Title
Time-adaptive Lagrangian variational integrators for accelerated optimization

Journal
The Journal of Geometric Mechanics, 15(1)

ISSN
1941-4889

Authors
Duruisseaux, Valentin
Leok, Melvin

Publication Date
2023

DOI
10.3934/jgm.2023010

Peer reviewed

TIME-ADAPTIVE LAGRANGIAN VARIATIONAL INTEGRATORS FOR ACCELERATED OPTIMIZATION ON MANIFOLDS

VALENTIN DURUISSEAUX AND MELVIN LEOK

Abstract. A variational framework for accelerated optimization was recently introduced on normed vector spaces and Riemannian manifolds in Wibisono et al. [65] and Duruisseaux and Leok [19]. It was observed that a careful combination of time-adaptivity and symplecticity in the numerical integration can result in a significant gain in computational efficiency. It is however well known that symplectic integrators lose their near energy preservation properties when variable time-steps are used. The most common approach to circumvent this problem involves the Poincaré transformation on the Hamiltonian side, and was used in Duruisseaux et al. [20] to construct efficient explicit algorithms for symplectic accelerated optimization. However, the current formulations of Hamiltonian variational integrators do not make intrinsic sense on more general spaces such as Riemannian manifolds and Lie groups. In contrast, Lagrangian variational integrators are well-defined on manifolds, so we develop here a framework for time-adaptivity in Lagrangian variational integrators and use the resulting geometric integrators to solve optimization problems on vector spaces and Lie groups.

1. Introduction

Many machine learning algorithms are designed around the minimization of a loss function or the maximization of a likelihood function. Due to the ever-growing scale of data sets, there has been a lot of focus on first-order optimization algorithms because of their low cost per iteration. In 1983, Nesterov's accelerated gradient method was introduced in [54], and was shown to converge in $\mathcal{O}(1/k^2)$ to the minimum of the convex objective function f, improving on the $\mathcal{O}(1/k)$ convergence rate exhibited by the standard gradient descent methods. This $\mathcal{O}(1/k^2)$ convergence rate was shown in [55] to be optimal among first-order methods using only information about ∇f at consecutive iterates. This phenomenon in which an algorithm displays this improved rate of convergence is referred to as acceleration, and other accelerated algorithms have been derived since Nesterov's algorithm. More recently, it was shown in [59] that Nesterov's accelerated gradient method limits, as the time-step goes to 0, to a second-order differential equation and that the objective function $f(x(t))$ converges to its optimal value at a rate of $\mathcal{O}(1/t^2)$ along the trajectories of this ordinary differential equation. It was later shown in [65] that in continuous time, the convergence rate of $f(x(t))$ can be accelerated to an arbitrary convergence rate $\mathcal{O}(1/t^p)$ in normed spaces, by considering flow maps generated by a family of time-dependent Bregman Lagrangian and Hamiltonian systems which is closed under time-rescaling. This framework for accelerated optimization in normed vector spaces has been studied and exploited using geometric numerical integrators in [3; 8; 20–22; 31]. In [20], time-adaptive geometric integrators have been proposed to take advantage of the time-rescaling property of the Bregman family and design efficient explicit algorithms for symplectic accelerated optimization. It was observed that a careful use of adaptivity and symplecticity could result in a significant gain in computational efficiency, by simulating higher-order Bregman dynamics using the computationally efficient lower-order Bregman integrators applied to the time-rescaled dynamics.

More generally, symplectic integrators form a class of geometric integrators of interest since, when applied to Hamiltonian systems, they yield discrete approximations of the flow that preserve the symplectic 2-form and as a result also preserve many qualitative aspects of the underlying dynamical system (see [26]). In particular, when applied to conservative Hamiltonian systems, symplectic integrators exhibit excellent long-time near-energy preservation. However, when symplectic integrators were first used in combination with variable time-steps, the near-energy preservation was lost and the integrators performed poorly (see [7; 24]). There has been a great effort to circumvent this problem, and there have been many successes, including methods based on the Poincaré transformation [25; 69]: a Poincaré transformed Hamiltonian in extended phase space is constructed which allows the use of variable time-steps in symplectic integrators without losing the nice conservation properties associated to these integrators. In [20], the Poincaré transformation was incorporated in the Hamiltonian variational integrator framework which provides a systematic method for constructing symplectic integrators of arbitrarily high-order based on the discretization of Hamilton's principle [27; 49], or equivalently, by the approximation of the generating function of the symplectic flow map. The Poincaré transformation was at the heart of the construction of time-adaptive geometric integrators for Bregman Hamiltonian systems which resulted in efficient, explicit algorithms for accelerated optimization in [20].

In [42; 60], accelerated optimization algorithms were proposed in the Lie group setting for specific choices of parameters in the Bregman family, and [2] provided a first example of Bregman dynamics on Riemannian manifolds. The entire variational framework was later generalized to the Riemannian manifold setting in [19], and time-adaptive geometric integrators taking advantage of the time-rescaling property of the Bregman family have been proposed in the Riemannian manifold setting as well using discrete variational integrators incorporating holonomic constraints [16] and projection-based variational integrators [18]. Note that both these strategies relied on exploiting the structure of the Euclidean spaces in which the Riemannian manifolds are embedded. Although the Whitney and Nash Embedding Theorems [53; 63; 64] imply that there is no loss of generality when studying Riemannian manifolds only as submanifolds of Euclidean spaces, designing intrinsic methods that would exploit and preserve the symmetries and geometric properties of the manifold could have advantages both in terms of computational efficiency and in terms of improving our understanding of the acceleration phenomenon on Riemannian manifolds. Developing an intrinsic extension of Hamiltonian variational integrators to manifolds would require some additional work, since the current approach involves Type II/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. On the other hand, Lagrangian variational integrators involve a Type I generating function $L_d(q_k, q_{k+1})$ which only depends on the position at the boundary points and is therefore well-defined on manifolds, and many Lagrangian variational integrators have been derived on Riemannian manifolds, especially in the Lie group [5; 27; 28; 36–39; 43; 56] and homogeneous space [40] settings. This gives an incentive to construct a mechanism on the Lagrangian side which mimics the Poincaré transformation, since it is more natural and easier to work on the Lagrangian side on more general spaces than on the Hamiltonian side. However, a first difficulty is that the Poincaré transformed Hamiltonian is degenerate and therefore does not have a corresponding Type I Lagrangian formulation. As a result, we cannot exploit the usual correspondence between Hamiltonian and Lagrangian dynamics and need to come up with a different strategy. A second difficulty is that all the literature to this day on the Poincaré transformation have constructed the Poincaré transformed system by reverse-engineering, which does not provide a lot of insight into the origin of the mechanism and how it can be extended to different systems.

Outline. We first review the basics of variational integration of Lagrangian and Hamiltonian systems, and the Poincaré transformation in Section 2. We then introduce a simple but novel derivation of the Poincaré transformation from a variational principle in Section 3.1. This gives additional insight into the transformation mechanism and provides natural candidates for time-adaptivity on the Lagrangian side, which we then construct both in continuous and discrete time in Sections 3.2 and 3.3. We then compare the performance of the resulting time-adaptive Lagrangian accelerated optimization algorithms to their Poincaré Hamiltonian analogues in Section 4. Finally, we demonstrate in Section 5 that our time-adaptive Lagrangian approach extends naturally to more general spaces without having to face the obstructions experienced on the Hamiltonian side.

Contributions. In summary, the main contributions of this paper are:

- A novel derivation of the Poincaré transformation from a variational principle, in Section 3.1
- New frameworks for variable time-stepping in Lagrangian integrators, in Sections 3.2 and 3.3
- Discrete variational formulations of continuous Lagrangian mechanics with the new variable time-stepping mechanisms, in Sections 3.2 and 3.3
- New explicit symplectic accelerated optimization algorithms on normed vector spaces
- New intrinsic symplectic accelerated optimization algorithms on Riemannian manifolds

2. Background

2.1. Lagrangian and Hamiltonian Mechanics.

Given a n-dimensional manifold $\mathcal{Q}$, a Lagrangian is a function $L : T\mathcal{Q} \to \mathbb{R}$. The corresponding action integral $\mathcal{S}$ is the functional

$$\mathcal{S}(q) = \int_0^T L(q,\dot{q})dt, \tag{2.1}$$

over the space of smooth curves $q : [0,T] \to \mathcal{Q}$. Hamilton's variational principle states that $\delta\mathcal{S} = 0$ where the variation $\delta\mathcal{S}$ is induced by an infinitesimal variation δq of the trajectory q that vanishes at the endpoints. Given local coordinates $(q^1, \dots, q^n)$ on the manifold $\mathcal{Q}$, Hamilton's variational principle can be shown to be equivalent to the Euler–Lagrange equations,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) = \frac{\partial L}{\partial q^k}, \qquad \text{for } k = 1, \dots, n. \tag{2.2}$$

The Legendre transform $\mathbb{F}L : T\mathcal{Q} \to T^*\mathcal{Q}$ of L is defined fiberwise by $\mathbb{F}L : (q^i, \dot{q}^i) \mapsto \left(q^i, \frac{\partial L}{\partial \dot{q}^i}\right)$, and we say that a Lagrangian L is regular or nondegenerate if the Hessian matrix $\frac{\partial^2 L}{\partial \dot{q}^2}$ is invertible for every q and $\dot{q}$, and hyperregular if the Legendre transform $\mathbb{F}L$ is a diffeomorphism. A hyperregular Lagrangian on $T\mathcal{Q}$ induces a Hamiltonian system on $T^*\mathcal{Q}$ via

$$H(q,p) = \langle \mathbb{F}L(q,\dot{q}), \dot{q}\rangle - L(q,\dot{q}) = \sum_{j=1}^n p_j \dot{q}^j - L(q,\dot{q})\bigg|_{p_i = \frac{\partial L}{\partial \dot{q}^i}}, \tag{2.3}$$

where $p_i = \frac{\partial L}{\partial \dot{q}^i} \in T^*\mathcal{Q}$ is the conjugate momentum of q^i. A Hamiltonian H is called hyperregular if $\mathbb{F}H : T^*\mathcal{Q} \to T\mathcal{Q}$ defined by $\mathbb{F}H(\alpha)\cdot\beta = \frac{d}{ds}\big|_{s=0} H(\alpha + s\beta)$, is a diffeomorphism. Hyperregularity of the Hamiltonian H implies invertibility of the Hessian matrix $\frac{\partial^2 H}{\partial p^2}$ and thus nondegeneracy of H. Theorem 7.4.3 in [48] states that hyperregular Lagrangians and hyperregular Hamiltonians correspond in a bijective manner. We can also define a Hamiltonian variational principle on the Hamiltonian side in momentum phase space which is equivalent to Hamilton's equations,

$$\dot{p}_k = -\frac{\partial H}{\partial q^k}(p,q), \qquad \dot{q}^k = \frac{\partial H}{\partial p_k}(p,q), \qquad \text{for } k = 1, \dots, n. \tag{2.4}$$

These equations can also be shown to be equivalent to the Euler–Lagrange equations (2.2), provided that the Lagrangian is hyperregular.

2.2. Variational Integrators.

Variational integrators are derived by discretizing Hamilton's principle, instead of discretizing Hamilton's equations directly. As a result, variational integrators are symplectic, preserve many invariants and momentum maps, and have excellent long-time near-energy preservation (see [49]).

Traditionally, variational integrators have been designed based on the Type I generating function known as the discrete Lagrangian, $L_d : Q \times Q \to \mathbb{R}$. The exact discrete Lagrangian that generates the time-h flow of Hamilton's equations can be represented in both a variational form and in a boundary-value form. The latter is given by

$$L_d^E(q_0, q_1; h) = \int_0^h L(q(t), \dot{q}(t))dt, \tag{2.5}$$

where $q(0) = q_0$, $q(h) = q_1$, and q satisfies the Euler–Lagrange equations over the time interval $[0, h]$. A variational integrator is defined by constructing an approximation $L_d : Q \times Q \to \mathbb{R}$ to L_d^E, and then applying the discrete Euler–Lagrange equations,

$$p_k = -D_1 L_d(q_k, q_{k+1}), \qquad p_{k+1} = D_2 L_d(q_k, q_{k+1}), \tag{2.6}$$

where D_i denotes a partial derivative with respect to the i-th argument, and these equations implicitly define the integrator $\tilde{F}_{L_d} : (q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$. The error analysis is greatly simplified via Theorem 2.3.1 of [49], which states that if a discrete Lagrangian, $L_d : Q \times Q \to \mathbb{R}$, approximates the exact discrete Lagrangian $L_d^E : Q \times Q \to \mathbb{R}$ to order r, i.e.,

$$L_d(q_0, q_1; h) = L_d^E(q_0, q_1; h) + \mathcal{O}(h^{r+1}), \tag{2.7}$$

then the discrete Hamiltonian map $\tilde{F}_{L_d} : (q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$, viewed as a one-step method, has order of accuracy r. Many other properties of the integrator, such as momentum conservation properties of the method, can be determined by analyzing the associated discrete Lagrangian, as opposed to analyzing the integrator directly.

Variational integrators have been extended to the framework of Type II and Type III generating functions, commonly referred to as discrete Hamiltonians (see [34; 46; 57]). Hamiltonian variational integrators are derived by discretizing Hamilton's phase space principle. The boundary-value formulation of the exact Type II generating function of the time-h flow of Hamilton's equations is given by the exact discrete right Hamiltonian,

$$H_d^{+,E}(q_0, p_1; h) = p_1^\top q_1 - \int_0^h \left[p(t)^\top \dot{q}(t) - H(q(t), p(t)) \right] dt, \tag{2.8}$$

where (q, p) satisfies Hamilton's equations with boundary conditions $q(0) = q_0$ and $p(h) = p_1$. A Type II Hamiltonian variational integrator is constructed by using an approximate discrete Hamiltonian H_d^+, and applying the discrete right Hamilton's equations,

$$p_0 = D_1 H_d^+(q_0, p_1), \qquad q_1 = D_2 H_d^+(q_0, p_1), \tag{2.9}$$

which implicitly defines the integrator, $\tilde{F}_{H_d^+} : (q_0, p_0) \mapsto (q_1, p_1)$.

Theorem 2.3.1 of [49], which simplified the error analysis for Lagrangian variational integrators, has an analogue for Hamiltonian variational integrators. Theorem 2.2 in [57] states that if a discrete right Hamiltonian H_d^+ approximates the exact discrete right Hamiltonian $H_d^{+,E}$ to order r, i.e.,

$$H_d^+(q_0, p_1; h) = H_d^{+,E}(q_0, p_1; h) + \mathcal{O}(h^{r+1}), \tag{2.10}$$

then the discrete right Hamilton's map $\tilde{F}_{H_d^+} : (q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$, viewed as a one-step method, is order r accurate. Note that discrete left Hamiltonians and corresponding discrete left Hamilton's maps can also be constructed in the Type III case (see [20; 46]).

Examples of variational integrators include Galerkin variational integrators [46; 49], Prolongation-Collocation variational integrators [45], and Taylor variational integrators [58]. In many cases, the Type I and Type II/III approaches will produce equivalent integrators. This equivalence has been established in [58] for Taylor variational integrators provided the Lagrangian is hyperregular, and in [46] for generalized Galerkin variational integrators constructed using the same choices of basis functions and numerical quadrature formula provided the Hamiltonian is hyperregular. However, Hamiltonian and Lagrangian variational integrators are not always equivalent. In particular, it was shown in [57] that even when the Hamiltonian and Lagrangian integrators are analytically equivalent, they might still have different numerical properties because of numerical conditioning issues. Even more to the point, Lagrangian variational integrators cannot always be constructed when the underlying Hamiltonian is degenerate. This is particularly relevant in variational accelerated optimization since the time-adaptive Hamiltonian framework for accelerated optimization presented in [20] relies on a degenerate Hamiltonian which has no associated Lagrangian description. We will thus not be able to exploit the usual correspondence between Hamiltonian and Lagrangian dynamics and will have to come up with a different strategy to allow time-adaptivity on the Lagrangian side.

