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. [44] and Duruisseaux and Leok [8]. 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. [9] 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 [34], and was shown to converge in $O(1/k^2)$ to the minimum of the convex objective function $f$, improving on the $O(1/k)$ convergence rate exhibited by the standard gradient descent methods. This $O(1/k^2)$ convergence rate was shown in [35] to be optimal among first-order methods using only information about $\nabla f$ 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 [39] that Nesterov’s accelerated gradient method limits, as the time-step goes to 0, to a second-order ODE and that the objective function $f(x(t))$ converges to its optimal value at a rate of $O(1/t^2)$ along the trajectories of this ODE. It was later shown in [44] that in continuous time, the convergence rate of $f(x(t))$ can be accelerated to an arbitrary convergence rate $O(1/t^p)$ 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 variational framework was later extended to the Riemannian manifolds setting in [8]. This variational framework and the time-rescaling property of the Bregman family were exploited on vector spaces [9] and Riemannian manifolds [6; 7], by using time-adaptive geometric integrators to 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 [13]). 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 [4; 11]). There has been a great effort to circumvent this problem, and there have been many successes, including methods based on the Poincaré transformation [12; 47]: 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 [9], 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 [14; 32], 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 [9].

This time-adaptivity based approach relying on the Poincaré transformation was also extended to the Riemannian manifold setting in [8], and exploited using discrete variational integrators incorporating holonomic constraints [6] and projection-based variational integrators [7]. 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 [33; 42; 43] 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 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 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 [2; 14; 15; 20–23; 27; 36] and homogeneous space [24] 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.
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

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

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 $\delta q$ of the trajectory $q$ that vanishes at the endpoints. Given local coordinates $(q^1, \ldots, q^n)$ on the manifold $\mathcal{Q}$, Hamilton’s variational principle can be shown to be equivalent to the Euler–Lagrange equations,

$$ \sum_{j=1}^n p_j \dot{q}^j - L(q, \dot{q}) \bigg|_{p_k = \frac{\partial L}{\partial q^k}} = 0, \quad (2.2) $$

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

$$ H(q, p) = \langle \mathcal{L}(q, \dot{q}), \dot{q} \rangle - L(q, \dot{q}) \bigg|_{p_k = \frac{\partial L}{\partial q^k}} = 0, \quad (2.3) $$

where $p_k = \frac{\partial L}{\partial q^k} \in T^*\mathcal{Q}$ is the conjugate momentum of $q^k$. A Hamiltonian $H$ is called hyperregular if $\mathcal{H} : T^*\mathcal{Q} \to T\mathcal{Q}$ defined by $\mathcal{H}(\alpha) : \beta = \frac{d}{ds} |_{s=0} H(\alpha + s\beta)$, is a diffeomorphism. Hyperregularity of the Hamiltonian $H$ implies invertibility of the Hessian matrix $\frac{\partial^2 H}{\partial p^j \partial p^k}$ and thus nondegeneracy of $H$. Theorem 7.4.3 in [31] 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,

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

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-term near-energy preservation (see [32]).

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

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

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 $L^E_d$, and then applying the discrete Euler–Lagrange equations,

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

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
A Type II Hamiltonian variational integrator is constructed by using an approximate discrete Lagrangian, \( L^E_d : Q \times Q \to \mathbb{R} \), approximates the exact discrete Lagrangian \( L^E : Q \times Q \to \mathbb{R} \) to order \( r \), i.e.,

\[
L_d(q_0, q_1; h) = L^E_d(q_0, q_1; h) + O(h^{r+1}),
\]

then the discrete Hamiltonian map \( \tilde{F}_d : (q_k, p_k) \to (q_{k+1}, p_{k+1}) \), 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 \cite{30, 32}). 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,

\[
H^+_{d,E}(q_0, p_1; h) = p_1^T q_1 - \int_0^h \left[ p(t)^T \dot{q}(t) - H(q(t), p(t)) \right] dt,
\]

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 \( H^+_{d,E} \), and applying the discrete right Hamilton’s equations,

\[
p_0 = D_1 H^+_{d,E}(q_0, p_1), \quad q_1 = D_2 H^+_{d,E}(q_0, p_1),
\]

which implicitly defines the integrator, \( \tilde{F}_{H^+_{d,E}} : (q_0, p_0) \to (q_1, p_1) \).

Theorem 2.3.1 of \cite{32}, which simplified the error analysis for Lagrangian variational integrators, has an analogue for Hamiltonian variational integrators. Theorem 2.2 in \cite{37} states that if a discrete right Hamiltonian \( H^+_{d,E} \) approximates the exact discrete right Hamiltonian \( H^+_{d,E} \) to order \( r \), i.e.,

\[
H^+_{d,E}(q_0, p_1; h) = H^+_{d,E}(q_0, p_1; h) + O(h^{r+1}),
\]

then the discrete right Hamilton’s map \( \tilde{F}_{H^+_{d,E}} : (q_k, p_k) \to (q_{k+1}, p_{k+1}) \), 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 \cite{9, 30}).

Examples of variational integrators include Galerkin variational integrators \cite{30, 32}, Prolongation-Collocation variational integrators \cite{29}, and Taylor variational integrators \cite{38}. In many cases, the Type I and Type II/III approaches will produce equivalent integrators. This equivalence has been established in \cite{38} for Taylor variational integrators provided the Lagrangian is hyperregular, and in \cite{30} 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 \cite{37} that even if 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 \cite{9} relies on a degenerate Hamiltonian which has no associated Lagrangian description. We will therefore 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 from the Lagrangian perspective.

We now describe the construction of Taylor variational integrators as introduced in \cite{38} 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 the ρ-order and (ρ + 1)-order Taylor methods \( \Psi_h^{(\rho)} \) and \( \Psi_h^{(\rho+1)} \) approximating the exact time-\( h \) flow map \( \Phi_h : T^*Q \to T^*Q \) corresponding to the Euler–Lagrange equation or Hamilton’s equation. 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 \( i = 1, \ldots, m \), 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 \( (q_{c_i}, v_{c_i}) = \Psi_h^{(\rho)}(q_{c_i h}) \).

(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,

\[
p_0 = -D_1 L_d(q_0, q_1), \quad p_1 = D_2 L_d(q_0, q_1).
\]

**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 \( (q_{c_i}, p_{c_i}) \approx (q(c_i h), \dot{p}(c_i h)) \) via \( (q_{c_i}, p_{c_i}) = \Psi_h^{(\rho)}(q_{c_i h}) \).

(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 \( \dot{c}_i = \frac{\partial H}{\partial p_{c_i}} \).

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

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

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

\[
q_1 = D_2 H_d^*(q_0, p_1), \quad p_0 = D_1 H_d^*(q_0, p_1).
\]

The following error analysis result concerning the order of accuracy of LTVis and HTVis were derived in [38] and [9]:

**Theorem 2.1.** If the Lagrangian \( L \) is Lipschitz continuous in both variables, 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 [32], the associated discrete Hamiltonian map has the same order of accuracy.

