ACCELERATED OPTIMIZATION ON RIEMANNIAN MANIFOLDS VIA DISCRETE CONSTRAINED VARIATIONAL INTEGRATORS

VALENTIN DURUISSEAUX AND MELVIN LEOK

Abstract. A variational formulation for accelerated optimization on normed spaces was introduced in Wibisono et al. [34], and later generalized to Riemannian manifolds in Duruisseaux and Leok [8]. This variational framework was exploited on normed spaces in Duruisseaux et al. [10] using time-adaptive geometric integrators to design efficient explicit algorithms for symplectic accelerated optimization, and it was observed that geometric discretizations which respect the time-rescaling invariance and symplecticity of the Lagrangian and Hamiltonian flows were substantially less prone to stability issues, and were therefore more robust, reliable, and computationally efficient. As such, it is natural to develop time-adaptive Hamiltonian variational integrators for accelerated optimization on Riemannian manifolds. In this paper, we consider the case of Riemannian manifolds embedded in a Euclidean space that can be characterized as the level set of a submersion. We will explore how holonomic constraints can be incorporated in discrete variational integrators to constrain the numerical discretization of the Riemannian Hamiltonian system to the Riemannian manifold, and we will test the performance of the resulting algorithms by solving eigenvalue and Procrustes problems formulated as optimization problems on the unit sphere and Stiefel manifold.

1. Introduction

Many data analysis algorithms are designed around the minimization of a loss function or the maximization of a likelihood function. Due to the ever-growing scale of the data sets and size of the problems, 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 [25] 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 standard gradient descent methods. This $\mathcal{O}(1/k^2)$ convergence rate was shown in [26] 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, which was shown in [30] to limit to a second-order ODE, as the timestep goes to 0, and that $f(x(t))$ converges to its optimal value at a rate of $\mathcal{O}(1/t^2)$ along the trajectories of this ODE. It was then shown in [34] that in continuous time, an arbitrary convergence rate $\mathcal{O}(1/t^p)$ can be achieved 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 and the time-rescaling property of this family was then exploited in [10] 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. More generally, when applied to Hamiltonian systems, symplectic integrators yield discrete approximations of the flow that preserve the symplectic 2-form [13]. The preservation of symplecticity results in the preservation of many qualitative aspects of the underlying dynamical system. In particular, when applied to conservative Hamiltonian systems, symplectic integrators exhibit excellent long-time near-energy preservation [7; 27]. Variational integrators provide a systematic method for constructing symplectic integrators of arbitrarily high-order based on the numerical discretization of Hamilton's principle [14; 23], or equivalently,

by the approximation of Jacobi's solution of the Hamilton–Jacobi equation, which is a generating function for the exact symplectic flow map.

In the past few years, there has been some effort to derive accelerated optimization algorithms in the Riemannian manifold setting [3–5; 8; 21; 35; 36]. In [8], it was shown that in continuous time, the convergence rate of $f(x(t))$ to its optimal value can be accelerated to an arbitrary convergence rate $\mathcal{O}(1/t^p)$ on Riemannian manifolds. This was achieved by considering a family of time-dependent Bregman Lagrangian and Hamiltonian systems on Riemannian manifolds which is closed under time-rescaling, thereby generalizing the variational framework for accelerated optimization of [34] to Riemannian manifolds. The time-adaptivity based approach relying on a Poincaré transformation from [10] was also extended to the Riemannian manifold setting in [8]. Now, the Whitney Embedding Theorems [32; 33] state that any smooth manifold of dimension $n \geq 2$ can be embedded in $\mathbb{R}^{2n}$ and immersed in $\mathbb{R}^{2n-1}$, and is thus diffeomorphic to a submanifold of $\mathbb{R}^{2n}$. Furthermore, the Nash Embedding Theorems [24] state that any Riemannian manifold can be globally isometrically embedded into some Euclidean space. As a consequence of these embedding theorems, the study of Riemannian manifolds can in principle be reduced to the study of submanifolds of Euclidean spaces. Altogether, this motivates the introduction of time-adaptive variational integrators on Riemannian manifolds that exploit the structure of the embedding Euclidean space, and in this paper we will study how holonomic constraints can be incorporated into Lagrangian and Hamiltonian variational integrators to constrain the numerical solutions of the Riemannian Hamiltonian system to the Riemannian manifold. Incorporating holonomic constraints in geometric integrators has been studied extensively in the past (see [13; 15; 22; 23] for instance), and some work has been done from the variational perspective for the Type I Lagrangian formulation in [23] via augmented Lagrangians.

Outline of the paper. We will first show in Section 2 the equivalence between constrained variational principles and constrained Euler–Lagrange equations, both in continuous and discrete time, before deriving analogous results for the Type II and Type III Hamiltonian formulations of mechanics in Section 3. In Section 4, we will exploit error analysis theorems for unconstrained mechanics from [23; 28] to obtain variational error analysis results for the maps defined implicitly by the discrete constrained Euler–Lagrange and Hamilton's equations. Finally, in Section 5, we will exploit these constrained variational integrators and the variational formulation of accelerated optimization on Riemannian manifolds from [8] to solve numerically generalized eigenvalue problems and Procrustes problems on the unit sphere and Stiefel manifold.

2. Constrained Variational Lagrangian Mechanics

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

$$L_d^E(q_0, q_h) = \int_0^h L(q(t), \dot{q}(t))dt, \tag{2.1}$$

where $q(0) = q_0$, $q(h) = q_h$, and q satisfies the Euler–Lagrange equations over the time interval $[0, h]$. This is closely related to Jacobi's solution of the Hamilton–Jacobi equation. A variational integrator is defined by constructing an approximation $L_d : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ to the exact discrete Lagrangian L_d^E, and then applying the **implicit discrete Euler–Lagrange equations**,

$$p_0 = -D_1 L_d(q_0, q_1), \qquad p_1 = D_2 L_d(q_0, q_1), \tag{2.2}$$

which implicitly define a numerical integrator, referred to as the **discrete Hamiltonian map** $\tilde{F}_{L_d} : (q_0, p_0) \mapsto (q_1, p_1)$, where D_i denotes a partial derivative with respect to the i-th argument. These equations define the **discrete Legendre transforms**, $\mathbb{F}^{\pm} L_d : \mathcal{Q} \times \mathcal{Q} \to T^*\mathcal{Q}$:

$$\mathbb{F}^+ L_d : (q_0, q_1) \mapsto (q_1, p_1) = (q_1, D_2 L_d(q_0, q_1)), \tag{2.3}$$

$$\mathbb{F}^- L_d : (q_0, q_1) \mapsto (q_0, p_0) = (q_0, -D_1 L_d(q_0, q_1)), \tag{2.4}$$

in which case the discrete Hamiltonian map can be expressed as $\tilde{F}_{L_d} \equiv (\mathbb{F}^+ L_d) \circ (\mathbb{F}^- L_d)^{-1}$. Such numerical methods are called variational integrators as they can derived from a **discrete Hamilton's principle**, which involves extremizing a discrete action sum $S_d(\{q_k\}_{k=0}^N) \equiv \sum_{k=0}^{N-1} L_d(q_k, q_{k+1})$, subject to fixed boundary conditions on q_0, q_N.

Now, suppose we are given a configuration manifold $\mathcal{M}$, and a holonomic constraint function $\mathcal{C} : \mathcal{M} \to \mathbb{R}^d$. Assuming that $0 \in \mathbb{R}^d$ is a regular point of $\mathcal{C}$, we can constrain the dynamics to the constraint submanifold $\mathcal{Q} = \mathcal{C}^{-1}(0)$, which is truly a submanifold of $\mathcal{M}$ (see [1; 23]). We will now consider variational Lagrangian mechanics with holonomic constraints $\mathcal{C}(q)$ using Lagrange multipliers $\lambda : [0, T] \to \Lambda$.

2.1. **Continuous Constrained Variational Lagrangian Mechanics.** We begin by presenting an equivalence between the continuous constrained variational principle and the continuous constrained Euler–Lagrange equations:

Theorem 2.1. *Consider the* ***constrained action functional*** $\mathfrak{S} : C^2([0, T], \mathcal{Q} \times \Lambda) \to \mathbb{R}$ *given by*

$$\mathfrak{S}(q(\cdot), \lambda(\cdot)) = \int_0^T \left[L(q(t), \dot{q}(t)) - \langle \lambda(t), \mathcal{C}(q(t)) \rangle \right] dt. \tag{2.5}$$

The condition that $\mathfrak{S}(q(\cdot), \lambda(\cdot))$ *is stationary with respect to the boundary conditions* $\delta q(0) = 0$ *and* $\delta q(T) = 0$ *is equivalent to* $(q(\cdot), \lambda(\cdot))$ *satisfying the* ***constrained Euler–Lagrange equations***

$$\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} = \langle \lambda, \nabla \mathcal{C}(q) \rangle, \qquad \mathcal{C}(q) = 0. \tag{2.6}$$

Proof. See Appendix A.1. □

Remark 2.1. *These constrained Euler–Lagrange equations can be thought of as the Euler–Lagrange equations coming from the augmented Lagrangian* $\bar{L}\left(q, \lambda, \dot{q}, \dot{\lambda}\right) = L(q, \dot{q}) - \langle \lambda, \mathcal{C}(q) \rangle$.

Consider the function $\mathcal{S}(q_0, q_T)$ given by the extremal value of the constrained action functional $\mathfrak{S}$ over the family of curves $(q(\cdot), \lambda(\cdot))$ satisfying the boundary conditions $q(0) = q_0$ and $q(T) = q_T$:

$$\mathcal{S}(q_0, q_T) = \operatorname*{ext}_{\substack{(q,\lambda) \in C^2([0,T], \mathcal{Q} \times \Lambda) \\ q(0) = q_0, \quad q(T) = q_T}} \mathfrak{S}(q(\cdot), \lambda(\cdot)). \tag{2.7}$$

The following theorem shows that $\mathcal{S}(q_0, q_T)$ is a generating function for the flow of the continuous constrained Euler–Lagrange equations:

Theorem 2.2. *The exact time-T flow map of Hamilton's equations* $(q_0, p_0) \mapsto (q_T, p_T)$ *is implicitly given by the following relations:*

$$D_1 \mathcal{S}(q_0, q_T) = -\frac{\partial L}{\partial \dot{q}}(q_0, \dot{q}(0)), \qquad D_2 \mathcal{S}(q_0, q_T) = \frac{\partial L}{\partial \dot{q}}(q_T, \dot{q}(T)). \tag{2.8}$$

In particular, $\mathcal{S}(q_0, q_T)$ *is a Type I generating function that generates the exact flow of the constrained Euler–Lagrange equations* (2.6).

