g,\mu] = \int_0^T \Big(\langle \mu, g^{-1}\cdot \dot{g}\rangle - h(g,\mu)\Big)dt.$$

We will construct discrete Hamiltonian variational integrators for the Lie–Poisson system (2.2a)-(2.2d) by discretizing the Type II d'Alembert variational principle (Theorem 2.1).

Partition $[0,T]$ into $\cup_{k=0}^{N-1}[t_k, t_{k+1}]$ where

$$0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T,$$

with uniformly spaced intervals $t_{k+1} - t_k = \Delta t = T/N$. To discretize the variational principle, we need a sequence of points $\{g_k \in G\}_{k=0}^{N-1}$ which interpolates a curve $g(t) \in G$. A simple way to do this is to utilize a retraction to relate a curve on G to a curve on $\mathfrak{g}$. Let τ be a retraction $\tau : \mathfrak{g} \to G$, which is a C^2-diffeomorphism about the origin such that $\tau(0) = e$. Let $d\tau_\xi : \mathfrak{g} \to \mathfrak{g}$ denote the right-trivialized tangent map of τ and $d\tau_\xi^{-1}$ its inverse (for a definition, see [6]). Using the retraction, we can approximate the velocity $g^{-1}\cdot \dot{g} \in \mathfrak{g}$ by

$$\xi_{k+1} = \tau^{-1}(g_k^{-1} g_{k+1})/\Delta t. \tag{2.3}$$

This defines the desired interpolated curve $\{g_k\}$ on G through elements $\{\xi_k\}$ on $\mathfrak{g}$ via $g_{k+1} = g_k \tau(\Delta t \xi_{k+1})$. We approximate the action as

$$s_d[\{g_k\},\{m_k\}] = \sum_{k=0}^{N-1} \Delta t\Big(\langle m_{k+1}, \xi_{k+1}\rangle - h(g_k, m_{k+1})\Big),$$

where again $\{\xi_k\}$ and $\{g_k\}$ are related by (2.3). Note that, by (2.3), the variations in ξ are related to the variations of g; this is explicitly given by ([23])

$$\delta\xi_{k+1} = \delta\tau^{-1}(g_k^{-1}g_{k+1})/\Delta t = d\tau^{-1}_{\Delta t\xi_{k+1}}(-g_k^{-1}\delta g_k + \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1}\delta g_{k+1})/\Delta t. \tag{2.4}$$

We now derive a variational integrator from a discrete approximation of the Type II d'Alembert variational principle.

Theorem 2.2 (Discrete Type II d'Alembert Variational Principle)**.** *The following are equivalent*

(i) *The discrete Type II d'Alembert variational principle holds*

$$\delta s_d[\{g_k\},\{m_k\}] = \langle (d\tau^{-1}_{-\Delta t\xi_N})^* m_N, g_N^{-1}\delta g_N\rangle,$$

subject to variations $\delta g_k, \delta m_k$ *satisfying* $\delta g_0 = 0, \delta m_N = 0$*, corresponding to Type II boundary conditions which prescribe* $g_0 = g(0), m_N = m(T)$*.*

(ii) *The discrete Lie–Poisson equations hold*

$$(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m_k = -\Delta t g_k^* \cdot D_g h(g_k, m_{k+1}), \tag{2.5a}$$

$$\xi_{k+1} = D_\mu h(g_k, m_{k+1}), \tag{2.5b}$$

$$g_{k+1} = g_k \tau(\Delta t\xi_{k+1}), \tag{2.5c}$$

with the above boundary conditions.

Proof. Compute the variation of s_d,

$$\delta s_d = \underbrace{\sum_{k=0}^{N-1} \Delta t\Big[\langle \delta m_{k+1}, \xi_{k+1}\rangle - \langle \delta m_{k+1}, D_\mu h(g_k, m_{k+1})\rangle\Big]}_{\equiv(a)}$$

$$+ \underbrace{\sum_{k=0}^{N-1} \Delta t\Big[\langle m_{k+1}, \delta\xi_{k+1}\rangle - \langle D_g h(g_k, m_{k+1}), \delta g_k\rangle\Big]}_{\equiv(b)}.$$

We will simplify the expressions (a) and (b) individually.

For (a), note that the $k = N-1$ term vanishes since $\delta m_N = 0$. Thus, the sum runs 0 to $N-2$. We re-index $k \to k-1$ so that (a) becomes

$$\text{(a)} = \sum_{k=1}^{N-1} \Delta t\Big[\langle \delta m_k, \xi_k - D_\mu h(g_{k-1}, m_k)\rangle\Big].$$

For (b), we rewrite the variation in ξ in terms of the variation of g,

$$\begin{aligned}
\text{(b)} &= \sum_{k=0}^{N-1} \Delta t\Big[\langle m_{k+1}, d\tau^{-1}_{\Delta t\xi_{k+1}}(-g_k^{-1}\delta g_k + \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}\delta g_{k+1})/\Delta t\rangle - \langle D_g h(g_k, m_{k+1}), \delta g_k\rangle\Big] \\
&= \sum_{k=0}^{N-1} \Delta t\Big[\langle -(g_k^{-1})^*(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}/\Delta t - D_g h(g_k, m_{k+1}), \delta g_k\rangle\Big] \\
&\quad + \sum_{k=0}^{N-1} \Delta t\langle (g_{k+1}^{-1})^* \mathrm{Ad}^*_{\tau(\Delta t\xi_{k+1})}(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}/\Delta t, \delta g_{k+1}\rangle.
\end{aligned}$$

In the first sum above, note that the $k = 0$ vanishes since $\delta g_0 = 0$. In the second sum above, we explicitly write the $k = N-1$ term and re-index the resulting sum $k \to k-1$. This gives

$$\begin{aligned}
\text{(b)} &= \sum_{k=1}^{N-1} \Delta t\Big[\langle -(g_k^{-1})^*(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}/\Delta t - D_g h(g_k, m_{k+1}), \delta g_k\rangle\Big] \\
&\quad + \sum_{k=1}^{N-1} \Delta t\langle (g_k^{-1})^* \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t_{\xi_k}})^* m_k/\Delta t, \delta g_k\rangle \\
&\quad + \Delta t\langle (g_N^{-1})^* \mathrm{Ad}^*_{\tau(\Delta t\xi_N)}(d\tau^{-1}_{\Delta t\xi_N})^* m_N/\Delta t, \delta g_N\rangle.
\end{aligned}$$

Note that, since $\mathrm{Ad}^*_{\tau(\Delta t\xi_N)}(d\tau^{-1}_{\Delta t\xi_N})^* = (d\tau^{-1}_{-\Delta t\xi_N})^*$ [6], the last term equals

$$\langle (d\tau^{-1}_{-\Delta t\xi_N})^* m_N, g_N^{-1}\delta g_N\rangle,$$

which is precisely the virtual work term in the discrete Type II d'Alembert variational principle. Putting everything together, we have

$$\begin{aligned}
&\delta s_d[\{g_k\},\{m_k\}] - \langle (d\tau^{-1}_{-\Delta t\xi_N})^* m_N, g_N^{-1}\delta g_N\rangle \\
&\quad = \sum_{k=1}^{N-1} \Delta t \Big[\langle \delta m_k, \xi_k - D_\mu h(g_{k-1}, m_k)\rangle \\
&\qquad\qquad + \langle -(g_k^{-1})^*(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}/\Delta t - D_g h(g_k, m_{k+1}) \\
&\qquad\qquad + (g_k^{-1})^* \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t_{\xi_k}})^* m_k/\Delta t, \delta g_k\rangle \Big].
\end{aligned}$$

Clearly, if the discrete Lie–Poisson equations hold, then the above vanishes, i.e., the discrete Type II d'Alembert variational principle holds, $\delta s_d[\{g_k\},\{m_k\}] = \langle (d\tau^{-1}_{-\Delta t\xi_N})^* m_N, g_N^{-1}\delta g_N\rangle$. Conversely, if the discrete Type II d'Alembert variational principle holds, the above vanishes, which gives

$$\begin{aligned}
0 = \sum_{k=1}^{N-1} \Delta t \Big[&\langle \delta m_k, \xi_k - D_\mu h(g_{k-1}, m_k)\rangle \\
&+ \langle -(g_k^{-1})^*(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}/\Delta t - D_g h(g_k, m_{k+1}) \\
&+ (g_k^{-1})^* \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t_{\xi_k}})^* m_k/\Delta t, \delta g_k\rangle \Big].
\end{aligned}$$

Since the variations δm_k and δg_k are arbitrary and independent for $k = 1, \dots, N-1$, this gives the discrete Lie–Poisson equations (2.5a)-(2.5c). □

Note that this integrator is similar to (and in some cases, the same as) various variational Lie group integrators in the literature, but there are some important distinctions.

For example, in [6], variational integrators for dynamics on TG are derived through a discrete Hamilton–Pontryagin principle. This can be related to our integrator, given that the Hamiltonian arises from a regular Lagrangian. In particular, if one is given a regular left-trivialized Lagrangian $l(g,\xi)$, from which $h(g,\mu)$ arises via the Legendre transform, then we can invert (2.5b) to obtain $m_{k+1} = D_\xi l(g_k, \xi_{k+1})$ and we have the relation $\partial l/\partial g = -\partial h/\partial g$. Substituting these into the equations (2.5a)-(2.5c) produces equation (4.19) of [6]. Of course, this equivalence does not hold when the Hamiltonian is not regular. In particular, note that the Legendre transform with respect to the Hamiltonian, equation (2.5b), appears in the discrete Lie–Poisson equations derived from the Hamiltonian side, as opposed to the Legendre transform with respect to the Lagrangian. Thus, our method is applicable to degenerate Hamiltonian systems, such as adjoint systems, which we will discuss further in Section 3.

As previously noted, an important distinction with the above integrator is that it is defined and derived from a variational principle entirely on the Hamiltonian side; this is particularly important when the Hamiltonian is not regular, as in the case of adjoint systems. Furthermore, the variational integrators in the literature make use of fixed endpoint boundary conditions, $\delta g_0 = 0 = \delta g_N$, in the variational principle ([6; 23; 34; 37]). As previously discussed, these boundary conditions are incompatible with adjoint systems. By utilizing a discrete Type II d'Alembert variational principle, we were able to derive an integrator on the Hamiltonian side which does not assume that the Hamiltonian is regular, nor assume fixed endpoint boundary conditions. Thus, as we will see in Section 3, we will be able to apply our integrator to adjoint systems to develop a structure-preserving integrator which preserves the quadratic adjoint sensitivity conservation law.

It is also interesting to note the virtual work term arising at the terminal point in the discrete variational principle,

$$\langle (d\tau^{-1}_{-\Delta t \xi_N})^* m_N, g_N^{-1}\delta g_N\rangle$$

is different than what one might expect from the continuous variational principle, $\langle \mu_N, g_N^{-1}\delta g_N\rangle$. This is due to the fact that the retraction relates the dynamics on G to dynamics on $\mathfrak{g}$, and so the pairing $\langle \mu, g^{-1}\dot{g}\rangle$ compared to the pairing $\langle m_k, \xi_k\rangle$ should not be identified, but rather, are related by a coordinate change. In fact, the coordinate change is given by the cotangent lift of τ^{-1}, which is precisely $(d\tau^{-1}_{-\Delta t \xi_N})^*$. As we will see below, this also induces a coordinate change in the expression for the symplectic form, which is the exterior derivative of the one-form corresponding to the above boundary term; the expression for the one-form and its exterior derivative is also derived in [6] through a discrete Hamilton–Pontryagin principle.

2.2.1. *Reduction for Left-invariant Hamiltonians.* A particularly important class of Hamiltonians are the left-invariant Hamiltonians, which are functions $H : T^*G \to \mathbb{R}$ that are invariant under the cotangent lift of left-multiplication by any $x \in G$, i.e.,

$$H \circ T^*L_x = H \text{ for all } x \in G.$$

In terms of our notation, that is

$$H(xg, x^{*-1}p) = H(g,p) \text{ for all } x \in G,\ (g,p) \in T^*G.$$

For such left-invariant Hamiltonians, the dynamics on T^*G reduce to dynamics on $\mathfrak{g}^*$ [35].

Given a left-invariant Hamiltonian H, we define the reduced Hamiltonian $\tilde{H} : \mathfrak{g}^* \to \mathbb{R}$ by

$$\tilde{H}(\mu) = H(e,\mu).$$

Then, equation (2.2b) for $\dot{\mu}$ reduces to

$$\dot{\mu} = \mathrm{ad}^*_{D_\mu \tilde{H}(\mu)}\mu, \tag{2.6}$$

where we used that

$$\tilde{H}(\mu) = H(e,\mu) = H(g, g^{*-1}\mu) = h(g,\mu)$$

and hence also, $D_g h(g,\mu) = 0$. Thus, as can be seen from equation (2.6), the momentum equation decouples from the dynamics on G: (2.6) can be solved independently and subsequently, (2.2a) can be used to reconstruct the dynamics on G. Hence, for a left-invariant system, the full dynamics on T^*G is completely encoded by the reduced dynamics on $\mathfrak{g}^*$.