**Theorem 2.2.** If the Hamiltonian \( H \) and its partial derivative \( \frac{\partial H}{\partial p} \) are Lipschitz continuous in both variables, 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 [37], 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 [9; 38] 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 [13]), 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 [4; 11]). Backward error
We can use a symplectic integrator with constant time-step in fictive time \( \tau \) with the same initial values, so they must be the same. Note that the Hessian is given by \( \bar{H} \), which will be singular in many cases. The degeneracy of the Hamiltonian \( \bar{H} \) will be referred to as the fictive time, \( \tau \), in the extended phase space defined by \( \bar{q} = \begin{bmatrix} q \\ \theta \end{bmatrix} \in \mathbb{Q} \) and \( \bar{p} = \begin{bmatrix} p \\ \sigma \end{bmatrix} \) where \( p \) is the conjugate momentum for \( q = t \) with \( p(0) = -\bar{H}(q(0),0,p(0)) \). The corresponding equations of motion in the extended phase space are then given by

\[
\dot{\bar{q}}' = \frac{\partial \bar{H}}{\partial \bar{p}}, \quad \dot{\bar{p}}' = -\frac{\partial \bar{H}}{\partial \bar{q}}.
\]

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 \( \bar{H} \). Then

\[
\bar{H}(\bar{Q}(\tau), \bar{P}(\tau)) = \bar{H}(\bar{Q}(0), \bar{P}(0)) = 0.
\]

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

\[
H(Q(\tau), \tau, P(\tau)) = -p(\tau), \quad H(Q(0),0,P(0)) = -p(0) = H(q(0),0,p(0)) \tag{2.15}
\]

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

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

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 \( \tau \) 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 [47], and extended to the case where \( g \) can also depend on time based on ideas from [12]. 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 a reverse-engineering strategy for the Poincaré transformation and present a new way to think about the transformation.

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

\[
\frac{dt}{d\tau} = g(q, t, p),
\]

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

\[
\mathcal{S}(\bar{q}(\cdot), \bar{p}(\cdot)) = \bar{p}(T) \bar{q}(T) - \int_0^T \left[ \bar{p}(t) \dot{\bar{q}}(t) - H(q(t), t, p(t)) - \bar{p}(t) \right] dt
\]

\[
= \bar{p}(T) \bar{q}(T) - \int_{\tau(t=0)}^{\tau(t=T)} \left[ \bar{p}(\tau) \frac{d\tau}{dt} \dot{\bar{q}}(\tau) - H(q(\tau), \bar{q}(\tau), \bar{p}(\tau)) - \bar{p}(\tau) \right] \frac{dt}{d\tau} d\tau
\]

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

\[
\mathcal{S}(\bar{q}(\cdot), \bar{p}(\cdot)) = \bar{p}(T) \bar{q}(T) - \int_{\tau(t=0)}^{\tau(t=T)} \{ \bar{p}(\tau) \dot{\bar{q}}(\tau) - g(q(\tau), q(\tau), p(\tau))[H(q(\tau), q(\tau), p(\tau)) + \bar{p}(\tau)] \} d\tau.
\]

Computing the variation of \(\mathcal{S}\) yields

\[
\delta \mathcal{S} = \bar{q}(T) \delta \bar{p}(T) + \bar{p}(T) \delta \bar{q}(T) - \int_{\tau(t=0)}^{\tau(t=T)} \left[ q' \delta p + p \delta q' - \left( g \frac{\partial H}{\partial q} + \frac{\partial q}{\partial q} (H + q) \right) \delta q - g \delta p \right] d\tau
\]

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

\[
\delta \mathcal{S} = \int_{\tau(t=0)}^{\tau(t=T)} \left[ \bar{p}' + g \frac{\partial H}{\partial q} + \frac{\partial q}{\partial q} (H + p) \right] \delta q d\tau + \int_{\tau(t=0)}^{\tau(t=T)} \left[ \frac{\partial H}{\partial p} + \frac{\partial q}{\partial p} (H + p) - q' \right] \delta p d\tau
\]

Thus, the condition that \(\mathcal{S}(\bar{q}(\cdot), \bar{p}(\cdot))\) 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,

\[
q' = g(q, q, p),
\]

\[
p' = -g(q, q, p) \frac{\partial H}{\partial q} (q, q, p) - \frac{\partial q}{\partial q} (q, q, p) [H(q, q, p) + p],
\]

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,

\[
g(q(\tau), q(\tau), p(\tau))[H(q(\tau), q(\tau), p(\tau)) + p(\tau)],
\]

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 $\tilde{q} = (q, q, \lambda) \in \tilde{Q}$ where $q = t$ and $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Consider the action functional $\mathcal{S}: C^2([0, T], T\tilde{Q}) \to \mathbb{R}$ given by

$$\mathcal{S}(\tilde{q}(\cdot), \dot{\tilde{q}}(\cdot)) = \int_0^T \left[ L(q(t), \dot{q}(t), q(t)) - \lambda(t) \left( \frac{dq}{dt} - g(q(t)) \right) \right] dt,$$

where $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Using the equation \[\dot{q} = g(q)\]

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics.

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 $\tilde{q} = (q, q, \lambda) \in \tilde{Q}$ where $q = t$ and $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Consider the action functional $\mathcal{S}: C^2([0, T], T\tilde{Q}) \to \mathbb{R}$ given by

$$\mathcal{S}(\tilde{q}(\cdot), \dot{\tilde{q}}(\cdot)) = \int_0^T \left[ L(q(t), \dot{q}(t), q(t)) - \lambda(t) \left( \frac{dq}{dt} - g(q(t)) \right) \right] dt,$$

where $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Using the equation \[\dot{q} = g(q)\]

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics.

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 $\tilde{q} = (q, q, \lambda) \in \tilde{Q}$ where $q = t$ and $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Consider the action functional $\mathcal{S}: C^2([0, T], T\tilde{Q}) \to \mathbb{R}$ given by

$$\mathcal{S}(\tilde{q}(\cdot), \dot{\tilde{q}}(\cdot)) = \int_0^T \left[ L(q(t), \dot{q}(t), q(t)) - \lambda(t) \left( \frac{dq}{dt} - g(q(t)) \right) \right] dt,$$

where $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Using the equation \[\dot{q} = g(q)\]

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics.

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 $\tilde{q} = (q, q, \lambda) \in \tilde{Q}$ where $q = t$ and $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Consider the action functional $\mathcal{S}: C^2([0, T], T\tilde{Q}) \to \mathbb{R}$ given by

$$\mathcal{S}(\tilde{q}(\cdot), \dot{\tilde{q}}(\cdot)) = \int_0^T \left[ L(q(t), \dot{q}(t), q(t)) - \lambda(t) \left( \frac{dq}{dt} - g(q(t)) \right) \right] dt,$$

where $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Using the equation \[\dot{q} = g(q)\]

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics.

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 $\tilde{q} = (q, q, \lambda) \in \tilde{Q}$ where $q = t$ and $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Consider the action functional $\mathcal{S}: C^2([0, T], T\tilde{Q}) \to \mathbb{R}$ given by

$$\mathcal{S}(\tilde{q}(\cdot), \dot{\tilde{q}}(\cdot)) = \int_0^T \left[ L(q(t), \dot{q}(t), q(t)) - \lambda(t) \left( \frac{dq}{dt} - g(q(t)) \right) \right] dt,$$

where $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Using the equation \[\dot{q} = g(q)\]

We now introduce a discrete variational formulation of these continuous Lagrangian mechanics.