Proof. See Appendix A.4. □

2.2. **Discrete Constrained Variational Lagrangian Mechanics.** We will now introduce a discrete variational formulation of Lagrangian mechanics that includes holonomic constraints. Suppose we are given a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the interval $[0, T]$, and a discrete curve in $\mathcal{Q} \times \Lambda$ denoted by $\{(q_k, \lambda_k)\}_{k=0}^N$ such that $q_k \approx q(t_k)$ and $\lambda_k \approx \lambda(t_k)$. We will formulate discrete

constrained variational Lagrangian mechanics in terms of the following discrete analogues of the constrained action functional $\mathfrak{S}$ given by equation (2.5):

$$\mathfrak{S}_d^+\left(\{(q_k,\lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1}\left[L_d(q_k,q_{k+1}) - \langle\lambda_{k+1}, C(q_{k+1})\rangle\right], \tag{2.9}$$

$$\mathfrak{S}_d^-\left(\{(q_k,\lambda_k)\}_{k=0}^N\right) = \sum_{k=0}^{N-1}\left[L_d(q_k,q_{k+1}) - \langle\lambda_k, C(q_k)\rangle\right], \tag{2.10}$$

where

$$L_d(q_k,q_{k+1}) \approx \underset{\substack{(q,p,\lambda)\in C^2([t_k,t_{k+1}],\mathcal{Q}\times\Lambda)\\ q(t_k)=q_k,\quad q(t_{k+1})=q_{k+1}}}{\operatorname{ext}} \int_{t_k}^{t_{k+1}} L(q(t),\dot{q}(t))dt. \tag{2.11}$$

We can now derive discrete analogues to Theorem 2.1 relating discrete Type I variational principles to discrete Euler–Lagrange equations:

Theorem 2.3. *The Type I discrete Hamilton's variational principles*

$$\delta\mathfrak{S}_d^\pm\left(\{(q_k,\lambda_k)\}_{k=0}^N\right) = 0, \tag{2.12}$$

are equivalent to the ***discrete constrained Euler–Lagrange equations***

$$D_1L_d(q_k,q_{k+1}) + D_2L_d(q_{k-1},q_k) = \langle\lambda_k, \nabla C(q_k)\rangle, \quad C(q_k) = 0, \tag{2.13}$$

where $L_d(q_k,q_{k+1})$ is defined via equation (2.11).

Proof. See Appendix A.7. □

Remark 2.2. *These discrete constrained Euler–Lagrange equations can be thought of as the discrete Euler–Lagrange equations coming from the augmented discrete Lagrangians*

$$\bar{L}_d^+(q_k,\lambda_k,q_{k+1},\lambda_{k+1}) = L_d(q_k,q_{k+1}) - \langle\lambda_{k+1}, C(q_{k+1})\rangle, \tag{2.14}$$

$$\bar{L}_d^-(q_k,\lambda_k,q_{k+1},\lambda_{k+1}) = L_d(q_k,q_{k+1}) - \langle\lambda_k, C(q_k)\rangle. \tag{2.15}$$

3. Constrained Variational Hamiltonian Mechanics

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,

$$H_d^{+,E}(q_0,p_h) = p_hq_h - \int_0^h \left[p(t)\dot{q}(t) - H(q(t),p(t))\right]dt, \tag{3.1}$$

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

$$p_0 = D_1H_d^+(q_0,p_1), \qquad q_1 = D_2H_d^+(q_0,p_1), \tag{3.2}$$

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

Similarly, the boundary-value formulation of the exact Type III generating function of the time-h flow of Hamilton's equations is given by the exact discrete left Hamiltonian,

$$H_d^{-,E}(q_h,p_0) = -p_0q_0 - \int_0^h \left[p(t)\dot{q}(t) - H(q(t),p(t))\right]dt, \tag{3.3}$$

where (q,p) satisfies Hamilton's equations with boundary conditions $q(h) = q_h$ and $p(0) = p_0$. A Type III Hamiltonian variational integrator is constructed by using an approximate discrete left Hamiltonian H_d^-, and applying the **discrete left Hamilton's equations**,

$$p_1 = -D_1H_d^-(q_1,p_0), \qquad q_0 = -D_2H_d^-(q_1,p_0), \tag{3.4}$$

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

We now derive analogous results to those of Section 2 from the Hamiltonian perspective. As in the Lagrangian case, we will assume we have a configuration manifold $\mathcal{M}$, a holonomic constraint function $C : \mathcal{M} \to \mathbb{R}^d$, and that the dynamics are constrained to the submanifold $\mathcal{Q} = C^{-1}(0)$.

3.1. **Continuous Constrained Variational Hamiltonian Mechanics.** The following theorem presents the equivalence between a continuous constrained variational principle and continuous constrained Hamilton's equations in the Type II case, generalizing Lemma 2.1 from [20] to include holonomic constraints:

Theorem 3.1. *Consider the Type II* ***constrained action functional*** $\mathfrak{S} : C^2([0,T], T^*\mathcal{Q}\times\Lambda) \to \mathbb{R}$

$$\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot)) = p(T)q(T) - \int_0^T \left[p(t)\dot{q}(t) - H(q(t),p(t)) - \langle\lambda(t), C(q(t))\rangle\right] dt. \tag{3.5}$$

The condition that $\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot))$ *is stationary with respect to the boundary conditions* $\delta q(0) = 0$ *and* $\delta p(T) = 0$ *is equivalent to* $(q(\cdot),p(\cdot),\lambda(\cdot))$ *satisfying* ***Hamilton's constrained equations***

$$\dot{q} = \frac{\partial H}{\partial p}(q,p), \qquad \dot{p} = -\frac{\partial H}{\partial q}(q,p) - \langle\lambda, \nabla C(q)\rangle, \qquad C(q) = 0. \tag{3.6}$$

Proof. See Appendix A.2. □

As in the Type II case, we can derive a theorem relating a continuous constrained variational principle and continuous constrained Hamilton's equations in the Type III case:

Theorem 3.2. *Consider the Type III* ***constrained action functional*** $\mathfrak{S} : C^2([0,T], T^*\mathcal{Q}\times\Lambda) \to \mathbb{R}$

$$\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot)) = -p(0)q(0) - \int_0^T \left[p(t)\dot{q}(t) - H(q(t),p(t)) - \langle\lambda(t), C(q(t))\rangle\right] dt. \tag{3.7}$$

The condition that $\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot))$ *is stationary with respect to the boundary conditions* $\delta q(T) = 0$ *and* $\delta p(0) = 0$ *is equivalent to* $(q(\cdot),p(\cdot),\lambda(\cdot))$ *satisfying* ***Hamilton's constrained equations***

$$\dot{q} = \frac{\partial H}{\partial p}(q,p), \qquad \dot{p} = -\frac{\partial H}{\partial q}(q,p) - \langle\lambda, \nabla C(q)\rangle, \qquad C(q) = 0. \tag{3.8}$$

Proof. See Appendix A.3. □

Remark 3.1. *Hamilton's constrained equations are the same in the Type II and Type III formulations of Hamiltonian mechanics, and they can be thought of as Hamilton's equations coming from the augmented Hamiltonian*

$$\bar{H}(q,\lambda,p,p) = H(q,p) + \langle\lambda, C(q)\rangle, \tag{3.9}$$

where p *is the conjugate variable to* λ*. Furthermore, they are equivalent to the constrained Euler–Lagrange equations (2.6), provided that the Lagrangian* L *is hyperregular.*

Remark 3.2. *It is sometimes beneficial to augment the continuous equations with the equation* $\langle\frac{\partial H}{\partial p}(q,p), \nabla C(q)\rangle = 0$ *(and analogously for the discrete case) to ensure that the momentum* p *lies in the cotangent space to the manifold, as explained and illustrated in [13, Chapter VII].*

We will now generalize Theorem 2.2 and its Type III analogue from [20] to include holonomic constraints $C(q)$ using Lagrange multipliers $\lambda : [0,T] \to \Lambda$.

In the Type II case, consider the function $\mathcal{S}(q_0,p_T)$ given by the extremal value of the constrained action functional $\mathfrak{S}$ over the family of curves $(q(\cdot),p(\cdot),\lambda(\cdot))$ satisfying the boundary conditions $q(0) = q_0$ and $p(T) = p_T$:

$$\mathcal{S}(q_0,p_T) = \operatorname*{ext}_{\substack{(q,p,\lambda)\in C^2([0,T],T^*\mathcal{Q}\times\Lambda)\\ q(0)=q_0,\quad p(T)=p_T}} \mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot)). \tag{3.10}$$

The following theorem shows that $\mathcal{S}(q_0,p_T)$ is a generating function for the flow of the continuous constrained Hamilton's equations:

Theorem 3.3. *The exact time-T flow map of Hamilton's equations $(q_0, p_0) \mapsto (q_T, p_T)$ is implicitly given by the following relations:*

$$q_T = D_2\mathcal{S}(q_0, p_T), \qquad p_0 = D_1\mathcal{S}(q_0, p_T). \tag{3.11}$$

In particular, $\mathcal{S}(q_0, p_T)$ is a Type II generating function that generates the exact flow of the constrained Hamilton's equations (3.6).

Proof. See Appendix A.5. □

In the Type III case, consider the function $\mathcal{S}(q_T, p_0)$ given by the extremal value of the constrained action functional $\mathfrak{S}$ over the family of curves $(q(\cdot), p(\cdot), \lambda(\cdot))$ satisfying the boundary conditions $q(T) = q_T$ and $p(0) = p_0$:

$$\mathcal{S}(q_T, p_0) = \operatorname*{ext}_{\substack{(q,p,\lambda)\in C^2([0,T],T^*\mathcal{Q}\times\Lambda)\\ q(T)=q_T,\quad p(0)=p_0}} \mathfrak{S}(q(\cdot), p(\cdot), \lambda(\cdot)). \tag{3.12}$$

The following theorem shows that $\mathcal{S}(q_T, p_0)$ is a generating function for the flow of the continuous constrained Hamilton's equations:

Theorem 3.4. *The exact time-T flow map of Hamilton's equations $(q_0, p_0) \mapsto (q_T, p_T)$ is implicitly given by the following relations:*

$$q_0 = -D_2\mathcal{S}(q_T, p_0), \qquad p_T = -D_1\mathcal{S}(q_T, p_0). \tag{3.13}$$

In particular, $\mathcal{S}(q_T, p_0)$ is a Type III generating function that generates the exact flow of the constrained Hamilton's equations (3.8).

Proof. See Appendix A.6. □

3.2. **Discrete Constrained Variational Hamiltonian Mechanics.** Let us now extend the results of Section 3 from [20] to introduce a discrete formulation of variational Hamiltonian mechanics that includes holonomic constraints. Suppose we are given a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the interval $[0, T]$, and a discrete curve in $T^*\mathcal{Q}\times\Lambda$ denoted by $\{(q_k, p_k, \lambda_k)\}_{k=0}^N$ such that $q_k \approx q(t_k)$, $p_k \approx p(t_k)$ and $\lambda_k \approx \lambda(t_k)$.