We now describe the construction of Taylor variational integrators as introduced in [58] as we will use them in our numerical experiments. A discrete approximate Lagrangian or Hamiltonian is constructed by approximating the flow map and the trajectory associated with the boundary values using a Taylor method, and approximating the integral by a quadrature rule. The Taylor variational integrator is generated by the implicit discrete Euler–Lagrange equations associated to the discrete Lagrangian or by the Hamilton's equations associated with the discrete Hamiltonian. More explicitly, we first construct ρ-order and $(\rho+1)$-order Taylor methods $\Psi_h^{(\rho)}$ and $\Psi_h^{(\rho+1)}$ approximating the exact time-h flow map $\Phi_h : TQ \to TQ$ corresponding to the Euler–Lagrange equation in the Type I case or the exact time-h flow map $\Phi_h : T^*Q \to T^*Q$ corresponding to Hamilton's equation in the Type II case. Let $\pi_Q : (q,p) \mapsto q$ and $\pi_{T^*Q} : (q,p) \mapsto p$. Given a quadrature rule of order s with weights and nodes (b_i, c_i) for $i = 1, ..., m$, the Taylor variational integrators are then constructed as follows:

Type I Lagrangian Taylor Variational Integrator (LTVI):

(i) Approximate $\dot{q}(0) = v_0$ by the solution $\tilde{v}_0$ of the problem $q_1 = \pi_Q \circ \Psi_h^{(\rho+1)}(q_0, \tilde{v}_0)$.

(ii) Generate approximations $(q_{c_i}, v_{c_i}) \approx (q(c_i h), \dot{q}(c_i h))$ via $(q_{c_i}, v_{c_i}) = \Psi_{c_i h}^{(\rho)}(q_0, \tilde{v}_0)$.

(iii) Apply the quadrature rule to obtain the associated discrete Lagrangian,

$$L_d(q_0, q_1; h) = h \sum_{i=1}^{m} b_i L(q_{c_i}, v_{c_i}).$$

(iv) The variational integrator is then defined by the implicit discrete Euler–Lagrange equations,

$$p_0 = -D_1 L_d(q_0, q_1), \qquad p_1 = D_2 L_d(q_0, q_1).$$

Type II Hamiltonian Taylor Variational Integrator (HTVI):

(i) Approximate $p(0) = p_0$ by the solution $\tilde{p}_0$ of the problem $p_1 = \pi_{T^*Q} \circ \Psi_h^{(\rho)}(q_0, \tilde{p}_0)$.

(ii) Generate approximations $(q_{c_i}, p_{c_i}) \approx (q(c_i h), p(c_i h))$ via $(q_{c_i}, p_{c_i}) = \Psi_{c_i h}^{(\rho)}(q_0, \tilde{p}_0)$.

(iii) Approximate q_1 via $\tilde{q}_1 = \pi_Q \circ \Psi_h^{(\rho+1)}(q_0, \tilde{p}_0)$.

(iv) Use the continuous Legendre transform to obtain $\dot{q}_{c_i} = \frac{\partial H}{\partial p_{c_i}}$.

(v) Apply the quadrature rule to obtain the associated discrete right Hamiltonian,

$$H_d^+(q_0, p_1; h) = p_1^\top \tilde{q}_1 - h \sum_{i=1}^{m} b_i \left[p_{c_i}^\top \dot{q}_{c_i} - H(q_{c_i}, p_{c_i}) \right].$$

(vi) The variational integrator is then defined by the implicit discrete right Hamilton's equations,

$$q_1 = D_2 H_d^+(q_0, p_1), \qquad p_0 = D_1 H_d^+(q_0, p_1).$$

The following error analysis results were derived in [58] and [20]:

Theorem 2.1. *Suppose the Lagrangian L is Lipschitz continuous in both variables, and is sufficiently regular for the Taylor method $\Psi_h^{(\rho+1)}$ to be well-defined.*
Then $L_d(q_0, q_1)$ approximates $L_d^E(q_0, q_1)$ with at least order of accuracy $\min(\rho+1, s)$.
By Theorem 2.3.1 in [49], the associated discrete Hamiltonian map has the same order of accuracy.

Theorem 2.2. *Suppose the Hamiltonian H and its partial derivative $\frac{\partial H}{\partial p}$ are Lipschitz continuous in both variables, and H is sufficiently regular for the Taylor method $\Psi_h^{(\rho+1)}$ to be well-defined. Then $H_d^+(q_0,p_1)$ approximates $H_d^{+,E}(q_0,p_1)$ with at least order of accuracy $\min(\rho+1,s)$. By Theorem 2.2 in [57], the associated discrete right Hamilton's map has the same order of accuracy.*

Note that analogous constructions and error analysis results have been derived in [20; 58] for discrete left Hamiltonians in the Type III case.

2.3. **Time-adaptive Hamiltonian integrators via the Poincaré transformation.** Symplectic integrators form a class of geometric numerical integrators of interest since, when applied to conservative Hamiltonian systems, they yield discrete approximations of the flow that preserve the symplectic 2-form (see [26]), which results in the preservation of many qualitative aspects of the underlying system and exhibit excellent long-time near-energy preservation. However, when symplectic integrators were first used in combination with variable time-steps, the near-energy preservation was lost and the integrators performed poorly (see [7; 24]). Backward error analysis provided justification both for the excellent long-time near-energy preservation of symplectic integrators and for the poor performance experienced when using variable time-steps (see Chapter IX of [26]). Backward error analysis shows that symplectic integrators can be associated with a modified Hamiltonian in the form of a formal power series in terms of the time-step. The use of a variable time-step results in a different modified Hamiltonian at every iteration, which is the source of the poor energy conservation. The Poincaré transformation is one way to incorporate variable time-steps in geometric integrators without losing the nice conservation properties associated with these integrators.

Given a Hamiltonian $H(q,t,p)$, consider a desired transformation of time $t \mapsto \tau$ described by the monitor function $g(q,t,p)$ via

$$\frac{dt}{d\tau} = g(q,t,p). \tag{2.11}$$

The time t shall be referred to as the physical time, while τ will be referred to as the fictive time, and we will denote derivatives with respect to t and τ by dots and apostrophes, respectively. A new Hamiltonian system is constructed using the Poincaré transformation,

$$\bar{H}(\bar{q},\bar{p}) = g(q,\mathfrak{q},p)\left(H(q,\mathfrak{q},p) + \mathfrak{p}\right), \tag{2.12}$$

in the extended phase space defined by $\bar{q} = \left[\begin{smallmatrix} q \\ \mathfrak{q} \end{smallmatrix}\right] \in \bar{\mathcal{Q}}$ and $\bar{p} = \left[\begin{smallmatrix} p \\ \mathfrak{p} \end{smallmatrix}\right]$ where $\mathfrak{p}$ is the conjugate momentum for $\mathfrak{q} = t$ with $\mathfrak{p}(0) = -H(q(0),0,p(0))$. The corresponding equations of motion in the extended phase space are then given by

$$\bar{q}' = \frac{\partial \bar{H}}{\partial \bar{p}}, \qquad \bar{p}' = -\frac{\partial \bar{H}}{\partial \bar{q}}. \tag{2.13}$$

Suppose $(\bar{Q}(\tau),\bar{P}(\tau))$ are solutions to these extended equations of motion, and let $(q(t),p(t))$ solve Hamilton's equations for the original Hamiltonian H. Then

$$\bar{H}(\bar{Q}(\tau),\bar{P}(\tau)) = \bar{H}(\bar{Q}(0),\bar{P}(0)) = 0. \tag{2.14}$$

Therefore, the components $(Q(\tau),P(\tau))$ in the original phase space of the augmented solutions $(\bar{Q}(\tau),\bar{P}(\tau))$ satisfy

$$H(Q(\tau),\tau,P(\tau)) = -\mathfrak{p}(\tau), \qquad H(Q(0),0,P(0)) = -\mathfrak{p}(0) = H(q(0),0,p(0)). \tag{2.15}$$

Then, $(Q(\tau),P(\tau))$ and $(q(t),p(t))$ both satisfy Hamilton's equations for the original Hamiltonian H with the same initial values, so they must be the same. Note that the Hessian is given by

$$\frac{\partial^2 \bar{H}}{\partial \bar{p}^2} = \begin{bmatrix} \frac{\partial H}{\partial p}\nabla_p g(\bar{q},p)^\top + g(\bar{q},p)\frac{\partial^2 H}{\partial p^2} + \nabla_p g(\bar{q},p)\frac{\partial H}{\partial p}^\top & \nabla_p g(\bar{q},p) \\ \nabla_p g(\bar{q},p)^\top & 0 \end{bmatrix}, \tag{2.16}$$

which will be singular in many cases. The degeneracy of the Hamiltonian $\bar{H}$ implies that there is no corresponding Type I Lagrangian formulation. This approach works seamlessly with the existing methods and theorems for Hamiltonian variational integrators, but where the system under consideration is the transformed Hamiltonian system resulting from the Poincaré transformation. We can use a symplectic integrator with constant time-step in fictive time τ on the Poincaré transformed system, which will have the effect of integrating the original system with the desired variable time-step in physical time t via the relation $\frac{dt}{d\tau} = g(q,t,p)$.

3. Time-adaptive Lagrangian Integrators

The Poincaré transformation for time-adaptive symplectic integrators on the Hamiltonian side presented in Section 2.3 with autonomous monitor function $g(q,p)$ was first introduced in [69], and extended to the case where g can also depend on time based on ideas from [25]. All the literature to date on the Poincaré transformation have constructed the Poincaré transformed system by reverse-engineering: the Poincaré transformed Hamiltonian is chosen in such a way that the corresponding component dynamics satisfy Hamilton's equations in the original space.

3.1. **Variational Derivation of the Poincaré Hamiltonian.**

We now depart from the traditional reverse-engineering strategy for the Poincaré transformation and present a new way to think about the Poincaré transformed Hamiltonian by deriving it from a variational principle. This simple derivation gives additional insight into the transformation mechanism and provides natural candidates for time-adaptivity on the Lagrangian side and for more general frameworks.

As before, we work in the extended space $(q,\mathfrak{q},p,\mathfrak{p})$ where $\mathfrak{q}=t$ and $\mathfrak{p}$ is the corresponding conjugate momentum, and consider a time transformation $t\to\tau$ given by

$$\frac{dt}{d\tau} = g(q,t,p). \tag{3.1}$$

We define an extended action functional $\mathfrak{S}: C^2([0,T],T^*\bar{\mathcal{Q}})\to\mathbb{R}$ by

$$\begin{aligned}
\mathfrak{S}(\bar{q}(\cdot),\bar{p}(\cdot)) &= \bar{p}(T)\bar{q}(T) - \int_0^T \left[\bar{p}(t)\dot{\bar{q}}(t) - H(q(t),t,p(t)) - \mathfrak{p}(t)\right]dt & (3.2)\\
&= \bar{p}(T)\bar{q}(T) - \int_{\tau(t=0)}^{\tau(t=T)} \left[\bar{p}(\tau)\frac{d\tau}{dt}\bar{q}'(\tau) - H(q(\tau),\mathfrak{q}(\tau),p(\tau)) - \mathfrak{p}(\tau)\right]\frac{dt}{d\tau}d\tau & (3.3)\\
&= \bar{p}(T)\bar{q}(T) - \int_{\tau(t=0)}^{\tau(t=T)} \left\{\bar{p}(\tau)\bar{q}'(\tau) - \frac{dt}{d\tau}\left[H(q(\tau),\mathfrak{q}(\tau),p(\tau)) + \mathfrak{p}(\tau)\right]\right\}d\tau, & (3.4)
\end{aligned}$$

where we have performed a change of variables in the integral. Then,

$$\mathfrak{S}(\bar{q}(\cdot),\bar{p}(\cdot)) = \bar{p}(T)\bar{q}(T) - \int_{\tau(t=0)}^{\tau(t=T)} \left\{\bar{p}(\tau)\bar{q}'(\tau) - g(q(\tau),\mathfrak{q}(\tau),p(\tau))\left[H(q(\tau),\mathfrak{q}(\tau),p(\tau)) + \mathfrak{p}(\tau)\right]\right\}d\tau. \tag{3.5}$$

Computing the variation of $\mathfrak{S}$ yields

$$\begin{aligned}
\delta\mathfrak{S} = \bar{q}(T)\delta\bar{p}(T) + \bar{p}(T)\delta\bar{q}(T) &- \int_{\tau(t=0)}^{\tau(t=T)} \left[\mathfrak{q}'\delta\mathfrak{p} + \mathfrak{p}\delta\mathfrak{q}' - \left(g\frac{\partial H}{\partial\mathfrak{q}} + \frac{\partial g}{\partial\mathfrak{q}}(H+\mathfrak{q})\right)\delta\mathfrak{q} - g\delta\mathfrak{p}\right]d\tau\\
&- \int_{\tau(t=0)}^{\tau(t=T)} \left[q'\delta p + p\delta q' - \left(g\frac{\partial H}{\partial q} + \frac{\partial g}{\partial q}(H+\mathfrak{p})\right)\delta q - \left(g\frac{\partial H}{\partial p} + \frac{\partial g}{\partial p}(H+\mathfrak{p})\right)\delta p\right]d\tau,
\end{aligned}$$

and using integration by parts and the boundary conditions $\delta\bar{q}(0)=0$ and $\delta\bar{p}(T)=0$, gives

$$\begin{aligned}
\delta\mathfrak{S} = &\int_{\tau(t=0)}^{\tau(t=T)} \left[p' + g\frac{\partial H}{\partial q} + \frac{\partial g}{\partial q}(H+\mathfrak{p})\right]\delta q\, d\tau + \int_{\tau(t=0)}^{\tau(t=T)} \left[g\frac{\partial H}{\partial p} + \frac{\partial g}{\partial p}(H+\mathfrak{p}) - q'\right]\delta p\, d\tau\\
&+ \int_{\tau(t=0)}^{\tau(t=T)} \left[\mathfrak{p}' + g\frac{\partial H}{\partial\mathfrak{q}} + \frac{\partial g}{\partial\mathfrak{q}}(H+\mathfrak{p})\right]\delta\mathfrak{q}\, d\tau - \int_{\tau(t=0)}^{\tau(t=T)} \left[\mathfrak{q}' - g\right]\delta\mathfrak{p}\, d\tau.
\end{aligned}$$

Thus, the condition that $\mathfrak{S}(\bar{q}(\cdot),\bar{p}(\cdot))$ is stationary with respect to the boundary conditions $\delta\bar{q}(0)=0$ and $\delta\bar{p}(T)=0$ is equivalent to $(\bar{q}(\cdot),\bar{p}(\cdot))$ satisfying Hamilton's canonical equations corresponding to the Poincaré transformed Hamiltonian,

$$\mathfrak{q}' = g(q,\mathfrak{q},p), \tag{3.6}$$

$$q' = g(q,\mathfrak{q},p)\frac{\partial H}{\partial p}(q,\mathfrak{q},p) + \frac{\partial g}{\partial p}(q,\mathfrak{q},p)\left[H(q,\mathfrak{q},p)+\mathfrak{p}\right], \tag{3.7}$$

$$p' = -g(q,\mathfrak{q},p)\frac{\partial H}{\partial q}(q,\mathfrak{q},p) - \frac{\partial g}{\partial q}(q,\mathfrak{q},p)\left[H(q,\mathfrak{q},p)+\mathfrak{p}\right], \tag{3.8}$$

$$\mathfrak{p}' = -g(q,\mathfrak{q},p)\frac{\partial H}{\partial \mathfrak{q}}(q,\mathfrak{q},p) - \frac{\partial g}{\partial \mathfrak{q}}(q,\mathfrak{q},p)\left[H(q,\mathfrak{q},p)+\mathfrak{p}\right]. \tag{3.9}$$

An alternative way to reach the same conclusion is by interpreting equation (3.5) as the usual Type II action functional for the modified Hamiltonian,

$$g(q(\tau),\mathfrak{q}(\tau),p(\tau))\left[H(q(\tau),\mathfrak{q}(\tau),p(\tau))+\mathfrak{p}(\tau)\right], \tag{3.10}$$

which coincides with the Poincaré transformed Hamiltonian.