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 $\tilde{q} = (q, q, \lambda) \in \tilde{Q}$ where $q = t$ and $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Consider the action functional $\mathcal{S}: C^2([0, T], T\tilde{Q}) \to \mathbb{R}$ given by

$$\mathcal{S}(\tilde{q}(\cdot), \dot{\tilde{q}}(\cdot)) = \int_0^T \left[ L(q(t), \dot{q}(t), q(t)) - \lambda(t) \left( \frac{dq}{dt} - g(q(t)) \right) \right] dt,$$

where $\lambda$ is a Lagrange multiplier used to enforce the time rescaling $\frac{dt}{d\tau} = g(t)$. Using the equation \[\dot{q} = g(q)\]
This discrete functional $\tilde{S}_d$ is a discrete analogue of the action functional $\tilde{S} : C^2([0, T], \mathbb{R} \times \mathbb{R} \times \mathbb{R}) \rightarrow \mathbb{R}$ given by

$$\tilde{S}(q(\cdot), q(\cdot), \lambda(\cdot)) = \int_0^T L(q(\tau), q(\tau), \lambda(\tau), q'(\tau), q'(\tau), \lambda'(\tau)) \, d\tau$$  \hspace{1cm} (3.18)

$$= \int_0^T \left[ L(q, g(q), q) - \lambda q' + \lambda g(q) \right] q' \, d\tau. \hspace{1cm} (3.19)$$

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,

$$\delta \tilde{S}_d \left( \{ (q_k, q_k, \lambda_k) \}^{N}_{k=0} \right) = 0,$$  \hspace{1cm} (3.20)

is equivalent to the discrete extended Euler–Lagrange equations,

$$q_{k+1} = q_k + (\tau_{k+1} - \tau_k) g(q_k),$$  \hspace{1cm} (3.21)

$$\frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, q_{k+1}) - \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, q_k) = 0,$$  \hspace{1cm} (3.22)

$$\left[ D_2 L_{d_k} + \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} - \frac{1}{\tau_{k+1} - \tau_k} \left[ L_{d_k} - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k g(q_k) \right]$$

$$+ \left[ D_4 L_{d_k} - \lambda_{k-1} \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} + \lambda_{k-1} g(q_{k-1}) \right] = 0,$$  \hspace{1cm} (3.23)

where $L_{d_k}$ denotes $L_d(q_k, q_{k+1}, q_k, q_{k+1}).$

**Proof.** See Appendix A.1. \hfill $\square$

Defining the discrete momenta via the discrete Legendre transformations,

$$p_k = -D_1 L_d(q_k, q_{k+1}, q_{k+1}), \hspace{1cm} (3.24)$$

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

$$p_k = -D_1 L_d(q_k, q_{k+1}, q_{k+1}),$$  \hspace{1cm} (3.25)

$$p_k = -D_2 L_d(q_k, q_{k+1}, q_{k+1}),$$  \hspace{1cm} (3.26)

$$q_{k+1} = q_k + h g(q_k),$$  \hspace{1cm} (3.27)

$$p_{k+1} = \frac{g(q_k)}{g(q_{k+1})} D_3 L_d(q_k, q_{k+1}, q_{k+1}),$$  \hspace{1cm} (3.28)

$$p_{k+1} = \frac{L_{d_k} - L_{d_{k+1}}}{h g(q_{k+1})} + \frac{\lambda_{k+1}}{h} + \lambda_{k+1} \nabla g(q_{k+1}) + \frac{g(q_k)}{g(q_{k+1})} \left[ D_4 L_{d_k} - \frac{\lambda_k}{h} \right].$$  \hspace{1cm} (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 $t \rightarrow \tau,$

$$\frac{dt}{d\tau} = g(t),$$  \hspace{1cm} (3.30)
Theorem 3.3. If \((\bar{q}(\tau), \bar{q}'(\tau))\) satisfies the Euler–Lagrange equations corresponding to the Lagrangian \(L\), then its components satisfy \(\frac{d}{d \tau} \frac{\partial L}{\partial \bar{q}} (q, \bar{q}, t) = \frac{\partial L}{\partial q} (q, \bar{q}, t)\).

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

\[ q' = g(q), \]

and

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

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

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

\[ \tilde{\mathcal{S}}_d \left( \{ (q_k, q_k, \lambda_k) \}_{k=0}^{N} \right) = \sum_{k=0}^{N-1} \left\{ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \left[ L_d(q_k, q_k, q_{k+1}, q_{k+1}) - \lambda_k \right] + \lambda_k g(q_k) \right\}, \quad (3.33) \]

where,

\[ L_d(q_k, q_k, q_{k+1}, q_{k+1}) \approx \max_{(q, q') \in \mathcal{C}^2([\tau_k, \tau_{k+1}], \Omega \times \mathbb{R})} \int_{\tau_k}^{\tau_{k+1}} L(q, q', \lambda) d\tau. \quad (3.34) \]

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

\[ \tilde{\mathcal{S}}(q(\cdot), q(\cdot), \lambda(\cdot)) = \int_0^T \tilde{L}(q(\tau), q(\tau), \lambda(\tau), \dot{q}(\tau), \dot{q}'(\tau), \lambda'(\tau)) d\tau \]

\[ = \int_0^T \left\{ \dot{q}' \left[ L(q, \frac{q'}{\lambda(q)}), q \right] - \lambda g(q) \right\} d\tau. \quad (3.35) \]

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,
\[ \delta \hat{S}_d \left( \{(q_k, q_k, \lambda_k)\}_{k=0}^{N} \right) = 0, \]  
(3.37)
is equivalent to the discrete extended Euler–Lagrange equations,
\[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, q_k, q_{k+1}, q_{k+1}) + \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} D_3 L_d(q_{k-1}, q_{k-1}, q_k, q_k) = 0, \]  
(3.38)
\[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_2 L_{d_k} = \frac{L_{d_k}}{\tau_{k+1} - \tau_k} + \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} D_4 L_{d_{k-1}} + \frac{L_{d_{k-1}}}{\tau_k - \tau_{k-1}} = \frac{\lambda_{k-1}}{\tau_k - \tau_{k-1}} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} - \lambda_k \nabla g(q_k), \]  
(3.40)
where \( L_{d_k} \) denotes \( L_d(q_k, q_k, q_{k+1}, q_{k+1}) \).

Proof. See Appendix A.2. \(\square\)

Defining the discrete momenta via the discrete Legendre transformations,
\[ p_k = -D_1 L_d(q_k, q_k, q_{k+1}, q_{k+1}), \quad p_k = -D_2 L_d(q_k, q_k, q_{k+1}, q_{k+1}), \]  
(3.41)
and using a constant time-step \( h \) in \( \tau \), the discrete Euler–Lagrange equations can be rewritten as
\[ p_k = -D_1 L_d(q_k, q_k, q_{k+1}, q_{k+1}); \]  
(3.42)
\[ p_k = -D_2 L_d(q_k, q_k, q_{k+1}, q_{k+1}); \]  
(3.43)
\[ q_{k+1} = q_k + h g(q_k), \]  
(3.44)
\[ p_{k+1} = \frac{g(q_k)}{g(q_{k+1})} D_3 L_d(q_k, q_k, q_{k+1}, q_{k+1}); \]  
(3.45)
\[ p_{k+1} = \frac{L_{d_k} - L_{d_{k-1}} + \lambda_{k+1} - \lambda_k + h \lambda_{k+1} \nabla g(q_{k+1}) + h g(q_k) D_4 L_{d_k}}{h g(q_{k+1})}. \]  
(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 \( q' = g(q, q, p) \) 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 [6; 32]).

The Poincaré transformation on the Hamiltonian side was presented in [9; 12; 47] 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 \( (q, q, p, t) \). Unlike the case where \( g = g(t) \), the corresponding Euler–Lagrange equation \( \frac{d}{dt} \frac{\partial L}{\partial q'} = \frac{\partial L}{\partial q'} \frac{\partial q'}{\partial q} (q, q, t) \) in the original phase-space,
\[ \frac{d}{dt} \frac{\partial L}{\partial q'} (q, \dot{q}, t) - \frac{\partial L}{\partial q'} (q, \dot{q}, t) = \lambda(t) \frac{\partial g}{\partial q'} (q, t). \]  
(3.47)
The discrete Euler–Lagrange equations become more complicated and involve terms with partial derivatives \( D_1 g(q_k, q_k) \) of \( g \) with respect to \( q \). Furthermore, when \( g = g(q, t) \), the discrete Euler–Lagrange equations involve \( \lambda_k \) but the time-evolution of the Lagrange multiplier \( \lambda \) is not well-defined,
so the discrete Hamiltonian map corresponding to the discrete Lagrangian $L_d$ is not well-defined, as explained in [32, 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 [9]). 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 [28; 45; 46] on the Pontryagin bundle $(q, v, p) \in TQ \oplus T^*Q$. Dirac dynamics are described by the Hamilton-Pontryagin variational principle where the momentum $p$ acts as a Lagrange multiplier to impose the kinematic equation $\dot{q} = v,$