We will formulate discrete constrained variational Hamiltonian mechanics in terms of the following discrete analogues of the constrained action functional $\mathfrak{S}$ given by equation (3.5):

$$\mathfrak{S}_d^+\left(\{(q_k, p_k, \lambda_k)\}_{k=0}^N\right) = p_N q_N - \sum_{k=0}^{N-1} \left[p_{k+1}q_{k+1} - H_d^+(q_k, p_{k+1}) - \langle \lambda_k, \mathcal{C}(q_k)\rangle\right], \tag{3.14}$$

$$\mathfrak{S}_d^-\left(\{(q_k, p_k, \lambda_k)\}_{k=0}^N\right) = -p_0 q_0 - \sum_{k=0}^{N-1} \left[-p_k q_k - H_d^-(q_{k+1}, p_k) - \langle \lambda_{k+1}, \mathcal{C}(q_{k+1})\rangle\right], \tag{3.15}$$

where

$$H_d^+(q_k, p_{k+1}) \approx \operatorname*{ext}_{\substack{(q,p,\lambda)\in C^2([t_k,t_{k+1}],T^*\mathcal{Q}\times\Lambda)\\ q(t_k)=q_k,\quad p(t_{k+1})=p_{k+1}}} p(t_{k+1})q(t_{k+1}) - \int_{t_k}^{t_{k+1}} \left[p(t)\dot{q}(t) - H(q(t), p(t))\right] dt \tag{3.16}$$

$$H_d^-(q_{k+1}, p_k) \approx \operatorname*{ext}_{\substack{(q,p,\lambda)\in C^2([t_k,t_{k+1}],T^*\mathcal{Q}\times\Lambda)\\ q(t_{k+1})=q_{k+1},\quad p(t_k)=p_k}} -p(t_k)q(t_k) - \int_{t_k}^{t_{k+1}} \left[p(t)\dot{q}(t) - H(q(t), p(t))\right] dt. \tag{3.17}$$

We can now present discrete analogues of Theorems 3.1 and 3.2 relating discrete variational principles to discrete constrained Hamilton's equations, generalizing Lemma 3.1 from [20]:

Theorem 3.5. *The Type II discrete Hamilton's phase space variational principle*

$$\delta\mathfrak{S}_d^+\left(\{(q_k,p_k,\lambda_k)\}_{k=0}^N\right)=0 \tag{3.18}$$

is equivalent to the ***discrete constrained right Hamilton's equations***

$$q_{k+1}=D_2H_d^+(q_k,p_{k+1}),\qquad p_k=D_1H_d^+(q_k,p_{k+1})+\langle\lambda_k,\nabla\mathcal{C}(q_k)\rangle,\qquad \mathcal{C}(q_k)=0, \tag{3.19}$$

where $H_d^+(q_k,p_{k+1})$ *is defined via equation* (3.16).

Proof. See Appendix A.8. □

Theorem 3.6. *The Type III discrete Hamilton's phase space variational principle*

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

is equivalent to the ***discrete constrained left Hamilton's equations***

$$q_k=-D_2H_d^-(q_{k+1},p_k),\qquad p_{k+1}=-D_1H_d^-(q_{k+1},p_k)-\langle\lambda_{k+1},\nabla\mathcal{C}(q_{k+1})\rangle,\qquad \mathcal{C}(q_k)=0, \tag{3.21}$$

where $H_d^-(q_{k+1},p_k)$ *is defined via equation* (3.17).

Proof. See Appendix A.9. □

Remark 3.3. *These discrete constrained Hamilton's equations can be thought of as the discrete Hamilton's equations coming from the augmented discrete Hamiltonians*

$$\bar{H}_d^+\left((q_k,\lambda_k),(p_{k+1},p_{k+1})\right)=H_d^+(q_k,p_{k+1})+\langle\lambda_k,\mathcal{C}(q_k)\rangle, \tag{3.22}$$

$$\bar{H}_d^-\left((q_{k+1},\lambda_{k+1}),(p_k,p_k)\right)=H_d^-(q_k,p_{k+1})+\langle\lambda_{k+1},\mathcal{C}(q_{k+1})\rangle. \tag{3.23}$$

This augmented Hamiltonian perspective together with the augmented Lagrangian perspective from Remark 2.2 imply that the constrained $\bar{H}_d^+$ *variational integrator is equivalent to the constrained* $\bar{L}_d^+$ *variational integrator whenever the* H_d^+ *variational integrator is equivalent to the* L_d^+ *variational integrator (and similarly for the integrators coming from* $\bar{H}_d^-$ *and* $\bar{L}_d^-$*). Examples where this happens are presented in [29] for Taylor variational integrators provided the Lagrangian is hyperregular, and in [20] for generalized Galerkin variational integrators constructed using the same choices of basis functions and numerical quadrature formula provided the Hamiltonian is hyperregular.*

4. Error Analysis for Variational Integrators

4.1. **Unconstrained Error Analysis.** Theorem 2.3.1 of [23] states that if a discrete Lagrangian, $L_d:\mathcal{Q}\times\mathcal{Q}\to\mathbb{R}$, approximates the exact discrete Lagrangian $L_d^E:\mathcal{Q}\times\mathcal{Q}\to\mathbb{R}$ to order r, i.e.,

$$L_d(q_0,q_h)=L_d^E(q_0,q_h)+\mathcal{O}(h^{r+1}), \tag{4.1}$$

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 defined implicitly from the discrete Euler–Lagrange equations

$$D_1L_d(q_k,q_{k+1})+D_2L_d(q_{k-1},q_k)=0, \tag{4.2}$$

or equivalently in terms of the implicit discrete Euler–Lagrange equations, which involve the corresponding discrete momenta via the discrete Legendre transforms,

$$p_k=-D_1L_d(q_k,q_{k+1}),\qquad p_{k+1}=D_2L_d(q_k,q_{k+1}), \tag{4.3}$$

has order of accuracy r.

Theorem 2.3.1 of [23] has an analogue for Hamiltonian variational integrators. Theorem 2.2 in [28] states that if a discrete right Hamiltonian H_d^+ approximates the exact discrete right Hamiltonian $H_d^{+,E}$ to order r, i.e.,

$$H_d^+(q_0,p_h)=H_d^{+,E}(q_0,p_h)+\mathcal{O}(h^{r+1}), \tag{4.4}$$

then the discrete right Hamiltonian map $\tilde{F}_{H_d^+} : (q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$, viewed as a one-step method defined implicitly by the discrete right Hamilton's equations

$$p_k = D_1 H_d^+(q_k, p_{k+1}), \qquad q_{k+1} = D_2 H_d^+(q_k, p_{k+1}), \tag{4.5}$$

is order r accurate. As mentioned in [28], the proof of Theorem 2.2 in [28] can be easily adjusted to prove an equivalent theorem for the discrete left Hamiltonian case, which states that if a discrete left Hamiltonian H_d^- approximates the exact discrete left Hamiltonian $H_d^{-,E}$ to order r, i.e.,

$$H_d^-(q_1, p_0) = H_d^{-,E}(q_1, p_0) + \mathcal{O}(h^{r+1}), \tag{4.6}$$

then the discrete left Hamiltonian map $\tilde{F}_{H_d^-} : (q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$, viewed as a one-step method defined implicitly by the discrete left Hamilton's equations

$$p_{k+1} = -D_1 H_d^-(q_{k+1}, p_k), \qquad q_k = -D_2 H_d^-(q_{k+1}, p_k), \tag{4.7}$$

is order r accurate. Many other properties of the integrator, such as momentum conservation properties of the method, can be determined by analyzing the associated discrete Lagrangian or Hamiltonian, as opposed to analyzing the integrator directly. We will exploit these error analysis results to derive analogous results for the constrained versions discussed in Sections 2 and 3.

4.2. **Constrained Error Analysis.** For the Lagrangian case, we can think of the Lagrange multipliers λ as extra position coordinates and define an augmented Lagrangian $\bar{L}$ given by

$$\bar{L}\left((q, \lambda), (\dot{q}, \dot{\lambda})\right) = L(q, \dot{q}) - \langle \lambda, \mathcal{C}(q) \rangle. \tag{4.8}$$

A corresponding augmented discrete Lagrangian is given by

$$\bar{L}_d\left((q_k, \lambda_k), (q_{k+1}, \lambda_{k+1})\right) = L_d(q_k, q_{k+1}) - \langle \lambda_k, \mathcal{C}(q_k) \rangle, \tag{4.9}$$

and the discrete Euler–Lagrange equations (4.2)

$$D_1 \bar{L}_d\left((q_k, \lambda_k), (q_{k+1}, \lambda_{k+1})\right) + D_2 \bar{L}_d\left((q_{k-1}, \lambda_{k-1}), (q_k, \lambda_k)\right) = 0, \tag{4.10}$$

yield the discrete constrained Euler–Lagrange equations

$$D_1 L_d(q_k, q_{k+1}) + D_2 L_d(q_{k-1}, q_k) = \langle \lambda_k, \nabla \mathcal{C}(q_k) \rangle, \qquad \mathcal{C}(q_k) = 0, \tag{4.11}$$

derived in Section 2.2. As a consequence, we can apply Theorem 2.3.1 of [23] to the augmented Lagrangian (4.8) and obtain the following result:

Theorem 4.1. *Suppose that for an exact discrete Lagrangian L_d^E and a discrete Lagrangian L_d,*

$$L_d(q_0, q_h) - \langle \lambda_0, \mathcal{C}(q_0) \rangle = L_d^E(q_0, q_h) - \int_0^h \langle \lambda(t), \mathcal{C}(q(t)) \rangle dt + \mathcal{O}(h^{r+1}). \tag{4.12}$$

Then, the discrete map $(q_k, p_k, \lambda_k) \mapsto (q_{k+1}, p_{k+1}, \lambda_{k+1})$, viewed as a one-step method defined implicitly by the discrete constrained Euler–Lagrange equations, has order of accuracy r.

For the Hamiltonian case, we can think of the Lagrange multipliers λ as extra position coordinates and define conjugate momenta p, which are constants of motion since the time-derivative of λ does not appear anywhere, and are constrained to be zero. The augmented Hamiltonian $\bar{H}$, given by

$$\bar{H}\left((q, \lambda), (p, p)\right) = H(q, p) + \langle \lambda, \mathcal{C}(q) \rangle, \tag{4.13}$$

yields the following augmented left and right discrete Hamiltonians

$$\bar{H}_d^-\left((q_{k+1}, \lambda_{k+1}), (p_k, p_k)\right) = H_d^-(q_{k+1}, p_k) + \langle \lambda_{k+1}, \mathcal{C}(q_{k+1}) \rangle, \tag{4.14}$$

$$\bar{H}_d^+\left((q_k, \lambda_k), (p_{k+1}, p_{k+1})\right) = H_d^+(q_k, p_{k+1}) + \langle \lambda_k, \mathcal{C}(q_k) \rangle, \tag{4.15}$$

and the discrete left and right Hamilton's equations

$$(p_{k+1}, p_{k+1}) = -D_1 \bar{H}_d^-\left((q_{k+1}, \lambda_{k+1}), (p_k, p_k)\right), \qquad (q_k, \lambda_k) = -D_2 \bar{H}_d^-\left((q_{k+1}, \lambda_{k+1}), (p_k, p_k)\right), \tag{4.16}$$

$$(p_k, p_k) = D_1 \bar{H}_d^+\left((q_k, \lambda_k), (p_{k+1}, p_{k+1})\right), \quad (q_{k+1}, \lambda_{k+1}) = D_2 \bar{H}_d^+\left((q_k, \lambda_k), (p_{k+1}, p_{k+1})\right), \tag{4.17}$$

yield the discrete constrained left Hamilton's equations

$$q_k = -D_2H_d^-(q_{k+1},p_k), \qquad p_{k+1} = -D_1H_d^-(q_{k+1},p_k) - \langle \lambda_{k+1}, \nabla C(q_{k+1})\rangle, \quad C(q_k) = 0, \tag{4.18}$$

and the discrete constrained right Hamilton's equations

$$q_{k+1} = D_2H_d^+(q_k,p_{k+1}), \qquad p_k = D_1H_d^+(q_k,p_{k+1}) + \langle \lambda_k, \nabla C(q_k)\rangle, \quad C(q_k) = 0, \tag{4.19}$$

derived in Section 3. As a consequence, we can apply Theorem 2.2 in [28] and its Type III analogue to the augmented Hamiltonians and obtain the following results

Theorem 4.2. *Suppose that given an exact discrete right Hamiltonian $H_d^{+,E}$ and a discrete right Hamiltonian H_d^+, we have*

$$H_d^+(q_0,p_h) + \langle \lambda_0, C(q_0)\rangle = H_d^{+,E}(q_0,p_h) + \int_0^h \langle \lambda(t), C(q(t))\rangle dt + \mathcal{O}(h^{r+1}). \tag{4.20}$$

Then, the discrete map $(q_k,p_k,\lambda_k) \mapsto (q_{k+1},p_{k+1},\lambda_{k+1})$, viewed as a one-step method defined implicitly by the discrete constrained right Hamilton's equations, has order of accuracy r.