Now, we develop a discrete analogue of the Type II variational principle in the left-invariant setting. Define the reduced discrete action

$$\tilde{s}_d[\{g_k\},\{m_k\}] = \sum_{k=0}^{N-1} \Delta t\Big(\langle m_{k+1}, \xi_{k+1}\rangle - \tilde{H}(m_{k+1})\Big).$$

Theorem 2.3 (Discrete Type II d'Alembert Reduced Variational Principle)**.** *Let $H : T^*G \to \mathbb{R}$ be left-invariant and let $h, \tilde{H}, s_d, \tilde{s}_d$ be defined as above. The following are equivalent*

(i) *The discrete Type II d'Alembert variational principle holds*

$$\delta s_d[\{g_k\},\{m_k\}] = \langle (d\tau^{-1}_{-\Delta t \xi_N})^* m_N, g_N^{-1}\delta g_N\rangle,$$

subject to variations $\delta g_k, \delta m_k$ satisfying $\delta g_0 = 0, \delta m_N = 0$, corresponding to Type II boundary conditions which prescribe $g_0 = g(0), m_N = m(T)$.

(ii) *The discrete Lie–Poisson equations hold*

$$(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m_k = -\Delta t g_k^* \cdot D_g h(g_k, m_{k+1}),$$
$$\xi_{k+1} = D_\mu h(g_k, m_{k+1}),$$
$$g_{k+1} = g_k \tau(\Delta t\xi_{k+1}),$$

with the above boundary conditions.

(iii) *The discrete reduced Type II d'Alembert variational principle holds*

$$\delta\tilde{s}_d[\{g_k\},\{m_k\}] = \langle (d\tau^{-1}_{-\Delta t\xi_N})^* m_N, g_N^{-1}\delta g_N\rangle,$$

subject to variations $\delta g_k, \delta m_k$ satisfying $\delta g_0 = 0, \delta m_N = 0$, corresponding to Type II boundary conditions which prescribe $g_0 = g(0), m_N = m(T)$.

(iv) *The discrete reduced Lie–Poisson equations hold*

$$(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m_k = 0, \tag{2.7a}$$
$$\xi_{k+1} = D_\mu \tilde{H}(m_{k+1}), \tag{2.7b}$$
$$g_{k+1} = g_k \tau(\Delta t\xi_{k+1}), \tag{2.7c}$$

with the above boundary conditions.

Proof. We already know that (i) is equivalent to (ii) by Theorem 2.2. Furthermore, (i) is clearly equivalent to (iii), since $s_d = \tilde{s}_d$. To see this, it suffices to show that $h(g, m_{k+1}) = \tilde{H}(m_{k+1})$. By the definition of h, we have

$$h(g, m_{k+1}) = H(g, g^{*-1} m_{k+1}) = H(e, m_{k+1}) = \tilde{H}(m_{k+1}),$$

where in the second equality, we used left-invariance of H.

Finally, we show that (ii) is equivalent to (iv). Clearly (2.5c) is the same as (2.7c). Since $\tilde{H}(m_{k+1}) = H(g_k, m_{k+1})$, we have that

$$D_\mu h(g_k, m_{k+1}) = D_\mu \tilde{H}(m_{k+1}),$$

so that (2.5b) is equivalent to (2.7b). Finally, we have

$$D_g h(g_k, m_{k+1}) = D_g \tilde{H}(m_{k+1}) = 0,$$

so that (2.5a) is equivalent to (2.7a). □

In practice, many Hamiltonian systems on T^*G arise from left-invariant Hamiltonians. The practical importance of the reduced formulation is that the dynamics on T^*G (or, equivalently, on $G\times\mathfrak{g}^*$) can be reduced to dynamics on $\mathfrak{g}^*$. To see this, note that we can eliminate $\{\xi_k\}$ in equation (2.7a) by using equation (2.7b) to obtain

$$(d\tau^{-1}_{\Delta t D_\mu \tilde{H}(m_{k+1})})^* m_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t D_\mu \tilde{H}(m_k))}(d\tau^{-1}_{\Delta t D_\mu \tilde{H}(m_k)})^* m_k = 0,$$

which only involves $\{m_k\}$. In the literature, this is often presented as a discrete coadjoint flow (see, for example, [34; 37]), which we can see by making the definition $\mu_k = (d\tau^{-1}_{\Delta t D_\mu \tilde{H}(m_k)})^* m_k$, so that the above becomes

$$\mu_{k+1} = \mathrm{Ad}^*_{\tau(\Delta t D_\mu \tilde{H}(m_k))}\mu_k. \tag{2.8}$$

In the following section, we will derive symplecticity and momentum conservation of the discrete Lie–Poisson equations (2.5a)-(2.5c). Subsequently, we will derive the discrete reduced Lie–Poisson equation (2.8) from a different perspective, by viewing it as a consequence of momentum conservation associated to the left-invariance symmetry.

2.2.2. *Discrete Conservation Properties.* In this section, we will show that the integrator (2.5a)-(2.5c) is both symplectic and momentum-preserving. Such symplectic-momentum schemes also enjoy long-term energy stability [36].

Discrete Symplecticity. We now show that the integrator (2.5a)-(2.5c) is symplectic. In essence, symplecticity of the integrator follows from the fact that the integrator was derived from a discrete variational principle, but we will show it explicitly. We perform the computation explicitly for two reasons. First, the proof of symplecticity for variational integrators is traditionally derived from the boundary term in the variational principle [36] or through the use of a generating function [31]. However, in our case, we utilize a modified d'Alembert Type II variational principle which involves a virtual work term. Thus, we cannot appeal directly to the previous methods. Furthermore, the setup for the proof will introduce the concept of variational equations which will be useful for discussing adjoint systems in Section 3; additionally, the computation for symplecticity will be similar to the computation for the quadratic conservation law for discrete adjoint systems.

From the boundary term arising from the variation of the discrete action s_d, we see that the discrete canonical form has the expression

$$\Theta_k = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, g_k^{-1} dg_k \rangle, \tag{2.9}$$

whose action on a vector $\delta_k = \delta m_k \partial/\partial m_k + \delta g_k \partial/\partial g_k$ is given by

$$\Theta_k \cdot \delta_k = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, g_k^{-1} \delta g_k \rangle.$$

Then, the corresponding discrete symplectic form $\Omega_k \equiv d\Theta_k$ has the expression

$$\Omega_k = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge g_k^{-1} dg_k \rangle. \tag{2.10}$$

Its action on vectors $\delta_k^i = \delta m_k^i \partial/\partial m_k + \delta g_k^i \partial/\partial g_k$ is given by

$$\Omega_k \cdot (\delta_k^1, \delta_k^2) = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* \delta m_k^1, g_k^{-1} \delta g_k^2 \rangle - \langle (d\tau^{-1}_{-\Delta t\xi_k})^* \delta m_k^2, g_k^{-1} \delta g_k^1 \rangle.$$

Symplecticity of the integrator (2.5a)-(2.5c) is the statement that $\Omega_{k+1} = \Omega_k$ when the discrete Lie–Poisson equations (2.5a)-(2.5c) hold, where the symplectic forms are evaluated on first variations of the discrete Lie–Poisson equations, i.e., variations whose flow preserves solutions of the discrete Lie–Poisson equations. Equivalently, such first variations are those which preserve (2.5a)-(2.5c) to linear order. By linearizing these equations, we obtain

$$\begin{aligned} (d\tau^{-1}_{\Delta t\xi_{k+1}})^* dm_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t\xi_k)} (d\tau^{-1}_{\Delta t\xi_k})^* dm_k & \\ = -\Delta t g_k^* \cdot D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} &+ \text{variation in } g_k, \end{aligned} \tag{2.11a}$$

$$d\xi_{k+1} = D^2_{g\mu} h(g_k, m_{k+1}) dg_k + \text{variation in } m_{k+1}, \tag{2.11b}$$

$$0 = d(g_{k+1} - g_k \tau(\Delta t \xi_{k+1})). \tag{2.11c}$$

In equation (2.11a) above, we omitted the terms involving the variation of (2.5a) with respect to g_k. Similarly, in equation (2.11b), we omitted the terms involving the variation of (2.5b) with respect to m_{k+1}. This is because it is difficult to express the former explicitly but we will write them as follows. Observe that equation (2.5a) and (2.5b) can respectively be expressed as

$$\frac{\delta}{\delta \eta_k} (\Delta t^{-1} s_d) = 0,$$

$$\frac{\delta}{\delta m_{k+1}} (\Delta t^{-1} s_d) = 0,$$

where $\eta_k = g_k^{-1}\delta g_k$, and the variations in g and ξ are related by the identity (2.4). Additionally, the variational derivatives are defined by

$$\delta s_d = \sum_k \left[\left\langle \delta m_{k+1}, \frac{\delta}{\delta m_{k+1}} s_d \right\rangle + \left\langle \frac{\delta}{\delta \eta_k} s_d, \eta_k \right\rangle \right].$$

Thus, the omitted variations in (2.11a), (2.11b) have the expressions,

$$\frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 \eta_k} g_k^{-1} dg_k,$$
$$\frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1},$$

respectively. Additionally, we will combine (2.11b)-(2.11c). Analogous to the identity (2.4), (2.11c) can be expressed as

$$d\xi_{k+1} = d\tau^{-1}_{\Delta t \xi_{k+1}}(-g_k^{-1} dg_k + \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1})/\Delta t.$$

Thus, we can combine (2.11b)-(2.11c) to yield

$$d\tau^{-1}_{\Delta t \xi_{k+1}}(-g_k^{-1} dg_k + \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1})/\Delta t = D^2_{g\mu} h(g_k, m_{k+1}) dg_k + \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1}.$$

We will additionally multiply both sides by Δt and act on both sides by $d\tau_{\Delta t \xi_{k+1}}$. Thus, we have the equations for the first variations of the discrete Lie–Poisson equations,

$$\begin{aligned} (d\tau^{-1}_{\Delta t \xi_{k+1}})^* dm_{k+1} &- \mathrm{Ad}^*_{\tau(\Delta t \xi_k)} (d\tau^{-1}_{\Delta t \xi_k})^* dm_k \\ &= -\Delta t g_k^* \cdot D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} + \frac{\delta^2(\Delta t^{-1} s_d))}{\delta^2 \eta_k} g_k^{-1} dg_k, \end{aligned} \tag{2.12a}$$

$$\begin{aligned} -g_k^{-1} dg_k &+ \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \\ &= \Delta t d\tau_{\Delta t \xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k + \Delta t d\tau_{\Delta t \xi_{k+1}} \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1}. \end{aligned} \tag{2.12b}$$

Remark 2.4. *In the subsequent proof of symplecticity of the discrete Lie–Poisson equations, some notation and manipulations for the computations will be useful.*

First, note that although for notational simplicity we will work at the level of differential forms, we will always implicitly understand that the differential forms will be evaluated on vectors. Because of this, we can manipulate expressions involving differential forms and the duality pairing with a wedge product $\langle \cdot \wedge \cdot \rangle$ *as we would an expression of the ordinary duality pairing. For example, given the expression for* $\Omega_k \cdot (\delta_k^1, \delta_k^2)$ *above, we can manipulate the expression as follows*

$$\Omega_k = \langle (d\tau^{-1}_{-\Delta t \xi_k})^* dm_k \wedge g_k^{-1} dg_k \rangle = \langle dm_k \wedge d\tau^{-1}_{-\Delta t \xi_k} g_k^{-1} dg_k \rangle,$$

i.e., we can move $(d\tau^{-1}_{-\Delta t \xi_k})^*$ *across the duality pairing by taking the adjoint, because we can do so in the expression* $\Omega_k \cdot (\delta_k^1, \delta_k^2)$*, when the differential form is evaluated on vectors.*

Furthermore, for some parts of the subsequent computation, it will be useful to used indexed coordinates. Let g_k^A*,* $A = 1, \dots, \dim(G)$ *be coordinates for* g_k *on* G *and let* m_{kA} *be coordinates for* m_k *on* $\mathfrak{g}^*$*. Then, for example, a typical duality pairing* $\langle m_k, g_k^{-1}\delta g_k \rangle$ *can be expressed as*

$$\langle m_k, g_k^{-1} \delta g_k \rangle = m_{kA} (g_k^{-1} \delta g_k)^A,$$

where we are using the Einstein summation convention that repeated indices, one raised and one lowered, are implicitly summed over. Similarly, an expression involving differential forms paired *with the duality pairing and wedge product, such as $\langle dm_k \wedge g_k^{-1} dg_k \rangle$, can be expressed as*

$$\langle dm_k \wedge g_k^{-1} dg_k \rangle = dm_{kA} \wedge (g_k^{-1} \delta g_k)^A = dm_{kA} \wedge (g_k^{-1})^A_B \delta g_k^B = (g_k^{-1})^A_B dm_{kA} \wedge \delta g_k^B.$$

In the last equality above, we used bilinearity of the wedge product and the fact that the quantity $(g_k^{-1})^A_B$, for each index A, B, is simply a number. In particular, indexed coordinates will be useful for quantities involving the second variations of s_d above, which can be expressed as

$$\begin{aligned} \left(\frac{\delta^2 s_d}{\delta^2 \eta_k} g_k^{-1} dg_k \right)_A &= \frac{\delta^2 s_d}{\delta \eta^A \delta \eta^B} (g_k^{-1} dg_k)^B, \\ \left(\frac{\delta^2 s_d}{\delta^2 m_{k+1}} dm_{k+1} \right)^A &= \frac{\delta^2 s_d}{\delta m_{(k+1)A} \delta m_{(k+1)B}} dm_{(k+1)B}. \end{aligned}$$

Similarly, the derivatives of the Hamiltonian in indexed coordinates become partial derivatives, e.g.,

$$(D_g h(g_k, m_{k+1}))_A = \frac{\partial}{\partial g_k^A} h(g_k, m_{k+1}).$$

We are now ready to prove the integrator (2.5a)-(2.5c) is symplectic.

Theorem 2.4. *The integrator* (2.5a)-(2.5c) *is symplectic, i.e., the symplectic form is preserved,*

$$\Omega_{k+1} = \Omega_k,$$

subject to first variations of the discrete Lie–Poisson equations. We will prove this by computing expressions for Ω_{k+1} and Ω_k separately and subsequently showing that their expressions are equivalent.