3.2. **Time-adaptivity from a Variational Principle on the Lagrangian side.** We will now derive a mechanism for time-adaptivity on the Lagrangian side by mimicking the derivation of the Poincaré Hamiltonian. We will work in the extended space $\bar{q}=(q,\mathfrak{q},\lambda)^\top \in \bar{\mathcal{Q}}$ where $\mathfrak{q}=t$ and λ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau}=g(t)$. Consider the action functional $\mathfrak{S}: C^2([0,T],T\bar{\mathcal{Q}})\to\mathbb{R}$ given by

$$\begin{aligned}
\mathfrak{S}(\bar{q}(\cdot),\dot{\bar{q}}(\cdot)) &= \int_0^T \left[L(q(t),\dot{q}(t),\mathfrak{q}(t)) - \lambda(t)\left(\frac{d\mathfrak{q}}{d\tau}-g(\mathfrak{q}(t))\right)\right]dt && (3.11)\\
&= \int_{\tau(t=0)}^{\tau(t=T)}\left[\frac{dt}{d\tau}L\left(q(\tau),\frac{d\tau}{dt}q'(\tau),\mathfrak{q}(\tau)\right) - \lambda(\tau)\frac{dt}{d\tau}\left(\frac{d\mathfrak{q}}{d\tau}-g(\mathfrak{q}(\tau))\right)\right]d\tau && (3.12)\\
&= \int_{\tau(t=0)}^{\tau(t=T)}\left[\mathfrak{q}'(\tau)L\left(q(\tau),\frac{d\tau}{dt}q'(\tau),\mathfrak{q}(\tau)\right) - \lambda(\tau)\mathfrak{q}'(\tau)\left[\mathfrak{q}'(\tau)-g(\mathfrak{q}(\tau))\right]\right]d\tau, && (3.13)
\end{aligned}$$

where, as before, we have performed a change of variables in the integral. This is the usual Type I action functional for the extended autonomous Lagrangian,

$$\bar{L}(\bar{q}(\tau),\bar{q}'(\tau)) = \mathfrak{q}'(\tau)L\left(q(\tau),\frac{d\tau}{dt}q'(\tau),\mathfrak{q}(\tau)\right) - \lambda(\tau)\mathfrak{q}'(\tau)\left[\mathfrak{q}'(\tau)-g(\mathfrak{q}(\tau))\right]. \tag{3.14}$$

Theorem 3.1. *If $(\bar{q}(\tau),\bar{q}'(\tau))$ satisfies the Euler–Lagrange equations corresponding to the Lagrangian $\bar{L}$, then its components satisfy $\frac{dt}{d\tau}=g(t)$ and the original Euler–Lagrange equations*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q,\dot{q},t) = \frac{\partial L}{\partial q}(q,\dot{q},t). \tag{3.15}$$

Proof. Substituting the expression for $\bar{L}$ into the Euler–Lagrange equations, $\frac{d}{d\tau}\frac{\partial\bar{L}}{\partial\lambda'}=\frac{\partial\bar{L}}{\partial\lambda}$, and $\frac{d}{d\tau}\frac{\partial\bar{L}}{\partial q'}=\frac{\partial\bar{L}}{\partial q}$, gives

$$\mathfrak{q}'\left[\mathfrak{q}'-g(\mathfrak{q})\right]=0,$$

and

$$\frac{d\mathfrak{q}}{d\tau}\frac{d}{d\mathfrak{q}}\left[\mathfrak{q}'\frac{\partial L\left(q,\frac{d\tau}{d\mathfrak{q}}q',\mathfrak{q}\right)}{\partial q'}\right] = \mathfrak{q}'\frac{\partial L\left(q,\frac{d\tau}{d\mathfrak{q}}q',\mathfrak{q}\right)}{\partial q}.$$

Now, $\mathfrak{q}' = g(\mathfrak{q}) > 0$ so $\mathfrak{q}' = g(\mathfrak{q})$, and the chain rule gives

$$\frac{d}{d\mathfrak{q}}\frac{\partial L}{\partial \dot{q}}\left(q, \frac{d\tau}{d\mathfrak{q}}q', \mathfrak{q}\right) = \frac{\partial L}{\partial q}\left(q, \frac{d\tau}{d\mathfrak{q}}q', \mathfrak{q}\right).$$

Using the equation $\dot{q} = \frac{d\tau}{d\mathfrak{q}}q'$ and replacing $\mathfrak{q}$ by t recovers the original Euler–Lagrange equations. □

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics. Suppose we are given a partition $0 = \tau_0 < \tau_1 < \ldots < \tau_N = \mathcal{T}$ of the interval $[0, \mathcal{T}]$, and a discrete curve in $\mathcal{Q} \times \mathbb{R} \times \mathbb{R}$ denoted by $\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N$ such that $q_k \approx q(\tau_k)$, $\mathfrak{q}_k \approx \mathfrak{q}(\tau_k)$, and $\lambda_k \approx \lambda(\tau_k)$. Consider the discrete action functional,

$$\bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1}\left[L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k}, \tag{3.16}$$

where $L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$ is obtained by approximating the exact discrete Lagrangian, which is related to Jacobi's solution of the Hamilton–Jacobi equation and is the generating function for the exact time-h flow map. It is given by the extremum of the action integral from τ_k to τ_{k+1} over twice continuously differentiable curves $(q, \mathfrak{q}) \in \mathcal{Q} \times \mathbb{R}$ satisfying the boundary conditions $(q(\tau_k), \mathfrak{q}(\tau_k)) = (q_k, \mathfrak{q}_k)$, and $(q(\tau_{k+1}), \mathfrak{q}(\tau_{k+1})) = (q_{k+1}, \mathfrak{q}_{k+1})$:

$$L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) \approx \underset{\substack{(q,\mathfrak{q}) \in C^2([\tau_k, \tau_{k+1}], \mathcal{Q}\times\mathbb{R}) \\ (q,\mathfrak{q})(\tau_k) = (q_k, \mathfrak{q}_k),\ (q,\mathfrak{q})(\tau_{k+1}) = (q_{k+1}, \mathfrak{q}_{k+1})}}{\operatorname{ext}} \int_{\tau_k}^{\tau_{k+1}} L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right) d\tau. \tag{3.17}$$

In practice, we can obtain an approximation by replacing the integral with a quadrature rule, and extremizing over a finite-dimensional function space instead of $C^2([\tau_k, \tau_{k+1}], \mathcal{Q} \times \mathbb{R})$. This discrete functional $\bar{\mathfrak{S}}_d$ is a discrete analogue of the action functional $\bar{\mathfrak{S}} : C^2([0, T], \mathcal{Q} \times \mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ given by

$$\begin{aligned}\bar{\mathfrak{S}}(q(\cdot), \mathfrak{q}(\cdot), \lambda(\cdot)) &= \int_0^{\mathcal{T}} \bar{L}\left(q(\tau), \mathfrak{q}(\tau), \lambda(\tau), q'(\tau), \mathfrak{q}'(\tau), \lambda'(\tau)\right) d\tau && (3.18)\\ &= \int_0^{\mathcal{T}} \left[L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right) - \lambda\mathfrak{q}' + \lambda g(\mathfrak{q})\right]\mathfrak{q}' d\tau. && (3.19)\end{aligned}$$

We can derive the following result which relates a discrete Type I variational principle to a set of discrete Euler–Lagrange equations:

Theorem 3.2. *The Type I discrete Hamilton's variational principle,*

$$\delta\bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = 0, \tag{3.20}$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1} - \tau_k) g(\mathfrak{q}_k), \tag{3.21}$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, \mathfrak{q}_{k-1}, q_k, \mathfrak{q}_k) = 0, \tag{3.22}$$

$$\begin{aligned}&\left[D_2 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} - \frac{1}{\tau_{k+1} - \tau_k}\left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k)\right] \\ &+ \left[D_4 L_{d_{k-1}} - \lambda_{k-1}\frac{1}{\tau_k - \tau_{k-1}}\right]\frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \frac{1}{\tau_k - \tau_{k-1}}\left[L_{d_{k-1}} - \lambda_{k-1}\frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \lambda_{k-1} g(\mathfrak{q}_{k-1})\right] = 0,\end{aligned} \tag{3.23}$$

where L_{d_k} denotes $L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$.

Proof. See Appendix A.1. □

Defining the discrete momenta via the discrete Legendre transformations,

$$p_k = -D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \qquad \mathfrak{p}_k = -D_2 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.24}$$

and using a constant time-step h in τ, the discrete Euler–Lagrange equations can be rewritten as

$$p_k = -D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.25}$$

$$\mathfrak{p}_k = -D_2 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.26}$$

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + hg(\mathfrak{q}_k), \tag{3.27}$$

$$p_{k+1} = \frac{g(\mathfrak{q}_k)}{g(\mathfrak{q}_{k+1})} D_3 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.28}$$

$$\mathfrak{p}_{k+1} = \frac{L_{d_k} - L_{d_{k+1}}}{hg(\mathfrak{q}_{k+1})} + \frac{\lambda_{k+1}}{h} + \lambda_{k+1} \nabla g(\mathfrak{q}_{k+1}) + \frac{g(\mathfrak{q}_k)}{g(\mathfrak{q}_{k+1})} \left[D_4 L_{d_k} - \frac{\lambda_k}{h} \right]. \tag{3.29}$$

3.3. A Second Time-Adaptive Framework obtained by Reverse-Engineering.

As mentioned earlier, all the literature to date on the Poincaré transformation have constructed the Poincaré transformed system by reverse-engineering. The Poincaré transformed Hamiltonian is chosen in such a way that the corresponding component dynamics satisfy the Hamilton's equations in the original space. We will follow a similar strategy to derive a second framework for time-adaptivity from the Lagrangian perspective.

Given a time-dependent Lagrangian $L(q(t), \dot{q}(t), t)$ consider a transformation of time $t \to \tau$,

$$\frac{dt}{d\tau} = g(t), \tag{3.30}$$

described by the monitor function $g(t)$. The time t shall be referred to as the physical time, while τ will be referred to as the fictive time, and we will denote derivatives with respect to t and τ by dots and apostrophes, respectively. We define the autonomous Lagrangian,

$$\bar{L}(\bar{q}(\tau), \bar{q}'(\tau)) = \mathfrak{q}' L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right) - \lambda\left(\mathfrak{q}' - g(\mathfrak{q})\right), \tag{3.31}$$

in the extended space with $\bar{q} = (q, \mathfrak{q}, \lambda)^\top$ where $\mathfrak{q} = t$, and where λ is a multiplier used to impose the constraint that the time evolution is guided by the monitor function $g(t)$. Note that in contrast to the earlier framework, the Lagrange multiplier term lacks an extra multiplicative factor of $\mathfrak{q}'$.

Theorem 3.3. *If $(\bar{q}(\tau), \bar{q}'(\tau))$ satisfies the Euler–Lagrange equations corresponding to the Lagrangian $\bar{L}$, then its components satisfy $\frac{dt}{d\tau} = g(t)$ and the original Euler–Lagrange equations*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q, \dot{q}, t) = \frac{\partial L}{\partial q}(q, \dot{q}, t). \tag{3.32}$$

Proof. Substituting the expression for $\bar{L}$ in the Euler–Lagrange equations $\frac{d}{d\tau}\frac{\partial \bar{L}}{\partial \lambda'} = \frac{\partial \bar{L}}{\partial \lambda}$ and $\frac{d}{d\tau}\frac{\partial \bar{L}}{\partial q'} = \frac{\partial \bar{L}}{\partial q}$ gives

$$\mathfrak{q}' = g(\mathfrak{q}),$$

and

$$\frac{d\mathfrak{q}}{d\tau}\frac{d}{d\mathfrak{q}}\left[\frac{d\mathfrak{q}}{d\tau}\frac{\partial L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right)}{\partial q'}\right] = \frac{d\mathfrak{q}}{d\tau}\frac{\partial L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right)}{\partial q}.$$

We can divide by $\frac{d\mathfrak{q}}{d\tau}$ and use the chain rule to get $\mathfrak{q}' = g(\mathfrak{q})$ and

$$\frac{d}{d\mathfrak{q}}\frac{\partial L}{\partial \dot{q}}\left(q, \frac{d\tau}{d\mathfrak{q}}q', \mathfrak{q}\right) = \frac{\partial L}{\partial q}\left(q, \frac{d\tau}{d\mathfrak{q}}q', \mathfrak{q}\right).$$

Using the equations $\dot{q} = \frac{d\tau}{d\mathfrak{q}} q'$ and $\mathfrak{q}' = g(\mathfrak{q})$, and replacing $\mathfrak{q}$ by t recovers the desired equations. □

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics. Suppose we are given a partition $0 = \tau_0 < \tau_1 < \ldots < \tau_N = \mathcal{T}$ of the interval $[0, \mathcal{T}]$, and a discrete curve in $\mathcal{Q} \times \mathbb{R} \times \mathbb{R}$ denoted by $\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N$ such that $q_k \approx q(\tau_k)$, $\mathfrak{q}_k \approx \mathfrak{q}(\tau_k)$, and $\lambda_k \approx \lambda(\tau_k)$. Consider the discrete action functional,

$$\bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1} \left\{ \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \left[L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) - \lambda_k \right] + \lambda_k g(\mathfrak{q}_k) \right\}, \tag{3.33}$$

where,

$$L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) \approx \underset{\substack{(q,\mathfrak{q}) \in C^2([\tau_k, \tau_{k+1}], \mathcal{Q} \times \mathbb{R}) \\ (q,\mathfrak{q})(\tau_k) = (q_k, \mathfrak{q}_k),\ (q,\mathfrak{q})(\tau_{k+1}) = (q_{k+1}, \mathfrak{q}_{k+1})}}{\operatorname{ext}} \int_{\tau_k}^{\tau_{k+1}} L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right) d\tau. \tag{3.34}$$

This discrete functional $\bar{\mathfrak{S}}_d$ is a discrete analogue of the action functional $\bar{\mathfrak{S}} : C^2([0, T], \mathcal{Q} \times \mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ given by

$$\bar{\mathfrak{S}}(q(\cdot), \mathfrak{q}(\cdot), \lambda(\cdot)) = \int_0^{\mathcal{T}} \bar{L}\left(q(\tau), \mathfrak{q}(\tau), \lambda(\tau), q'(\tau), \mathfrak{q}'(\tau), \lambda'(\tau)\right) d\tau \tag{3.35}$$

$$= \int_0^{\mathcal{T}} \left\{ \mathfrak{q}' \left[L\left(q, \frac{q'}{g(\mathfrak{q})}, \mathfrak{q}\right) - \lambda \right] + \lambda g(\mathfrak{q}) \right\} d\tau. \tag{3.36}$$

We can derive the following result which relates a discrete Type I variational principle to a set of discrete Euler–Lagrange equations:

Theorem 3.4. *The Type I discrete Hamilton's variational principle,*

$$\delta \bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = 0, \tag{3.37}$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1} - \tau_k) g(\mathfrak{q}_k), \tag{3.38}$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, \mathfrak{q}_{k-1}, q_k, \mathfrak{q}_k) = 0, \tag{3.39}$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_2 L_{d_k} - \frac{L_{d_k}}{\tau_{k+1} - \tau_k} + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_4 L_{d_{k-1}} + \frac{L_{d_{k-1}}}{\tau_k - \tau_{k-1}} = \frac{\lambda_{k-1}}{\tau_k - \tau_{k-1}} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} - \lambda_k \nabla g(\mathfrak{q}_k), \tag{3.40}$$

where L_{d_k} denotes $L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$.

Proof. See Appendix A.2. □

Defining the discrete momenta via the discrete Legendre transformations,

$$p_k = -D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \qquad \mathfrak{p}_k = -D_2 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.41}$$

and using a constant time-step h in τ, the discrete Euler–Lagrange equations can be rewritten as

$$p_k = -D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.42}$$

$$\mathfrak{p}_k = -D_2 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.43}$$

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + h g(\mathfrak{q}_k), \tag{3.44}$$

$$p_{k+1} = \frac{g(\mathfrak{q}_k)}{g(\mathfrak{q}_{k+1})} D_3 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{3.45}$$

$$\mathfrak{p}_{k+1} = \frac{L_{d_k} - L_{d_{k+1}} + \lambda_{k+1} - \lambda_k + h \lambda_{k+1} \nabla g(\mathfrak{q}_{k+1}) + h g(\mathfrak{q}_k) D_4 L_{d_k}}{h g(\mathfrak{q}_{k+1})}. \tag{3.46}$$

3.4. **Remarks on the Framework for Time-Adaptivity.** Time-adaptivity comes more naturally on the Hamiltonian side through the Poincaré transformation. Indeed, in the Hamiltonian case, the time-rescaling equation $\mathfrak{q}' = g(q, \mathfrak{q}, p)$ emerged naturally through the change of time variable inside the extended action functional. By contrast, in the Lagrangian case, we need to impose the time-rescaling equation as a constraint via a multiplier, which we then consider as an extra position coordinate. This strategy can be thought of as being part of the more general framework for constrained variational integrators (see [16; 49]).