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

(3.48)

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

$$\dot{q} = v, \quad \dot{p} = \frac{\partial L}{\partial q}, \quad p = \frac{\partial L}{\partial v},$$

(3.49)

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

$$E(q, v, p, t) = pv - L(q, v, t),$$

(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 [44] for accelerated optimization on normed vector spaces. The $p$-Bregman Lagrangians and Hamiltonians are defined to be

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

(4.1)

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

(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 of time $t$. In [44], it was shown that solutions to the $p$-Bregman Euler–Lagrange equations converge to a minimizer of $f$ at a convergence rate of $O(1/t^p)$. Furthermore, this family of Bregman dynamics is closed under time dilation: time-rescaling a solution to the $p$-Bregman Euler–Lagrange equations via $\tau(t) = t^{p/p}$ yields a solution to the $\tilde{p}$-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 [9], 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 $\tilde{p}$-Bregman Hamiltonian in extended phase-space, where $\tilde{p} < p$. This allowed integrating the higher-order $p$-Bregman dynamics while benefiting from the computational efficiency of integrating the lower-order $\tilde{p}$-Bregman dynamics.
Explicitly, using the time rescaling $\tau(t) = \frac{\dot{p}}{p}$ within the Poincaré transformation framework yields the adaptive approach $p \to \dot{p}$-Bregman Hamiltonian,

$$\bar{H}_{p\to p}(\bar{q}, \bar{r}) = \frac{p^2}{2p\bar{q}^2+\dot{p}^2}(r, r) + C\frac{p^2}{\bar{p}} q^{2p-\dot{p}} f(q) + \frac{p}{\bar{p}} q^{1-\dot{p}} f,$$

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

$$\bar{H}_{p}(\bar{q}, \bar{r}) = \frac{p}{2q^{p+1}}(r, r) + Cpq^{2p-1} f(q) + r.$$  \hspace{1cm} (4.4)

In [9], 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,

$$\bar{L}_p(q, q', q) = \frac{q^{p+1}}{2p} (q', q') - Cpq^{2p-1} f(q),$$

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,

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

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

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

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 $\rho$ 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 $\dot{q}(0) = \ddot{v}_0$ by the solution $\ddot{v}_0$ of the problem $\ddot{q}_1 = \pi_Q o \Psi_h^{(1)}(\ddot{q}_0, \ddot{v}_0) = \ddot{q}_0 + \dot{h} \ddot{v}_0$, that is $\ddot{v}_0 = \ddot{q}_1 - \ddot{q}_0$. Then, applying the quadrature rule gives the associated discrete Lagrangian,

$$L_d(\ddot{q}_0, \ddot{q}_1) = hL_p \left(q_0, \frac{1}{g(q_0)} \ddot{v}_0 \right) = \frac{q_0^{p+1}}{2p(g(q_0))^2} h(\ddot{v}_0, \ddot{v}_0) - Chpq_0^{2p-1} f(q_0).$$

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 $r$, and since it will not appear in the updates for the other variables, the discrete equations from Sections 3.2 and 3.3 both reduce to the same updates,

$$r_k = -D_1 L_d(q_k, q_k, q_k, q_k)$$

$$r_{k+1} = g(q_k) D_3 L_d(q_k, q_k, q_k, q_k)$$

$$q_{k+1} = q_k + h g(q_k).$$

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

$$L_d(q_k, q_k, q_k, q_k) = \frac{\tilde{p}^2}{2h^p q_k} q_k^{p-1+2p/p} (q_k - q_k - q_k) - Chp q_k^{2p-1} f(q_k),$$

yields the adaptive LTVI algorithm,

$$q_{k+1} = q_k + h \frac{p}{\tilde{p}} q_k^{1-\tilde{p}/p}$$

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

$$r_{k+1} = \frac{p^2 q_k^{p+4/p}}{h^p q_k^{1-p/p}} (q_k - q_k).$$

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

$$q_{k+1} = q_k + h,$$

$$q_{k+1} = q_k + \frac{hp^3}{q_k^{p-1+2p/p}} r_k - Ch^2 p^2 q_k^{p-2} \nabla f(q_k),$$

$$r_{k+1} = \frac{q_k^{p+1}}{h^p} (q_k - q_k).$$

### 4.2.2. A Hamiltonian Taylor Variational Integrator (HTVI)

In [9], a Type II Hamiltonian Taylor Variational Integrator (HTVI) was derived following the strategy from Section 2.2 with $\rho = 0$ for the adaptive approach $p \rightarrow \tilde{p}$-Bregman Hamiltonian,

$$\tilde{H}_{p \rightarrow \tilde{p}}(\tilde{q}, \tilde{r}) = \frac{p^2}{2hpq^{p+1/p}} (r, r) + \frac{Cp^2}{\tilde{p}} q^{2p-\tilde{p}/\tilde{p}} f(q) + \frac{p}{\tilde{p}} \tilde{r} q^{1-\tilde{p}/\tilde{p}}.$$

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

$$q_{k+1} = q_k + h \frac{p}{\tilde{p}} q_k^{1-\tilde{p}/p}$$

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

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

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

$$q_{k+1} = q_k + h,$$

$$r_{k+1} = r_k - Ch \tilde{q}_k^{2p-1} \nabla f(q_k),$$

$$q_{k+1} = q_k + h \tilde{p} q_k^{p-1} r_{k+1}.$$
4.3. Numerical Experiments. Numerical experiments using the numerical methods presented in the previous section have been conducted to minimize the quartic function,

\[ f(x) = \left( (x - 1)^\top \Sigma(x - 1) \right)^2, \] (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 [9] 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 \( t = \tau^{p/\beta} \) 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 [9].

![Figure 1. Comparison of the direct and adaptive approaches for the LTVI algorithm, when applied to the quartic function (4.26).](image)

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 [9] for the HTVI algorithm extends to the LTVI algorithm. In particular, this means that the LTVI algorithm is much more efficient than non-adaptive non-symplectic and adaptive non-symplectic integrators for the Bregman dynamics. It also means that the adaptive LTVI method can be a competitive first-order explicit algorithm since it can outperform certain popular optimization algorithms such as Nesterov’s Accelerated Gradient [34], Trust Region Steepest Descent, ADAM [17], AdaGrad [5], and RMSprop [40], for certain choices of objective functions.
5. ACCELERATED OPTIMIZATION ON MORE GENERAL SPACES

5.1. Motivation and Prior Work. The variational framework for accelerated optimization on normed vector spaces from [9; 44] was extended to the Riemannian manifold setting in [8] via a Riemannian $p$-Bregman Lagrangian $\mathcal{L}_p : T^*Q \times \mathbb{R} \rightarrow \mathbb{R}$ and a corresponding Riemannian $p$-Bregman Hamiltonian $\mathcal{H}_p : T^*Q \times \mathbb{R} \rightarrow \mathbb{R}$, for $p > 0$, of the form

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

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

where $\xi$ and $\lambda$ 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,

$$\nabla_X \dot{X} + \frac{\xi p + \lambda}{\lambda t} \dot{X} + Cpt^{2p^{-2}} \text{grad} f(X) = 0. \quad (5.3)$$

Here, grad$f$ denotes the Riemannian gradient of $f$, $\nabla_X Y$ is the covariant derivative of $Y$ along $X$, and $\langle \cdot, \cdot \rangle$ is the fiber metric on $T^*Q$ induced by the Riemannian metric $\langle \cdot, \cdot \rangle$ on $Q$ whose local coordinates representation is the inverse of the local representation of $\langle \cdot, \cdot \rangle$. See [1; 3; 8; 16; 19; 31] for a more detailed description of these notions from Riemannian geometry and of this Riemannian variational framework for accelerated optimization.