Proof. We will start with computing an expression for

$$\Omega_{k+1} = \langle (d\tau^{-1}_{-\Delta t \xi_{k+1}})^* dm_{k+1} \wedge g_{k+1}^{-1} dg_{k+1} \rangle.$$

Using the identity $\mathrm{Ad}^*_{\tau(\Delta t \xi_j)} (d\tau^{-1}_{\Delta t \xi_j})^* = (d\tau^{-1}_{-\Delta t \xi_j})^*$ and subsequently, equation (2.12a), we have

$$\begin{aligned} \Omega_{k+1} &= \langle \mathrm{Ad}^*_{\tau(\Delta t \xi_{k+1})} (d\tau^{-1}_{\Delta t \xi_{k+1}})^* dm_{k+1} \wedge g_{k+1}^{-1} dg_{k+1} \rangle \\ &= \langle (d\tau^{-1}_{\Delta t \xi_{k+1}})^* dm_{k+1} \wedge \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \rangle \\ &= \langle \mathrm{Ad}^*_{\tau(\Delta t \xi_k)} (d\tau^{-1}_{\Delta t \xi_k})^* dm_k \wedge \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \rangle \\ &\quad - \Delta t \langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \rangle \\ &\quad + \left\langle \frac{\delta^2 (\Delta t^{-1} s_d)}{\delta \eta_k^2} g_k^{-1} dg_k \wedge \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \right\rangle. \end{aligned}$$

Using equation (2.12b) in the third term above, this becomes

$$\begin{aligned} \Omega_{k+1} = &\langle \mathrm{Ad}^*_{\tau(\Delta t \xi_k)} (d\tau^{-1}_{\Delta t \xi_k})^* dm_k \wedge \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \rangle \\ &- \Delta t \langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge \mathrm{Ad}_{\tau(\Delta t \xi_{k+1})} g_{k+1}^{-1} dg_{k+1} \rangle \\ &+ \left\langle \frac{\delta^2 (\Delta t^{-1} s_d)}{\delta^2 \eta_k} g_k^{-1} dg_k \wedge g_k^{-1} dg_k \right\rangle \\ &+ \Delta t \left\langle \frac{\delta^2 (\Delta t^{-1} s_d)}{\delta^2 \eta_k} g_k^{-1} dg_k \wedge d\tau_{\Delta t \xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle \\ &+ \Delta t \left\langle \frac{\delta^2 (\Delta t^{-1} s_d)}{\delta^2 \eta_k} g_k^{-1} dg_k \wedge d\tau_{\Delta t \xi_{k+1}} \frac{\delta^2 (\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1} \right\rangle. \end{aligned}$$

The third term above vanishes by the symmetry of the second variation and the asymmetry of the wedge product. To see this, in coordinates, the third term above can be expressed as

$$\Delta t^{-1}\left\langle \frac{\delta^2 s_d}{\delta^2\eta_k} g_k^{-1}dg_k \wedge g_k^{-1}dg_k \right\rangle = \Delta t^{-1}\frac{\delta^2 s_d}{\delta\eta_k^A\delta\eta_k^B}(g_k^{-1}dg_k)^A \wedge (g_k^{-1}dg_k)^B.$$

The second variation of s_d above is symmetric under the interchange $A \leftrightarrow B$ while the wedge product above is antisymmetric under the interchange $A \leftrightarrow B$; hence, this term vanishes. Thus, we have the expression

$$\begin{aligned}\Omega_{k+1} = &\underbrace{\langle \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* dm_k \wedge \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}dg_{k+1}\rangle}_{\equiv(\mathrm{a1})}\\ &\underbrace{-\Delta t\langle g_k^* D^2_{\mu g}h(g_k,m_{k+1})dm_{k+1} \wedge \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}dg_{k+1}\rangle}_{\equiv(\mathrm{a2})}\\ &\underbrace{+\Delta t\left\langle \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2\eta_k} g_k^{-1}dg_k \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu}h(g_k,m_{k+1})dg_k\right\rangle}_{\equiv(\mathrm{a3})}\\ &\underbrace{+\Delta t\left\langle \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2\eta_k} g_k^{-1}dg_k \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle}_{\equiv(\mathrm{a4})}..\end{aligned}$$

Now, we will determine an expression for

$$\Omega_k = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge g_k^{-1}dg_k\rangle.$$

Using equation (2.12b), we have

$$\begin{aligned}\Omega_k &= \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge g_k^{-1}dg_k\rangle\\ &= \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}dg_{k+1}\rangle\\ &\quad - \Delta t\langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu}h(g_k,m_{k+1})dg_k\rangle\\ &\quad - \Delta t\left\langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle.\end{aligned}$$

Using equation (2.12a) in the third term above, this becomes

$$\begin{aligned}\Omega_k &= \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge g_k^{-1}dg_k\rangle\\ &= \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}dg_{k+1}\rangle\\ &\quad - \Delta t\langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu}h(g_k,m_{k+1})dg_k\rangle\\ &\quad - \Delta t\left\langle (d\tau^{-1}_{\Delta t\xi_{k+1}})^* dm_{k+1} \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle\\ &\quad - \Delta t^2\left\langle g_k^* D^2_{\mu g}h(g_k,m_{k+1})dm_{k+1} \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle\\ &\quad - \Delta t\left\langle \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2\eta_k} g_k^{-1}dg_k \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1}s_d)}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle.\end{aligned}$$

The third term above can be expressed as

$$-\left\langle (d\tau^{-1}_{\Delta t\xi_{k+1}})^* dm_{k+1} \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2 s_d}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle = -\left\langle dm_{k+1} \wedge \frac{\delta^2 s_d}{\delta^2 m_{k+1}}dm_{k+1}\right\rangle.$$

By applying an analogous symmetry and antisymmetry argument to the term in Ω_{k+1}, this term vanishes. Thus, we have the expression

$$\begin{aligned}\Omega_k = &\underbrace{\langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1} dg_{k+1}\rangle}_{\equiv(\mathrm{b1})} \\ &\underbrace{-\Delta t \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k\rangle}_{\equiv(\mathrm{b2})} \\ &\underbrace{-\Delta t^2 \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1} \right\rangle}_{\equiv(\mathrm{b3})} \\ &\underbrace{-\Delta t \left\langle \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 \eta_k} g_k^{-1} dg_k \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1} \right\rangle}_{\equiv(\mathrm{b4})}.\end{aligned}$$

Comparing the expressions for Ω_{k+1} and Ω_k above, we see that (a1) = (b1) (using the identity $\mathrm{Ad}^*_{\tau(\Delta t\xi_j)}(d\tau^{-1}_{\Delta t\xi_j})^* = (d\tau^{-1}_{-\Delta t\xi_j})^*$); additionally, we see that (a4) = (b4). Thus, we have left to show that (a2) + (a3) = (b2) + (b3). Equivalently, we have to show that (a2) − (b3) = (b2) − (a3). We will compute both sides of this expression.

Starting with the left hand side, we have

$$\begin{aligned}(\mathrm{a2}) - (\mathrm{b3}) &= -\Delta t \langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1} dg_{k+1}\rangle \\ &\quad + \Delta t^2 \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1} \right\rangle \\ &= -\Delta t \Big\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \\ &\qquad \wedge \left(\mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1} dg_{k+1} - \Delta t d\tau_{\Delta t\xi_{k+1}} \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 m_{k+1}} dm_{k+1} \right) \Big\rangle \\ &= -\Delta t \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge \left(g_k^{-1} dg_k + \Delta t d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right) \right\rangle,\end{aligned}$$

where in the last equality, we used (2.12b). We split this expression into two terms

$$\begin{aligned}(\mathrm{a2}) - (\mathrm{b3}) = &\underbrace{-\Delta t \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge g_k^{-1} dg_k \right\rangle}_{\equiv(\mathrm{x1})} \\ &\underbrace{-\Delta t^2 \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle}_{\equiv(\mathrm{x2})}.\end{aligned}$$

For the right hand side, we have

$$\begin{aligned}(\mathrm{b2}) - (\mathrm{a3}) &= -\Delta t \langle (d\tau^{-1}_{-\Delta t\xi_k})^* dm_k \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k\rangle \\ &\quad - \Delta t \left\langle \frac{\delta^2 s_k}{\delta^2 \eta_k} g_k^{-1} dg_k \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle \\ &= -\Delta t \left\langle \left((d\tau^{-1}_{-\Delta t\xi_k})^* dm_k + \frac{\delta^2(\Delta t^{-1} s_d)}{\delta^2 \eta_k} g_k^{-1} dg_k \right) \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle \\ &= -\Delta t \Big\langle \left((d\tau^{-1}_{\Delta t\xi_{k+1}})^* dm_{k+1} + \Delta t g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \right) \\ &\qquad \wedge d\tau_{\Delta t\xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \Big\rangle\end{aligned}$$

We split this expression into two terms

$$(\mathrm{b2}) - (\mathrm{a3}) = \underbrace{-\Delta t \left\langle (d\tau^{-1}_{\Delta t \xi_{k+1}})^* dm_{k+1} \wedge d\tau_{\Delta t \xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle}_{\equiv(\mathrm{y1})}$$
$$\underbrace{-\Delta t^2 \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge d\tau_{\Delta t \xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle}_{\equiv(\mathrm{y2})}.$$

Clearly, (x2) = (y2). Furthermore, we express (y1) as

$$\begin{aligned}(\mathrm{y1}) &= -\Delta t \left\langle (d\tau^{-1}_{\Delta t \xi_{k+1}})^* dm_{k+1} \wedge d\tau_{\Delta t \xi_{k+1}} D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle \\ &= -\Delta t \left\langle dm_{k+1} \wedge D^2_{g\mu} h(g_k, m_{k+1}) dg_k \right\rangle \\ &= -\Delta t dm_{(k+1)A} \wedge \frac{\partial^2}{\partial g^B \partial \mu_A} h(g_k, m_{k+1}) dg_k^B \\ &= -\Delta t \frac{\partial^2}{\partial g^B \partial \mu_A} h(g_k, m_{k+1}) dm_{(k+1)A} \wedge dg_k^B.\end{aligned}$$

We express (x1) as

$$\begin{aligned}(\mathrm{x1}) &= -\Delta t \left\langle g_k^* D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge g_k^{-1} dg_k \right\rangle \\ &= -\Delta t \left\langle D^2_{\mu g} h(g_k, m_{k+1}) dm_{k+1} \wedge dg_k \right\rangle \\ &= -\Delta t \frac{\partial^2}{\partial g^B \partial \mu_A} h(g_k, m_{k+1}) dm_{(k+1)A} \wedge dg_k^B.\end{aligned}$$

Thus, (x1) = (y1) and so we have shown (a2) − (b3) = (b2) − (a3). Thus, $\Omega_{k+1} = \Omega_k$ as claimed. □

Discrete Noether's Theorem. We will now show that the integrator (2.5a)-(2.5c) preserves the momentum map associated with a symmetry of the discrete action.

Let $\{g_k^\epsilon, m_k^\epsilon\}$ be a one-parameter family of discrete time curves with $g_k^0 = g_k$ and $m_k^0 = m_k$. Let

$$\begin{aligned}\delta g_k &= \frac{d}{d\epsilon} g_k^\epsilon \Big|_{\epsilon=0}, \\ \delta m_k &= \frac{d}{d\epsilon} m_k^\epsilon \Big|_{\epsilon=0},\end{aligned}$$

denote the variations associated to the one-parameter family of discrete time curves. Furthermore, let $s_k = \langle m_{k+1}, \xi_{k+1} \rangle - h(g_k, m_{k+1})$ denote the k^{th} discrete action density. Then, we have the following momentum preservation property of (2.5a)-(2.5c).

Theorem 2.5 (Discrete Noether's Theorem)**.** *Suppose that* (2.5a)-(2.5c) *hold and furthermore, suppose that the k^{th} discrete action density is invariant under the above variations,*

$$\delta s_k = 0.$$

Then, for any time indices $I < J$,

$$\Theta_I \cdot \delta g_I = \Theta_J \cdot \delta g_J, \tag{2.13}$$

where Θ_k is the discrete canonical form (2.9).