The Poincaré transformation on the Hamiltonian side was presented in [20; 25; 69] for the general case where the monitor function can depend on position, time and momentum, $g = g(q, t, p)$. For the accelerated optimization application which was our main motivation to develop a time-adaptive framework for geometric integrators, the monitor function only depends on time, $g = g(t)$. For the sake of simplicity and clarity, we have decided to only present the theory for time-adaptive Lagrangian integrators for monitor functions of the form $g = g(t)$ in this paper. Note however that this time-adaptivity framework on the Lagrangian side can be extended to the case where the monitor function also depends on position, $g = g(q, t)$. The action integral remains the same with the exception that g is now a function of $(q, \mathfrak{q})$. Unlike the case where $g = g(t)$, the corresponding Euler–Lagrange equation $\frac{d}{d\tau}\frac{\partial \bar{L}}{\partial q'} = \frac{\partial L}{\partial q}$ yields an extra term $\lambda(t)\frac{\partial g}{\partial q}(q, t)$ in the original phase-space,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q, \dot{q}, t) - \frac{\partial L}{\partial q}(q, \dot{q}, t) = \lambda(t)\frac{\partial g}{\partial q}(q, t). \tag{3.47}$$

The discrete Euler–Lagrange equations become more complicated and involve terms with partial derivatives $D_1 g(q_k, \mathfrak{q}_k)$ of g with respect to q. Furthermore, when $g = g(q, t)$, the discrete Euler–Lagrange equations involve λ_k but the time-evolution of the Lagrange multiplier λ is not well-defined, so the discrete Hamiltonian map corresponding to the discrete Lagrangian L_d is not well-defined, as explained in [49, page 440]. Although there are ways to circumvent this problem in practice, this adds some difficulty and makes the time-adaptive Lagrangian approach with $g = g(q, t)$ less natural and desirable than the corresponding Poincaré transformation on the Hamiltonian side. It might also be tempting to generalize further and consider the case where $g = g(q, \dot{q}, t)$. However, in this case, the time-rescaling equation $\frac{dt}{d\tau} = g(q, \dot{q}, t)$ becomes implicit and it becomes less clear how to generalize the variational derivation presented in this paper. There are examples where time-adaptivity with these more general monitor functions proved advantageous (see for instance Kepler's problem in [20]). This motivates further effort towards developing a better framework for time-adaptivity on the Lagrangian side with more general monitor functions.

It might be more natural to consider these time-rescaled Lagrangian and Hamiltonian dynamics as Dirac mechanics [44; 66; 67] on the Pontryagin bundle $(q, v, p) \in T\mathcal{Q} \oplus T^*\mathcal{Q}$. Dirac dynamics are described by the Hamilton-Pontryagin variational principle where the momentum p acts as a Lagrange multiplier to impose the kinematic equation $\dot{q} = v$,

$$\delta \int_0^T \left[L(q, v, t) + p(\dot{q} - v) \right] dt = 0. \tag{3.48}$$

This provides a variational description of both Lagrangian and Hamiltonian mechanics, yields the implicit Euler–Lagrange equations

$$\dot{q} = v, \qquad \dot{p} = \frac{\partial L}{\partial q}, \qquad p = \frac{\partial L}{\partial v}, \tag{3.49}$$

and suggests the introduction of a more general quantity, the generalized energy

$$E(q, v, p, t) = pv - L(q, v, t), \tag{3.50}$$

as an alternative to the Hamiltonian.

4. Application to Accelerated Optimization on Vector Spaces

4.1. **A Variational Framework for Accelerated Optimization.** A variational framework was introduced in [65] for accelerated optimization on normed vector spaces. The p-Bregman Lagrangians and Hamiltonians are defined to be

$$\mathcal{L}_p(x,v,t) = \frac{t^{p+1}}{2p}\langle v,v\rangle - Cpt^{2p-1}f(x), \tag{4.1}$$

$$\mathcal{H}_p(x,r,t) = \frac{p}{2t^{p+1}}\langle r,r\rangle + Cpt^{2p-1}f(x), \tag{4.2}$$

which are scalar-valued functions of position $x \in \mathcal{X}$, velocity $v \in \mathbb{R}^d$ or momentum $r \in \mathbb{R}^d$, and time t. In [65], it was shown that solutions to the p-Bregman Euler–Lagrange equations converge to a minimizer of f at a convergence rate of $\mathcal{O}(1/t^p)$. Furthermore, this family of Bregman dynamics is closed under time dilation: time-rescaling a solution to the p-Bregman Euler–Lagrange equations via $\tau(t) = t^{\mathring{p}/p}$ yields a solution to the $\mathring{p}$-Bregman Euler–Lagrange equations. Thus, the entire subfamily of Bregman trajectories indexed by the parameter p can be obtained by speeding up or slowing down along the Bregman curve corresponding to any value of p. In [20], the time-rescaling property of the Bregman dynamics was exploited together with a carefully chosen Poincaré transformation to transform the p-Bregman Hamiltonian into an autonomous version of the $\mathring{p}$-Bregman Hamiltonian in extended phase-space, where $\mathring{p} < p$. This strategy allowed us to achieve the faster rate of convergence associated with the higher-order p-Bregman dynamics, but with the computational efficiency of integrating the lower-order $\mathring{p}$-Bregman dynamics. Explicitly, using the time rescaling $\tau(t) = t^{\mathring{p}/p}$ within the Poincaré transformation framework yields the adaptive approach $p \to \mathring{p}$-Bregman Hamiltonian,

$$\bar{H}_{p\to\mathring{p}}(\bar{q},\bar{r}) = \frac{p^2}{2\mathring{p}\mathfrak{q}^{p+\mathring{p}/p}}\langle r,r\rangle + \frac{Cp^2}{\mathring{p}}\mathfrak{q}^{2p-\mathring{p}/p}f(q) + \frac{p}{\mathring{p}}\mathfrak{r}\mathfrak{q}^{1-\mathring{p}/p}, \tag{4.3}$$

and when $\mathring{p} = p$, the direct approach p-Bregman Hamiltonian,

$$\bar{H}_p(\bar{q},\bar{r}) = \frac{p}{2\mathfrak{q}^{p+1}}\langle r,r\rangle + Cp\mathfrak{q}^{2p-1}f(q) + \mathfrak{r}. \tag{4.4}$$

In [20], a careful computational study was performed on how time-adaptivity and symplecticity of the numerical scheme improve the performance of the resulting optimization algorithm. In particular, it was observed that time-adaptive Hamiltonian variational discretizations, which are automatically symplectic, with adaptive time-steps informed by the time-rescaling of the family of p-Bregman Hamiltonians yielded the most robust and computationally efficient optimization algorithms, outperforming fixed-timestep symplectic discretizations, adaptive-timestep non-symplectic discretizations, and Nesterov's accelerated gradient algorithm which is neither time-adaptive nor symplectic.

4.2. Numerical Methods.

4.2.1. *A Lagrangian Taylor Variational Integrator (LTVI).* We will now construct a time-adaptive Lagrangian Taylor variational integrator (LTVI) for the p-Bregman Lagrangian,

$$\bar{L}_p\left(q,q',\mathfrak{q}\right) = \frac{\mathfrak{q}^{p+1}}{2p}\langle q',q'\rangle - Cp\mathfrak{q}^{2p-1}f(q), \tag{4.5}$$

using the strategy outlined in Section 2.2 together with the discrete Euler–Lagrange equations derived in Sections 3.2 and 3.3.

Looking at the form of the continuous p-Bregman Euler–Lagrange equations,

$$\ddot{q} + \frac{p+1}{\mathfrak{q}}\dot{q} + Cp^2\mathfrak{q}^{p-2}\nabla f(q) = 0, \tag{4.6}$$

we can note that ∇f appears in the expression for $\ddot{q}$. Now, the construction of a LTVI as presented in Section 2.2 requires ρ-order and $(\rho+1)$-order Taylor approximations of q. This means that if we take $\rho \geq 1$, then ∇f and higher-order derivatives of f will appear in the resulting discrete Lagrangian L_d, and as a consequence, the discrete Euler–Lagrange equations,

$$p_0 = -D_1 L_d(q_0, q_1), \qquad p_1 = D_2 L_d(q_0, q_1), \tag{4.7}$$

will yield an integrator which is not gradient-based. Keeping in mind the machine learning applications where data sets are very large, we will restrict ourselves to explicit first-order optimization algorithms, and therefore the highest value of ρ that we can choose to obtain a gradient-based algorithm is $\rho = 0$.

With $\rho = 0$, the choice of quadrature rule does not matter, so we can take the rectangular quadrature rule about the initial point ($c_1 = 0$ and $b_1 = 1$). We first approximate $\dot{\bar{q}}(0) = \bar{v}_0$ by the solution $\tilde{\bar{v}}_0$ of the problem $\bar{q}_1 = \pi_Q \circ \Psi_h^{(1)}(\bar{q}_0, \tilde{\bar{v}}_0) = \bar{q}_0 + h\tilde{\bar{v}}_0$, that is $\tilde{\bar{v}}_0 = \frac{\bar{q}_1 - \bar{q}_0}{h}$. Then, applying the quadrature rule gives the associated discrete Lagrangian,

$$L_d(\bar{q}_0, \bar{q}_1) = h\bar{L}_p\left(\bar{q}_0, \frac{1}{g(\mathfrak{q}_0)}\tilde{\bar{v}}_0\right) = \frac{\mathfrak{q}_0^{p+1}}{2p(g(\mathfrak{q}_0))^2} h\langle \tilde{v}_0, \tilde{v}_0\rangle - Chp\mathfrak{q}_0^{2p-1} f(q_0). \tag{4.8}$$

The variational integrator is then defined by the discrete extended Euler–Lagrange equations derived in Sections 3.2 and 3.3. In practice, we are not interested in the evolution of the conjugate momentum $\mathfrak{r}$, and since it will not appear in the updates for the other variables, the discrete equations of motion from Sections 3.2 and 3.3 both reduce to the same updates,

$$r_k = -D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{4.9}$$

$$r_{k+1} = \frac{g(\mathfrak{q}_k)}{g(\mathfrak{q}_{k+1})} D_3 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}), \tag{4.10}$$

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + hg(\mathfrak{q}_k). \tag{4.11}$$

Now, for the adaptive approach, substituting $g(\mathfrak{q}) = \frac{p}{\mathring{p}}\mathfrak{q}^{1-\frac{\mathring{p}}{p}}$ and

$$L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) = \frac{\mathring{p}^2}{2hp^3}\mathfrak{q}_k^{p-1+2\mathring{p}/p}\langle q_{k+1} - q_k, q_{k+1} - q_k\rangle - Chp\mathfrak{q}_0^{2p-1} f(q_k), \tag{4.12}$$

yields the adaptive LTVI algorithm,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + h\frac{p}{\mathring{p}}\mathfrak{q}_k^{1-\mathring{p}/p}, \tag{4.13}$$

$$q_{k+1} = q_k + \frac{hp^3}{\mathring{p}^2\mathfrak{q}_k^{p-1+2\mathring{p}/p}} r_k - \frac{Ch^2p^4}{\mathring{p}^2}\mathfrak{q}_k^{p-2\mathring{p}/p}\nabla f(q_k), \tag{4.14}$$

$$r_{k+1} = \frac{\mathring{p}^2\mathfrak{q}_k^{p+\mathring{p}/p}}{hp^3\mathfrak{q}_{k+1}^{1-\mathring{p}/p}}(q_{k+1} - q_k). \tag{4.15}$$

In the direct approach, $\mathring{p} = p$ so $g(\mathfrak{q}) = 1$ and we obtain the direct LTVI algorithm,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + h, \tag{4.16}$$

$$q_{k+1} = q_k + \frac{hp}{\mathfrak{q}_k^{p+1}} r_k - Ch^2p^2\mathfrak{q}_k^{p-2}\nabla f(q_k), \tag{4.17}$$

$$r_{k+1} = \frac{\mathfrak{q}_k^{p+1}}{hp}(q_{k+1} - q_k). \tag{4.18}$$

4.2.2. *A Hamiltonian Taylor Variational Integrator (HTVI).* In [20], a Type II Hamiltonian Taylor Variational Integrator (HTVI) was derived following the strategy from Section 2.2 with $\rho = 0$ for the adaptive approach $p \to \mathring{p}$-Bregman Hamiltonian,

$$\bar{H}_{p\to\mathring{p}}(\bar{q},\bar{r}) = \frac{p^2}{2\mathring{p}\mathfrak{q}^{p+\mathring{p}/p}}\langle r,r\rangle + \frac{Cp^2}{\mathring{p}}\mathfrak{q}^{2p-\mathring{p}/p}f(q) + \frac{p}{\mathring{p}}\mathfrak{r}\mathfrak{q}^{1-\mathring{p}/p}. \tag{4.19}$$

This adaptive HTVI is the most natural Hamiltonian analogue of the LTVI described in Section 4.2.1, and its updates are given by

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + h\frac{p}{\mathring{p}}\mathfrak{q}_k^{1-\mathring{p}/p}, \tag{4.20}$$

$$r_{k+1} = r_k - \frac{p^2}{\mathring{p}}Ch\mathfrak{q}_k^{2p-\mathring{p}/p}\nabla f(q_k), \tag{4.21}$$

$$q_{k+1} = q_k + \frac{p^2}{\mathring{p}}h\mathfrak{q}_k^{-p-\mathring{p}/p}r_{k+1}. \tag{4.22}$$

When $\mathring{p} = p$, it reduces to the direct HTVI,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + h, \tag{4.23}$$

$$r_{k+1} = r_k - hCp\mathfrak{q}_k^{2p-1}\nabla f(q_k), \tag{4.24}$$

$$q_{k+1} = q_k + hp\mathfrak{q}_k^{-p-1}r_{k+1}. \tag{4.25}$$

4.3. **Numerical Experiments.** Numerical experiments using the numerical methods presented in the previous section have been conducted to minimize the quartic function,

$$f(x) = \left[(x-1)^\top\Sigma(x-1)\right]^2, \tag{4.26}$$

where $x \in \mathbb{R}^d$ and $\Sigma_{ij} = 0.9^{|i-j|}$. This convex function achieves its minimum value 0 at $x^* = 1$.

As was observed in [20] for the HTVI algorithm, the numerical experiments showed that a carefully tuned adaptive approach algorithm enjoyed a significantly better rate of convergence and required a much smaller number of steps to achieve convergence than the direct approach, as can be seen in Figure 1 for the LTVI methods. Although the adaptive approach requires a smaller fictive time-step h than the direct approach, the physical time-steps resulting from $t = \tau^{p/\mathring{p}}$ in the adaptive approach grow rapidly to values larger than the physical time-step of the direct approach. The results of Figure 1 also show that the adaptive and direct LTVI methods become more and more efficient as p is increased, which was also the case for the HTVI algorithm in [20].

The LTVI and HTVI algorithms presented in Section 4.2 perform empirically almost exactly in the same way for the same parameters and time-step, as can be seen for instance in Figure 2. As a result, the computational analysis carried in [20] for the HTVI algorithm extends to the LTVI algorithm. In particular, it was shown in [20] that the HTVI algorithm is much more efficient than non-adaptive non-symplectic and adaptive non-symplectic integrators for the Bregman dynamics, and that it can be a competitive first-order explicit algorithm which can outperform certain popular optimization algorithms such as Nesterov's Accelerated Gradient [54], Trust Region Steepest Descent, ADAM [33], AdaGrad [15], and RMSprop [61], for certain choices of objective functions. Since the computational performance of the LTVI algorithm is almost exactly the same as that of the HTVI algorithm, this means that the LTVI algorithm is also much more efficient than non-symplectic integrators for the Bregman dynamics and can also be very competitive as a first-order explicit optimization algorithm.

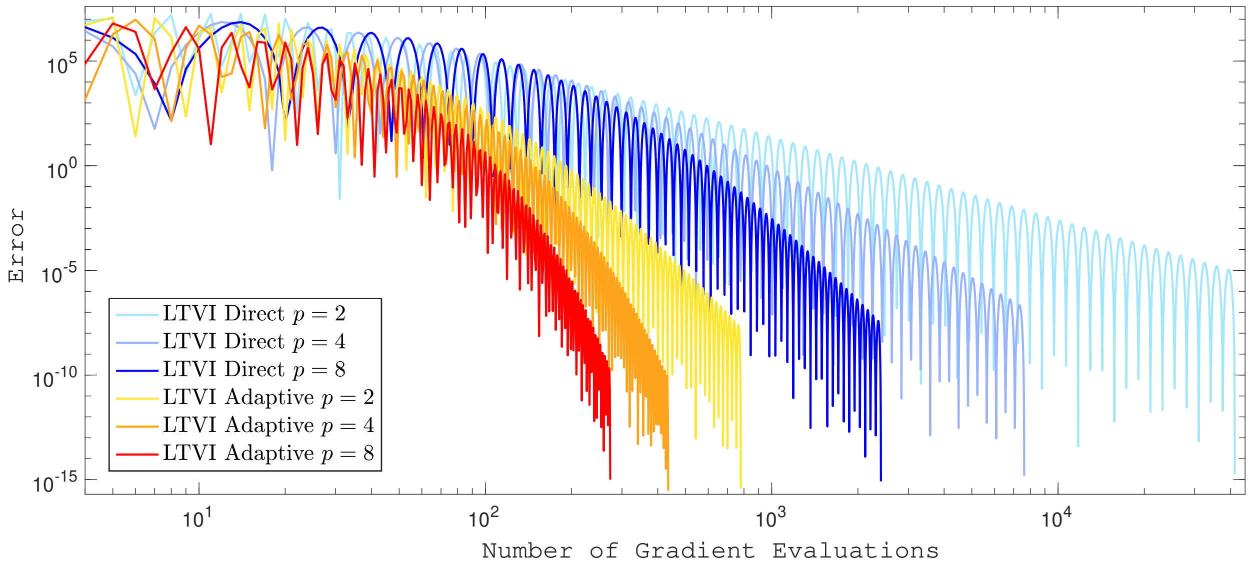


FIGURE 1. Comparison of the direct and adaptive approaches for the LTVI algorithm, when applied to the quartic function (4.26).