It was shown in [8] that solutions to the $p$-Bregman Euler–Lagrange equations converge to a minimizer of $f$ at a convergence rate of $O(1/t^p)$, under suitable assumptions, and proven that time-rescaling a solution to the $p$-Bregman Euler–Lagrange equations via $\tau(t) = t^{\frac{1}{p} + 1}$ yields a solution to the $p$-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 $p \rightarrow \bar{p}$ Bregman Hamiltonian,

$$\mathcal{H}_{p \rightarrow \bar{p}}(\bar{Q}, \bar{R}) = \frac{p^2}{2\bar{p}\bar{\Omega}^{\frac{2p+2}{p}}} \langle \bar{R}, \bar{R} \rangle + \frac{Cp^2}{\bar{p}} \bar{\Omega}^{\xi + 1}p^{-\xi} f(\bar{Q}) + \frac{p}{\bar{p}} \bar{\Omega}^{1-\xi} \mathcal{R}. \quad (5.4)$$

This adaptive framework was exploited using discrete variational integrators incorporating holonomic constraints [6] and projection-based variational integrators [7]. Both these strategies

![Figure 2. Comparison of the HTVI and LTVI algorithms with the same parameters.](image-url)
relied on embedding the Riemannian manifolds into an ambient Euclidean space. Although the Whitney and Nash Embedding Theorems \[33; 42; 43]\] 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 $H^\tau_d(q_k,p_{k+1}),\ H^\tau_d(p_k,q_{k+1})$, 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 \[28; 45; 46]\].

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 \[2; 14; 15; 20–23; 27; 36\] and homogeneous space \[24\] 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. More precisely, we will work in the setting of \[26\]. Consider a $n$-dimensional Lie group $G$ with associated Lie algebra $\mathfrak{g} = T_1G$, and a left-trivialization of the tangent bundle of the group $TG := G \times \mathfrak{g}$, via $(q,q) \mapsto (q, L_{q^{-1}} q) \equiv (q, \xi)$, where $L : G \times G \rightarrow G$ is the left action defined by $L_q h = q h$ 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,

$$(v \cdot w)_{T_qG} = (T_q L_{q^{-1}} v \cdot T_q L_{q^{-1}} w)_{\mathfrak{g}}, \quad \forall v, w \in T_qG.$$  

With this inner product, we identify $\mathfrak{g} \cong \mathfrak{g}^*$ and $T_qG \cong T_q^*G \cong G \times \mathfrak{g}^*$ via the Riesz representation. Let $J : \mathfrak{g} \rightarrow \mathfrak{g}^*$ be chosen such that $(J(\xi) \cdot \zeta) = \delta(\xi, \zeta)$, where $\delta(\cdot, \cdot) : \mathfrak{g} 
\times \mathfrak{g} \rightarrow \mathbb{R}$ with $\delta(\xi, \zeta) = (J(\xi) \cdot \zeta)$ serves as a left-invariant Riemannian metric on $G$. The adjoint and ad operators are denoted by $Ad_q : \mathfrak{g} \rightarrow \mathfrak{g}$ and $ad_\xi : \mathfrak{g} \rightarrow \mathfrak{g}$, respectively. See \[10; 25; 31\] for a more detailed description of Lie group theory and mechanics on Lie groups.

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

$$q_{k+1} = q_k f_k,$$  

(5.5)

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

Consider the discrete action functional,

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

(5.6)

where,

$$L_d(q_k, f_k, q_k, q_{k+1}) \equiv \frac{\text{ext}}{(q,q) \in \mathfrak{C}^2([\tau_k,\tau_{k+1}],G \times \mathbb{R})} \int_{\tau_k}^{\tau_{k+1}} L \left( q, \frac{\xi(q)}{g(q)}, q \right) d\tau.$$  

(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,

$$\delta \mathcal{G}_d \left( \{(q_k, p_k, \lambda_k)\}_{k=0}^N \right) = 0,$$

where,

$$\mathcal{G}_d \left( \{(q_k, p_k, \lambda_k)\}_{k=0}^N \right) = \sum_{k=0}^{N-1} L_d(q_k, f_k, q_k, q_{k+1}) - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k g(q_k),$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$q_{k+1} = q_k + (\tau_{k+1} - \tau_k)g(q_k),$$

$$\text{Ad}^*_f \left( T^*_e L^r D_2 L_{d_k} \right) = T^*_e L^r D_1 L_{d_k} + \frac{\tau_{k+1} - \tau_k}{\tau_{k+1} - \tau_k} \left( q_{k+1} - q_k, \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k g(q_k) \right) \left( L_{d_k} - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k g(q_k) \right) + \frac{1}{\tau_{k+1} - \tau_k} \left( L_{d_{k+1}} - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k g(q_k) \right) = 0,$$

where $L_{d_k}$ denotes $L_d(q_k, f_k, q_k, q_{k+1})$.

*Proof.* See Appendix A.3. □

Now, define

$$p_k = -D_3 L_d(q_k, f_k, q_k, q_{k+1}),$$

$$\mu_k = \text{Ad}^*_f \left( T^*_e L^r D_2 L_d(q_k, f_k, q_k, q_{k+1}) \right) - T^*_e L^r D_1 L_d(q_k, f_k, q_k, q_{k+1}),$$

and thus,

$$\mu_{k+1} = \frac{\tau_{k+2} - \tau_{k+1}}{\tau_{k+2} - \tau_{k+1}} \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} T^*_e L^r D_2 L_d(q_k, f_k, q_k, q_{k+1}).$$

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

$$\mu_k = \text{Ad}^*_f \left( T^*_e L^r f_k D_2 L_d(q_k, f_k, q_k, q_{k+1}) \right) - T^*_e L^r q_k D_1 L_d(q_k, f_k, q_k, q_{k+1}),$$

$$\mu_{k+1} = \frac{1}{\bar{p}} T^*_e L^r f_k D_2 L_d(q_k, f_k, q_k, q_{k+1}),$$

$$q_{k+1} = q_k + \frac{1}{h^p} \frac{1-\bar{p}/p}{\bar{p}} q_k,$$

$$p_{k+1} = \bar{p} \left[ \lambda_{k+1} - \lambda_k + L_{d_k} - L_{d_{k+1}} \right] + \frac{\lambda_{k+1}}{q_{k+1}} \left( 1 - \frac{\bar{p}}{p} \right).$$

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

$$\mathcal{L}_p(g, \xi, t) = \frac{\mu^{p+1}}{2p} (\xi, \xi) - C p t \xi^{p-1} f(q),$$

with corresponding Euler–Lagrange equation,

$$\frac{d J(\xi)}{dt} + \frac{\kappa p + 1}{t} J(\xi) - \text{ad}^*_p J(\xi) + C p t^{p-2} \nabla_t f(q) = 0,$$
where $\nabla_L f$ is the left-trivialized derivative of $f$, given by $\nabla_L f(q) = T_e^* L_q (D_q f(q))$. We then consider the discrete Lagrangian,

$$L_d(q_k, f_k, q_k, q_{k+1}) = \frac{q_k^{p+1}}{h^p g(q_k)} T_d(f_k) - Chp q_k^{(k+1)p-1} f(q_k),$$

(5.22)

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

$$L_d(q_k, f_k, q_k, q_{k+1}) \approx \max_{(q, \xi) \in C^2((\tau_k, \tau_{k+1}) \times \mathbb{F})} \int_{\tau_k}^{\tau_{k+1}} L \left( q, \frac{\xi}{g(q)}, q \right) \, d\tau. \quad (5.23)$$

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

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

Its Lie algebra is

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

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 $\hat{\cdot} : \mathbb{R}^3 \rightarrow \mathfrak{so}(3)$, 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 $(\cdot)^\vee : \mathfrak{so}(3) \rightarrow \mathbb{R}^3$. The inner product on $\mathfrak{so}(3)$ is given by

$$\langle \hat{\eta} \cdot \hat{\xi} \rangle_{\mathfrak{so}(3)} = \frac{1}{2} \text{Trace} \left( \hat{\eta}^T \hat{\xi} \right) = \eta^T \xi, \quad (5.26)$$

and the metric is chosen so that

$$\langle \hat{\eta}, \hat{\xi} \rangle = (J(\hat{\eta}) \cdot \hat{\xi})_{\mathfrak{so}(3)} = \text{Trace} \left( \hat{\eta}^T J_d \hat{\xi} \right) = \eta^T J_d \xi; \quad (5.27)$$

where $J \in \mathbb{R}^{3 \times 3}$ is a symmetric positive-definite matrix and $J_d = \frac{1}{2} \text{Trace}(J) I_{3 \times 3} - J$.