Proof. Define the IJ-partial discrete action sum as

$$s_d^{IJ} \equiv \sum_{k=I}^{J-1} \Delta t s_k.$$

By assumption, $\delta s_d^{IJ} = 0$ subject to the above variations. We compute the variation explicitly

$$\begin{aligned} 0 = \delta s_d^{IJ} &= \sum_{k=I}^{J-1} \Delta t \delta s_k \\ &= \sum_{k=I}^{J-1} \Delta t \langle \delta m_{k+1}, \xi_{k+1} - D_\mu h(g_k, m_{k+1}) \rangle \\ &\quad + \sum_{k=I}^{J-1} \Big[\langle m_{k+1}, \delta \xi_{k+1} \rangle - \langle D_g h(g_k, m_{k+1}), \delta g_k \rangle \Big]. \end{aligned}$$

The first sum above vanishes by (2.5b). Analogous to the proof of Theorem 2.2, we rewrite the second sum by rewriting the variations $\{\delta \xi_k\}$ in terms of $\{\delta g_k\}$. This gives

$$\begin{aligned} 0 &= \delta s_d^{IJ} \\ &= \sum_{k=I}^{J-1} \Delta t \langle -(g_k^{-1})^* (d\tau^{-1}_{\Delta t \xi_{k+1}})^* m_{k+1}/\Delta t - D_g h(g_k, m_{k+1}), \delta g_k \rangle \\ &\quad + \sum_{k=I}^{J-1} \Delta t \langle (g_{k+1})^* \mathrm{Ad}^*_{\tau(\Delta t \xi_{k+1})} (d\tau^{-1}_{\Delta t \xi_{k+1}})^* m_{k+1}/\Delta t, \delta g_{k+1} \rangle. \end{aligned}$$

We explicitly write the $k = I$ term in the first sum above and the $k = J - 1$ term in the second sum, and subsequently, reindex the second sum from $k \to k - 1$. This gives

$$\begin{aligned} 0 &= \delta s_d^{IJ} \\ &= \Delta t \langle -(g_I^{-1})^* (d\tau^{-1}_{\Delta t \xi_{I+1}})^* m_{I+1}/\Delta t - D_g h(g_I, m_{I+1}), \delta g_I \rangle \\ &\quad + \Delta t \langle (g_J)^* \mathrm{Ad}^*_{\tau(\Delta t \xi_J)} (d\tau^{-1}_{\Delta t \xi_J})^* m_J/\Delta t, \delta g_J \rangle + \\ &\quad + \sum_{k=I+1}^{J-1} \Delta t \Big[\langle -(g_k^{-1})^* (d\tau^{-1}_{\Delta t \xi_{k+1}})^* m_{k+1}/\Delta t - D_g h(g_k, m_{k+1}) \\ &\qquad\qquad + (g_k)^* \mathrm{Ad}^*_{\tau(\Delta t \xi_k)} (d\tau^{-1}_{\Delta t \xi_k})^* m_k/\Delta t, \delta g_k \rangle \Big]. \end{aligned}$$

The summation above vanishes by (2.5a). Additionally, we rewrite the first term above using (2.5a) and we rewrite the second term above using the identity $\mathrm{Ad}^*_{\tau(\Delta t \xi_J)} (d\tau^{-1}_{\Delta t \xi_J})^* = (d\tau^{-1}_{-\Delta t \xi_J})^*$. Hence,

$$\begin{aligned} 0 &= \delta s_d^{IJ} \\ &= -\langle (d\tau^{-1}_{-\Delta t \xi_I})^* m_I, g_I^{-1} \delta g_I \rangle + \langle (d\tau^{-1}_{-\Delta t \xi_J})^* m_J, g_J^{-1} \delta g_J \rangle \\ &= -\Theta_I \cdot \delta g_I + \Theta_J \cdot \delta g_J. \end{aligned}$$

□

As an application of Theorem 2.5, we will re-derive the discrete reduced Lie–Poisson equation (2.8), interpreted as momentum conservation associated to left-invariance symmetry. Let H be a left-invariant Hamiltonian, let X be a right-invariant vector field on G with $X(e) = \chi \in \mathfrak{g}$, and let φ_ϵ denote the time-ϵ flow of X. We choose X to be a right-invariant vector field, since its flow is given by left translations

$$\varphi_\epsilon(g) = e^{\epsilon \chi} g.$$

We define a one-parameter family of discrete time curves $\{g_k^\epsilon, m_k^\epsilon\}$ as

$$\begin{aligned} g_k^\epsilon &= \varphi_\epsilon(g_k) = e^{\epsilon \chi} g, \\ m_k^\epsilon &= m_k, \end{aligned}$$

i.e., the one-parameter family of discrete time curves is defined by flowing g_k by φ_ϵ, whereas m_k^ϵ remains constant with ϵ. To see why we defined m_k^ϵ this way, recall that the left-trivialized momenta m_k corresponds to a momenta $p_k = (g_k)^{*-1}m_k$ or equivalently, $m_k = g_k^* p_k$. For a given $x \in G$, the point (g_k, p_k) transforms under the cotangent lift of left-multiplication by x as $(g_k, p_k) \mapsto (xg_k, x^{*-1}p_k)$. Thus, m_k transforms as

$$m_k = g_k^* p_k \mapsto (xg_k)^* x^{*-1} p_k = g_k^* x^* x^{*-1} p_k = g_k^* p_k = m_k,$$

i.e., m_k is invariant under this transformation; thus, we define m_k^ϵ to be constant in ϵ. Additionally, observe that the variations associated to this one-parameter family of discrete time curves can be expressed as

$$\begin{aligned}\delta g_k &= \frac{d}{d\epsilon}\Big|_0 g_k^\epsilon = \frac{d}{d\epsilon}\Big|_0 \varphi_\epsilon(g_k) = X(g_k),\\ \delta m_k &= \frac{d}{d\epsilon}\Big|_0 m_k^\epsilon = \frac{d}{d\epsilon}\Big|_0 m_k = 0.\end{aligned}$$

Now, we will verify the assumption of Theorem 2.5. The k^{th} discrete action density is

$$s_k = \langle m_{k+1}, \xi_{k+1}\rangle - h(g_k, m_{k+1}) = \langle m_{k+1}, \xi_{k+1}\rangle - \tilde{H}(m_{k+1}),$$

where in the second equality, we used that $h(g_k, m_{k+1}) = \tilde{H}(m_{k+1})$ for a left-invariant Hamiltonian (see Section 2.2.1). As stated above, $m_j^\epsilon = m_j$ is invariant under this transformation. Furthermore, since $\xi_j = \tau^{-1}(g_k^{-1}g_{k+1})/\Delta t$, the corresponding transformation for ξ_j is given by

$$\begin{aligned}\xi_j^\epsilon &= \tau^{-1}((g_k^\epsilon)^{-1}g_{k+1}^\epsilon)/\Delta = \tau^{-1}((e^{\epsilon\chi}g_k)^{-1}e^{\epsilon\chi}g_{k+1})/\Delta t\\ &= \tau^{-1}(g_k^{-1}(e^{\epsilon\chi})^{-1}e^{\epsilon\chi}g_{k+1})/\Delta = \tau^{-1}(g_k^{-1}g_{k+1})/\Delta t = \xi_j,\end{aligned}$$

i.e., ξ_j is also invariant under this transformation. Hence, s_k is invariant under the above variation, so Theorem 2.5 applies. We thus have $\Theta_{k+1}\cdot \delta g_{k+1} = \Theta_k \cdot \delta g_k$, i.e.,

$$\langle (d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1}, g_{k+1}^{-1}\delta g_{k+1}\rangle = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, g_k^{-1}\delta g_k\rangle.$$

Equivalently, this can be expressed as

$$\begin{aligned}&\implies \langle g_{k+1}^{*-1}(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1}, \delta g_{k+1}\rangle = \langle g_k^{*-1}(d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \delta g_k\rangle,\\ &\implies \langle g_{k+1}^{*-1}(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1} g_{k+1}^*, \delta g_{k+1} g_{k+1}^{-1}\rangle = \langle g_k^{*-1}(d\tau^{-1}_{-\Delta t\xi_k})^* m_k g_k^*, \delta g_k g_k^{-1}\rangle.\end{aligned}$$

Now, observe that since X is right-invariant,

$$\delta g_j g_j^{-1} = X(g_j)g_j^{-1} = X(g_j g_j^{-1}) = X(e) = \chi.$$

Hence, we have

$$\langle g_{k+1}^{*-1}(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1} g_{k+1}^*, \chi\rangle = \langle g_k^{*-1}(d\tau^{-1}_{-\Delta t\xi_k})^* m_k g_k^*, \chi\rangle.$$

In particular, $\chi \in \mathfrak{g}$ was arbitrary, so we have

$$g_{k+1}^{*-1}(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1} g_{k+1}^* = g_k^{*-1}(d\tau^{-1}_{-\Delta t\xi_k})^* m_k g_k^*.$$

Multiplying on the left by g_k^* and on the right by g_k^{*-1} gives

$$\mathrm{Ad}^*_{g_k}\mathrm{Ad}^*_{g_{k+1}^{-1}}(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1} = (d\tau^{-1}_{-\Delta t\xi_k})^* m_k.$$

Since for any $x, y \in G$, $\mathrm{Ad}_x\mathrm{Ad}_y = \mathrm{Ad}_{xy}$ and $\mathrm{Ad}_{x^{-1}} = \mathrm{Ad}_x^{-1}$, we have

$$\mathrm{Ad}^*_{g_k}\mathrm{Ad}^*_{g_{k+1}^{-1}} = \mathrm{Ad}^*_{g_{k+1}^{-1}g_k} = \mathrm{Ad}^*_{(g_k^{-1}g_{k+1})^{-1}} = \mathrm{Ad}^*_{\tau(\Delta t\xi_{k+1})^{-1}} = \mathrm{Ad}^{*-1}_{\tau(\Delta t\xi_{k+1})}.$$

Using this in the equation above yields

$$\mathrm{Ad}^{*-1}_{\tau(\Delta t\xi_{k+1})}(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1} = (d\tau^{-1}_{-\Delta t\xi_k})^* m_k.$$

From the identity $\mathrm{Ad}^*_{\tau(\Delta t\xi_j)}(d\tau^{-1}_{\Delta t\xi_j})^* = (d\tau^{-1}_{-\Delta t\xi_j})^*$, we can rewrite the left and right hand sides as

$$(d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1} = \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m_k,$$

which is precisely the discrete reduced Lie–Poisson equation (2.7a).

3. Adjoint Systems on Lie Groups

The aim of this section is to develop the geometric theory of adjoint sensitivity analysis on Lie groups, in both the continuous and discrete settings. We thus focus on the case where the Hamiltonian system on T^*G is an adjoint system, as introduced in Example 2.1.

Let F be a vector field on G and consider the differential equation $\dot{g} = F(g)$. We define the adjoint Hamiltonian associated to F as

$$\begin{aligned} H &: T^*G \to \mathbb{R}, \\ (g,p) &\mapsto H(g,p) \equiv \langle p, F(g)\rangle. \end{aligned}$$

In canonical coordinates (g,p) on T^*G, the adjoint system (2.1a)-(2.1b) has the form

$$\begin{aligned} \dot{g} &= F(g), \\ \dot{p} &= -[DF(g)]^* p. \end{aligned}$$

We begin by computing the Lie–Poisson equations (2.2a)-(2.2b) for this particular class of adjoint Hamiltonian systems. We denote by f the left-trivialization of F,

$$\begin{aligned} f &: G \to \mathfrak{g}, \\ g &\mapsto f(g) \equiv g^{-1} \cdot F(g). \end{aligned}$$

Then, the left-trivialized Hamiltonian $h : G \times \mathfrak{g}^* \to \mathbb{R}$ has the form

$$h(g,\mu) = \langle \mu, f(g)\rangle.$$

Computing the functional derivatives of h yields

$$\begin{aligned} D_\mu h(g,\mu) &= f(g), \\ D_g h(g,\mu) &= [Df(g)]^* \mu. \end{aligned}$$

In particular, the Lie–Poisson system (2.2a)-(2.2b) for the adjoint Hamiltonian has the form

$$\dot{g} = F(g), \tag{3.1a}$$

$$\dot{\mu} = -g^* \cdot [Df(g)]^* \mu + \mathrm{ad}^*_{f(g)}\mu. \tag{3.1b}$$

We now address the question of existence and uniqueness for solutions of the Type II system (2.2a)-(2.2d). For general Hamiltonians on T^*G, this is a complicated question which is dependent on the particular Hamiltonian. In particular, since the system has Type II boundary conditions $g(0) = g_0, \mu(T) = \mu_1$, even a local solution theory cannot be stated generally, as opposed to systems with initial-value conditions $g(0) = g_0, \mu(0) = \mu_0$. A simple way to see this is that we can think of a Hamiltonian system on $G \times \mathfrak{g}^*$ with Type II boundary conditions as a fixed-time, free-position-endpoint, fixed-fiber-endpoint shooting control problem: given $g(0) = g_0 \in G$ and $T > 0$, find $\mu(0) = \mu_0$ such that $\mu(T) = \mu_1$ subject to the Hamiltonian dynamics. This is in general a tricky problem that is dependent on the Hamiltonian under consideration.

However, for adjoint systems in particular, we can provide a global solution theory which utilizes the fact that the adjoint system covers an ODE on G; assuming the ODE on G behaves nicely, we will have unique solutions for the adjoint system on T^*G. We make this more precise in the following proposition.

Proposition 3.1 (Global Existence and Uniqueness of Solutions to Adjoint Systems on T^*G)**.** *Let $T > 0, g_0 \in G, \mu_1 \in \mathfrak{g}^*$. Let F be a complete vector on field on G, i.e., it generates a global flow $\Phi_F : \mathbb{R} \times G \to G$.*

Then, there exists a unique curve $(g, \mu) : [0, T] \to G \times \mathfrak{g}^*$ satisfying the Lie–Poisson system with Type II boundary conditions (2.2a)-(2.2d)*, where h is the left-trivialized adjoint Hamiltonian associated to F.*

Furthermore, there exists a unique curve $(g, p) : [0, T] \to T^*G$ satisfying Hamilton's equations with Type II boundary conditions (2.1a)-(2.1d)*, where H is the adjoint Hamiltonian associated to F.*

Proof. By the fundamental theorem on flows [25], there exists a unique curve $g : \mathbb{R} \to G$ satisfying $\dot{g} = F(g)$ and $g(0) = g_0$, given by the flow of F on g_0, $g(t) = \Phi_F(t, g_0)$. In particular, g is a smooth function of t, since F is smooth. Recall that we assume all maps and manifolds are smooth, unless otherwise stated.