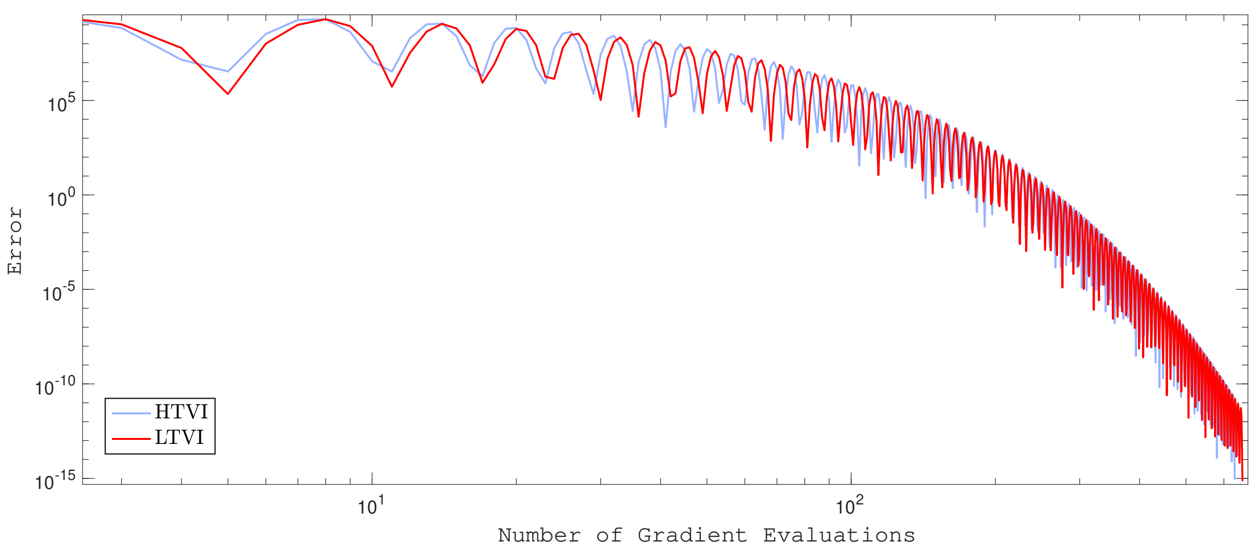


FIGURE 2. Comparison of the HTVI and LTVI algorithms with the same parameters.

5. Accelerated Optimization on More General Spaces

5.1. Motivation and Prior Work. The variational framework for accelerated optimization on normed vector spaces from [20; 65] was extended to the Riemannian manifold setting in [19] via a Riemannian p-Bregman Lagrangian $\mathcal{L}_p : T\mathcal{Q} \times \mathbb{R} \to \mathbb{R}$ and a corresponding Riemannian p-Bregman Hamiltonian $\mathcal{H}_p : T^*\mathcal{Q} \times \mathbb{R} \to \mathbb{R}$, for $p > 0$, of the form

$$\mathcal{L}_p(X,V,t) = \frac{t^{\frac{\zeta}{\lambda}p+1}}{2p}\langle V,V\rangle - Cpt^{(\frac{\zeta}{\lambda}+1)p-1}f(X), \tag{5.1}$$

$$\mathcal{H}_p(X,R,t) = \frac{p}{2t^{\frac{\zeta}{\lambda}p+1}}\langle\!\langle R,R\rangle\!\rangle + Cpt^{(\frac{\zeta}{\lambda}+1)p-1}f(X), \tag{5.2}$$

where ζ and λ are constants having to do with the curvature of the manifold and the convexity of the objective function f. These yield the associated p-Bregman Euler–Lagrange equations,

$$\nabla_{\dot{X}}\dot{X} + \frac{\zeta p + \lambda}{\lambda t}\dot{X} + Cp^2 t^{p-2}\text{grad}f(X) = 0. \tag{5.3}$$

Here, $\text{grad}f$ denotes the Riemannian gradient of f, $\nabla_X Y$ is the covariant derivative of Y along X, and $\langle\langle \cdot,\cdot \rangle\rangle$ is the fiber metric on $T^*\mathcal{Q}$ induced by the Riemannian metric $\langle \cdot,\cdot \rangle$ on $\mathcal{Q}$ whose local coordinates representation is the inverse of the local representation of $\langle \cdot,\cdot \rangle$. See [1; 6; 19; 32; 35; 48] for a more detailed description of these notions from Riemannian geometry and of this Riemannian variational framework for accelerated optimization. Note that some work was done on accelerated optimization via numerical integration of Bregman dynamics in the Lie group setting [42; 60] before the theory for more general Bregman families on Riemannian manifolds was established in [19].

It was shown in [19] that solutions to the p-Bregman Euler–Lagrange equations converge to a minimizer of f at a convergence rate of $\mathcal{O}(1/t^p)$, under suitable assumptions, and proven that time-rescaling a solution to the p-Bregman Euler–Lagrange equations via $\tau(t) = t^{\mathring{p}/p}$ yields a solution to the $\mathring{p}$-Bregman Euler–Lagrange equations. As a result, the adaptive approach involving the Poincaré transformation was extended to the Riemannian manifold setting via the adaptive approach Riemannian $p \to \mathring{p}$ Bregman Hamiltonian,

$$\bar{\mathcal{H}}_{p\to\mathring{p}}\left(\bar{Q},\bar{R}\right) = \frac{p^2}{2\mathring{p}\mathfrak{Q}^{\frac{\zeta}{\lambda}p+\frac{\mathring{p}}{p}}}\langle\langle R,R\rangle\rangle + \frac{Cp^2}{\mathring{p}}\mathfrak{Q}^{\left(\frac{\zeta}{\lambda}+1\right)p-\frac{\mathring{p}}{p}}f(Q) + \frac{p}{\mathring{p}}\mathfrak{Q}^{1-\frac{\mathring{p}}{p}}\mathfrak{R}. \tag{5.4}$$

This adaptive framework was exploited using discrete variational integrators incorporating holonomic constraints [16] and projection-based variational integrators [18]. Both these strategies relied on embedding the Riemannian manifolds into an ambient Euclidean space. Although the Whitney and Nash Embedding Theorems [53; 63; 64] imply that there is no loss of generality when studying Riemannian manifolds only as submanifolds of Euclidean spaces, designing intrinsic methods that would exploit and preserve the symmetries and geometric properties of the Riemannian manifold and of the problem at hand could have advantages, both in terms of computation and in terms of improving our understanding of the acceleration phenomenon on Riemannian manifolds.

Developing an intrinsic extension of Hamiltonian variational integrators to manifolds would require some additional work, since the current approach involves Type II/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 on 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, and one should ideally consider a Hamiltonian variational integrator construction based on discrete Dirac mechanics [44; 66; 67].

On the other hand, Lagrangian variational integrators involve a Type I generating function $L_d(q_k,q_{k+1})$ which only depends on the position at the boundary points and is therefore well-defined on manifolds, and many Lagrangian variational integrators have been derived on Riemannian manifolds, especially in the Lie group [5; 27; 28; 36–39; 43; 56] and homogeneous space [40] settings. The time-adaptive framework developed in this paper makes it now possible to design time-adaptive Lagrangian integrators for accelerated optimization on these more general spaces, where it is more natural and easier to work on the Lagrangian side than on the Hamiltonian side.

5.2. **Accelerated Optimization on Lie Groups.** Although it is possible to work on Riemannian manifolds, we will restrict ourselves to Lie groups for simplicity of exposition since there is more literature available on Lie group integrators than Riemannian integrators. Note as well that prior work is available on accelerated optimization via numerical integration of Bregman dynamics in the Lie group setting [42; 60].

Here, we will work in the setting introduced in [42]. The setting of [60] can be thought of as a special case of the more general Lie group framework for accelerated optimization presented here. Consider a n-dimensional Lie group G with associated Lie algebra $\mathfrak{g} = T_eG$, and a left-trivialization of the tangent bundle of the group $TG \simeq G \times \mathfrak{g}$, via $(q, \dot{q}) \mapsto (q, \mathrm{L}_{q^{-1}}\dot{q}) \equiv (q, \xi)$, where $\mathrm{L} : G \times G \to G$ is the left action defined by $\mathrm{L}_q h = qh$ for all $q, h \in G$. Suppose that $\mathfrak{g}$ is equipped with an inner product which induces an inner product on T_qG via left-trivialization,

$$(v \bullet w)_{T_qG} = (\mathrm{T}_q\mathrm{L}_{q^{-1}} v \bullet \mathrm{T}_q\mathrm{L}_{q^{-1}} w)_{\mathfrak{g}}, \qquad \forall v, w \in T_qG.$$

With this inner product, we identify $\mathfrak{g} \simeq \mathfrak{g}^*$ and $T_qG \simeq T_q^*G \simeq G \times \mathfrak{g}^*$ via the Riesz representation. Let $\mathbf{J} : \mathfrak{g} \to \mathfrak{g}^*$ be chosen such that $(\mathbf{J}(\xi) \bullet \zeta)$ is positive-definite and symmetric as a bilinear form of $\xi, \zeta \in \mathfrak{g}$. Then, the metric $\langle \cdot, \cdot \rangle : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ with $\langle \xi, \zeta \rangle = (\mathbf{J}(\xi) \bullet \zeta)$ serves as a left-invariant Riemannian metric on G. The adjoint and ad operators are denoted by $\mathrm{Ad}_q : \mathfrak{g} \to \mathfrak{g}$ and $\mathrm{ad}_\xi : \mathfrak{g} \to \mathfrak{g}$, respectively. We refer the reader to [23; 41; 48] for a more detailed description of Lie group theory and mechanics on Lie groups.

As mentioned earlier, there is a lot of literature available on Lie group integrators. We refer the reader to [11–13; 30] for very thorough surveys of the literature on Lie group methods, which acknowledge all the foundational contributions leading to the current state of Lie group integrator theory. In particular, the Crouch and Grossman approach [14], the Lewis and Simo approach [47], Runge–Kutta–Munthe–Kaas methods [9; 50–52], Magnus and Fer expansions [4; 29; 68], and commutator-free Lie group methods [10] are outlined in these surveys. Variational integrators have also been derived on the Lagrangian side in the Lie group setting [5; 27; 28; 36–39; 43; 56].

We now introduce a discrete variational formulation of time-adaptive Lagrangian mechanics on Lie groups. Suppose we are given a partition $0 = \tau_0 < \tau_1 < \ldots < \tau_N = \mathcal{T}$ of the interval $[0, \mathcal{T}]$, and a discrete curve in $G \times \mathbb{R} \times \mathbb{R}$ denoted by $\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N$ such that $q_k \approx q(\tau_k)$, $\mathfrak{q}_k \approx \mathfrak{q}(\tau_k)$, and $\lambda_k \approx \lambda(\tau_k)$. The discrete kinematics equation is chosen to be

$$q_{k+1} = q_k f_k, \tag{5.5}$$

where $f_k \in G$ represents the relative update over a single step.

Consider the discrete action functional,

$$\bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1}\left[L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}) - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k}, \tag{5.6}$$

where,

$$L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}) \approx \underset{\substack{(q,\mathfrak{q}) \in C^2([\tau_k, \tau_{k+1}], G \times \mathbb{R}) \\ (q,\mathfrak{q})(\tau_k) = (q_k, \mathfrak{q}_k),\ (q,\mathfrak{q})(\tau_{k+1}) = (q_k f_k, \mathfrak{q}_{k+1})}}{\mathrm{ext}} \int_{\tau_k}^{\tau_{k+1}} L\left(q, \frac{\xi}{g(\mathfrak{q})}, \mathfrak{q}\right) d\tau. \tag{5.7}$$

We can derive the following result which relates a discrete Type I variational principle to a set of discrete Euler–Lagrange equations:

Theorem 5.1. *The Type I discrete Hamilton's variational principle,*

$$\delta \bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = 0, \tag{5.8}$$

where,

$$\bar{\mathfrak{S}}_d\left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1}\left[L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}) - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k}, \tag{5.9}$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1} - \tau_k) g(\mathfrak{q}_k), \tag{5.10}$$

$$\mathrm{Ad}^*_{f_k^{-1}}\left(\mathrm{T}_e^*\mathrm{L}_{f_k} D_2 L_{d_k}\right) = \mathrm{T}_e^*\mathrm{L}_{q_k} D_1 L_{d_k} + \frac{\tau_{k+1}-\tau_k}{\mathfrak{q}_{k+1}-\mathfrak{q}_k}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}\mathrm{T}_e^*\mathrm{L}_{f_{k-1}} D_2 L_{d_{k-1}}, \tag{5.11}$$

$$\begin{aligned}\left[D_3 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1}-\tau_k} + \lambda_k \nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} - \frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right] \\ + \left[D_4 L_{d_k} - \lambda_{k-1}\frac{1}{\tau_k-\tau_{k-1}}\right]\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}} + \frac{1}{\tau_k-\tau_{k-1}}\left[L_{d_{k-1}} - \lambda_{k-1}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}} + \lambda_{k-1} g(\mathfrak{q}_{k-1})\right] = 0,\end{aligned} \tag{5.12}$$

where L_{d_k} *denotes* $L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1})$.

Proof. See Appendix A.3. □

Now, define

$$\mathfrak{p}_k = -D_3 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}) \tag{5.13}$$

and

$$\mu_k = \mathrm{Ad}^*_{f_k^{-1}}\left(\mathrm{T}_e^*\mathrm{L}_{f_k} D_2 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1})\right) - \mathrm{T}_e^*\mathrm{L}_{q_k} D_1 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}). \tag{5.14}$$

Then,

$$\mu_{k+1} = \frac{\tau_{k+2}-\tau_{k+1}}{\mathfrak{q}_{k+2}-\mathfrak{q}_{k+1}}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\mathrm{T}_e^*\mathrm{L}_{f_k} D_2 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}). \tag{5.15}$$

With these definitions, if we use a constant time-step h in τ and substitute $g(\mathfrak{q}) = \frac{p}{\mathring{p}}\mathfrak{q}^{1-\mathring{p}/p}$, the discrete Euler–Lagrange equations can be rewritten as

$$\mu_k = \mathrm{Ad}^*_{f_k^{-1}}\left(\mathrm{T}_e^*\mathrm{L}_{f_k} D_2 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1})\right) - \mathrm{T}_e^*\mathrm{L}_{q_k} D_1 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}), \tag{5.16}$$

$$\mu_{k+1} = \frac{\mathfrak{q}_k^{1-\mathring{p}/p}}{\mathfrak{q}_{k+1}^{1-\mathring{p}/p}}\mathrm{T}_e^*\mathrm{L}_{f_k} D_2 L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}), \tag{5.17}$$

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + h\frac{p}{\mathring{p}}\mathfrak{q}_k^{1-\mathring{p}/p}, \tag{5.18}$$

$$\mathfrak{p}_{k+1} = \frac{\mathring{p}\left[\lambda_{k+1} - \lambda_k + L_{d_k} - L_{d_{k+1}}\right]}{hp\mathfrak{q}_{k+1}^{1-\mathring{p}/p}} + D_4 L_{d_k} + \frac{\lambda_{k+1}}{\mathfrak{q}_{k+1}}\left(1 - \frac{\mathring{p}}{p}\right). \tag{5.19}$$