On $SO(3)$, for any $u, v \in \mathbb{R}^3$ and $F \in SO(3)$,

$$\text{ad}_a \hat{v} = [\hat{u}, \hat{v}] = \hat{u} v - \hat{v} \hat{u} = \hat{u} \times v, \quad \text{Ad}_F \hat{u} = F \hat{u} F^T = \hat{F} u. \quad (5.28)$$

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

$$\text{ad}_u \hat{v} = \hat{u} \hat{v} = u \times v, \quad \text{ad}_u^* \hat{v} = -\hat{u} \hat{v} = v \times u, \quad \text{Ad}_F \hat{u} = F \hat{u}, \quad \text{Ad}_F^* \hat{u} = F^T \hat{u}. \quad (5.29)$$

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

$$\mathcal{L}_p(R, \Omega, t) = \frac{t^{p+1}}{2p} \Omega^T J_d \Omega - C p t^{2p-1} f(R), \quad (5.30)$$

and the corresponding Euler–Lagrange equations are given by

$$J \hat{\Omega} + \frac{p+1}{t} J \hat{\Omega} + \hat{\Omega} J \hat{\Omega} + C p^2 t^{p-2} \nabla_L f(R) = 0, \quad \hat{R} = R \hat{\Omega}. \quad (5.31)$$

The discrete kinematics equations is written as

$$R_{k+1} = R_k F_k, \quad (5.32)$$

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

$$L_d(R_k, F_k, R_k, R_{k+1}) = \frac{p^2}{h^{2p}} R_k^{p+2p/2} T_d(F_k) - Chp R_k^{2p-1} f(R_k). \quad (5.33)$$

As in [22; 26], the angular velocity is approximated with $\hat{\Omega}_k \approx \frac{1}{h} R_k^T (R_{k+1} - R_k) = \frac{1}{h} (F_k - I_{3 \times 3})$ so we can take

$$T_d(F_k) = \text{Trace} \left( [I_{3 \times 3} - F_k] J_d \right). \quad (5.34)$$
Differentiating this equation and using the identity Trace\((-\dot{x}A) = (A - A^\top)\cdot x\) yields
\[
T_f^* L F_k \left( D_{F_k} T_d (F_k) \right) = \left( J_d F_k - F_k^\top J_d \right)^\top.
\]
(5.35)
Then, the discrete Euler–Lagrange equations for \(\mu_k\) and \(\mu_{k+1}\) become
\[
\mu_k = \frac{\dot{p}^2}{p^2} \mathfrak{R}_k^{1+2\hat{p}/p} \left( F_k J_d - J_d F_k^\top \right)^\top + Chp \mathfrak{R}_k^{2p-1} \nabla_L f(R_k),
\]
(5.36)
\[
\mu_{k+1} = \frac{q_k^{1-p/p}}{q_{k+1}^{1-p/p}} F_k^\top \left[ \mu_k - Chp \mathfrak{R}_k^{2p-1} \nabla_L f(R_k) \right].
\]
(5.37)
Now, equation (5.36) can be solved explicitly when \(J = I_{3\times3}\) as described in [26]:
\[
F_k = \exp \left( \frac{\sin^{-1} \|a_k\|}{\|a_k\|} \hat{a}_k \right), \quad \text{where} \quad a_k = \frac{hp^3}{\dot{p}^2} \mathfrak{R}_k^{1-p-2\hat{p}/p} \left[ \mu_k - Chp \mathfrak{R}_k^{2p-1} \nabla_L f(R_k) \right].
\]
(5.38)
Therefore, we get the **Adaptive LLGVI** (Adaptive Lagrangian Lie Group Variational Integrator)
\[
F_k = \exp \left( \frac{\sin^{-1} \|a_k\|}{\|a_k\|} \hat{a}_k \right), \quad \text{where} \quad a_k = \frac{hp^3}{\dot{p}^2} \mathfrak{R}_k^{1-p-2\hat{p}/p} \left[ \mu_k - Chp \mathfrak{R}_k^{2p-1} \nabla_L f(R_k) \right],
\]
(5.39)
\[
\mathfrak{R}_{k+1} = \mathfrak{R}_k + \frac{p}{p} \mathfrak{R}_k^{1-p/p},
\]
(5.40)
\[
\mu_{k+1} = \frac{q_k^{1-p/p}}{q_{k+1}^{1-p/p}} F_k^\top \left[ \mu_k - Chp \mathfrak{R}_k^{2p-1} \nabla_L f(R_k) \right],
\]
(5.41)
\[
R_{k+1} = R_k F_k.
\]
(5.42)
We will use this integrator to solve the problem of minimizing the objective function,
\[
f(R) = \frac{1}{2} \|A - R\|_F^2 = \frac{1}{2} \left( \|A\|_F^2 + 3 \right) - \text{Trace}(A^\top R),
\]
(5.43)
over \(R \in SO(3)\), where \(\| \cdot \|_F\) denotes the Frobenius norm. Its left-trivialized gradient is given by
\[
\nabla_L f(R) = \left( A^\top R - R^\top A \right)^\top.
\]
(5.44)
Minimizing this objective function appears in the least-squares estimation of attitude, referred to as Wahba’s problem [41]. The optimal attitude is explicitly given by
\[
R^* = U \text{diag} \left( 1, 1, \det(UV) \right) V^\top,
\]
(5.45)
where \(A = U S V^\top\) is the singular value decomposition of \(A\) with \(U, V \in O(3)\) and \(S\) diagonal.

We have tested the Adaptive LLGVI integrator on Wahba’s problem against the Implicit Lie Group Variational Integrator (**Implicit LGVI**) from [26]. The Implicit LGVI is a Lagrangian Lie group variational integrator which adaptively adjusts the stepsize 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 [26] 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
\[
|f(R_k) - f(R^*)| < \delta \quad \text{and} \quad |f(R_k) - f(R_{k-1})| < \delta.
\]
(5.46)
We can see from Figure 3 that both algorithms preserve the orthogonality condition $R_k^T R_k = I_{3 \times 3}$ 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.
22 VALENTIN DURUISSEAUX AND MELVIN LEOK

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

### Table

<table>
<thead>
<tr>
<th>Termination Criterion $\delta$</th>
<th>$10^{-3}$</th>
<th>$10^{-4}$</th>
<th>$10^{-5}$</th>
<th>$10^{-6}$</th>
<th>$10^{-7}$</th>
<th>$10^{-8}$</th>
<th>$10^{-9}$</th>
<th>$10^{-10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive LGVI: Iterations</td>
<td>350</td>
<td>510</td>
<td>648</td>
<td>722</td>
<td>724</td>
<td>1181</td>
<td>1604</td>
<td>2237</td>
</tr>
<tr>
<td>Implicit LGVI: Iterations</td>
<td>179</td>
<td>474</td>
<td>970</td>
<td>1380</td>
<td>1392</td>
<td>4506</td>
<td>11992</td>
<td>16758</td>
</tr>
<tr>
<td>Adaptive LGVI: Time (in seconds)</td>
<td>0.007</td>
<td>0.009</td>
<td>0.011</td>
<td>0.012</td>
<td>0.012</td>
<td>0.016</td>
<td>0.027</td>
<td>0.034</td>
</tr>
<tr>
<td>Implicit LGVI: Time (in seconds)</td>
<td>0.15</td>
<td>0.36</td>
<td>0.71</td>
<td>1.00</td>
<td>1.01</td>
<td>3.24</td>
<td>9.22</td>
<td>17.55</td>
</tr>
</tbody>
</table>

6. Conclusion

A variational framework for accelerated optimization on vector spaces was introduced [44] by considering a family of time-dependent Bregman Lagrangian and Hamiltonian systems which is closed under time-rescaling. This variational framework was exploited in [9] 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.