Now, with this curve $g(t)$ fixed, we substitute this into the differential equation for μ (2.2b), to obtain

$$\dot{\mu} = -g(t)^* \cdot [Df(g(t))]^* \mu + \mathrm{ad}^*_{f(g(t))}\mu.$$

In particular, this equation has the form of a time-dependent linear differential equation on $\mathfrak{g}^*$,

$$\dot{\mu} = L(t)\mu,$$

where we define the time-dependent linear operator $L : \mathbb{R} \to \mathrm{End}(\mathfrak{g}^*)$ by

$$L(t) = -g(t)^* \cdot [Df(g(t))]^* + \mathrm{ad}^*_{f(g(t))}. \tag{3.2}$$

Since g is a smooth function of t, L is a smooth, and in particular continuous, function of t. Hence, by the standard solution theory for linear differential equations, there exists a unique curve $\mu : [0, T] \to \mathfrak{g}^*$ satisfying $\dot{\mu} = L(t)\mu$ and $\mu(T) = \mu_1$.

For the second statement of the proposition, note that solution curves $(g, p) : [0, T] \to T^*G$ of (2.1a)-(2.1d) are in one-to-one correspondence with solution curves $(g, \mu) : [0, T] \to G \times \mathfrak{g}^*$ of (2.2a)-(2.2d) via left-translation. □

By the above proposition, we know that there exists a unique solution to the adjoint system on T^*G with Type II boundary conditions, under the assumption that F is complete. For Lie groups, there are two particularly important cases where this assumption is satisfied.

Corollary 3.1. *If G is a compact Lie group, then the above proposition holds for any vector field F on G.*

If F is a left-invariant vector field on a (not necessarily compact) Lie group G, then the above proposition holds.

Proof. The first statement follows from the fact that any vector field on a compact manifold is complete. The second statement follows from the fact that any left-invariant vector field on a Lie group is complete. See [25]. □

The Variational System. An important property of adjoint systems is that they satisfy a quadratic conservation law, which is at the heart of the method of adjoint sensitivity analysis [44].

To state this conservation law, we introduce the variational equation associated to an ODE $\dot{g} = F(g)$ on a Lie group G, which is defined to be the linearization of the ODE,

$$\frac{d}{dt}\delta g = DF(g)\delta g.$$

We refer to the combined system

$$\frac{d}{dt}g = F(g), \tag{3.3a}$$

$$\frac{d}{dt}\delta g = DF(g)\delta g, \tag{3.3b}$$

as the variational system, which is interpreted as an ODE on TG.

As with the adjoint system, it will be useful to left-trivialize this system, which will give an ODE on $G \times \mathfrak{g}$. As before, let $f(g) = g^{-1} \cdot F(g)$ be the left-trivialization of F. Let $\eta = g^{-1} \cdot \delta g$ and let $\xi = g^{-1} \cdot \dot{g}$. As is well-known (see, for example, [35]), we have the relation

$$\dot{\eta} = \delta\xi - [\xi, \eta].$$

In particular, since $\xi = g^{-1} \cdot \dot{g} = f(g)$, we have $\delta\xi = Df(g)\delta g = Df(g)g \cdot \eta$, so that the above relation becomes

$$\dot{\eta} = Df(g)g \cdot \eta - [f(g), \eta],$$

which we refer to as the left-trivialized variational equation. We refer to the combined system

$$\dot{g} = F(g), \tag{3.4a}$$

$$\dot{\eta} = Df(g)g \cdot \eta - \mathrm{ad}_{f(g)}\eta, \tag{3.4b}$$

as the left-trivialized variational system on $G \times \mathfrak{g}$. Analogous to the existence and uniqueness result for adjoint systems, Proposition 3.1, we have the following result.

Proposition 3.2 (Global Existence and Uniqueness of Solutions to Variational Systems on TG)**.** *Let $T > 0, g_0 \in G, \eta_0 \in \mathfrak{g}$. Let F be a complete vector on field on G, i.e., it generates a global flow $\Phi_F : \mathbb{R} \times G \to G$.*

Then, there exists a unique curve $(g, \eta) : [0, T] \to G \times \mathfrak{g}$ satisfying the left-trivialized variational system (3.4a)-(3.4b) *with initial conditions $g(0) = g_0, \eta(0) = \eta_0$.*

Furthermore, there exists a unique curve $(g, \delta g) : [0, T] \to TG$ satisfying the variational system (3.3a)-(3.3b) *with initial conditions $g(0) = g_0, \delta g(0) = g_0 \cdot \eta_0$.*

Proof. The proof is almost identical to the proof of Proposition 3.1, noting that once the solution curve $g(t)$ of $\dot{g} = F(g), g(0) = g_0$ is fixed, the variational equation can be expressed as a time-dependent linear equation on $\mathfrak{g}$,

$$\dot{\eta} = M(t)\eta,$$

where the time-dependent linear operator $M : \mathbb{R} \to \mathrm{End}(\mathfrak{g})$ is smooth. In fact, it is easily verified that $M(t) = -L(t)^*$, where L is the time-dependent linear operator (3.2) defined in the proof of Proposition 3.1.

Furthermore, by left-translation, solutions to the left-trivialized variational system and the variational system, with the above respective initial conditions, are in one-to-one correspondence. □

We can now state the quadratic conservation law enjoyed by solutions of the adjoint and variational systems.

Theorem 3.1. *Let (g, μ) be a solution curve of the left-trivialized adjoint system and let (g, η) be a solution curve of the left-trivialized variational system, both covering the same base curve g. Let (g, p) and $(g, \delta g)$ be the respective solution curves for the adjoint system and variational system obtained by left-translation. Then,*

$$\frac{d}{dt}\langle \mu(t), \eta(t)\rangle = 0, \tag{3.5a}$$

$$\frac{d}{dt}\langle p(t), \delta q(t)\rangle = 0. \tag{3.5b}$$

Proof. Note that it suffices to prove either (3.5a) or (3.5b), since left-translation preserves the duality pairing,

$$\langle \mu(t), \eta(t)\rangle = \langle g(t)^{-1} \cdot p(t), g(t) \cdot \delta q(t)\rangle = \langle p(t), \delta g(t)\rangle.$$

We will prove (3.5a). Compute

$$\begin{aligned}
&\frac{d}{dt}\langle \mu(t), \eta(t)\rangle \\
&\quad = \langle \dot\mu(t), \eta(t)\rangle + \langle \mu(t), \dot\eta(t)\rangle \\
&\quad = \langle -g(t)^* \cdot [Df(g(t))]^*\mu(t) + \mathrm{ad}^*_{f(g(t))}\mu(t), \eta(t)\rangle + \langle \mu(t), Df(g(t))g(t) \cdot \eta(t) - \mathrm{ad}_{f(g(t))}\eta(t)\rangle \\
&\quad = -\langle \mu(t), Df(g(t))g(t) \cdot \eta(t)\rangle + \langle \mu(t), Df(g(t))g(t) \cdot \eta(t)\rangle \\
&\qquad + \langle \mu(t), \mathrm{ad}_{f(g(t))}\eta(t)\rangle - \langle \mu(t), \mathrm{ad}_{f(g(t))}\eta(t)\rangle \\
&\quad = 0.
\end{aligned}$$

□

In particular, we have the following corollary of Propositions 3.1 and 3.2 and Theorem 3.1.

Corollary 3.2. *Let $T > 0, g_0 \in G, \mu_1 \in \mathfrak{g}^*, \eta_0 \in \mathfrak{g}$. Let F be a complete vector field on G. Then, the solution curves of the adjoint and variational systems from Propositions 3.1 and 3.2 satisfy the quadratic conservation law*

$$\langle \mu(0), \eta_0\rangle = \langle \mu_1, \eta(T)\rangle.$$

As we will see in Section 3.3, this conservation law will be the basis for adjoint sensitivity analysis on Lie groups.

3.1. **Reduction of Adjoint Systems for Left-invariant Vector Fields.** In practice, many interesting mechanical systems arise from the flow of left-invariant vector fields on Lie groups. As such, we will consider adjoint systems in the particular case where the vector field is left-invariant. First, we will show that left-invariant vector fields are in one-to-one correspondence with left-invariant adjoint Hamiltonians. Subsequently, we will state the adjoint equations in this particular case.

Proposition 3.3. *Let F be a vector field on G. Then the adjoint Hamiltonian $H(g, p) = \langle p, F(g)\rangle$ associated to F is left-invariant if and only if F is left-invariant.*

Proof. Assume that F is left-invariant, i.e., $F(xg) = xF(g)$ for all $x, g \in G$. Then, for any $x, g \in G, p \in T^*_g G$,

$$H(xg, x^{*-1}p) = \langle x^{*-1}p, F(xg)\rangle = \langle x^{*-1}p, xF(g)\rangle = \langle x^* x^{*-1}p, F(g)\rangle = \langle p, F(g)\rangle = H(g, p),$$

i.e., H is left-invariant.

Conversely, assume that H is left-invariant, i.e., $H(g,p) = H(xg, x^{*-1}p)$ for all $x, g \in G, p \in T_g^*G$. Then, for any $x, g \in G, p \in T_g^*G$,

$$\langle p, F(g)\rangle = H(g,p) = H(xg, x^{*-1}p) = \langle x^{*-1}p, F(xg)\rangle = \langle p, x^{-1}F(xg)\rangle.$$

Since $p \in T_g^*G$ is arbitrary, we have for all $x, g \in G$,

$$F(g) = x^{-1}F(xg),$$

i.e., $xF(g) = F(xg)$, so F is left-invariant. □

Since a left-invariant vector field corresponds to a left-invariant adjoint Hamiltonian, the reduction theory discussed in Section 2.2.1 applies. Thus, the adjoint equation for the momenta μ, from equation (2.6), is given by

$$\dot{\mu} = \mathrm{ad}^*_{F(e)}\mu,$$

since $\tilde{H}(\mu) = H(e,\mu) = \langle \mu, F(e)\rangle$ and hence, $D_\mu \tilde{H}(\mu) = F(e)$.

3.2. **Type II Variational Discretization of Adjoint Systems.** In this section, we apply the Type II variational integrators developed in Section 2.2 to the particular case of adjoint systems. We will show explicitly that these integrators preserve the adjoint-variational quadratic conservation law which is key to adjoint sensitivity analysis, and thus, these methods are geometric structure-preserving methods for adjoint sensitivity analysis on Lie groups.

Consider the variational integrators that we derived in Section 2.2, applied to the adjoint system (3.1a)-(3.1b). Substituting $h(g,\mu) = \langle \mu, f(g)\rangle$ into the discrete Lie–Poisson equations (2.5a)-(2.5c), we have the discrete Lie–Poisson adjoint equations

$$(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1} - \mathrm{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m_k = -\Delta t g_k^*[Df(g_k)]^* m_{k+1}, \tag{3.6a}$$

$$\xi_{k+1} = f(g_k) \tag{3.6b}$$

$$g_{k+1} = g_k\tau(\Delta t\xi_{k+1}) = g_k\tau(\Delta t f(g_k)). \tag{3.6c}$$

In order to derive a discrete analogue of the adjoint conservation law, we consider the discrete variational equation, which is a discretization of the continuous variational equation (3.4b). To derive the discrete variational equation, note that as mentioned in Section 2.2, the variation of equation (3.6c) can be expressed

$$\delta\xi_{k+1} = d\tau^{-1}_{\Delta t\xi_{k+1}}(-g_k^{-1}\delta g_k + \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}\delta g_{k+1})/\Delta t.$$

Furthermore, by taking the variation of equation (3.6b), we have

$$\delta\xi_{k+1} = Df(g_k)\delta g_k.$$

Combining these two equations yields

$$Df(g_k)\delta g_k = d\tau^{-1}_{\Delta t\xi_{k+1}}(-g_k^{-1}\delta g_k + \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})} g_{k+1}^{-1}\delta g_{k+1})/\Delta t.$$

Defining the left-trivialized variation $\eta_k = g_k^{-1}\delta g_k$, the above can be expressed as

$$\Delta t d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k \eta_k = -\eta_k + \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})}\eta_{k+1}, \tag{3.7}$$

which we refer to as the discrete variational equation.