In the Lie group setting, the Riemannian p-Bregman Lagrangian becomes

$$\mathcal{L}_p(q, \xi, t) = \frac{t^{\kappa p+1}}{2p}\langle \xi, \xi \rangle - Cpt^{(\kappa+1)p-1} f(q), \tag{5.20}$$

with corresponding Euler–Lagrange equation,

$$\frac{d\mathbf{J}(\xi)}{dt} + \frac{\kappa p + 1}{t}\mathbf{J}(\xi) - \mathrm{ad}^*_\xi \mathbf{J}(\xi) + Cp^2 t^{p-2}\nabla_{\mathrm{L}} f(q) = 0, \tag{5.21}$$

where $\nabla_{\mathrm{L}} f$ is the left-trivialized derivative of f, given by $\nabla_{\mathrm{L}} f(q) = \mathrm{T}_e^*\mathrm{L}_q(\mathrm{D}_q f(q))$. We then consider the discrete Lagrangian,

$$L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}) = \frac{\mathfrak{q}_k^{\kappa p+1}}{hp(g(\mathfrak{q}_k))^2} T_d(f_k) - Chp\mathfrak{q}_k^{(\kappa+1)p-1} f(q_k), \tag{5.22}$$

where $T_d(f_k) \approx \frac{1}{2}\langle h\xi_k, h\xi_k \rangle$, which approximates

$$L_d(q_k, f_k, \mathfrak{q}_k, \mathfrak{q}_{k+1}) \approx \underset{\substack{(q,\mathfrak{q}) \in C^2([\tau_k, \tau_{k+1}], G\times\mathbb{R}) \\ (q,\mathfrak{q})(\tau_k) = (q_k, \mathfrak{q}_k),\ (q,\mathfrak{q})(\tau_{k+1}) = (q_k f_k, \mathfrak{q}_{k+1})}}{\mathrm{ext}} \int_{\tau_k}^{\tau_{k+1}} L\left(q, \frac{\xi}{g(\mathfrak{q})}, \mathfrak{q}\right) d\tau. \tag{5.23}$$

5.3. **Numerical Experiment on SO(3).** We work on the 3-dimensional Special Orthogonal group,

$$\mathrm{SO}(3)=\{R\in\mathbb{R}^{3\times 3}|R^\top R=I_{3\times 3},\det(R)=1\}. \tag{5.24}$$

Its Lie algebra is

$$\mathfrak{so}(3)=\{S\in\mathbb{R}^{3\times 3}|S^\top=-S\}, \tag{5.25}$$

with the matrix commutator as the Lie bracket. We have an identification between $\mathbb{R}^3$ and $\mathfrak{so}(3)$ given by the hat map $\hat{\cdot}:\mathbb{R}^3\to\mathfrak{so}(3)$, defined such that $\hat{x}y=x\times y$ for any $x,y\in\mathbb{R}^3$. The inverse of the hat map is the vee map $(\cdot)^\vee:\mathfrak{so}(3)\to\mathbb{R}^3$. The inner product on $\mathfrak{so}(3)$ is given by

$$\left(\hat{\eta}\bullet\hat{\xi}\right)_{\mathfrak{so}(3)}=\frac{1}{2}\mathrm{Trace}\left(\hat{\eta}^\top\hat{\xi}\right)=\eta^\top\xi, \tag{5.26}$$

and the metric is chosen so that

$$\langle\hat{\eta},\hat{\xi}\rangle=\left(\mathbf{J}(\hat{\eta})\bullet\hat{\xi}\right)_{\mathfrak{so}(3)}=\mathrm{Trace}\left(\hat{\eta}^\top J_d\hat{\xi}\right)=\eta^\top J\xi, \tag{5.27}$$

where $J\in\mathbb{R}^{3\times 3}$ is a symmetric positive-definite matrix and $J_d=\frac{1}{2}\mathrm{Trace}(J)I_{3\times 3}-J$. On SO(3), for any $u,v\in\mathbb{R}^3$ and $F\in\mathrm{SO}(3)$,

$$\mathrm{ad}_{\hat{u}}\hat{v}=[\hat{u},\hat{v}]=\hat{u}\hat{v}-\hat{v}\hat{u}=\widehat{u\times v},\qquad \mathrm{Ad}_F\hat{u}=F\hat{u}F^\top=\widehat{Fu}. \tag{5.28}$$

Identifying $\mathfrak{so}(3)^*\simeq\mathfrak{so}(3)\simeq\mathbb{R}^3$, we have for any $u,v\in\mathbb{R}^3$ and $F\in\mathrm{SO}(3)$ that

$$\mathrm{ad}_u v=\hat{u}v=u\times v,\qquad \mathrm{ad}_u^* v=-\hat{u}v=v\times u,\qquad \mathrm{Ad}_F u=Fu,\qquad \mathrm{Ad}_F^* u=F^\top u. \tag{5.29}$$

On SO(3), the Riemannian p-Bregman Lagrangian becomes

$$\mathcal{L}_p(R,\Omega,t)=\frac{t^{p+1}}{2p}\Omega^\top J\Omega-Cpt^{2p-1}f(R), \tag{5.30}$$

and the corresponding Euler–Lagrange equations are given by

$$J\dot{\Omega}+\frac{p+1}{t}J\Omega+\hat{\Omega}J\Omega+Cp^2t^{p-2}\nabla_{\mathrm{L}}f(R)=0,\qquad \dot{R}=R\hat{\Omega}. \tag{5.31}$$

The discrete kinematics equations is written as

$$R_{k+1}=R_kF_k, \tag{5.32}$$

where $F_k\in\mathrm{SO}(3)$, and $\kappa=1$ so we get the discrete Lagrangian,

$$L_d(R_k,F_k,\mathfrak{R}_k,\mathfrak{R}_{k+1})=\frac{\mathring{p}^2}{hp^3}\mathfrak{R}_k^{p-1+2\mathring{p}/p}T_d(F_k)-Chp\mathfrak{R}_k^{2p-1}f(R_k). \tag{5.33}$$

As in [38; 42], the angular velocity is approximated with $\hat{\Omega}_k\approx\frac{1}{h}R_k^\top(R_{k+1}-R_k)=\frac{1}{h}(F_k-I_{3\times 3})$ so we can take

$$T_d(F_k)=\mathrm{Trace}\left([I_{3\times 3}-F_k]J_d\right). \tag{5.34}$$

Differentiating this equation and using the identity $\mathrm{Trace}(-\hat{x}A)=(A-A^\top)^\vee\cdot x$ yields

$$\mathrm{T}_I^*\mathrm{L}_{F_k}\left(D_{F_k}T_d(F_k)\right)=\left(J_dF_k-F_k^\top J_d\right)^\vee. \tag{5.35}$$

Then, the discrete Euler–Lagrange equations for μ_k and μ_{k+1} become

$$\mu_k=\frac{\mathring{p}^2}{hp^3}\mathfrak{R}_k^{p-1+2\mathring{p}/p}\left(F_kJ_d-J_dF_k^\top\right)^\vee+Chp\mathfrak{R}_k^{2p-1}\nabla_{\mathrm{L}}f(R_k), \tag{5.36}$$

$$\mu_{k+1}=\frac{\mathfrak{q}_k^{1-\mathring{p}/p}}{\mathfrak{q}_{k+1}^{1-\mathring{p}/p}}F_k^\top\left[\mu_k-Chp\mathfrak{R}_k^{2p-1}\nabla_{\mathrm{L}}f(R_k)\right]. \tag{5.37}$$

Now, equation (5.36) can be solved explicitly when $J = I_{3\times 3}$ as described in [42]:

$$F_k = \exp\left(\frac{\sin^{-1}\|a_k\|}{\|a_k\|}\hat{a}_k\right), \qquad \text{where} \quad a_k = \frac{hp^3}{\mathring{p}^2}\mathfrak{R}_k^{1-p-2\mathring{p}/p}\left[\mu_k - Chp\mathfrak{R}_k^{2p-1}\nabla_{\mathrm{L}} f(R_k)\right]. \tag{5.38}$$

Therefore, we get the **Adaptive LLGVI** (Adaptive Lagrangian Lie Group Variational Integrator)

$$F_k = \exp\left(\frac{\sin^{-1}\|a_k\|}{\|a_k\|}\hat{a}_k\right), \quad \text{where} \quad a_k = \frac{hp^3}{\mathring{p}^2}\mathfrak{R}_k^{1-p-2\mathring{p}/p}\left[\mu_k - Chp\mathfrak{R}_k^{2p-1}\nabla_{\mathrm{L}} f(R_k)\right], \tag{5.39}$$

$$\mathfrak{R}_{k+1} = \mathfrak{R}_k + h\frac{p}{\mathring{p}}\mathfrak{R}_k^{1-\mathring{p}/p}, \tag{5.40}$$

$$\mu_{k+1} = \frac{\mathfrak{R}_k^{1-\mathring{p}/p}}{\mathfrak{R}_{k+1}^{1-\mathring{p}/p}}F_k^\top\left[\mu_k - Chp\mathfrak{R}_k^{2p-1}\nabla_{\mathrm{L}} f(R_k)\right], \tag{5.41}$$

$$R_{k+1} = R_k F_k. \tag{5.42}$$

We will use this integrator to solve the problem of minimizing the objective function,

$$f(R) = \frac{1}{2}\|A - R\|_F^2 = \frac{1}{2}\left(\|A\|_F^2 + 3\right) - \mathrm{Trace}(A^\top R), \tag{5.43}$$

over $R \in \mathrm{SO}(3)$, where $\|\cdot\|_F$ denotes the Frobenius norm. Its left-trivialized gradient is given by

$$\nabla_{\mathrm{L}} f(R) = \left(A^\top R - R^\top A\right)^\vee. \tag{5.44}$$

Minimizing this objective function appears in the least-squares estimation of attitude, referred to as Wahba's problem [62]. The optimal attitude is explicitly given by

$$R^* = U\mathrm{diag}\left[1, 1, \det(UV)\right]V^\top, \tag{5.45}$$

where $A = USV^\top$ is the singular value decomposition of A with $U, V \in \mathrm{O}(3)$ and S diagonal.

We have tested the Adaptive LLGVI integrator on Wahba's problem against the Implicit Lie Group Variational Integrator (**Implicit LGVI**) from [42]. The Implicit LGVI is a Lagrangian Lie group variational integrator which adaptively adjusts the step size at every step. It should be noted that these two adaptive approaches use adaptivity in two fundamentally different ways: our Adaptive LLGVI method uses *a priori* adaptivity based on known global properties of the family of differential equations considered (i.e. the time-rescaling symmetry of the family of Bregman dynamics), while the implicit method from [42] adapts the time-steps in an *a posteriori* way, by solving a system of nonlinear equations coming from an extended variational principle. The results of our numerical experiments are presented in Figures 3 and 4. In these numerical experiments, we have used the termination criteria

$$|f(R_k) - f(R^*)| < \delta \qquad \text{and} \qquad |f(R_k) - f(R_{k-1})| < \delta. \tag{5.46}$$

We can see from Figure 3 that both algorithms preserve the orthogonality condition $R_k^\top R_k = I_{3\times 3}$ very well. Now, we can observe from Figure 3 that although both algorithms follow the same curve in time t, they do not travel along this curve at the same speed. Despite the fact that the Adaptive LLGVI algorithm initially takes smaller time-steps, those time-steps eventually become much larger than for the Implicit LGVI algorithm, and as a result, the Adaptive LLGVI algorithm achieves the termination criteria in a smaller number of iterations, which can also be seen more explicitly in the table from Figure 4. Unlike the Implicit LGVI algorithm, the Adaptive LLGVI algorithm is explicit, so each iteration is much cheaper and is therefore significantly faster, as can be seen from the running times displayed in Figure 4. Furthermore, the Adaptive LLGVI algorithm is significantly easier to implement.

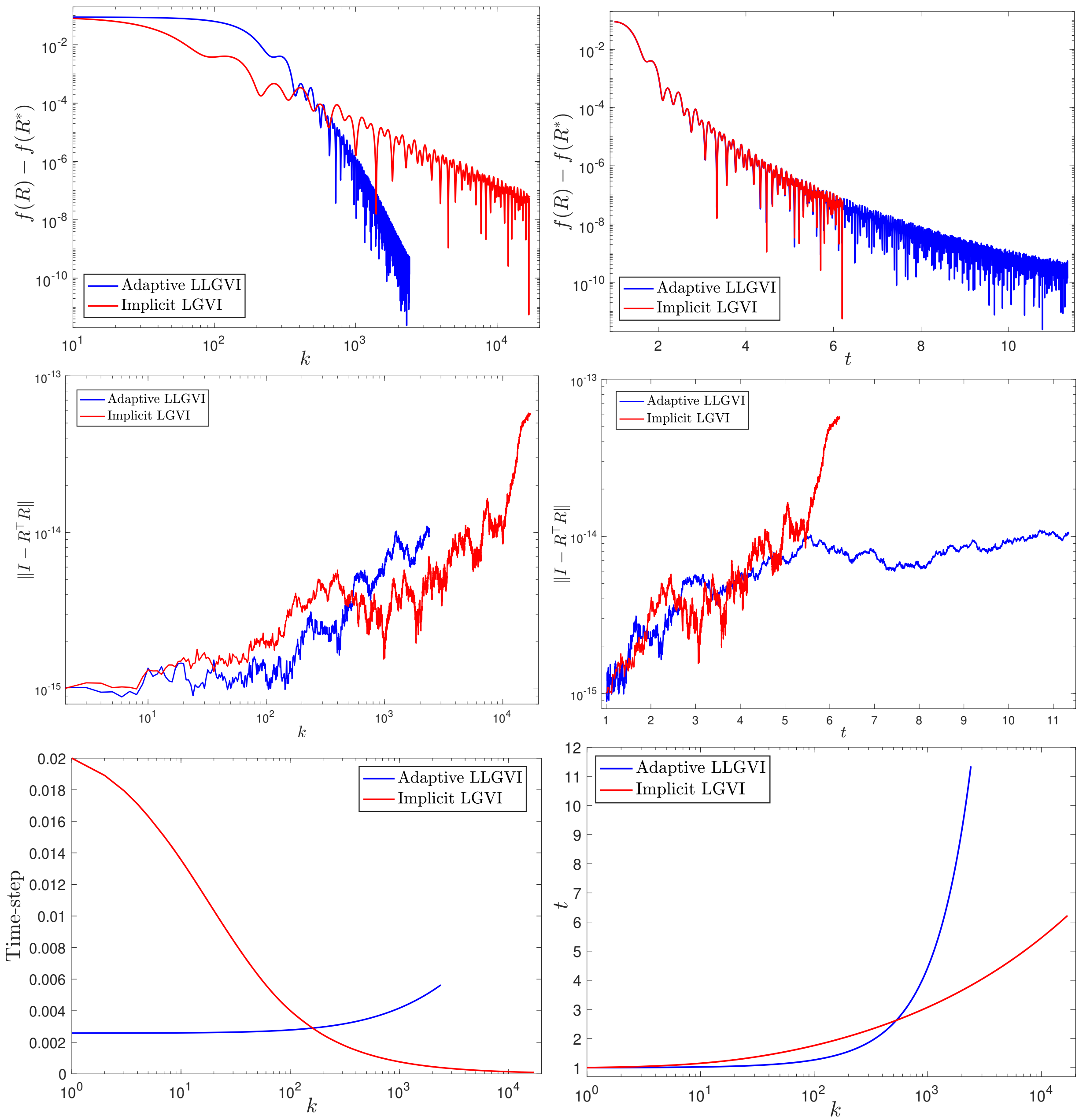
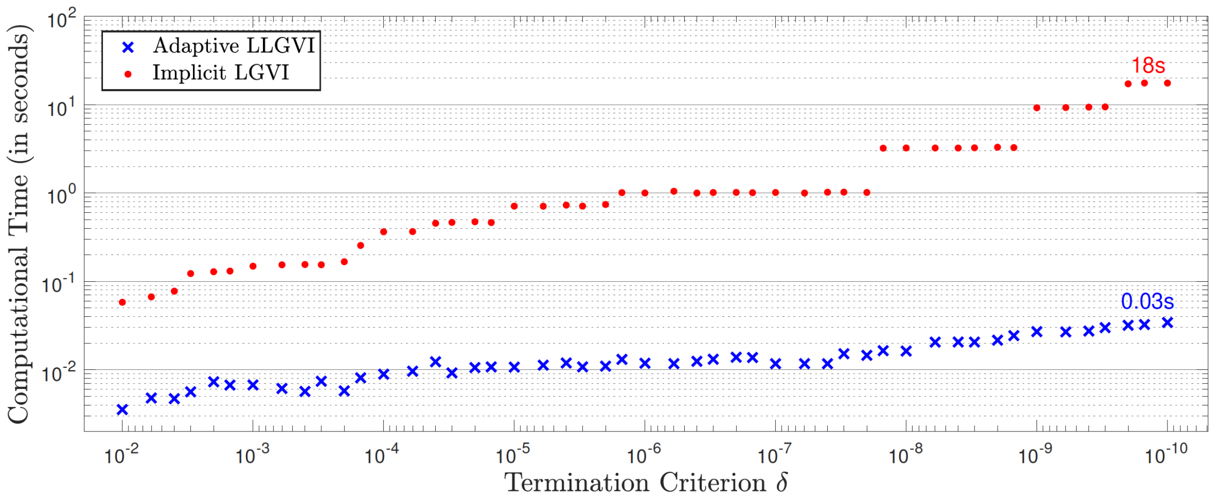


FIGURE 3. Comparison of the Adaptive LLGVI algorithm and of the Implicit LGVI algorithm from [42] with $p = 6$, to solve Wahba's problem (5.43).