Theorem 4.3. *Suppose that given an exact discrete left Hamiltonian $H_d^{-,E}$ and a discrete left Hamiltonian H_d^-, we have*

$$H_d^-(q_h,p_0) + \langle \lambda_h, C(q_h)\rangle = H_d^{-,E}(q_h,p_0) + \int_0^h \langle \lambda(t), C(q(t))\rangle dt + \mathcal{O}(h^{r+1}). \tag{4.21}$$

Then, the discrete map $(q_k,p_k,\lambda_k) \mapsto (q_{k+1},p_{k+1},\lambda_{k+1})$, viewed as a one-step method defined implicitly by the discrete constrained left Hamilton's equations, has order of accuracy r.

5. Variational Riemannian Accelerated Optimization

5.1. **Riemannian Geometry.** We first introduce the main notions from Riemannian geometry that will be used throughout this section (see [2; 5; 8; 16–18] for more details).

Definition 5.1. *Suppose we have a Riemannian manifold $\mathcal{Q}$ with Riemannian metric $g(\cdot,\cdot) = \langle\cdot,\cdot\rangle$, represented by the positive-definite symmetric matrix (g_{ij}) in local coordinates. Then, we define the* ***musical isomorphism*** *$g^\flat : T\mathcal{Q} \to T^*\mathcal{Q}$ via*

$$g^\flat(u)(v) = g_q(u,v) \quad \forall q \in \mathcal{Q} \text{ and } \forall u,v \in T_q\mathcal{Q},$$

and its ***inverse musical isomorphism*** *$g^\sharp : T^*\mathcal{Q} \to T\mathcal{Q}$. The Riemannian metric $g(\cdot,\cdot) = \langle\cdot,\cdot\rangle$ induces a* ***fiber metric*** *$g^*(\cdot,\cdot) = \langle\!\langle\cdot,\cdot\rangle\!\rangle$ on $T^*\mathcal{Q}$ via*

$$\langle\!\langle u,v\rangle\!\rangle = \langle g^\sharp(u), g^\sharp(v)\rangle \quad \forall u,v \in T^*\mathcal{Q},$$

represented by the positive-definite symmetric matrix (g^{ij}) in local coordinates, which is the inverse of the Riemannian metric matrix (g_{ij}).

Definition 5.2. *Denoting the differential of f by df, the* ***Riemannian gradient*** $\mathrm{grad} f(q) \in T_q\mathcal{Q}$ *at a point $q \in \mathcal{Q}$ of a smooth function $f : \mathcal{Q} \to \mathbb{R}$ is the tangent vector at q such that*

$$\langle \mathrm{grad} f(q), u\rangle = df(q)u \qquad \forall u \in T_q\mathcal{Q}.$$

This can also be expressed in terms of the inverse musical isomorphism, $\mathrm{grad} f(q) = g^\sharp(df(q))$.

Definition 5.3. *A* ***geodesic*** *in a Riemannian manifold $\mathcal{Q}$ is a parametrized curve $\gamma : [0,1] \to \mathcal{Q}$ which is of minimal local length, and is a generalization of the notion of straight line from Euclidean spaces to Riemannian manifolds. The other generalization of straight lines involves curves having zero "acceleration" or constant "speed", which requires the introduction of an affine connection. These two generalizations are equivalent if the Riemannian manifold is endowed with the Levi–Civita connection. Given two points $q,\tilde{q} \in \mathcal{Q}$, a vector in $T_q\mathcal{Q}$ can be transported to $T_{\tilde{q}}\mathcal{Q}$ along a geodesic γ by an operation $\Gamma_q^{\tilde{q}} : T_q\mathcal{Q} \to T_{\tilde{q}}\mathcal{Q}$ called the* ***parallel transport along*** *γ.*

Definition 5.4. *The* ***Riemannian Exponential map*** $\mathrm{Exp}_q : T_q\mathcal{Q} \to \mathcal{Q}$ *at* $q \in \mathcal{Q}$ *is defined via*

$$\mathrm{Exp}_q(v) = \gamma_v(1),$$

where γ_v *is the unique geodesic in* $\mathcal{Q}$ *such that* $\gamma_v(0) = q$ *and* $\gamma_v'(0) = v$*, for any* $v \in T_q\mathcal{Q}$*.* Exp_q *is a diffeomorphism in some neighborhood* $U \subset T_q\mathcal{Q}$ *containing 0, so we can define its inverse map, the* ***Riemannian Logarithm map*** $\mathrm{Log}_p : \mathrm{Exp}_q(U) \to T_q\mathcal{Q}$*.*

Definition 5.5. *A* ***retraction*** *on a manifold* $\mathcal{Q}$ *is a smooth mapping* $\mathcal{R} : T\mathcal{Q} \to \mathcal{Q}$ *such that for any* $q \in \mathcal{Q}$*, the restriction* $\mathcal{R}_q : T_q\mathcal{Q} \to \mathcal{Q}$ *of* $\mathcal{R}$ *to* $T_q\mathcal{Q}$ *satisfies*

- $\mathcal{R}_q(0_q) = q$*, where* 0_q *denotes the zero element of* $T_q\mathcal{Q}$*,*
- $T_{0_q}\mathcal{R}_q = \mathbb{I}_{T_q\mathcal{Q}}$ *with the canonical identification* $T_{0_q}T_q\mathcal{Q} \simeq T_q\mathcal{Q}$*, where* $T_{0_q}\mathcal{R}_q$ *is the tangent map of* $\mathcal{R}$ *at* $0_q \in T_q\mathcal{Q}$ *and* $\mathbb{I}_{T_q\mathcal{Q}}$ *is the identity map on* $T_q\mathcal{Q}$*.*

The Riemannian Exponential map is a natural example of a retraction on a Riemannian manifold.

Definition 5.6. *A subset* A *of a Riemannian manifold* $\mathcal{Q}$ *is called* ***geodesically uniquely convex*** *if every two points of* A *are connected by a unique geodesic in* A*. A function* $f : \mathcal{Q} \to \mathbb{R}$ *is called* ***geodesically convex*** *if for any two points* $q, \tilde{q} \in \mathcal{Q}$ *and a geodesic* γ *connecting them,*

$$f(\gamma(t)) \le (1-t)f(q) + tf(\tilde{q}) \qquad \forall t \in [0,1].$$

Note that if f *is a smooth geodesically convex function on a geodesically uniquely convex subset* A*,*

$$f(q) - f(\tilde{q}) \ge \langle \mathrm{grad} f(\tilde{q}), \mathrm{Log}_{\tilde{q}}(q) \rangle \qquad \forall q, \tilde{q} \in A.$$

A function $f : A \to \mathbb{R}$ *is called* ***geodesically*** α***-weakly-quasi-convex*** *(*α*-WQC) with respect to* $q \in \mathcal{Q}$ *for some* $\alpha \in (0,1]$ *if*

$$\alpha\left(f(q) - f(\tilde{q})\right) \ge \langle \mathrm{grad} f(\tilde{q}), \mathrm{Log}_{\tilde{q}}(q) \rangle \qquad \forall \tilde{q} \in A.$$

Note that a local minimum of a geodesically convex or α*-WQC function is also a global minimum.*

Definition 5.7. *Given a Riemannian manifold* $\mathcal{Q}$ *with sectional curvature bounded below by* $K_{\min}$*, and an upper bound* D *for the diameter of the domain of consideration, define*

$$\zeta = \begin{cases} \sqrt{-K_{\min}} D \coth\left(\sqrt{-K_{\min}} D\right) & \text{if } K_{\min} < 0 \\ 1 & \text{if } K_{\min} \ge 0 \end{cases}. \tag{5.1}$$

Note that $\zeta \ge 1$ *since* $x \coth x \ge 1$ *for all real values of* x*.*

5.2. **Hamiltonian Approach.** Our approach consists in integrating the Riemannian Bregman Hamiltonians derived in [8] which live on the Riemannian manifold $\mathcal{Q}$, via discrete constrained variational Hamiltonian integrators which enforce the numerical solution to lie on the Riemannian manifold $\mathcal{Q}$. With ζ given by equation (5.1), we know from [8] that if we let $\lambda = \zeta$ in the geodesically convex case, and $\lambda = \zeta/\alpha$ in the geodesically α-weakly-quasi-convex case, we obtain the Direct approach Riemannian p-Bregman Hamiltonian

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

and the Adaptive approach Riemannian $p \to \mathring{p}$ Bregman Hamiltonian

$$\bar{\mathcal{H}}_{p\to\mathring{p}}(\bar{Q}, \bar{R}) = \frac{p^2}{2\mathring{p}(Q^t)^{\lambda p+\mathring{p}/p}} \langle\!\langle R, R \rangle\!\rangle + \frac{Cp^2}{\mathring{p}}(Q^t)^{(\lambda+1)p-\mathring{p}/p} f(Q) + \frac{p}{\mathring{p}}(Q^t)^{1-\mathring{p}/p} R^t. \tag{5.3}$$

It was proven in [8] that along the trajectories of the Riemannian p-Bregman Hamiltonian dynamics, $f(Q(t))$ converges to its optimal value at a rate of $\mathcal{O}(1/t^p)$, under suitable assumptions on $\mathcal{Q}$.

Remark 5.1. *In the vector space setting, these Riemannian Bregman Hamiltonians reduce to the Direct and Adaptive approach Bregman Hamiltonians derived in [10] for convex functions:*

$$\bar{H}_p(\bar{q},\bar{r}) = \frac{p}{2(q^t)^{p+1}}\langle r,r\rangle + Cp(q^t)^{2p-1}f(q) + r^t, \tag{5.4}$$

$$\bar{H}_{p\to\mathring{p}}(\bar{q},\bar{r}) = \frac{p^2}{2\mathring{p}(q^t)^{p+\mathring{p}/p}}\langle r,r\rangle + \frac{Cp^2}{\mathring{p}}(q^t)^{2p-\mathring{p}/p}f(q) + \frac{p}{\mathring{p}}(q^t)^{1-\mathring{p}/p}r^t. \tag{5.5}$$

5.3. **Some Optimization Problems on Riemannian Manifolds.**

5.3.1. *Rayleigh Quotient Minimization on the Unit Sphere.* An eigenvector v corresponding to the largest eigenvalue of a symmetric $n\times n$ matrix A maximizes the Rayleigh quotient $\frac{v^\top Av}{v^\top v}$ over $\mathbb{R}^n$. Thus, a unit eigenvector corresponding to the largest eigenvalue of the matrix A is a minimizer of the function $f(v) = -v^\top Av$ over the unit sphere $\mathcal{Q} = \mathbb{S}^{n-1}$, which can be thought of as a Riemannian submanifold with constant positive curvature $K = 1$ of $\mathbb{R}^n$ endowed with the Riemannian metric inherited from the Euclidean inner product $g_v(u,w) = u^\top w$. Solving the Rayleigh quotient optimization problem efficiently is challenging when the given symmetric matrix A is ill-conditioned and high-dimensional. Note that an efficient algorithm that solves the above minimization problem can also be used to find eigenvectors corresponding to the smallest eigenvalue of A by using the fact that the eigenvalues of A are the negative of the eigenvalues of $-A$.