Theorem 3.2. *The discrete Lie–Poisson adjoint equations and the discrete variational equation satisfy the following quadratic conservation law,*

$$\langle (d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1}, \eta_{k+1}\rangle = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle. \tag{3.8}$$

Proof. Recall the identity $\mathrm{Ad}^*_{\tau(\Delta t\xi_j)}(d\tau^{-1}_{\Delta t\xi_j})^* = (d\tau^{-1}_{-\Delta t\xi_j})^*$. Starting from the left hand side of equation (3.8), we compute

$$\begin{aligned}\langle (d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1}, \eta_{k+1}\rangle &= \langle \mathrm{Ad}^*_{\tau(\Delta t\xi_{k+1})}(d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}, \eta_{k+1}\rangle \\ &= \langle (d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}, \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})}\eta_{k+1}\rangle.\end{aligned}$$

Substituting (3.6a) and (3.7) yields

$$\begin{aligned}&\langle (d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1}, \eta_{k+1}\rangle \\ &\qquad = \langle (d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}, \mathrm{Ad}_{\tau(\Delta t\xi_{k+1})}\eta_{k+1}\rangle \\ &\qquad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k - \Delta t g_k^*[Df(g_k)]^* m_{k+1}, \eta_k + \Delta t d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle \\ &\qquad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle + \Delta t\langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle \\ &\qquad\quad - \Delta t\langle g_k^*[Df(g_k)]^* m_{k+1}, \eta_k\rangle - \Delta t^2\langle g_k^*[Df(g_k)]^* m_{k+1}, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle.\end{aligned}$$

Substitute (3.6a) into the second term above,

$$\begin{aligned}&\langle (d\tau^{-1}_{-\Delta t\xi_{k+1}})^* m_{k+1}, \eta_{k+1}\rangle \\ &\quad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle + \Delta t\langle (d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1} + \Delta t g_k^*[Df(g_k)]^* m_{k+1}, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle \\ &\quad\quad - \Delta t\langle g_k^*[Df(g_k)]^* m_{k+1}, \eta_k\rangle - \Delta t^2\langle g_k^*[Df(g_k)]^* m_{k+1}, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle \\ &\quad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle + \Delta t\langle (d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{k+1}, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle \\ &\quad\quad + \cancel{\Delta t^2\langle g_k^*[Df(g_k)]^* m_{k+1}, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle} \\ &\quad\quad - \Delta t\langle g_k^*[Df(g_k)]^* m_{k+1}, \eta_k\rangle - \cancel{\Delta t^2\langle g_k^*[Df(g_k)]^* m_{k+1}, d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle} \\ &\quad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle + \Delta t\langle m_{k+1}, (d\tau^{-1}_{\Delta t\xi_{k+1}}) d\tau_{\Delta t\xi_{k+1}} Df(g_k) g_k\eta_k\rangle \\ &\quad\quad - \Delta t\langle g_k^*[Df(g_k)]^* m_{k+1}, \eta_k\rangle \\ &\quad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle + \Delta t\langle m_{k+1}, Df(g_k) g_k\eta_k\rangle - \Delta t\langle g_k^*[Df(g_k)]^* m_{k+1}, \eta_k\rangle \\ &\quad = \langle (d\tau^{-1}_{-\Delta t\xi_k})^* m_k, \eta_k\rangle. \qquad \square\end{aligned}$$

3.3. **Adjoint Sensitivity Analysis on Lie Groups.** In this section, we utilize the discrete methods for adjoint systems on Lie groups developed in the previous sections to address the following two types of optimization problems: an initial condition optimization problem,

$$\begin{gathered}\min_{g_0\in G} C(g(T)), \\ \text{such that } \dot g(t) = F(g(t)),\ t\in(0,T), \\ g(0) = g_0,\end{gathered} \tag{3.9}$$

and an optimal control problem,

$$\begin{gathered}\min_{u\in U} C(g(T)), \\ \text{such that } \dot g(t) = F(g(t),u),\ t\in(0,T), \\ g(0) = g_0,\end{gathered} \tag{3.10}$$

where in (3.10), we have introduced a parameter-dependent vector field F.

Initial Condition Sensitivity. We begin with problem (3.9). We refer to $C : G \to \mathbb{R}$ as the terminal cost function. Thus, the problem (3.9) is to find an initial condition $g_0 \in G$ which minimizes the cost function at the terminal-value $g(T)$, subject to the dynamics of the ODE $\dot{g} = F(g)$.

For gradient-based algorithms, one needs the derivative of the terminal cost function $C(g(T))$ with respect to the initial condition g_0; we refer to this derivative as the *initial condition sensitivity.* One cannot generally compute an expression for the sensitivity analytically, since such an expression would require knowing $g(T)$ explicitly as a function of g_0.

However, the adjoint system pulls back derivatives with respect to $g(T)$ into derivatives with respect to $g(0)$ [44]. In other words, the derivative of $C(g(T))$ with respect to g_0 can be computed by setting the terminal momenta to be $p(T) = dC(g(T))$ and evolving backwards in time to find the desired derivative $p(0)$. One cannot generally compute the curves $g(t)$ and $p(t)$ exactly, so we instead use the method (3.6a)-(3.6c) to approximately solve the ODE and its adjoint. By Theorem 3.2, the property that the adjoint system pulls back derivatives with respect to $g(T)$ to derivatives with respect to $g(0)$ is preserved by the method, so it exactly gives the desired derivative.

More specifically, to obtain the initial condition sensitivity, recall the quadratic conservation law

$$\langle (d\tau^{-1}_{-\Delta t \xi_N})^* m_N, g_N^{-1}\delta g_N\rangle = \langle (d\tau^{-1}_{-\Delta t \xi_0})^* m_0, g_0^{-1}\delta g_0\rangle.$$

We set $(d\tau^{-1}_{-\Delta t \xi_N})^* m_N = d_L C(g(T))$, where d_L denotes the left-trivialized derivative, $d_L C(g(T)) \equiv g_N^{-*} dC(g(T))$. This gives $m_N = (d\tau_{-\Delta t \xi_N})^* d_L C(g(T))$. Subsequently, evolve the momenta backward in time using (3.6a) to obtain m_0. Finally, the left-trivialized derivative of $C(g(T))$ with respect to g_0 is given by $(d\tau^{-1}_{-\Delta t \xi_0})^* m_0$. This is summarized in the following algorithm.

Algorithm 1 Left-Trivialized Initial Condition Sensitivity

Input: g_{init}
Initialize: $g_0 \leftarrow g_{\text{init}}, \{g_k\}_{k=1}^N, \{m_k\}_0^N$
Output: Left-Trivialized Derivative of $C(g(T))$ with respect to g_0
for k=1,...,N **do**
  $g_k \leftarrow g_{k-1}\tau(\Delta t f(g_{k-1}))$
end for
$m_N \leftarrow (d\tau_{-\Delta t f(g_N)})^* d_L C(g_N)$
for k=1,...,N **do**
  $m_{N-k} \leftarrow$ Solve $m : (d\tau^{-1}_{\Delta t \xi_{k+1}})^* m_{N-k+1} - \text{Ad}^*_{\tau(\Delta t \xi_k)}(d\tau^{-1}_{\Delta t \xi_k})^* m = -\Delta t g_k^* [Df(g_k)]^* m_{N-k+1}$,
end for
Return $(d\tau^{-1}_{-\Delta t f(g_0)})^* m_0$

This can be combined with a line-search algorithm to solve the optimization problem (3.9). More precisely, fixing an inner product on $\mathfrak{g}$, such as the Frobenius inner product

$$(A, B)_F \equiv \text{Tr}(A^* B),$$

we can identity $\mathfrak{g}^*$ with $\mathfrak{g}$ and hence, identify the output of Algorithm 1 with the left-trivialized gradient of $C(g(T))$ with respect to g_0, $\nabla_{g_0} C(g(T))$, which is an element of $\mathfrak{g}$. With this identification, the initial condition can be updated as $g_0 \leftarrow g_0\tau(-\gamma\nabla_{g_0} C(g(T)))$, for some line-search step size γ.

Remark 3.1. *The preceding discussion of adjoint sensitivity analysis for terminal cost functions can be easily adapted to a cost function consisting of both a terminal cost and a running cost,*

$$C(g(T)) + \int_0^T L(g(t))dt.$$

This is done by augmenting the adjoint Hamiltonian with the running cost Lagrangian L, i.e., by using the augmented adjoint Hamiltonian

$$H_L(g,p) \equiv H(g,p) + L(g) = \langle p, F(g)\rangle + L(g).$$

The only modification is an additional term in the momentum equation (3.6a) *corresponding to the derivative of L. For more details, see [49].*

The significance of this approach is that it is intrinsic; at any iteration in the line-search algorithm, the iterate g_0 is valued in G, to numerical precision. Furthermore, while this is also true of projection-based optimization algorithms, such methods generally no longer preserve the adjoint-variational quadratic conservation law and hence, may not capture the appropriate descent direction.

Parameter Sensitivity. We now consider problem (3.10). The problem (3.10) is to find a parameter $u \in U$ which minimizes the terminal cost function $C(g(T))$, subject to the dynamics of the parameter-dependent ODE $\dot{g} = F(g,u)$ with $g(0) = g_0$ fixed. We assume that F is continuously differentiable with respect to u.

For a gradient-based algorithm, we will need the derivative of the terminal cost function $C(g(T))$ with respect to the parameter u; we refer to this derivative as the *parameter sensitivity.*

We begin with a discussion of how the parameter sensitivity is obtained from the adjoint system in the continuous setting, since the derivation will be analogous in the discrete setting. Define the parameter-dependent action as

$$S[g,p;u] \equiv \int_0^T \langle p, \dot{g} - F(g,u)\rangle dt.$$

Consider the augmented cost function, given by subtracting the parameter-dependent action from the terminal cost function,

$$\begin{aligned} J &\equiv C(g(T)) - S[g,p;u] \\ &= C(g(T)) - \int_0^T \langle p, \dot{g} - F(g,u)\rangle dt. \end{aligned}$$

By left-trivialization, this is equivalent to

$$J = C(g(T)) - \int_0^T \langle \mu, \xi - f(g,u)\rangle dt, \tag{3.11}$$

where $\xi = g^{-1}\dot{g}$ and f is the left-trivialization of F. Observe that since the integral of (3.11) vanishes when $\dot{g} = f(g,u)$, we have that the derivative of J with respect to u equals the derivative of C with respect to u, subject to the variational equations, where the variation $\delta_u g$ is given by the variation of g induced by varying u. Thus,

$$\frac{d}{du}C(g(T)) = \frac{d}{du}J.$$

Proposition 3.4. *The (continuous) parameter sensitivity is given by*

$$\frac{d}{du}C(g(T)) = \frac{d}{du}J = \int_0^T \langle \mu, \frac{\partial}{\partial u} f(g,u)\rangle dt,$$

where μ is chosen to satisfy the adjoint equation $-\dot{\mu} + ad^*_\xi \mu - g^*[D_g f(g,u)]^*\mu = 0$ and the terminal condition $\mu(T) = g^{-*}dC(g(T))$.

Proof. We compute dJ/du explicitly,

$$\frac{d}{du}J(g(T)) = \langle dC(g(T)), \delta_u g(T)\rangle - \int_0^T \Big[\langle \mu, \delta_u \xi - [D_g f(g,u)]g\eta_u\rangle - \langle \mu, \frac{\partial}{\partial u} f(g,u)\rangle\Big] dt,$$

where we introduced the left-trivialized variation $\eta_u = g^{-1}\delta_u g$ and we have decomposed the total derivative of f with respect to u into its implicit dependence on u through g as well as its explicit dependence on u, i.e.,

$$\frac{d}{du}f(g,u) = [D_g f(g,u)]\delta_u g + \frac{\partial}{\partial u} f(g,u).$$

Using the relation $\dot{\eta}_u = \delta_u \xi - \mathrm{ad}_\xi \eta_u$, this becomes

$$\begin{aligned}\frac{d}{du}J(g(T)) &= \langle dC(g(T)), \delta_u g(T)\rangle - \int_0^T \Big[\langle \mu, \dot{\eta}_u + \mathrm{ad}_\xi \eta_u - [D_g f(g,u)]g\eta_u\rangle - \langle \mu, \frac{\partial}{\partial u} f(g,u)\rangle\Big] dt \\ &= \langle dC(g(T)), \delta_u g(T)\rangle - \langle \mu, \eta_u\rangle\Big|_0^T \\ &\quad - \int_0^T \Big[\langle -\dot{\mu} + \mathrm{ad}_\xi^* \mu - g^*[D_g f(g,u)]^*\mu, \eta_u\rangle - \langle \mu, \frac{\partial}{\partial u} f(g,u)\rangle\Big] dt,\end{aligned}$$

where we integrated the $\langle \mu, \dot{\eta}_u\rangle$ term by parts. Now, the first pairing in the integral vanishes if μ satisfies the adjoint equation. Furthermore, $\eta_u(0) = 0$ since the initial condition for problem (3.10) is fixed. Hence, we have

$$\frac{d}{du}J(g(T)) = \langle dC(g(T)), \delta_u g(T)\rangle - \langle \mu(T), \eta_u(T)\rangle + \int_0^T \langle \mu, \frac{\partial}{\partial u} f(g,u)\rangle dt.$$

If we choose the terminal condition $\mu(T) = g^{-*}dC(g(T))$, the first two terms on the right hand side cancel, which gives the expression for the desired parameter sensitivity

$$\frac{d}{du}C(g(T)) = \frac{d}{du}J(g(T)) = \int_0^T \langle \mu, \frac{\partial}{\partial u} f(g,u)\rangle dt,$$

where μ is chosen to satisfy the adjoint equation $-\dot{\mu} + \mathrm{ad}_\xi^*\mu - g^*[D_g f(g,u)]^*\mu = 0$ and the terminal condition $\mu(T) = g^{-*}dC(g(T))$. □

From here, the generalization to the discrete setting is straightforward. In analogy with the continuous case, we define the parameter-dependent left-trivialized discrete action

$$s_d[\{g_k\}, \{m_k\}; u] = \Delta t \sum_{k=0}^{N-1} \Big(\langle m_{k+1}, \xi_{k+1}\rangle - h(g_k, m_{k+1}l; u)\Big),$$

and form the discrete augmented cost function by subtracting the discrete action from the terminal cost, i.e.,

$$\begin{aligned} J_d &\equiv C(g_N) - s_d[\{g_k\}, \{m_k\}; u] \\ &= C(g_N) - \Delta t \sum_{j=0}^{N-1} \langle m_{j+1}, \xi_{j+1} - f(g_j, u)\rangle.\end{aligned}$$

We then have an analogous result to determine the parameter sensitivity by computing the derivative dJ_d/du.