6. CONCLUSION

A variational framework for accelerated optimization on vector spaces was introduced [65] by considering a family of time-dependent Bregman Lagrangian and Hamiltonian systems which is closed under time-rescaling. This variational framework was exploited in [20] by using time-adaptive geometric Hamiltonian integrators to design efficient, explicit algorithms for symplectic accelerated optimization. It was observed that a careful use of adaptivity and symplecticity, which was possible on the Hamiltonian side thanks to the Poincaré transformation, could result in a significant gain in computational efficiency, by simulating higher-order Bregman dynamics using the computationally efficient lower-order Bregman integrators applied to the time-rescaled dynamics.



Termination Criterion δ	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
Adaptive LLGVI: Iterations	350	510	648	722	724	1181	1604	2237
Implicit LGVI: Iterations	179	474	970	1380	1392	4506	11992	16758
Adaptive LLGVI: Time (in seconds)	0.007	0.009	0.011	0.012	0.012	0.016	0.027	0.034
Implicit LGVI: Time (in seconds)	0.15	0.36	0.71	1.00	1.01	3.24	9.22	17.55

FIGURE 4. Time and number of iterations needed by the Adaptive LLGVI and Implicit LGVI algorithms with $p = 6$, to satisfy the termination criterion (5.46) on Wahba's problem (5.43).

These variational framework and time-adaptive approach on the Hamiltonian side were later extended to the Riemannian manifolds setting in [19]. However, the current formulations of Hamiltonian variational integrators do not make sense intrinsically on manifolds, so this framework was only exploited using methods which take advantage of the structure of the Euclidean spaces in which the Riemannian manifolds are embedded [16; 18] instead of the structure of the Riemannian manifolds themselves. On the other hand, existing formulations of Lagrangian variational integrators are well-defined on manifolds, and many Lagrangian variational integrators have been derived on Riemannian manifolds, especially in the Lie group setting. This motivated exploring whether it is possible to construct a mechanism on the Lagrangian side which mimics the Poincaré transformation, since it is more natural and easier to work on the Lagrangian side on curved manifolds.

The usual correspondence between Hamiltonian and Lagrangian dynamics could not be exploited here since the Poincaré Hamiltonian is degenerate and therefore does not have a corresponding Lagrangian formulation. Instead, we introduced a novel derivation of the Poincaré transformation from a variational principle which gave us additional insight into the transformation mechanism and provided natural candidates for a time-adaptive framework on the Lagrangian side. Based on these observations, we constructed a theory of time-adaptive Lagrangian mechanics both in continuous and discrete time, and tested the resulting time-adaptive Lagrangian variational integrators to solve optimization problems by simulating Bregman dynamics, within the variational framework introduced in [65]. We observed empirically that our time-adaptive Lagrangian variational integrators performed almost exactly in the same way as the time-adaptive Hamiltonian variational integrators coming from the Poincaré framework of [20], whenever they are used with the same parameters and time-step. As a result, the computational analysis carried in [20] for the HTVI algorithm extends to the LTVI algorithm, and thus the LTVI algorithm is much more efficient than non-symplectic integrators for the Bregman dynamics and can be a competitive first-order explicit algorithm since it can outperform commonly used optimization algorithms for certain objective functions.

Finally, we showed that our time-adaptive Lagrangian approach extends naturally to more general spaces such as Riemannian manifolds and Lie groups without having to face the difficulties experienced on the Hamiltonian side, and we applied time-adaptive Lie group Lagrangian variational integrators to solve optimization problems on the three-dimensional Special Orthogonal group SO(3). In particular, the resulting algorithms were significantly faster and easier to implement than other recently proposed time-adaptive Lie group variational integrators for accelerated optimization.

In future work, we will explore the issue of time-adaptive Lagrangian mechanics for more general monitor functions, using the primal-dual framework of Dirac mechanics. We will also study the convergence properties of the discrete-time algorithms, and try to better understand how to reconcile the Nesterov barrier theorem with the convergence properties of the continuous Bregman flows. It would also be useful to study the extent to which the practical considerations recently presented in [17], which significantly improved the computational performance of the symplectic optimization algorithms in the normed vector space setting, extend to the Riemannian manifold and Lie group settings with the Lagrangian Riemannian and Lie group variational integrators.

Acknowledgments

The authors would like to thank Taeyoung Lee and Molei Tao for sharing the code for the implicit Lie group variational integrator from [42], which was used in the numerical comparison with the method introduced in this paper. The authors would also like to thank the referees for their helpful comments and suggestions.

Funding

The authors were supported in part by NSF under grants DMS-1345013, DMS-1813635, CCF-2112665, by AFOSR under grant FA9550-18-1-0288, and by the DoD under grant HQ00342010023 (Newton Award for Transformative Ideas during the COVID-19 Pandemic).

Appendix A. Proofs of Theorems

A.1. **Proof of Theorem 3.2.**

Theorem A.1. *The Type I discrete Hamilton's variational principle,*

$$\delta \bar{\mathfrak{S}}_d \left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N \right) = 0,$$

where,

$$\bar{\mathfrak{S}}_d \left(\{(q_k, \mathfrak{q}_k, \lambda_k)\}_{k=0}^N \right) = \sum_{k=0}^{N-1} \left[L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k},$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1} - \tau_k) g(\mathfrak{q}_k),$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, \mathfrak{q}_{k-1}, q_k, \mathfrak{q}_k) = 0,$$

$$\begin{aligned} &\left[D_2 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(\mathfrak{q}_k) \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} - \frac{1}{\tau_{k+1} - \tau_k} \left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \\ &+ \left[D_4 L_{d_{k-1}} - \lambda_{k-1} \frac{1}{\tau_k - \tau_{k-1}} \right] \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \frac{1}{\tau_k - \tau_{k-1}} \left[L_{d_{k-1}} - \lambda_{k-1} \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \lambda_{k-1} g(\mathfrak{q}_{k-1}) \right] = 0, \end{aligned}$$

where L_{d_k} denotes $L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$.

Proof. We use the notation $L_{d_k} = L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$, and we will use the fact that $\delta q_0 = \delta q_N = \delta \mathfrak{q}_0 = \delta \mathfrak{q}_N = 0$ throughout the proof. We have

$$\begin{aligned}
\delta \bar{\mathfrak{S}}_d &= \delta \left(\sum_{k=0}^{N-1} \left[L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \right) \\
&= \sum_{k=1}^{N-1} \left[D_2 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(\mathfrak{q}_k) \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \delta \mathfrak{q}_k - \sum_{k=1}^{N-1} \frac{1}{\tau_{k+1} - \tau_k} \left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \delta \mathfrak{q}_k \\
&\quad + \sum_{k=0}^{N-2} \left[D_4 L_{d_k} - \lambda_k \frac{1}{\tau_{k+1} - \tau_k} \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \delta \mathfrak{q}_{k+1} + \sum_{k=0}^{N-2} \frac{1}{\tau_{k+1} - \tau_k} \left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \delta \mathfrak{q}_{k+1} \\
&\quad + \sum_{k=1}^{N-1} \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_{d_k} \delta q_k + \sum_{k=0}^{N-2} \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_3 L_{d_k} \delta q_{k+1} + \sum_{k=0}^{N-1} \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \left(g(\mathfrak{q}_k) - \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \right) \delta \lambda_k .
\end{aligned}$$

Thus,

$$\begin{aligned}
\delta \bar{\mathfrak{S}}_d = &\sum_{k=1}^{N-1} \left[D_2 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(\mathfrak{q}_k) \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \delta \mathfrak{q}_k - \sum_{k=1}^{N-1} \frac{1}{\tau_{k+1} - \tau_k} \left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \delta \mathfrak{q}_k \\
&+ \sum_{k=1}^{N-1} \left[D_4 L_{d_{k-1}} - \lambda_{k-1} \frac{1}{\tau_k - \tau_{k-1}} \right] \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} \delta \mathfrak{q}_k + \sum_{k=1}^{N-1} \frac{1}{\tau_k - \tau_{k-1}} \left[L_{d_{k-1}} - \lambda_{k-1} \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \lambda_{k-1} g(\mathfrak{q}_{k-1}) \right] \delta \mathfrak{q}_k \\
&+ \sum_{k=1}^{N-1} \left[\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_{d_k} + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_{d_{k-1}} \right] \delta q_k + \sum_{k=0}^{N-1} \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \left(g(\mathfrak{q}_k) - \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \right) \delta \lambda_k .
\end{aligned}$$

As a consequence, if

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1} - \tau_k) g(\mathfrak{q}_k),$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, \mathfrak{q}_{k-1}, q_k, \mathfrak{q}_k) = 0,$$

$$\begin{aligned}
&\left[D_2 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(\mathfrak{q}_k) \right] \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} - \frac{1}{\tau_{k+1} - \tau_k} \left[L_{d_k} - \lambda_k \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} + \lambda_k g(\mathfrak{q}_k) \right] \\
&\quad + \left[D_4 L_{d_{k-1}} - \lambda_{k-1} \frac{1}{\tau_k - \tau_{k-1}} \right] \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \frac{1}{\tau_k - \tau_{k-1}} \left[L_{d_{k-1}} - \lambda_{k-1} \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} + \lambda_{k-1} g(\mathfrak{q}_{k-1}) \right] = 0,
\end{aligned}$$

then $\delta \bar{\mathfrak{S}}_d \left(\{ (q_k, \mathfrak{q}_k, \lambda_k) \}_{k=0}^N \right) = 0$. Conversely, if $\delta \bar{\mathfrak{S}}_d \left(\{ (q_k, \mathfrak{q}_k, \lambda_k) \}_{k=0}^N \right) = 0$, then a discrete fundamental theorem of the calculus of variations yields the above equations. □

A.2. **Proof of Theorem 3.4.**

Theorem A.2. *The Type I discrete Hamilton's variational principle,*

$$\delta \bar{\mathfrak{S}}_d \left(\{ (q_k, \mathfrak{q}_k, \lambda_k) \}_{k=0}^N \right) = 0,$$

where,

$$\bar{\mathfrak{S}}_d \left(\{ (q_k, \mathfrak{q}_k, \lambda_k) \}_{k=0}^N \right) = \sum_{k=0}^{N-1} \left\{ \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \left[L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) - \lambda_k \right] + \lambda_k g(\mathfrak{q}_k) \right\},$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1} - \tau_k) g(\mathfrak{q}_k),$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, \mathfrak{q}_{k-1}, q_k, \mathfrak{q}_k) = 0,$$

$$\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_2 L_{d_k} - \frac{1}{\tau_{k+1} - \tau_k} L_{d_k} + \frac{\mathfrak{q}_k - \mathfrak{q}_{k-1}}{\tau_k - \tau_{k-1}} D_4 L_{d_{k-1}} + \frac{1}{\tau_k - \tau_{k-1}} L_{d_{k-1}} = \frac{\lambda_{k-1}}{\tau_k - \tau_{k-1}} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} - \lambda_k \nabla g(\mathfrak{q}_k),$$

where L_{d_k} *denotes* $L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$.

Proof. We use the notation $L_{d_k} = L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1})$, and we will use the fact that $\delta q_0 = \delta q_N = \delta \mathfrak{q}_0 = \delta \mathfrak{q}_N = 0$ throughout the proof. We have

$$\begin{aligned}
\delta \bar{\mathfrak{S}}_d &= \delta \left(\sum_{k=0}^{N-1} \left\{ \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \left[L_d(q_k, \mathfrak{q}_k, q_{k+1}, \mathfrak{q}_{k+1}) - \lambda_k \right] + \lambda_k g(\mathfrak{q}_k) \right\} \right) \\
&= \sum_{k=1}^{N-1} \left[\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_2 L_{d_k} - \frac{1}{\tau_{k+1} - \tau_k} L_{d_k} + \frac{\lambda_k}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(\mathfrak{q}_k) \right] \delta \mathfrak{q}_k + \sum_{k=0}^{N-2} \left[\frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_4 L_{d_k} + \frac{1}{\tau_{k+1} - \tau_k} L_{d_k} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} \right] \delta \mathfrak{q}_{k+1} \\
&\quad + \sum_{k=1}^{N-1} \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_1 L_{d_k} \delta q_k + \sum_{k=0}^{N-2} \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} D_3 L_{d_k} \delta q_{k+1} + \sum_{k=0}^{N-1} \left(g(\mathfrak{q}_k) - \frac{\mathfrak{q}_{k+1} - \mathfrak{q}_k}{\tau_{k+1} - \tau_k} \right) \delta \lambda_k .
\end{aligned}$$

Thus,

$$\delta\bar{\mathfrak{S}}_d = \sum_{k=1}^{N-1}\left[\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}D_2L_{d_k} - \frac{1}{\tau_{k+1}-\tau_k}L_{d_k} + \frac{\lambda_k}{\tau_{k+1}-\tau_k} + \lambda_k\nabla g(\mathfrak{q}_k) + \frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}D_4L_{d_{k-1}} + \frac{1}{\tau_k-\tau_{k-1}}L_{d_{k-1}} - \frac{\lambda_{k-1}}{\tau_k-\tau_{k-1}}\right]\delta\mathfrak{q}_k$$
$$+ \sum_{k=1}^{N-1}\left[\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}D_1L_{d_k} + \frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}D_3L_{d_{k-1}}\right]\delta q_k + \sum_{k=0}^{N-1}\left(g(\mathfrak{q}_k) - \frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\right)\delta\lambda_k.$$

As a consequence, if

$$\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}D_2L_{d_k} - \frac{1}{\tau_{k+1}-\tau_k}L_{d_k} + \frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}D_4L_{d_{k-1}} + \frac{1}{\tau_k-\tau_{k-1}}L_{d_{k-1}} = \frac{\lambda_{k-1}}{\tau_k-\tau_{k-1}} - \frac{\lambda_k}{\tau_{k+1}-\tau_k} - \lambda_k\nabla g(\mathfrak{q}_k),$$

$$\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}D_1L_d(q_k,\mathfrak{q}_k,q_{k+1},\mathfrak{q}_{k+1}) + \frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}D_3L_d(q_{k-1},\mathfrak{q}_{k-1},q_k,\mathfrak{q}_k) = 0,$$

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1}-\tau_k)g(\mathfrak{q}_k),$$

then $\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = 0$. Conversely, if $\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = 0$, then a discrete fundamental theorem of the calculus of variations yields the above equations. □

A.3. **Proof of Theorem 5.1.**

Theorem A.3. *The Type I discrete Hamilton's variational principle,*

$$\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = 0,$$

where,

$$\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1}\left[L_d(q_k,f_k,\mathfrak{q}_k,\mathfrak{q}_{k+1}) - \lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k},$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k + (\tau_{k+1}-\tau_k)g(\mathfrak{q}_k),$$

$$\mathrm{Ad}^*_{f_k^{-1}}\left(\mathrm{T}_e^*\mathrm{L}_{f_k}D_2L_{d_k}\right) = \mathrm{T}_e^*\mathrm{L}_{q_k}D_1L_{d_k} + \frac{\tau_{k+1}-\tau_k}{\mathfrak{q}_{k+1}-\mathfrak{q}_k}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}\mathrm{T}_e^*\mathrm{L}_{f_{k-1}}D_2L_{d_{k-1}},$$

$$\left[D_3L_{d_k} + \lambda_k\frac{1}{\tau_{k+1}-\tau_k} + \lambda_k\nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} - \frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k} - \lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right]$$
$$+\left[D_4L_{d_k} - \lambda_{k-1}\frac{1}{\tau_k-\tau_{k-1}}\right]\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}} + \frac{1}{\tau_k-\tau_{k-1}}\left[L_{d_{k-1}} - \lambda_{k-1}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}} + \lambda_{k-1}g(\mathfrak{q}_{k-1})\right] = 0,$$

where L_{d_k} denotes $L_d(q_k,f_k,\mathfrak{q}_k,\mathfrak{q}_{k+1})$.