5.3.2. *Eigenvalue and Procrustes Problems on the Stiefel Manifold.* When endowed with the Riemannian metric $g_X(A,B) = \mathrm{Trace}(A^\top B)$, the Stiefel manifold

$$\mathrm{St}(m,n) = \{X\in\mathbb{R}^{n\times m} | X^\top X = I_m\} \tag{5.6}$$

is a Riemannian submanifold of $\mathbb{R}^{n\times m}$. The tangent space at any $X\in\mathrm{St}(m,n)$ is given by $T_X\mathrm{St}(m,n) = \{Z\in\mathbb{R}^{n\times m} | X^\top Z + Z^\top X = 0\}$, and the orthogonal projection P_X onto $T_X\mathrm{St}(m,n)$ is given by $P_XZ = Z - \frac{1}{2}X(X^\top Z + Z^\top X)$. A retraction on $\mathrm{St}(m,n)$ is given by $\mathcal{R}_X(\xi) = \mathrm{qf}(X+\xi)$, where $\mathrm{qf}(A)$ denotes the Q factor of the QR factorization of the matrix $A\in\mathbb{R}^{n\times m}$ as $A = QR$ where $Q\in\mathrm{St}(m,n)$ and R is an upper triangular $n\times m$ matrix with strictly positive diagonal elements [2].

A generalized eigenvector problem consists of finding the m smallest eigenvalues of a $n\times n$ symmetric matrix A and corresponding eigenvectors. This problem can be formulated as a Riemannian optimization problem on the Stiefel manifold $\mathrm{St}(m,n)$ via the Brockett cost function

$$f:\mathrm{St}(m,n)\to\mathbb{R},\quad X\mapsto f(X) = \mathrm{Trace}(X^\top AXN), \tag{5.7}$$

where $N = \mathrm{diag}(\mu_1,\ldots,\mu_m)$ for arbitrary $0\le\mu_1\le\ldots\le\mu_m$. The columns of a global minimizer of f are eigenvectors corresponding to the m smallest eigenvalues of A (see [2]). If we define $\bar{f}:\mathbb{R}^{n\times m}\to\mathbb{R}$ via $X\mapsto\bar{f}(X) = \mathrm{Trace}(X^\top AXN)$, then f is the restriction of $\bar{f}$ to $\mathrm{St}(m,n)$ so

$$\mathrm{grad}f(X) = P_X\mathrm{grad}\bar{f}(X),\quad \text{where}\ \ \mathrm{grad}\bar{f}(X) = 2AXN. \tag{5.8}$$

The unbalanced orthogonal Procrustes problem consists of minimizing the function

$$f:\mathrm{St}(m,n)\to\mathbb{R},\quad X\mapsto f(X) = \|AX-B\|_F^2, \tag{5.9}$$

on the Stiefel manifold $\mathrm{St}(m,n)$, for given matrices $A\in\mathbb{R}^{l\times n}$ and $B\in\mathbb{R}^{l\times m}$ with $l\ge n$ and $l > m$, where $\|\cdot\|_F$ is the Frobenius norm $\|X\|_F^2 = \mathrm{Trace}(X^\top X) = \sum_{ij}X_{ij}^2$. If we define $\bar{f}:\mathbb{R}^{n\times m}\to\mathbb{R}$ via $X\mapsto\bar{f}(X) = \|AX-B\|_F^2$, then f is the restriction of $\bar{f}$ to $\mathrm{St}(m,n)$ so

$$\mathrm{grad}f(X) = P_X\mathrm{grad}\bar{f}(X),\quad \text{where}\ \ \mathrm{grad}\bar{f}(X) = 2A^\top(AX-B). \tag{5.10}$$

Note that the special case where $n = m$ is the balanced orthogonal Procrustes problem. In this case, $\mathrm{St}(m,n) = O(n)$ so $\|AX\|_F^2 = \|A\|_F^2$ and minimizing the function $f(X) = \|AX-B\|_F^2$ is replaced by the problem of maximizing $\mathrm{Trace}(X^\top A^\top B)$ over $X\in O(n)$. A solution is then given by $X^* = UV^\top$ where $B^\top A = U\Sigma V^\top$ is the Singular Value Decomposition of $B^\top A$ with square orthogonal matrices U and V, and the solution is unique provided $B^\top A$ is nonsingular (see [11; 12]).

5.4. **Numerical Methods.**

5.4.1. *Hamiltonian Taylor Variational Integrators (HTVIs).* HTVIs were first introduced in [29]. A discrete approximate 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 Hamiltonian Taylor variational integrator is then generated by the discrete Hamilton's equations. More explicitly, Type II HTVIs are constructed as follows:

(i) Construct the r-order and $(r+1)$-order Taylor methods $\Psi_h^{(r)}$ and $\Psi_h^{(r+1)}$ approximating the exact time-h flow map $\Phi_h : T^*Q \to T^*Q$.
(ii) Approximate $p(0) = p_0$ by the solution $\tilde{p}_0$ of $p_1 = \pi_{T^*Q} \circ \Psi_h^{(r)}(q_0, \tilde{p}_0)$, where $\pi_{T^*Q} : (q,p) \mapsto p$.
(iii) Choose a quadrature rule of order s with weights and nodes given by (b_i, c_i) for $i = 1, ..., m$ and generate approximations $(q_{c_i}, p_{c_i}) \approx (q(c_i h), p(c_i h))$ via $(q_{c_i}, p_{c_i}) = \Psi_{c_i h}^{(r)}(q_0, \tilde{p}_0)$.
(iv) Approximate q_1 via $\tilde{q}_1 = \pi_Q \circ \Psi_h^{(r+1)}(q_0, \tilde{p}_0)$, where $\pi_Q : (q,p) \mapsto q$.
(v) Use the continuous Legendre transform to obtain $\dot{q}_{c_i} = \frac{\partial H}{\partial p_{c_i}}$.
(vi) Apply the quadrature rule to obtain the associated discrete right Hamiltonian $H_d^+(q_0, p_1) = p_1 \tilde{q}_1 - h \sum_{i=1}^m b_i \left[p_{c_i} \dot{q}_{c_i} - H(q_{c_i}, p_{c_i}) \right]$.
(vii) The variational integrator is then defined by the discrete right Hamilton's equations.

Note that the following error analysis result concerning the order of accuracy of HTVIs was derived in [10] (it can be extended to the constrained case via the strategy and results of Section 4.2):

Theorem 5.1. *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 $\min(r+1, s)$. By Theorem 2.2 in [28], the associated discrete Hamiltonian map has the same order of accuracy.*

In this paper, we will use the Direct approach and Adaptive approach $r = 0$ Type II HTVIs constructed in [10] based on the Direct and Adaptive discrete right Hamiltonians (respectively)

$$H_d^+(\bar{q}_0, \bar{r}_1; h) = r_1^\top q_0 + r_1^t q_0^t + h \frac{p}{2(q_0^t)^{p+1}} r_1^\top r_1 + hCp(q_0^t)^{2p-1} f(q_0) + h r_1^t, \tag{5.11}$$

$$H_d^+(\bar{q}_0, \bar{r}_1; h) = r_1^\top q_0 + r_1^t q_0^t + h \frac{p^2}{2\mathring{p}(q_0^t)^{p+\frac{\mathring{p}}{p}}} r_1^\top r_1 + hC\frac{p^2}{\mathring{p}}(q_0^t)^{2p-\frac{\mathring{p}}{p}} f(q_0) + h\frac{p}{\mathring{p}}(q_0^t)^{1-\frac{\mathring{p}}{p}} r_1^t. \tag{5.12}$$

Algorithm 1: Direct and Adaptive Hamiltonian Taylor variational integrators (HTVIs)

Input: A function $f : \mathcal{Q} \to \mathbb{R}$, constants $C, h, p, \mathring{p} > 0$, $q_0^t, r_0^t \in \mathbb{R}$, and $(q_0, r_0, \lambda_0) \in T^*_{q_0}\mathcal{Q} \times \Lambda$.

while convergence criterion is not met, **solve** the following system of equations

Direct Approach	**Adaptive Approach**
$0 = r_{k+1} - r_k + hCp(q_k^t)^{2p-1} \nabla f(q_k) + \lambda_k^\top \nabla \mathcal{C}(q_k)$	$0 = r_{k+1} - r_k + \frac{hCp^2}{\mathring{p}}(q_k^t)^{2p-\frac{\mathring{p}}{p}} \nabla f(q_k) + \lambda_k^\top \nabla \mathcal{C}(q_k)$
$0 = r_{k+1}^t - r_k^t - h\frac{p(p+1)}{2(q_k^t)^{p+2}} r_{k+1}^\top r_{k+1} + hCp(2p-1)(q_k^t)^{2p-2} f(q_k)$	$0 = r_{k+1}^t - r_k^t + \frac{p\mathring{p} - 2p^3}{\mathring{p}} hC(q_k^t)^{2p-\frac{\mathring{p}}{p}-1} f(q_k) + h\frac{p^3 + p\mathring{p}}{2\mathring{p}(q_k^t)^{p+\frac{\mathring{p}}{p}+1}} r_{k+1}^\top r_{k+1} + \frac{\mathring{p} - p}{\mathring{p}(q_k^t)^{\frac{\mathring{p}}{p}}} h r_{k+1}^t$
$0 = q_{k+1} - q_k - h\frac{p}{(q_k^t)^{p+1}} r_{k+1}$	$0 = q_{k+1} - q_k - \frac{p^2}{\mathring{p}} h(q_k^t)^{-p-\frac{\mathring{p}}{p}} r_{k+1}$
$0 = q_{k+1}^t - q_k^t - h$	$0 = q_{k+1}^t - q_k^t - \frac{p}{\mathring{p}} h(q_k^t)^{1-\frac{\mathring{p}}{p}}$
$0 = \mathcal{C}(q_{k+1})$	$0 = \mathcal{C}(q_{k+1})$

5.4.2. *Euler–Lagrange Simple Discretization.* In [8], the p-Bregman Euler–Lagrange equations were rewritten as a first order system of differential equations, for which a Riemannian version of a semi-implicit Euler scheme was applied to obtain the following algorithm:

Algorithm 2: Semi-Implicit Euler Integration of the p-Bregman Euler–Lagrange Equations

Input: A geodesically-convex ($\lambda = \zeta$) or α-WQC ($\lambda = \zeta/\alpha$) function $f : \mathcal{Q} \to \mathbb{R}$.
A retraction $\mathcal{R}$ from $T\mathcal{Q}$ to $\mathcal{Q}$, constants $C, h, p > 0$, and $X_0 \in \mathcal{Q}$, $V_0 \in T_{X_0}\mathcal{Q}$.

while *convergence criterion is not met* **do**
- $b_k \leftarrow 1 - \frac{\lambda p+1}{k}, \quad c_k \leftarrow Cp^2(kh)^{p-2}$
- **Version I**: $a_k \leftarrow b_k V_k - hc_k \mathrm{grad} f(X_k)$
- **Version II**: $a_k \leftarrow b_k V_k - hc_k \mathrm{grad} f\left(\mathcal{R}_{X_k}(hb_k V_k)\right)$
- $X_{k+1} \leftarrow \mathcal{R}_{X_k}(ha_k), \quad V_{k+1} \leftarrow \Gamma_{X_k}^{X_{k+1}} a_k$

Version I of Algorithm 2 corresponds to the usual update for the semi-implicit Euler scheme, while Version II is inspired by the reformulation of Nesterov's method from [31] that uses a corrected gradient $\nabla f(X_k + hb_k V_k)$ instead of the traditional gradient $\nabla f(X_k)$.