These variational framework and time-adaptive approach on the Hamiltonian side were later extended to the Riemannian manifolds setting in [8]. 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 [6; 7] 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 [44]. 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 [9], whenever they are used with the same parameters and time-step. As a result, the computational analysis carried in [9] 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.

ACKNOWLEDGMENTS

The authors were supported in part by NSF under grants DMS-1411792, 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). The authors would also like to thank Taeyoung Lee and Molei Tao for sharing the code for the implicit Lie group variational integrator from [26], which we used in the numerical comparisons with our method.

APPENDIX A. PROOFS OF THEOREMS

A.1. Proof of Theorem 3.2.

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

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

where,

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

is equivalent to the discrete extended Euler–Lagrange equations,

$$ q_{k+1} = q_k + (\tau_{k+1} - \tau_k) g(q_k), $$

$$ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_1 L_d(q_k, q_k, q_{k+1}, q_{k+1}) + \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} D_2 L_d(q_{k-1}, q_{k-1}, q_k, q_k) = 0, $$

$$ \left[ D_2 L_{d_k} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} - \frac{1}{\tau_{k+1} - \tau_k} \left[ L_{d_k} - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k g(q_k) \right] $$

$$ + \left[ D_3 L_{d_{k-1}} - \lambda_{k-1} \frac{1}{\tau_k - \tau_{k-1}} \right] \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} + \frac{1}{\tau_k - \tau_{k-1}} \left[ L_{d_{k-1}} - \lambda_{k-1} \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} + \lambda_{k-1} g(q_{k-1}) \right] = 0, $$

where $L_{d_k}$ denotes $L_d(q_k, q_k, q_{k+1}, q_{k+1})$. 

Proof. We use the notation $L_{dk} = L_d(q_k, q_{k+1}, q_{k+2})$, and we will use the fact that $\delta q_0 = \delta q_N = \delta q_0 = \delta q_N = 0$ throughout the proof.

We have

$$\delta \bar{\mathcal{S}}_d = \sum_{k=0}^{N-1} \left[ L_d(q_k, q_{k+1}, q_{k+2}) - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \tau_{k+1} \tilde{g}(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \frac{1}{\tau_{k+1} - \tau_k}$$

$$= \sum_{k=0}^{N-1} \left[ D_2 L_{dk} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \tilde{g}(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \delta q_k$$

$$+ \sum_{k=0}^{N-2} \left[ D_4 L_{dk} - \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \tilde{g}(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \delta q_k + \sum_{k=0}^{N-1} \left[ L_{dk} - \lambda_k \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} + \lambda_k \tilde{g}(q_k) \right] \delta q_{k+1}$$

$$+ \sum_{k=0}^{N-2} \left[ q_{k+1} - q_k \right] D_4 L_{dk} \delta q_k + \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_3 L_{dk} \delta q_{k+1} + \sum_{k=0}^{N-1} \left( g(q_k) - \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \right) \delta \lambda_k \right]$$

Thus,

$$\delta \bar{\mathcal{S}}_d = \sum_{k=0}^{N-1} \left[ D_2 L_{dk} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \tilde{g}(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \delta q_k$$

$$+ \sum_{k=0}^{N-1} \left[ D_4 L_{dk} - \lambda_k \frac{1}{\tau_{k+1} - \tau_k} + \lambda_k \tilde{g}(q_k) \right] \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \delta q_k$$

$$+ \sum_{k=0}^{N-1} \left[ q_{k+1} - q_k \right] D_4 L_{dk} \delta q_k + \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_3 L_{dk} \delta q_{k+1} + \sum_{k=0}^{N-1} \left( g(q_k) - \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \right) \delta \lambda_k \right]$$

As a consequence, if

$$q_{k+1} = q_k + \left( \tau_{k+1} - \tau_k \right) \tilde{g}(q_k),$$

$$D_2 L_{dk} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} = 0,$$

$$D_4 L_{dk} - \lambda_k \frac{1}{\tau_{k+1} - \tau_k} \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} = 0,$$

then $\delta \bar{\mathcal{S}}_d \left( \{ (q_k, q_{k+1}) \} \right) = 0$. Conversely, if $\delta \bar{\mathcal{S}}_d \left( \{ (q_k, q_{k+1}) \} \right) = 0$, 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,

$$\delta \mathcal{S}_d \left( \{ (q_k, q_{k+1}) \} \right) = 0,$$

where,

$$\mathcal{S}_d \left( \{ (q_k, q_{k+1}, \lambda_k) \} \right) = \sum_{k=0}^{N-1} \left\{ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \right\} \left[ L_d(q_k, q_{k+1}, q_{k+2}) - \lambda_k \right] + \lambda_k \tilde{g}(q_k),$$

is equivalent to the discrete extended Euler–Lagrange equations,

$$q_{k+1} = q_k + \left( \tau_{k+1} - \tau_k \right) \tilde{g}(q_k),$$

$$D_2 L_{dk} + \lambda_k \frac{1}{\tau_{k+1} - \tau_k} \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} = 0,$$

$$D_4 L_{dk} - \lambda_k \frac{1}{\tau_{k+1} - \tau_k} \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} = 0,$$

where $L_{dk}$ denotes $L_d(q_k, q_{k+1}, q_{k+2})$.

Proof. We use the notation $L_{dk} = L_d(q_k, q_{k+1}, q_{k+2})$, and we will use the fact that $\delta q_0 = \delta q_N = \delta q_0 = \delta q_N = 0$ throughout the proof.

We have

$$\delta \mathcal{S}_d = \delta \left( \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \right] \left[ L_d(q_k, q_{k+1}, q_{k+2}) - \lambda_k \right] + \lambda_k \tilde{g}(q_k) \right)$$

$$= \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_2 L_{dk} - \frac{1}{\tau_{k+1} - \tau_k} \frac{L_{dk}}{\tau_{k+1} - \tau_k} + \lambda_k \tilde{g}(q_k) \right] \delta q_k$$

$$+ \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_4 L_{dk} \delta q_k + \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_3 L_{dk} \delta q_{k+1} + \sum_{k=0}^{N-1} \left( g(q_k) - \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} \right) \delta \lambda_k \right] \right.$$

$$\right.$$
Thus,
\[ \delta \mathcal{L}_d = \sum_{k=0}^{N-1} \left[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_k L_{d_k} - \frac{1}{\tau_{k+1} - \tau_k} L_{d_k} + \frac{\lambda_k}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(q_k) + \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} D_k L_{d_{k-1}} + \frac{1}{\tau_k - \tau_{k-1}} L_{d_{k-1}} - \frac{\lambda_{k-1}}{\tau_k - \tau_{k-1}} \right] \delta q_k + \sum_{k=1}^{N-1} \left[ q_{k+1} - q_k \right] \frac{\eta_k}{\tau_{k+1} - \tau_k} D_k L_{d_k} + \frac{q_k - q_{k-1}}{\tau_k - \tau_{k-1}} D_k L_{d_{k-1}} \right] \delta \lambda_k. \]

As a consequence, if
\[ \frac{q_{k+1} - q_k}{\tau_{k+1} - \tau_k} D_k L_{d_k} - \frac{1}{\tau_{k+1} - \tau_k} L_{d_k} + \frac{\lambda_k}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(q_k), \]
\[ q_{k+1} - q_k \frac{\left( \eta_{k+1} - \eta_k \right)}{\tau_{k+1} - \tau_k} D_k L_{d_{k-1}}, \]

then \[ \delta \mathcal{L}_d \left( (q_k, q_k, \lambda_k)_{k=0}^{N} \right) = 0. \] Conversely, if \[ \delta \mathcal{L}_d \left( (q_k, q_k, \lambda_k)_{k=0}^{N} \right) = 0, \] then a discrete fundamental theorem of the calculus of variations yields the above equations.