Proposition 3.5. *The (discrete) parameter sensitivity is given by*

$$\frac{d}{du}C(g_N) = \Delta t \sum_{j=0}^{N-1} \langle m_j, \frac{\partial}{\partial u} f(g_j, u)\rangle,$$

where m_j is chosen to satisfy the discrete Lie–Poisson adjoint equation (3.6a) *and the terminal condition $m_N = (d\tau_{-\Delta t\xi_N})^* d_L C(g_N)$.*

Proof. Analogous to the continuous setting, we have

$$\frac{d}{du}C(g_N) = \frac{d}{du}J_d.$$

Now, we calculate dJ_d/du explicitly,

$$\frac{d}{du}J_d = \langle dC(g_N), \delta_u g_N\rangle - \Delta t \sum_{j=0}^{N-1} \langle m_{j+1}, \delta_u \xi_{j+1} - D_g f(g_j, u)\delta_u g_j - \frac{\partial f}{\partial u}(g_j, u)\rangle.$$

Using the identity $\delta_u \xi_{j+1} = d\tau^{-1}_{\Delta t\xi_{j+1}}(-g_k^{-1}\delta_u g_k + \mathrm{Ad}_{\tau(\Delta t\xi_{j+1})} g_{j+1}^{-1}\delta_u g_{j+1})/\Delta t$, the above can be expressed as

$$\begin{aligned}\frac{d}{du}J_d = \langle dC(g_N), \delta_u g_N\rangle - \Delta t \sum_{j=0}^{N-1} \Big[&\Delta t^{-1}\langle (d\tau^{-1}_{\Delta t\xi_{j+1}})^* m_{j+1}, -g_j^{-1}\delta_u g_j\rangle \\ &+ \Delta t^{-1}\langle \mathrm{Ad}^*_{\tau(\Delta t\xi_{j+1})}(d\tau^{-1}_{\Delta t\xi_{j+1}})^* m_{j+1}, g_{j+1}^{-1}\delta_u g_{j+1}\rangle \\ &- \langle g_j^*[D_g f(g_j, u)]^* m_{j+1}, g_j^{-1}\delta_u g_j\rangle - \langle m_{j+1}, \frac{\partial f}{\partial u}(g_j, u)\rangle\Big].\end{aligned}$$

Now, we reindex $j \to j-1$ the second pairing inside the square brackets above; the sum for this term now runs from 1 to N. However, we explicitly write out the $j = N$ term and note that we can include the $j = 0$ term in the sum since $\delta_u g_0 = 0$, as the initial condition g_0 is fixed under the variation. Hence, we have

$$\begin{aligned}\frac{d}{du}J_d = \;&\langle dC(g_N), \delta_u g_N\rangle - \langle (d\tau^{-1}_{-\Delta t\xi_N})^* m_N, g_N^{-1}\delta_u g_N\rangle \\ &- \Delta t \sum_{j=0}^{N-1} \Big[\Delta t^{-1}\langle (d\tau^{-1}_{\Delta t\xi_{j+1}})^* m_{j+1}, -g_j^{-1}\delta_u g_j\rangle \\ &+ \Delta t^{-1}\langle \mathrm{Ad}^*_{\tau(\Delta t\xi_j)}(d\tau^{-1}_{\Delta t\xi_j})^* m_j, g_j^{-1}\delta_u g_j\rangle \\ &- \langle g_j^*[D_g f(g_j, u)]^* m_{j+1}, g_j^{-1}\delta_u g_j\rangle - \langle m_{j+1}, \frac{\partial f}{\partial u}(g_j, u)\rangle\Big].\end{aligned}$$

The first two terms above vanish if we set the terminal condition $(d\tau^{-1}_{-\Delta t\xi_N})^* m_N = g_N^{-*} dC(g_N)$, i.e., $m_N = (d\tau_{-\Delta t\xi_N})^* d_L C(g_N)$. Furthermore, the first three terms in the square brackets vanish if m_j satisfies the discrete Lie–Poisson adjoint equation (3.6a). Hence, we have the parameter sensitivity

$$\frac{d}{du}C(g_N) = \frac{d}{du}J_d = \Delta t \sum_{j=0}^{N-1} \langle m_{j+1}, \frac{\partial f}{\partial u}(g_j, u)\rangle. \qquad \square$$

Thus, assuming that we can calculate $\partial f/\partial u$ (which is generally known since we know how the parameter-dependent vector field varies with u), we can calculate the sensitivity $dC(g_N)/du$. This is summarized in the following algorithm.

This can be combined with a line-search algorithm to solve the optimization problem (3.10). Note that U could be a vector space, in which case a standard line-search algorithm could be used, or U could be a manifold, in which case a line-search algorithm on manifolds could be used (see, for example, [1]).

Algorithm 2 Parameter Sensitivity

Input: g_{init}
Initialize: $g_0 \leftarrow g_{\text{init}}, \{g_k\}_{k=1}^N, \{m_k\}_0^N$
Output: Derivative of $C(g(T))$ with respect to u
for k=1,...,N **do**
  $g_k \leftarrow g_{k-1}\tau(\Delta t f(g_{k-1}))$
end for
$m_N \leftarrow (d\tau_{-\Delta t f(g_N)})^* d_L C(g_N)$
for k=1,...,N **do**
  $m_{N-k} \leftarrow$ Solve $m : (d\tau^{-1}_{\Delta t\xi_{k+1}})^* m_{N-k+1} - \text{Ad}^*_{\tau(\Delta t\xi_k)}(d\tau^{-1}_{\Delta t\xi_k})^* m = -\Delta t g_k^*[D_g f(g_k, u)]^* m_{N-k+1}$,
end for
Return $\Delta t \sum_{j=0}^{N-1} \langle m_{j+1}, \frac{\partial}{\partial u} f(g_j, u)\rangle$

3.3.1. *Numerical Examples.* As examples of adjoint sensitivity analysis on Lie groups, we will solve an example of each of the problems (3.9) and (3.10).

Initial Condition Sensitivity Example. Fixing $g_{\text{target}} \in G$, find $g_0 \in G$ such that $g(T) = g_{\text{target}}$ subject to the initial-value problem $\dot{g} = F(g), g(0) = g_0$.

Using the Frobenius inner product $(\cdot,\cdot)_F$ and its induced norm $\|\cdot\|_F$, this can be cast as an optimization problem of the form (3.9),

$$\begin{aligned} \min_{g_0 \in G} C(g(T)) &\equiv \frac{1}{2}\|g(T) - g_{\text{target}}\|_F^2, \\ \text{such that } \dot{g}(t) &= F(g(t)),\ t \in (0,T), \\ g(0) &= g_0. \end{aligned}$$

This optimization problem is clearly equivalent to the above shooting problem because C is minimized at the unique minimizer $g(T) = g_{\text{target}}$, since $C(g_{\text{target}}) = 0$ and $C(g) > 0$ for any $g \neq g_{\text{target}}$ by nondegeneracy of the norm. We choose this simple problem because the analytic answer is known: g_0 should simply be chosen to be the reverse time-T flow of g_{target} under F.

For our numerical example, we take $G = \text{SO}(3)$, $F(g) = gX$ with

$$X = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ -1 & -2 & 0 \end{pmatrix} \in \mathfrak{g},\ g_{\text{target}} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in G,$$

and some initial iterate

$$g_0 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \in G.$$

The left-trivialized gradient can be computed to be $\nabla_{g(T)} C(g(T)) = g(T)^T g_1 - g_1^T g(T)$ [29] which is identified with the left-trivialized derivative through the inner product. This allows us to initialize the terminal momenta m_N as described in the previous section. Subsequently, we solve the optimization problem using Algorithm 1 and a line-search method. For simplicity, since this example is just to provide a demonstration of the theory laid out in the paper, we will use a fixed line-search step size $\gamma = 0.1$, although in practice one would likely use a more sophisticated method such as Armijo backtracking. We take $T = 1$ with $\Delta t = 0.01$. Finally, for the retraction, we use the Cayley transform and its derivatives given by

$$\tau(\xi) = \mathrm{cay}(\xi) \equiv (I_{3\times 3} + \frac{1}{2}\xi)(I_{3\times 3} - \frac{1}{2}\xi)^{-1},$$
$$d\tau_\xi(x) = (I_{3\times 3} - \xi/2)^{-1} x (I_{3\times 3} + \xi/2)^{-1},$$
$$d\tau_\xi^{-1}(x) = (I_{3\times 3} - \xi/2) x (I_{3\times 3} + \xi/2).$$

The cost function $C(g(T))$ over 50 iterations is shown in Figure 1. The $SO(3)$ manifold error of each iteration of g_0 is shown in Figure 2, where the manifold error is defined as

$$\mathrm{Error}(g) \equiv \frac{1}{2}\|g^T g - \mathrm{Id}_{3\times 3}\|_F^2;$$

as can be seen in the figure, each iterate lies on $SO(3)$ to machine precision.

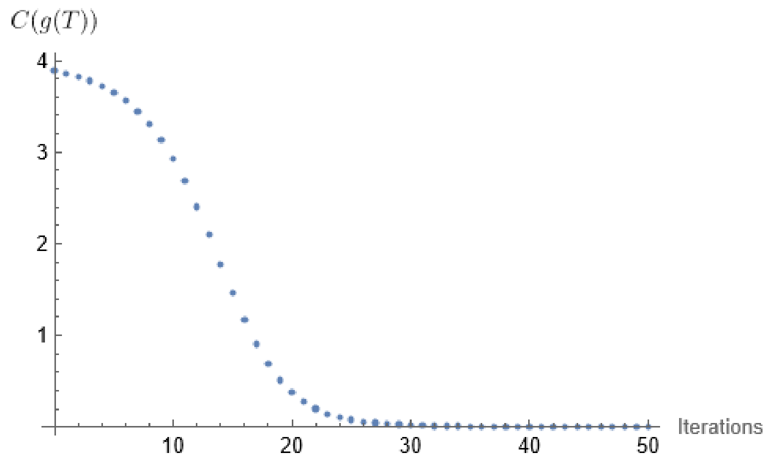

FIGURE 1. Cost function minimization via line-search algorithm for shooting problem on $SO(3)$

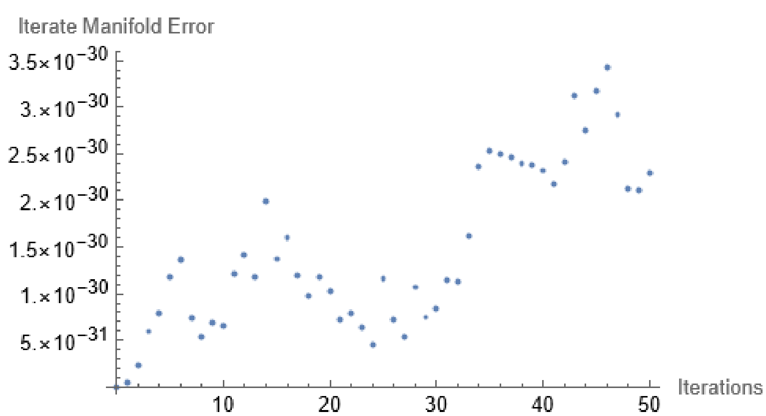

FIGURE 2. Iterate $SO(3)$ manifold error

This example can be visualized as follows. The objective is to find an element g_0 of $SO(3)$ such that under the time $T = 1$ flow of F, $g(1) = g_{\text{target}}$ where $g(0) = g_0$. For this example, g_{target} is chosen to be a $\pi/2$ counterclockwise rotation about the z-axis, i.e., in the xy plane. Thus, we can imagine some test mass located at $\vec{v} \in \mathbb{R}^3$ which is rotated by $g(t)$, which generates a curve $g(t)\vec{v}$.

In particular, choosing $\vec{v} = \hat{x}$, the unit vector in the x direction, then the curve produced from rotating the test mass should end at $g_{\text{target}}\hat{x} = \hat{y}$. Each iteration in the algorithm generates such a curve. In Figure 3, several such curves are shown, with the initial point in the curve $g_0\hat{x}$ marked. Additionally, the desired terminal point $g_{\text{target}}\hat{x}$ is marked.

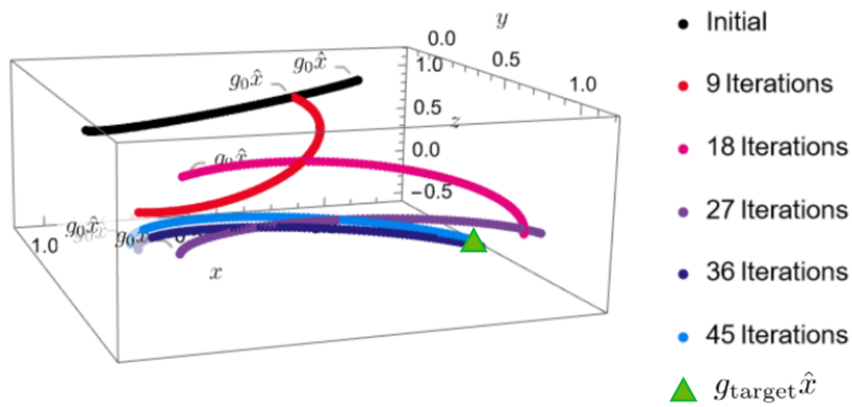

FIGURE 3. Visualization of the shooting problem on $SO(3)$ using a test mass over several iterations

Parameter Sensitivity Example. For our second example, we consider the following problem of the form (3.10),

$$\begin{gathered}\min_{u\in U} C(g(T)) \equiv \frac{1}{2}\|g(T) - g_{\text{target}}\|_F^2,\\ \text{such that } \dot{g}(t) = F(g(t), u),\ t \in (0,T),\\ g(0) = g_0.\end{gathered}$$

Thus, this optimal control problem is to find $u \in U$ such that the vector field $F(\cdot, u)$ steers the initial condition $g_0 \in G$ to some desired terminal-value $g_{\text{target}} \in G$.