5.4.3. *Riemannian Gradient Descent.* This is a generalization of Gradient Descent to the setting of Riemannian manifolds which involves the Riemannian gradient and a retraction.

Algorithm 3: Riemannian Gradient Descent (RGD)

Input: A function $f : \mathcal{Q} \to \mathbb{R}$, a retraction $\mathcal{R}$ from $T\mathcal{Q}$ to $\mathcal{Q}$, $h > 0$, and $X_0 \in \mathcal{Q}$.

while *convergence criterion is not met* **do**
- $X_{k+1} = \mathcal{R}_{X_k}(-h \ \mathrm{grad} f(X_k))$

5.5. **Numerical Results.** It was noted in [8] that although higher values of p in Algorithm 2 result in provably faster rates of convergence, they also appear to be more prone to stability issues under numerical discretization, which can cause the numerical optimization algorithm to diverge. Numerical experiments in [10] showed that in the normed vector space setting, geometric discretizations which respect the time-rescaling invariance and symplecticity of the Bregman Lagrangian and Hamiltonian flows were substantially less prone to these stability issues, and were therefore more robust, reliable, and computationally efficient. This was one of the motivations to develop time-adaptive Hamiltonian variational integrators for the Bregman Hamiltonians. Numerical experiments were conducted for the Rayleigh quotient minimization problem on $\mathbb{S}^{n-1}$, and for the generalized eigenvalue and Procrustes problems on the Stiefel manifold $\mathrm{St}(m, n)$.

The results from Figure 1 show how the Hamiltonian Taylor variational integrators compare to the Euler–Lagrange discretizations from [8] and the standard Riemannian gradient descent. Note that for certain instances of the Procrustes problem with certain initial values, all the algorithms converged to a local minimizer, and not the global minimizer, of the objective function. We can observe from Figure 1 that for the same value of the timestep h, the Adaptive Hamiltonian variational integrator clearly outperforms its Direct counterpart, Riemannian gradient descent and the Euler–Lagrange discretizations in terms of number of iterations required. Furthermore, unlike the Euler–Lagrange discretizations (Algorithm 2) and the Riemannian gradient descent (Algorithm 3), the HTVI methods (Algorithm 1) do not require the use of retractions or parallel transports. Note that the Rayleigh minimization results indicate that the Euler–Lagrange discretizations suffer from stability issues leading to a loss of convergence, as the polynomially growing unbounded coefficient $Cp^2(kh)^{p-2}$ is multiplied with $\mathrm{grad} f$, so for this product to be bounded, the gradient has to decay to zero, but due to finite numerical precision, the gradient remains bounded away from zero, thereby causing the product to grow without bound. This issue can be resolved by adding a suitable upper bound to the coefficient $Cp^2(kh)^{p-2}$ in the updates, as can be seen both

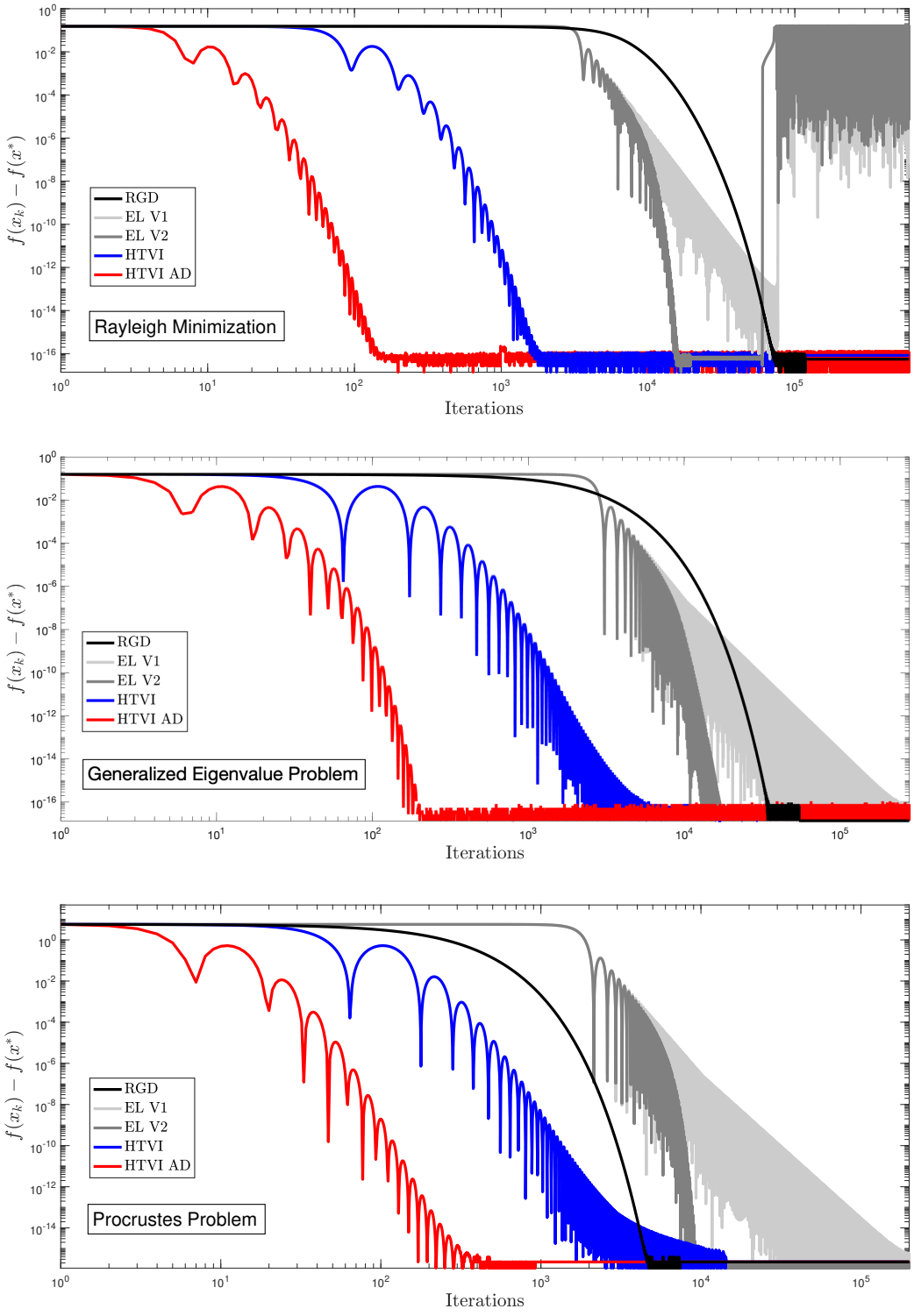

FIGURE 1. Comparison of the Direct and Adaptive (AD) Type II HTVIs with the Riemannian Gradient Descent (RGD) method and the Euler–Lagrange discretizations (EL V1 and EL V2) from [8] with $p = 6$ and the same timestep $h = 0.001$, for the Rayleigh quotient minimization problem on the unit sphere $\mathbb{S}^{n-1}$, and for the generalized eigenvalue and Procrustes problems on the Stiefel manifold $\mathrm{St}(m, n)$.

for the Euler–Lagrange discretizations and Hamiltonian variational integrators for the problems on $\mathrm{St}(m,n)$.

However, the algorithms generated by these constrained Hamiltonian variational integrators are implicit, which can significantly increase the cost per iteration as the dimension of the problem becomes very large. In this case, it might be beneficial to consider other options using the unconstrained explicit Hamiltonian Taylor variational integrator, such as incorporating the constraints within the objective function as a penalty, although this might not constrain the solution trajectory to lie exactly on the manifold, or using projections if they can be computed efficiently and accurately for the Riemannian manifold of interest [9]. Further, note that the implementation of the Hamiltonian variational integrators needs a very careful tuning of the various parameters at play, which may be challenging and thus also motivates the development of different methods.

6. Conclusion

Motivated by variational formulations of optimization problems on Riemannian manifolds, we first studied the relationship between the constrained Type I/II/III variational principles and the corresponding constrained Hamilton's or Euler–Lagrange equations both in continuous and discrete time, and derived variational error analysis results for the maps defined implicitly by the resulting discrete constrained equations. We then exploited these discrete constrained variational integrators and the variational formulation of accelerated optimization on Riemannian manifolds from [8] to numerically solve the generalized eigenvalue and Procrustes problems on $\mathbb{S}^{n-1}$ and $\mathrm{St}(m,n)$.

The numerical experiments conducted in this paper corroborated the observation made for the vector space setting in [10] that the Adaptive Hamiltonian variational integrator is significantly more efficient than the Direct Hamiltonian variational integrator, and that it can significantly outperform the Euler–Lagrange discretizations and Riemannian gradient descent, when its parameters are tuned carefully. Furthermore, it was noted that unlike the Euler–Lagrange discretizations and Riemannian gradient descent, the Hamiltonian algorithms did not require the use of retractions or parallel transports, which could be important when the problem considered lies on a Riemannian manifold for which it might not be possible to compute or approximate these objects efficiently.

We noted however that tuning the parameters of these discrete constrained variational integrators can be challenging, and also that the resulting algorithms are implicit, which may significantly increase the cost per iteration as the dimension of the problem becomes very large, in which case it might be beneficial to consider using the unconstrained explicit HTVIs with projections [9] or by incorporating the constraints within the objective function as a penalty. Moreover, although the Whitney and Nash Embedding Theorems [24; 32; 33] imply that there is no loss of generality when studying Riemannian manifolds only as submanifolds of Euclidean spaces, there are limitations to the constrained integration strategy based on embeddings presented in this paper, and an approach intrinsically defined on Riemannian manifolds would be desirable. Indeed, the embedding approach usually leads to higher-dimensional computations, and requires an effective way of constructing the embedding or a natural way of writing down equations that constrain the problem and the numerical solutions to the Riemannian manifold. Furthermore, most results in Riemannian geometry or results concerning specific Riemannian manifolds are proven from an intrinsic perspective because the embedding approach tends to flood intrinsic geometric properties of the manifold with superfluous information coming from the additional dimensions of the Euclidean space. This motivates the development of intrinsic methods that would exploit the symmetries and geometric properties of the manifold and of the problem at hand.