Proof. We use the notation $L_{d_k} = L_d(q_k,f_k,\mathfrak{q}_k,\mathfrak{q}_{k+1})$ and we will use the fact that $\delta q_0 = \delta q_N = \delta\mathfrak{q}_0 = \delta\mathfrak{q}_N = \eta_0 = \eta_N = 0$ throughout the proof. We have

$$\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = \delta\left(\sum_{k=0}^{N-1}\left[L_d(q_k,f_k,\mathfrak{q}_k,\mathfrak{q}_{k+1}) - \lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\right)$$
$$= \sum_{k=1}^{N-1}\left[D_3L_{d_k} + \lambda_k\frac{1}{\tau_{k+1}-\tau_k} + \lambda_k\nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\delta\mathfrak{q}_k - \sum_{k=1}^{N-1}\frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k} - \lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\delta\mathfrak{q}_k$$
$$+\sum_{k=0}^{N-2}\left[D_4L_{d_k} - \lambda_k\frac{1}{\tau_{k+1}-\tau_k}\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\delta\mathfrak{q}_{k+1} + \sum_{k=0}^{N-2}\frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k} - \lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\delta\mathfrak{q}_{k+1}$$
$$+\sum_{k=1}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}D_1L_{d_k}\delta q_k + \sum_{k=0}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}D_2L_{d_k}\delta f_k + \sum_{k=0}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(g(\mathfrak{q}_k) - \frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\right)\delta\lambda_k.$$

We can write δg_k as $\delta g_k = g_k\eta_k$ for some $\eta_k \in \mathfrak{g}$. Then, taking the variation of the discrete kinematics equation $q_{k+1} = q_kf_k$ gives the equation $\delta q_{k+1} = \delta q_kf_k + q_k\delta f_k$ and $f_k = q_k^{-1}q_{k+1}$. Therefore,

$$\delta f_k = q_k^{-1}\delta q_{k+1} - q_k^{-1}\delta q_kf_k = q_k^{-1}q_{k+1}\eta_{k+1} - q_k^{-1}q_k\eta_kf_k = f_k\eta_{k+1} - \eta_kf_k,$$

so

$$\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = \sum_{k=1}^{N-1}\left[D_3L_{d_k} + \lambda_k\frac{1}{\tau_{k+1}-\tau_k} + \lambda_k\nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\delta\mathfrak{q}_k - \sum_{k=1}^{N-1}\frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k} - \lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k} + \lambda_k g(\mathfrak{q}_k)\right]\delta\mathfrak{q}_k$$
$$+\sum_{k=1}^{N-1}\left[\left(D_4L_{d_{k-1}} - \lambda_{k-1}\frac{1}{\tau_k-\tau_{k-1}}\right)\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}} + \frac{1}{\tau_k-\tau_{k-1}}\left(L_{d_{k-1}} - \lambda_{k-1}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}} + \lambda_{k-1}g(\mathfrak{q}_{k-1})\right)\right]\delta\mathfrak{q}_k$$
$$+\sum_{k=1}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(\mathrm{T}_e^*\mathrm{L}_{q_k}D_1L_{d_k}\bullet\eta_k\right) + \sum_{k=0}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(\mathrm{T}_e^*\mathrm{L}_{f_k}D_2L_{d_k}\bullet\left[\eta_{k+1} - f_k^{-1}\eta_kf_k\right]\right)$$
$$+\sum_{k=0}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(g(\mathfrak{q}_k) - \frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\right)\delta\lambda_k.$$

Then,

$$
\begin{aligned}
\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right) = {} & \sum_{k=1}^{N-1}\left[D_3L_{d_k}+\lambda_k\frac{1}{\tau_{k+1}-\tau_k}+\lambda_k\nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\delta\mathfrak{q}_k-\sum_{k=1}^{N-1}\frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k}-\lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}+\lambda_k g(\mathfrak{q}_k)\right]\delta\mathfrak{q}_k \\
& +\sum_{k=1}^{N-1}\left[\left(D_4L_{d_{k-1}}-\lambda_{k-1}\frac{1}{\tau_k-\tau_{k-1}}\right)\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}+\frac{1}{\tau_k-\tau_{k-1}}\left(L_{d_{k-1}}-\lambda_{k-1}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}+\lambda_{k-1}g(\mathfrak{q}_{k-1})\right)\right]\delta\mathfrak{q}_k \\
& +\sum_{k=1}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(\mathrm{T}_e^*\mathrm{L}_{q_k}D_1L_{d_k}\bullet\eta_k\right)+\sum_{k=0}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(g(\mathfrak{q}_k)-\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\right)\delta\lambda_k \\
& +\sum_{k=0}^{N-1}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}\left(\mathrm{T}_e^*\mathrm{L}_{f_{k-1}}D_2L_{d_{k-1}}\bullet\eta_k\right)-\sum_{k=0}^{N-1}\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}\left(\mathrm{T}_e^*\mathrm{L}_{f_k}D_2L_{d_k}\bullet\mathrm{Ad}_{f_k^{-1}}\eta_k\right).
\end{aligned}
$$

As a consequence, if

$$\mathfrak{q}_{k+1}=\mathfrak{q}_k+(\tau_{k+1}-\tau_k)g(\mathfrak{q}_k),$$

$$Ad^*_{f_k^{-1}}\left(T_e^*L_{f_k}D_2L_{d_k}\right)=T_e^*L_{q_k}D_1L_{d_k}+\frac{\tau_{k+1}-\tau_k}{\mathfrak{q}_{k+1}-\mathfrak{q}_k}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}T_e^*L_{f_{k-1}}D_2L_{d_{k-1}},$$

$$
\begin{aligned}
& \left[D_3L_{d_k}+\lambda_k\frac{1}{\tau_{k+1}-\tau_k}+\lambda_k\nabla g(\mathfrak{q}_k)\right]\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}-\frac{1}{\tau_{k+1}-\tau_k}\left[L_{d_k}-\lambda_k\frac{\mathfrak{q}_{k+1}-\mathfrak{q}_k}{\tau_{k+1}-\tau_k}+\lambda_k g(\mathfrak{q}_k)\right] \\
& \quad +\left[D_4L_{d_k}-\lambda_{k-1}\frac{1}{\tau_k-\tau_{k-1}}\right]\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}+\frac{1}{\tau_k-\tau_{k-1}}\left[L_{d_{k-1}}-\lambda_{k-1}\frac{\mathfrak{q}_k-\mathfrak{q}_{k-1}}{\tau_k-\tau_{k-1}}+\lambda_{k-1}g(\mathfrak{q}_{k-1})\right]=0,
\end{aligned}
$$

then $\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right)=0$. Conversely, if $\delta\bar{\mathfrak{S}}_d\left(\{(q_k,\mathfrak{q}_k,\lambda_k)\}_{k=0}^N\right)=0$, then a discrete fundamental theorem of the calculus of variations yields the above equations. □

The authors have no competing interests to declare that are relevant to the content of this article.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13298-3.

[2] F. Alimisis, A. Orvieto, G. Bécigneul, and A. Lucchi. A continuous-time perspective for modeling acceleration in Riemannian optimization. In *Proceedings of the 23rd International AISTATS Conference*, volume 108 of *PMLR*, pages 1297–1307, 2020.

[3] M. Betancourt, M. I. Jordan, and A. Wilson. On symplectic optimization. 2018.

[4] S. Blanes, F. Casas, J. A. Oteo, and J. Ros. The Magnus expansion and some of its applications. *Physics Reports*, 470:151–238, 11 2008. doi: 10.1016/j.physrep.2008.11.001.

[5] N. Bou-Rabee and J. E. Marsden. Hamilton–Pontryagin integrators on Lie groups part I: Introduction and structure-preserving properties. *Foundations of Computational Mathematics*, 9(2):197–219, 2009.

[6] N. Boumal. An introduction to optimization on smooth manifolds, 2020. URL http://www.nicolasboumal.net/book.

[7] J. P. Calvo and J. M. Sanz-Serna. The development of variable-step symplectic integrators, with application to the two-body problem. *SIAM J. Sci. Comp.*, 14(4):936–952, 1993.

[8] C. M. Campos, A. Mahillo, and D. Martín de Diego. A discrete variational derivation of accelerated methods in optimization. 2021. URL https://arxiv.org/abs/2106.02700.

[9] F. Casas and B. Owren. Cost efficient Lie group integrators in the RKMK class. *BIT*, 43:723–742, 11 2003. doi: 10.1023/B:BITN.0000009959.29287.d4.

[10] E. Celledoni, A. Marthinsen, and B. Owren. Commutator-free Lie group methods. *Future Gener. Comput. Syst.*, 19:341–352, 2003.

[11] E. Celledoni, H. Marthinsen, and B. Owren. An introduction to Lie group integrators – basics, new developments and applications. *Journal of Computational Physics*, 257:1040–1061, 2014. ISSN 0021-9991. doi: 10.1016/j.jcp.2012.12.031.

[12] E. Celledoni, E. Çokaj, A. Leone, D. Murari, and B. Owren. Lie group integrators for mechanical systems. *International Journal of Computer Mathematics*, 99(1):58–88, 2022. doi: 10.1080/00207160.2021.1966772.

[13] S. H. Christiansen, H. Z. Munthe-Kaas, and B. Owren. Topics in structure-preserving discretization. *Acta Numerica*, 20:1–119, 2011. doi: 10.1017/S096249291100002X.

[14] P. E. Crouch and R. L. Grossman. Numerical integration of ordinary differential equations on manifolds. *Journal of Nonlinear Science*, 3:1–33, 1993.

[15] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011.

[16] V. Duruisseaux and M. Leok. Accelerated optimization on Riemannian manifolds via discrete constrained variational integrators. *Journal of Nonlinear Science*, 32(42), 2022. URL https://doi.org/10.1007/s00332-022-09795-9.

[17] V. Duruisseaux and M. Leok. Practical perspectives on symplectic accelerated optimization. 2022. URL https://arxiv.org/abs/2207.1146.

[18] V. Duruisseaux and M. Leok. Accelerated optimization on Riemannian manifolds via projected variational integrators. 2022. URL https://arxiv.org/abs/2201.02904.

[19] V. Duruisseaux and M. Leok. A variational formulation of accelerated optimization on Riemannian manifolds. *SIAM Journal on Mathematics of Data Science*, 4(2):649–674, 2022. URL https://doi.org/10.1137/21M1395648.

[20] V. Duruisseaux, J. Schmitt, and M. Leok. Adaptive Hamiltonian variational integrators and applications to symplectic accelerated optimization. *SIAM Journal on Scientific Computing*, 43(4):A2949–A2980, 2021. URL https://doi.org/10.1137/20M1383835.

[21] G. França, M. I. Jordan, and R. Vidal. On dissipative symplectic integration with applications to gradient-based optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2021 (4):043402. doi: 10.1088/1742-5468/abf5d4.

[22] G. França, A. Barp, M. Girolami, and M. I. Jordan. Optimization on manifolds: A symplectic approach, 07 2021.

[23] J. Gallier and J. Quaintance. *Differential Geometry and Lie Groups: A Computational Perspective*. Geometry and Computing. Springer International Publishing, 2020. ISBN 9783030460402.

[24] B. Gladman, M. Duncan, and J. Candy. Symplectic integrators for long-time integrations in celestial mechanics. *Celestial Mech. Dynamical Astronomy*, 52:221–240, 1991.

[25] E. Hairer. Variable time step integration with symplectic methods. *Applied Numerical Mathematics*, 25(2-3):219–227, 1997.

[26] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[27] J. Hall and M. Leok. Spectral Variational Integrators. *Numer. Math.*, 130(4):681–740, 2015.

[28] I. I. Hussein, M. Leok, A. K. Sanyal, and A. M. Bloch. A discrete variational integrator for optimal control problems on SO(3). *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 6636–6641, 2006.

[29] A. Iserles and S. P. Nørsett. On the solution of linear differential equations in Lie groups. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 357(1754):983–1019, 1999. ISSN 1364503X.

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

[31] M. I. Jordan. Dynamical, symplectic and stochastic perspectives on gradient-based optimization. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 523–549. doi: 10.1142/9789813272880_0022.

[32] J. Jost. *Riemannian Geometry and Geometric Analysis*. Universitext. Springer, Cham, 7th edition, 2017.
[33] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.
[34] S. Lall and M. West. Discrete variational Hamiltonian mechanics. *J. Phys. A*, 39(19):5509–5519, 2006.
[35] J. M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2018.
[36] T. Lee. Computational geometric mechanics and control of rigid bodies. *Ph.D. dissertation, University of Michigan*, 2008.
[37] 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. *Proceedings of 2005 IEEE Conference on Control Applications. (CCA 2005)*, pages 962–967.
[38] 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.
[39] 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.
[40] T. Lee, M. Leok, and N. H. McClamroch. Lagrangian mechanics and variational integrators on two-spheres. *International Journal for Numerical Methods in Engineering*, 79(9):1147–1174, 2009.
[41] T. Lee, M. Leok, and N. H. McClamroch. *Global Formulations of Lagrangian and Hamiltonian Dynamics on Manifolds: A Geometric Approach to Modeling and Analysis*. Interaction of Mechanics and Mathematics. Springer International Publishing, 2017. ISBN 9783319569536.
[42] T. Lee, M. Tao, and M. Leok. Variational symplectic accelerated optimization on Lie groups. 2021.
[43] M. Leok. Generalized Galerkin variational integrators: Lie group, multiscale, and pseudospectral methods. (preprint, arXiv:math.NA/0508360), 2004.
[44] M. Leok and T. Ohsawa. Variational and geometric structures of discrete Dirac mechanics. *Found. Comput. Math.*, 11(5):529–562, 2011.
[45] M. Leok and T. Shingel. Prolongation-collocation variational integrators. *IMA J. Numer. Anal.*, 32(3):1194–1216, 2012.
[46] M. Leok and J. Zhang. Discrete Hamiltonian variational integrators. *IMA Journal of Numerical Analysis*, 31(4):1497–1532, 2011.
[47] D. Lewis and J. C. Simo. Conserving algorithms for the dynamics of Hamiltonian systems on Lie groups. *Journal of Nonlinear Science*, 4(1):253–299, December 1994. doi: 10.1007/BF02430634.
[48] J. E. Marsden and T. S. Ratiu. *Introduction to mechanics and symmetry*, volume 17 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 1999.
[49] J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numer.*, 10: 357–514, 2001.
[50] H. Z. Munthe-Kaas. Lie-Butcher theory for Runge-Kutta methods. *BIT*, 35:572–587, 1995.
[51] H. Z. Munthe-Kaas. Runge-Kutta methods on Lie groups. *BIT Numerical Mathematics*, 38: 92–111, 1998.
[52] H. Z. Munthe-Kaas. High order Runge-Kutta methods on manifolds. *Applied Numerical Mathematics*, 29:115–127, 1999.
[53] J. Nash. The imbedding problem for Riemannian manifolds. *Annals of Mathematics*, 63(1): 20–63, 1956. ISSN 0003486X.
[54] Y. Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

[55] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004.
[56] N. Nordkvist and A. K. Sanyal. A Lie group variational integrator for rigid body motion in SE(3) with applications to underwater vehicle dynamics. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5414–5419, 2010. doi: 10.1109/CDC.2010.5717622.
[57] J. M. Schmitt and M. Leok. Properties of Hamiltonian variational integrators. *IMA Journal of Numerical Analysis*, 38(1):377–398, 03 2017.
[58] J. M. Schmitt, T. Shingel, and M. Leok. Lagrangian and Hamiltonian Taylor variational integrators. *BIT Numerical Mathematics*, 58:457–488, 2018. doi: 10.1007/s10543-017-0690-9.
[59] W. Su, S. Boyd, and E. Candes. A differential equation for modeling Nesterov's Accelerated Gradient method: theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.
[60] M. Tao and T. Ohsawa. Variational optimization on Lie groups, with examples of leading (generalized) eigenvalue problems. In *Proceedings of the 23rd International AISTATS Conference*, volume 108 of *PMLR*, 2020.
[61] T. Tieleman and G. Hinton. Coursera: Neural Networks for Machine Learning - Lecture 6.5: RMSprop. 2012.
[62] G. Wahba. A Least Squares Estimate of Satellite Attitude. *SIAM Review*, 7(3):409, July 1965. doi: 10.1137/1007077.
[63] H. Whitney. The singularities of a smooth n-manifold in $(2n-1)$-space. *Annals of Mathematics*, 45(2):247–293, 1944. ISSN 0003486X.
[64] H. Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *Annals of Mathematics*, 45(2):220–246, 1944. ISSN 0003486X.
[65] A. Wibisono, A. Wilson, and M. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.
[66] H. Yoshimura and J. E. Marsden. Dirac structures in Lagrangian mechanics Part I: Implicit Lagrangian systems. *Journal of Geometry and Physics*, 57(1):133–156, 2006.
[67] H. Yoshimura and J. E. Marsden. Dirac structures in Lagrangian mechanics Part II: Variational structures. *Journal of Geometry and Physics*, 57(1):209–250, 2006.
[68] A. Zanna. Collocation and relaxed collocation for the Fer and the Magnus expansions. *SIAM Journal on Numerical Analysis*, 36(4):1145–1182. ISSN 00361429.
[69] K. Zare and V. G. Szebehely. Time transformations in the extended phase-space. *Celestial mechanics*, 11:469–482, 1975.