**Theorem A.3.** The Type I discrete Hamilton’s variational principle,
\[ \delta \mathcal{S}_d \left( (q_k, q_k, \lambda_k)_{k=0}^{N} \right) = 0, \]
where
\[ \mathcal{S}_d \left( (q_k, q_k, \lambda_k)_{k=0}^{N} \right) = \sum_{k=0}^{N-1} \left[ L_d(q_k, f_k, q_k, q_{k+1}) - \frac{\lambda_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \lambda_k g(q_k) \right], \]
is equivalent to the discrete extended Euler–Lagrange equations,
\[ q_{k+1} = q_k + \left( \tau_{k+1} - \tau_k \right) g(q_k), \]
\[ A d \left[ T^*_{\tau_k} L_{f_k} D_k L_{d_k} \right] = T^*_{\tau_k} L_{q_k} D_k L_{d_k} + \frac{\tau_{k+1} - \tau_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \frac{\tau_{k-1} - \tau_k}{\tau_{k-1} - \tau_k} T^*_{\tau_k} L_{f_k} - D_k L_{d_{k-1}}, \]
\[ \left[ D_k L_{d_k} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(q_k), \right] q_{k+1} - q_k - \frac{1}{\tau_{k+1} - \tau_k} \left[ L_{d_k} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \lambda_k g(q_k) \right] \]
\[ + \left[ D_k L_{d_k} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} + \lambda_k \nabla g(q_k), \right] q_{k+1} - q_k - \frac{1}{\tau_{k+1} - \tau_k} \left[ L_{d_k} - \frac{\lambda_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \lambda_k g(q_k) \right] \delta \lambda_k. \]

**Proof.** We use the notation \( L_{d_k} = L_d(q_k, f_k, q_k, q_{k+1}) \) and we will use the fact that \( \delta q_0 = \delta q_N = \delta q_0 = \delta q_N = \eta_0 = \eta_N = 0 \) throughout the proof. We have
\[ \delta \mathcal{S}_d \left( (q_k, q_k, \lambda_k)_{k=0}^{N} \right) = \sum_{k=0}^{N-1} \left[ L_d(q_k, f_k, q_k, q_{k+1}) - \frac{\lambda_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \lambda_k g(q_k) \right] \delta q_k + \sum_{k=0}^{N-1} \left[ f_k - \frac{\lambda_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \lambda_k g(q_k) \right] \delta \lambda_k. \]

We can write \( \delta q_k \) as \( \delta q_k = \eta_k \eta_k \) for some \( \eta_k \in \mathfrak{g} \). Then, taking the variation of the discrete kinematics equation \( q_{k+1} = q_k f_k \) gives the equation \( \delta q_{k+1} = q_k f_k \delta f_k + \delta q_k \) and \( \delta f_k = q_k^* \delta q_k \). Therefore,
\[ f_k = q_k^* \delta q_k, \]
so
\[ \delta \mathcal{S}_d \left( (q_k, q_k, \lambda_k)_{k=0}^{N} \right) = \sum_{k=0}^{N-1} \left[ L_d(q_k, f_k, q_k, q_{k+1}) - \frac{\lambda_k}{\tau_{k+1} - \tau_k} q_k - q_{k+1} + \lambda_k g(q_k) \right] \delta q_k + \sum_{k=0}^{N-1} \left[ f_k \delta f_k + \delta q_k \right] \delta \lambda_k. \]
Then, \[
\delta \tilde{\mathcal{H}} \left( (q_k, q_k, \lambda_k)_{k=0}^N \right) = \sum_{k=1}^{N-1} \left[ D_3 L_{d_{k}} + \lambda_k \frac{1}{q_{k+1} - q_k} + \lambda_k \nabla g(q_k) \right] \frac{q_{k+1} - q_k}{q_{k+1} - q_k} \delta q_k - \sum_{k=0}^{N-1} \frac{1}{q_{k+1} - q_k} \left[ L_{d_{k}} - \lambda_k \frac{q_{k+1} - q_k}{q_{k+1} - q_k} + \lambda_k g(q_k) \right] \delta q_k
\]
\[+ \sum_{k=1}^{N-1} \left[ D_4 L_{d_{k-1}} - \lambda_k \frac{1}{q_{k+1} - q_k} + \lambda_k \nabla g(q_k) \right] \frac{q_k - q_{k-1}}{q_{k+1} - q_k} \frac{1}{q_{k+1} - q_k} \left( L_{d_{k-1}} - \lambda_k \frac{q_k - q_{k-1}}{q_{k+1} - q_k} + \lambda_k g(q_k) \right) \delta q_k
\]
\[+ \sum_{k=1}^{N-1} \frac{q_{k+1} - q_k}{q_{k+1} - q_k} \left( T^* L_{f_{k+1}} D_1 L_{d_{k}} \bullet \eta_k \right) + \sum_{k=0}^{N-1} \frac{q_{k+1} - q_k}{q_{k+1} - q_k} g(q_k) - \frac{q_{k+1} - q_k}{q_{k+1} - q_k} \delta \lambda_k
\]
\[+ \sum_{k=0}^{N-1} \frac{1}{q_{k+1} - q_k} \left[ T^* L_{f_{k+1}} D_2 L_{d_{k-1}} \bullet \eta_k \right] - \sum_{k=0}^{N-1} \frac{q_{k+1} - q_k}{q_{k+1} - q_k} \left( T^* L_{f_{k+1}} D_2 L_{d_{k}} \bullet \text{Ad}_{f_{k+1}} J_{k+1} \eta_k \right).
\]

As a consequence, if \( q_{k+1} = q_k + (\tau_{k+1} - \tau_k) g(q_k) \),
\[
\text{Ad}_{f_{k+1}} \left( T^* L_{f_{k+1}} D_2 L_{d_{k}} \right) = T^* L_{q_{k+1}} D_1 L_{d_{k}} \frac{\tau_{k+1} - \tau_k}{q_{k+1} - q_k} \frac{q_k - q_{k-1}}{q_{k+1} - q_k} T^* L_{f_{k+1}} D_2 L_{d_{k-1}},
\]
\[
\left[ D_3 L_{d_{k}} + \lambda_k \frac{1}{q_{k+1} - q_k} + \lambda_k \nabla g(q_k) \right] \frac{q_{k+1} - q_k}{q_{k+1} - q_k} - \frac{1}{q_{k+1} - q_k} \left[ L_{d_{k}} - \lambda_k \frac{q_{k+1} - q_k}{q_{k+1} - q_k} + \lambda_k g(q_k) \right] = 0.
\]
then \( \delta \tilde{\mathcal{H}} \left( (q_k, q_k, \lambda_k)_{k=0}^N \right) = 0 \). Conversely, if \( \delta \tilde{\mathcal{H}} \left( (q_k, q_k, \lambda_k)_{k=0}^N \right) = 0 \), then a discrete fundamental theorem of the calculus of variations yields the above equations. \( \square \)

References