Developing an intrinsic extension of Hamiltonian variational integrators to manifolds will 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, and one should

ideally consider a Hamiltonian variational integrator construction based on discrete Dirac mechanics [19], which would yield a generating function $E_d^+(q_k, q_{k+1}, p_{k+1})$, $E_d^-(q_k, p_k, q_{k+1})$, that depends on the position at both boundary points and the momentum at one of the boundary points. This approach can be viewed as a discretization of the generalized energy $E(q, v, p) = \langle p, v\rangle - L(q, v)$, in contrast to the Hamiltonian $H(q, p) = \mathrm{ext}_v\langle p, v\rangle - L(q, v) = \langle p, v\rangle - L(q, v)\big|_{p=\frac{\partial L}{\partial v}}$.

Acknowledgments

The authors were supported in part by NSF under grants DMS-1411792, DMS-1345013, DMS-1813635, by AFOSR under grant FA9550-18-1-0288, and by the DoD under grant 13106725 (Newton Award for Transformative Ideas during the COVID-19 Pandemic).

Appendix A. Proofs of Theorems for Constrained Variational Mechanics

A.1. Proof of Theorem 2.1.

Theorem A.1. *Consider the constrained action functional* $\mathfrak{S} : C^2([0,T], \mathcal{Q}\times\Lambda)\to\mathbb{R}$ *given by*

$$\mathfrak{S}(q(\cdot),\lambda(\cdot)) = \int_0^T [L(q(t),\dot q(t)) - \langle\lambda(t), \mathcal{C}(q(t))\rangle]\,dt. \tag{A.1}$$

The condition that $\mathfrak{S}(q(\cdot),\lambda(\cdot))$ *is stationary with respect to the boundary conditions* $\delta q(0) = 0$ *and* $\delta q(T) = 0$ *is equivalent to* $(q(\cdot),\lambda(\cdot))$ *satisfying the constrained Euler–Lagrange equations*

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot q} = \langle\lambda, \nabla\mathcal{C}(q)\rangle, \qquad \mathcal{C}(q) = 0. \tag{A.2}$$

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

$$\delta\mathfrak{S} = \int_0^T \left[\frac{\partial L}{\partial q}(q(t),\dot q(t))\delta q(t) + \frac{\partial L}{\partial \dot q}(q(t),\dot q(t))\delta\dot q(t)\right]dt - \int_0^T [\langle\lambda(t), \nabla\mathcal{C}(q(t))\delta q(t)\rangle + \langle\delta\lambda(t), \mathcal{C}(q(t))\rangle]\,dt.$$

Using integration by parts and the boundary conditions $\delta q(0) = 0$ and $\delta q(T) = 0$, we get

$$\delta\mathfrak{S} = \int_0^T \left[\frac{\partial L}{\partial q}(q(t),\dot q(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot q}(q(t),\dot q(t)) - \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right]\delta q(t)dt - \int_0^T \langle\delta\lambda(t), \mathcal{C}(q(t))\rangle dt.$$

Now, if $\delta\mathfrak{S} = 0$, then the fundamental theorem of the calculus of variations [6] yields the constrained Euler–Lagrange equations (A.2). Conversely, if (q,λ) satisfies the constrained Euler–Lagrange equations (A.2), then the integrand vanishes and $\delta\mathfrak{S} = 0$. □

A.2. Proof of Theorem 3.1.

Theorem A.2. *Consider the constrained action functional* $\mathfrak{S} : C^2([0,T], T^*\mathcal{Q}\times\Lambda)\to\mathbb{R}$ *given by*

$$\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot)) = p(T)q(T) - \int_0^T [p(t)\dot q(t) - H(q(t),p(t)) - \langle\lambda(t), \mathcal{C}(q(t))\rangle]\,dt. \tag{A.3}$$

The condition that $\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot))$ *is stationary with respect to the boundary conditions* $\delta q(0) = 0$ *and* $\delta p(T) = 0$ *is equivalent to* $(q(\cdot),p(\cdot),\lambda(\cdot))$ *satisfying Hamilton's canonical constrained equations*

$$\dot q = \frac{\partial H}{\partial p}(q,p), \qquad \dot p = -\frac{\partial H}{\partial q}(q,p) - \langle\lambda, \nabla\mathcal{C}(q)\rangle, \qquad \mathcal{C}(q) = 0. \tag{A.4}$$

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

$$\begin{aligned}\delta\mathfrak{S} = {} & q(T)\delta p(T) + p(T)\delta q(T) + \int_0^T [\langle\lambda(t), \nabla\mathcal{C}(q(t))\delta q(t)\rangle + \langle\delta\lambda(t), \mathcal{C}(q(t))\rangle]\,dt \\ & - \int_0^T \left[\dot q(t)\delta p(t) + p(t)\delta\dot q(t) - \frac{\partial H}{\partial q}(q(t),p(t))\delta q(t) - \frac{\partial H}{\partial p}(q(t),p(t))\delta p(t)\right]dt.\end{aligned} \tag{A.5}$$

Using integration by parts and the boundary conditions $\delta q(0) = 0$ and $\delta p(T) = 0$, we get

$$\begin{aligned}\delta\mathfrak{S} = {} & q(T)\delta p(T) + p(T)\delta q(T) - p(T)\delta q(T) + p(0)\delta q(0) + \int_0^T \langle\delta\lambda(t), \mathcal{C}(q(t))\rangle dt \\ & + \int_0^T \left[\dot p(t) + \frac{\partial H}{\partial q}(q(t),p(t)) + \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right]\delta q(t)dt + \int_0^T \left[\frac{\partial H}{\partial p}(q(t),p(t)) - \dot q(t)\right]\delta p(t)dt \\ = {} & \int_0^T \left[\dot p(t) + \frac{\partial H}{\partial q}(q(t),p(t)) + \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right]\delta q(t)dt + \int_0^T \left[\frac{\partial H}{\partial p}(q(t),p(t)) - \dot q(t)\right]\delta p(t)dt + \int_0^T \langle\delta\lambda(t), \mathcal{C}(q(t))\rangle dt.\end{aligned}$$

Now, if $\delta\mathfrak{S} = 0$, then the fundamental theorem of the calculus of variations [6] yields Hamilton's constrained equations (A.4). Conversely, if (q,p,λ) satisfies Hamilton's constrained equations (A.4), then the integrand vanishes and $\delta\mathfrak{S} = 0$. □

A.3. **Proof of Theorem 3.2.**

Theorem A.3. *Consider the constrained action functional* $\mathfrak{S}: C^2([0,T], T^*\mathcal{Q}\times\Lambda)\to\mathbb{R}$ *given by*

$$\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot)) = -p(0)q(0) - \int_0^T [p(t)\dot{q}(t) - H(q(t),p(t)) - \langle\lambda(t), \mathcal{C}(q(t))\rangle]\,dt. \tag{A.6}$$

The condition that $\mathfrak{S}(q(\cdot),p(\cdot),\lambda(\cdot))$ *is stationary with respect to the boundary conditions* $\delta q(T)=0$ *and* $\delta p(0)=0$ *is equivalent to* $(q(\cdot),p(\cdot),\lambda(\cdot))$ *satisfying Hamilton's canonical constrained equations*

$$\dot{q} = \frac{\partial H}{\partial p}(q,p), \qquad \dot{p} = -\frac{\partial H}{\partial q}(q,p) - \langle\lambda, \nabla\mathcal{C}(q)\rangle, \qquad \mathcal{C}(q)=0. \tag{A.7}$$

Proof. The proof is almost identical to that of Theorem 3.1. We compute the variation of $\mathfrak{S}$ as before and get equation (A.5) except that the term $(q(T)\delta p(T) + p(T)\delta q(T))$ is replaced by $(-q(0)\delta p(0) - p(0)\delta q(0))$. As before, integration by parts and the boundary conditions $\delta q(T)=0$ and $\delta p(0)=0$ yield

$$\delta\mathfrak{S} = \int_0^T \left[\dot{p}(t) + \frac{\partial H}{\partial q}(q(t),p(t)) + \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right]\delta q(t)dt + \int_0^T \left[\frac{\partial H}{\partial p}(q(t),p(t)) - \dot{q}(t)\right]\delta p(t)dt + \int_0^T \langle\delta\lambda(t), \mathcal{C}(q(t))\rangle dt.$$

Then, if $\delta\mathfrak{S}=0$, then the fundamental theorem of the calculus of variations [6] yields Hamilton's constrained equations (A.7). Conversely, if (q,p,λ) satisfies Hamilton's constrained equations (A.7), then the integrand vanishes and $\delta\mathfrak{S}=0$. □

A.4. **Proof of Theorem 2.2.**

Theorem A.4. *The exact time-T flow map of Hamilton's equations* $(q_0,p_0)\mapsto(q_T,p_T)$ *is implicitly given by the following relations:*

$$D_1\mathcal{S}(q_0,q_T) = -\frac{\partial L}{\partial\dot{q}}(q_0,\dot{q}(0)), \qquad D_2\mathcal{S}(q_0,q_T) = \frac{\partial L}{\partial\dot{q}}(q_T,\dot{q}(T)). \tag{A.8}$$

Thus, $\mathcal{S}(q_0,q_T)$ *is a Type I generating function that generates the exact flow of the constrained Euler–Lagrange equations (2.6).*

Proof. Using integration by parts and simplifying gives

$$\begin{aligned}\frac{\partial\mathcal{S}}{\partial q_0}(q_0,q_T) &= \int_0^T \left[\frac{\partial q(t)}{\partial q_0}\frac{\partial L}{\partial q}(q(t),\dot{q}(t)) + \frac{\partial\dot{q}(t)}{\partial q_0}\frac{\partial L}{\partial\dot{q}}(q(t),\dot{q}(t))\right]dt - \int_0^T \left[\langle\lambda(t), \frac{\partial q(t)}{\partial q_0}\nabla\mathcal{C}(q(t))\rangle + \langle\frac{\partial\lambda(t)}{\partial q_0}, \mathcal{C}(q(t))\rangle\right]dt\\ &= \int_0^T \frac{\partial q(t)}{\partial q_0}\left(\frac{\partial L}{\partial q}(q(t),\dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial\dot{q}}(q(t),\dot{q}(t)) - \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right)dt - \int_0^T \langle\frac{\partial\lambda(t)}{\partial q_0}, \mathcal{C}(q(t))\rangle dt - \frac{\partial L}{\partial\dot{q}}(q(0),\dot{q}(0)),\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{S}}{\partial q_T}(q_0,q_T) &= \int_0^T \left[\frac{\partial q(t)}{\partial q_T}\frac{\partial L}{\partial q}(q(t),\dot{q}(t)) + \frac{\partial\dot{q}(t)}{\partial q_T}\frac{\partial L}{\partial\dot{q}}(q(t),\dot{q}(t))\right]dt - \int_0^T \left[\langle\lambda(t), \frac{\partial q(t)}{\partial q_T}\nabla\mathcal{C}(q(t))\rangle + \langle\frac{\partial\lambda(t)}{\partial q_T}, \mathcal{C}(q(t))\rangle\right]dt\\ &= \int_0^T \frac{\partial q(t)}{\partial q_T}\left(\frac{\partial L}{\partial q}(q(t),\dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial\dot{q}}(q(t),\dot{q}(t)) - \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right)dt - \int_0^T \langle\frac{\partial\lambda(t)}{\partial q_T}, \mathcal{C}(q(t))\rangle dt + \frac{\partial L}{\partial\dot{q}}(q(T),\dot{q}(T)).\end{aligned}$$