We again take $G = SO(3)$. We will assume that F is a parameter-dependent left-invariant vector field $F(g,u) = gX(u)$, where $u \in \mathbb{R}^3$ parametrizes $\mathfrak{so}(3)$ as

$$X(u) = \begin{pmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{pmatrix}.$$

For simplicity, we take $g_0 = \text{Id}_{3\times 3} = g_{\text{target}}$ since the analytic answer is known: F should be the zero vector field, since $g_0 = g_{\text{target}}$ and hence, the optimal value of u is $u = (0,0,0)^T$. We take an initial guess of $u = (1,2,-1)^T$. We again take $T = 1$ with $\Delta t = 0.01$, using the same retraction as the previous example. We combine the parameter sensitivity, obtained from Algorithm 2, with a simple vector space line-search algorithm,

$$u \leftarrow u - \gamma \frac{d}{du} C,$$

with a fixed line-search step size $\gamma = 0.1$. The cost function $C(g(T))$ over 50 iterations is shown in Figure 4.

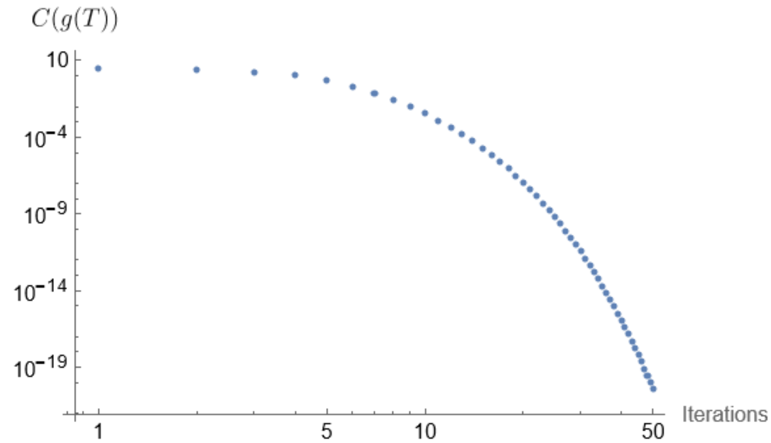

FIGURE 4. Cost function minimization via line-search algorithm for optimal control problem on $SO(3)$

4. Conclusion and Future Research Directions

In this paper, we developed continuous and discrete global Type II variational principles on the cotangent bundle of a Lie group G. In the discrete setting, the Type II variational principle leads to a structure-preserving variational integrator on T^*G which we showed to be symplectic and momentum-preserving. Subsequently, we applied these Type II variational principles to the class of adjoint Hamiltonian systems on T^*G. This results in a structure-preserving method to perform adjoint sensitivity analysis on Lie groups, allowing one to exactly compute sensitivities in optimization problems subject to the dynamics of an ODE on G.

One research direction which we are currently pursuing is to explore the geometry of adjoint sensitivity analysis in the infinite-dimensional setting, with the application of PDE-constrained optimization in mind. It would be interesting to synthesize this line of research with the ideas presented in this paper, to develop Hamiltonian integrators for PDEs where the solutions are valued in Lie groups, algebras, or more generally, solutions which are stationary sections over principal and fibre bundles associated to a structure group G, such as gauge field theories (see, for example, [19; 39]). It would be particularly interesting to extend the Type II multisymplectic Hamiltonian variational integrators developed in [48] to apply to the setting of Lie groups-valued fields, in order to investigate the role of multisymplectic integrators for adjoint sensitivity analysis in both space and time.

Another natural research direction would be to explore the applications of geometric structure-preserving adjoint sensitivity analysis on Lie groups. One such application is the training of neural networks via backpropagation. In particular, if a neural network is viewed as a discretization of a *neural ODE* [10], then backpropagation can be viewed as a discretization of the corresponding adjoint equation [38]. As is discussed in [38], utilizing symplectic methods to perform backpropagation leads to efficient methods for training neural networks. It would be interesting to utilize the methods presented in this paper to perform symplectic backpropagation of neural networks where the neural ODE evolves over a Lie group, which would arise in group-equivariant neural networks [11; 24] where a Lie group symmetry is a fundamental feature of the neural network. In particular, the reduction theory for adjoint systems on Lie groups that was developed in this paper would be relevant.

Acknowledgements

BT was supported by the NSF Graduate Research Fellowship DGE-2038238, and by NSF under grants DMS-1813635. ML was supported by NSF under grants DMS-1345013, DMS-1813635, and by AFOSR under grant FA9550-18-1-0288.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, 2008. (Page 34)

[2] M. Barbero-Liñán and D. Martín de Diego. Retraction maps: a seed of geometric integrators. *ArXiv*, abs/2106.00607, 2021.

[3] G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms. *J. Stat. Phys.*, 74: 1117–1143, 1994. (Page 3)

[4] M. Benning, E. Celledoni, M. J. Ehrhardt, B. Owren, and C.-B. Schönlieb. Deep learning as optimal control problems: Models and numerical methods. *J. Comput. Dyn.*, 6(2):171–198, 2019. (Page 6)

[5] A. I. Bobenko and Yu. B. Suris. Discrete time Lagrangian mechanics on Lie groups, with an application to the Lagrange top. *Commun. Math. Phys.*, 204:147–188, 1999. (Page 2, 5, 8)

[6] N. Bou-Rabee and J. E. Marsden. Hamilton–Pontryagin integrators on Lie groups part I: Introduction and structure-preserving properties. *Found. Comput. Math.*, 9:197–219, 2009. (Page 8, 12, 14, 15)

[7] D. G. Cacuci. Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. *J. Math. Phys.*, 22(12):2794–2802, 1981. (Page 6)

[8] Y. Cao, S. Li, L. Petzold, and R. Serban. Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution. *SIAM J. Sci. Comput.*, 24(3): 1076–1089 (14 pages), 2003. (Page 6)

[9] Elena Celledoni and Arieh Iserles. Approximating the exponential from a Lie algebra to a Lie group. *Math. Comp.*, 69(232):1457–1480, 2000. (Page 5)

[10] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 6572–6583, Red Hook, NY, USA, 2018. Curran Associates Inc. (Page 38)

[11] T. Cohen and M. Welling. Group equivariant convolutional networks. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. (Page 38)

[12] J. d'Alembert. *Traité de dynamique.* 1743.

[13] M. Delgado-Téllez and A. Ibort. A panorama of geometrical optimal control theory. *Extracta mathematicae*, 18(2):129–151, 2003.

[14] A. Echeverría-Enríquez, J. Marín-Solano, M.C. Muñoz-Lecanda, and N. Román-Roy. Geometric reduction in optimal control theory with symmetries. *Reports on Mathematical Physics*, 52 (1):89–113, 2003.

[15] E. G. Fahnestock, T. Lee, M. Leok, N. H. McClamroch, and D. J. Scheeres. Polyhedral potential and variational integrator computation of the full two body problem. *Proc. AIAA/AAS Astrodynamics Specialist Conf.*, AIAA-2006-6289, 2006. (Page 5)

[16] M. B. Giles and N. A. Pierce. An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion*, 65:393–415, 2000. (Page 6)

[17] A. Griewank. A mathematical view of automatic differentiation. In *Acta Numer.*, volume 12, pages 321–398. Cambridge University Press, 2003. (Page 6)

[18] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-preserving algorithms for ordinary differential equations*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. (Page 2)

[19] M. J. D. Hamilton. *Mathematical Gauge Theory*. Springer Cham, 2017. (Page 38)

[20] N. H. Ibragimov. Integrating factors, adjoint equations and Lagrangians. *Journal of Mathematical Analysis and Applications*, 318(2):742–757, 2006. (Page 6)

[21] N. H. Ibragimov. A new conservation theorem. *J. Math. Anal. Appl.*, 333(1):311–328, 2007. (Page 6)

[22] A. Iserles, H. Munthe-Kaas, S. P. Nørsett, and A. Zanna. Lie-group methods. In *Acta Numerica*, volume 9, pages 215–365. Cambridge University Press, 2000. (Page 5)

[23] M. B. Kobilarov and J. E. Marsden. Discrete geometric optimal control on Lie groups. *IEEE Transactions on Robotics*, 27(4):641–655, 2011. (Page 2, 5, 8, 12, 14)

[24] R. Kondor and S. Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2747–2755. PMLR, 10–15 Jul 2018. (Page 38)

[25] J. M. Lee. *Introduction to Smooth Manifolds*. Springer New York, NY, 2012. (Page 26)

[26] T. Lee, N. H. McClamroch, and M. Leok. A Lie group variational integrator for the attitude dynamics of a rigid body with applications to the 3D pendulum. *Proc. IEEE Conf. on Control Applications*, pages 962–967, 2005. (Page 2, 5)

[27] T. Lee, M. Leok, and N. H. McClamroch. Lie group variational integrators for the full body problem in orbital mechanics. *Celestial Mechanics and Dynamical Astronomy*, 98(2):121–144, 2007. (Page 5)

[28] T. Lee, M. Leok, and N. H. McClamroch. Lie group variational integrators for the full body problem. *Comput. Methods Appl. Mech. Engrg.*, 196(29-30):2907–2924, 2007. (Page 2, 5)

[29] T. Lee, T. Tao, and M. Leok. Variational symplectic accelerated optimization on Lie groups. *Proc. IEEE Conf. on Decision and Control*, pages 233–240, 2021. (Page 35)

[30] M. Leok and T. Ohsawa. Variational and geometric structures of discrete Dirac mechanics. *Found. Comput. Math.*, 11(5):529–562, 2011. (Page 2, 4)

[31] M. Leok and J. Zhang. Discrete Hamiltonian variational integrators. *IMA J. Numer. Anal.*, 31(4):1497–1532, 2011. (Page 4, 17)

[32] S. Li and L. Petzold. Adjoint sensitivity analysis for time-dependent partial differential equations with adaptive mesh refinement. *Journal of Computational Physics*, 198(1):310–325, 2004.

[33] S. Li and L. R. Petzold. Solution adapted mesh refinement and sensitivity analysis for parabolic partial differential equation systems. In L. T. Biegler, M. Heinkenschloss, O. Ghattas, and B. van Bloemen Waanders, editors, *Large-Scale PDE-Constrained Optimization*, pages 117–132. Springer, Berlin, 2003. (Page 6)

[34] Z. Ma and C. W. Rowley. Lie–Poisson integrators: A Hamiltonian, variational approach. *Int. J. Numer. Meth. Engng.*, 82:1609–1644, 2010. (Page 2, 5, 14, 16)

[35] J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry*. Springer New York, NY, 1999. (Page 15, 27)

[36] J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numer.*, 10: 317–514, 2001. (Page 2, 3, 10, 17)

[37] J. E. Marsden, S. Pekarsky, and S. Shkoller. Discrete Euler–Poincaré and Lie–Poisson equations. *Nonlinearity*, 12(6):1647, 1999. (Page 2, 5, 14, 16)

[38] T. Matsubara, Y. Miyatake, and T. Yaguchi. Symplectic adjoint method for exact gradient of neural ODE with minimal memory. In *Advances in Neural Information Processing Systems*,

2021. (Page 38)

[39] M. Nakahara. *Geometry, Topology, and Physics*. CRC Press, second edition, 2003. (Page 38)

[40] V. T. Nguyen, D. Georges, and G. Besançon. State and parameter estimation in 1-D hyperbolic PDEs based on an adjoint method. *Automatica*, 67(C):185–191, May 2016. ISSN 0005-1098. (Page 6)

[41] N. A. Pierce and M. B. Giles. Adjoint recovery of superconvergent functionals from PDE approximations. *SIAM Rev.*, 42(2):247–264, 2000. (Page 6)

[42] I. M. Ross. A roadmap for optimal control: The right way to commute. *Ann. NY Acad. Sci.*, 1065(1):210–231, 2005. (Page 6)

[43] M. Ross and F. Fahroo. A pseudospectral transformation of the convectors of optimal control systems. *IFAC Proc. Ser.*, 34(13):543–548, 2001. (Page 6)

[44] J. M. Sanz-Serna. Symplectic Runge–Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more. 58(1):3–33, 2016. (Page 2, 6, 26, 31)

[45] J. M. Schmitt and M. Leok. Properties of Hamiltonian variational integrators. *IMA Journal of Numerical Analysis*, 38(1):377–398, 2017. (Page 4)

[46] Z. Sirkes and E. Tziperman. Finite difference of adjoint or adjoint of finite difference? *Mon. Weather Rev.*, 125(12):3373–3378, 1997. (Page 6)

[47] Y.-F. Tang. Formal energy of a symplectic scheme for Hamiltonian systems and its applications (i). *Computers & Mathematics with Applications*, 27(7):31 – 39, 1994. (Page 3)

[48] B. Tran and M. Leok. Multisymplectic Hamiltonian variational integrators. *International Journal of Computer Mathematics (Special Issue on Geometric Numerical Integration, Twenty-Five Years Later)*, 99(1):113–157, 2022. (Page 38)

[49] B. Tran and M. Leok. Geometric methods for adjoint systems. (preprint, `arxiv/2205.02901 [math.OC]`), 2022. (Page 6, 32)

[50] J. Vankerschaver, H. Yoshimura, and M. Leok. The Hamilton–Pontryagin principle and multi-Dirac structures for classical field theories. *J. Math. Phys.*, 53(7):072903 (25 pages), 2012.

[51] Q. Wang, K. Duraisamy, J. J. Alonso, and G. Iaccarino. Risk assessment of scramjet unstart using adjoint-based sampling. *AIAA J.*, 50(3):581–592, 2012. (Page 6)

[52] H. Yoshimura and J. E. Marsden. Dirac structures in Lagrangian mechanics Part I: Implicit Lagrangian systems. *J. Geom. Phys.*, 57(1):133–156, 2006.

[53] H. Yoshimura and J. E. Marsden. Dirac structures in Lagrangian mechanics Part II: Variational structures. *J. Geom. Phys.*, 57(1):209–250, 2006.