By Theorem 2.1, the extremum of $\mathfrak{S}$ is achieved when (q,λ) satisfies the constrained Euler–Lagrange equations (2.6), so we get $D_1\mathcal{S}(q_0,q_T) = -\frac{\partial L}{\partial\dot{q}}(q_0,\dot{q}(0))$ and $D_2\mathcal{S}(q_0,q_T) = \frac{\partial L}{\partial\dot{q}}(q_T,\dot{q}(T))$. □

A.5. **Proof of Theorem 3.3.**

Theorem A.5. *The exact time-T flow map of Hamilton's equations* $(q_0,p_0)\mapsto(q_T,p_T)$ *is implicitly given by the following relations:*

$$q_T = D_2\mathcal{S}(q_0,p_T), \qquad p_0 = D_1\mathcal{S}(q_0,p_T). \tag{A.9}$$

In particular, $\mathcal{S}(q_0,p_T)$ *is a Type II generating function that generates the exact flow of Hamilton's constrained equations (3.6).*

Proof. Using integration by parts and simplifying gives

$$\begin{aligned}\frac{\partial\mathcal{S}}{\partial q_0}(q_0,p_T) &= \frac{\partial q_T}{\partial q_0}p_T + \int_0^T \left[\langle\lambda(t), \frac{\partial q(t)}{\partial q_0}\nabla\mathcal{C}(q(t))\rangle + \langle\frac{\partial\lambda(t)}{\partial q_0}, \mathcal{C}(q(t))\rangle\right]dt\\ &\quad - \int_0^T \left[\frac{\partial p(t)}{\partial q_0}\dot{q}(t) + \frac{\partial\dot{q}(t)}{\partial q_0}p(t) - \frac{\partial q(t)}{\partial q_0}\frac{\partial H}{\partial q}(q(t),p(t)) - \frac{\partial p(t)}{\partial q_0}\frac{\partial H}{\partial p}(q(t),p(t))\right]dt\\ &= p_0 + \int_0^T \frac{\partial q(t)}{\partial q_0}\left(\dot{p}(t) + \frac{\partial H}{\partial q}(q(t),p(t)) + \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right)dt\\ &\quad - \int_0^T \frac{\partial p(t)}{\partial q_0}\left(\dot{q}(t) - \frac{\partial H}{\partial p}(q(t),p(t))\right)dt + \int_0^T \langle\frac{\partial\lambda(t)}{\partial q_0}, \mathcal{C}(q(t))\rangle dt,\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{S}}{\partial p_T}(q_0,p_T) &= q_T + \frac{\partial q_T}{\partial p_T}p_T + \int_0^T \left[\langle\lambda(t), \frac{\partial q(t)}{\partial p_T}\nabla\mathcal{C}(q(t))\rangle + \langle\frac{\partial\lambda(t)}{\partial p_T}, \mathcal{C}(q(t))\rangle\right]dt\\ &\quad - \int_0^T \left[\frac{\partial p(t)}{\partial p_T}\dot{q}(t) + \frac{\partial\dot{q}(t)}{\partial p_T}p(t) - \frac{\partial q(t)}{\partial p_T}\frac{\partial H}{\partial q}(q(t),p(t)) - \frac{\partial p(t)}{\partial p_T}\frac{\partial H}{\partial p}(q(t),p(t))\right]dt\\ &= q_T + \int_0^T \frac{\partial q(t)}{\partial p_T}\left(\dot{p}(t) + \frac{\partial H}{\partial q}(q(t),p(t)) + \langle\lambda(t), \nabla\mathcal{C}(q(t))\rangle\right)dt\\ &\quad - \int_0^T \frac{\partial p(t)}{\partial p_T}\left(\dot{q}(t) - \frac{\partial H}{\partial p}(q(t),p(t))\right)dt + \int_0^T \langle\frac{\partial\lambda(t)}{\partial p_T}, \mathcal{C}(q(t))\rangle dt.\end{aligned}$$

By Theorem 3.1, the extremum of $\mathfrak{S}$ is achieved when the curve (q,p,λ) satisfies Hamilton's constrained equations (3.6), so the integrands vanish, and thus $p_0 = \frac{\partial\mathcal{S}}{\partial q_0}(q_0,p_T) = D_1\mathcal{S}(q_0,p_T)$ and $q_T = \frac{\partial\mathcal{S}}{\partial p_T}(q_0,p_T) = D_2\mathcal{S}(q_0,p_T)$. □

A.6. **Proof of Theorem 3.4.**

Theorem A.6. *The exact time-T flow map of Hamilton's equations $(q_0, p_0) \mapsto (q_T, p_T)$ is implicitly given by the following relations:*

$$q_0 = -D_2\mathcal{S}(q_T, p_0), \qquad p_T = -D_1\mathcal{S}(q_T, p_0). \tag{A.10}$$

In particular, $\mathcal{S}(q_T, p_0)$ is a Type III generating function that generates the exact flow of Hamilton's constrained equations (3.8).

Proof. Integrating by parts and simplifying yields

$$\begin{aligned}
\frac{\partial \mathcal{S}}{\partial q_T}(q_T, p_0) &= -\frac{\partial q_0}{\partial q_T} p_0 + \int_0^T \left[\langle \lambda(t), \frac{\partial q(t)}{\partial q_T} \nabla \mathcal{C}(q(t)) \rangle + \langle \frac{\partial \lambda(t)}{\partial q_T}, \mathcal{C}(q(t)) \rangle \right] dt \\
&\qquad - \int_0^T \left[\frac{\partial p(t)}{\partial q_T} \dot{q}(t) + \frac{\partial \dot{q}(t)}{\partial q_T} p(t) - \frac{\partial q(t)}{\partial q_T} \frac{\partial H}{\partial q}(q(t), p(t)) - \frac{\partial p(t)}{\partial q_T} \frac{\partial H}{\partial p}(q(t), p(t)) \right] dt \\
&= -p_T + \int_0^T \frac{\partial q(t)}{\partial q_T} \left(\dot{p}(t) + \frac{\partial H}{\partial q}(q(t), p(t)) + \langle \lambda(t), \nabla \mathcal{C}(q(t)) \rangle \right) dt \\
&\qquad - \int_0^T \frac{\partial p(t)}{\partial q_T} \left(\dot{q}(t) - \frac{\partial H}{\partial p}(q(t), p(t)) \right) dt + \int_0^T \langle \frac{\partial \lambda(t)}{\partial q_T}, \mathcal{C}(q(t)) \rangle dt,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{S}}{\partial p_0}(q_T, p_0) &= -q_0 - \frac{\partial q_0}{\partial p_0} p_0 + \int_0^T \left[\langle \lambda(t), \frac{\partial q(t)}{\partial p_0} \nabla \mathcal{C}(q(t)) \rangle + \langle \frac{\partial \lambda(t)}{\partial p_0}, \mathcal{C}(q(t)) \rangle \right] dt \\
&\qquad - \int_0^T \left[\frac{\partial p(t)}{\partial p_0} \dot{q}(t) + \frac{\partial \dot{q}(t)}{\partial p_0} p(t) - \frac{\partial q(t)}{\partial p_0} \frac{\partial H}{\partial q}(q(t), p(t)) - \frac{\partial p(t)}{\partial p_0} \frac{\partial H}{\partial p}(q(t), p(t)) \right] dt \\
&= -q_0 + \int_0^T \frac{\partial q(t)}{\partial p_0} \left(\dot{p}(t) + \frac{\partial H}{\partial q}(q(t), p(t)) + \langle \lambda(t), \nabla \mathcal{C}(q(t)) \rangle \right) dt \\
&\qquad - \int_0^T \frac{\partial p(t)}{\partial p_0} \left(\dot{q}(t) - \frac{\partial H}{\partial p}(q(t), p(t)) \right) dt + \int_0^T \langle \frac{\partial \lambda(t)}{\partial p_0}, \mathcal{C}(q(t)) \rangle dt.
\end{aligned}$$

By Theorem 3.2, the extremum of $\mathfrak{S}$ is achieved when the curve (q, p, λ) satisfies Hamilton's constrained equations (3.8), so the integrands vanish, and thus $p_T = -\frac{\partial \mathcal{S}}{\partial q_T}(q_T, p_0) = -D_1\mathcal{S}(q_T, p_0)$ and $q_0 = -\frac{\partial \mathcal{S}}{\partial p_0}(q_T, p_0) = -D_2\mathcal{S}(q_T, p_0)$. □

A.7. **Proof of Theorem 2.3.**

Theorem A.7. *The Type I discrete Hamilton's variational principles*

$$\delta \mathfrak{S}_d^{\pm} \left(\{(q_k, \lambda_k)\}_{k=0}^N \right) = 0 \tag{A.11}$$

are equivalent to the discrete constrained Euler–Lagrange equations

$$D_1 L_d(q_k, q_{k+1}) + D_2 L_d(q_{k-1}, q_k) = \langle \lambda_k, \nabla \mathcal{C}(q_k) \rangle, \quad \mathcal{C}(q_k) = 0, \tag{A.12}$$

where $L_d(q_k, q_{k+1})$ is defined via equation (2.11).

Proof. Using the fact that $\delta q_0 = 0$ and $\delta q_N = 0$, we have

$$\begin{aligned}
\delta \mathfrak{S}_d^- &= \delta \left(\sum_{k=0}^{N-1} \left[L_d(q_k, q_{k+1}) - \langle \lambda_k, \mathcal{C}(q_k) \rangle \right] \right) = \sum_{k=0}^{N-1} \left[D_1 L_d(q_k, q_{k+1}) \delta q_k + D_2 L_d(q_k, q_{k+1}) \delta q_{k+1} \right] - \sum_{k=0}^{N-1} \left(\langle \lambda_k, \nabla \mathcal{C}(q_k) \delta q_k \rangle + \langle \delta \lambda_k, \mathcal{C}(q_k) \rangle \right) \\
&= \sum_{k=1}^{N-1} \left[D_1 L_d(q_k, q_{k+1}) + D_2 L_d(q_{k-1}, q_k) - \langle \lambda_k, \nabla \mathcal{C}(q_k) \rangle \right] \delta q_k - \sum_{k=0}^{N-1} \langle \delta \lambda_k, \mathcal{C}(q_k) \rangle, \\
\delta \mathfrak{S}_d^+ &= \delta \left(\sum_{k=0}^{N-1} \left[L_d(q_k, q_{k+1}) - \langle \lambda_{k+1}, \mathcal{C}(q_{k+1}) \rangle \right] \right) \\
&= \sum_{k=0}^{N-1} \left[D_1 L_d(q_k, q_{k+1}) \delta q_k + D_2 L_d(q_k, q_{k+1}) \delta q_{k+1} \right] - \sum_{k=0}^{N-1} \left(\langle \lambda_{k+1}, \nabla \mathcal{C}(q_{k+1}) \delta q_{k+1} \rangle + \langle \delta \lambda_{k+1}, \mathcal{C}(q_{k+1}) \rangle \right) \\
&= \sum_{k=1}^{N-1} \left[D_1 L_d(q_k, q_{k+1}) + D_2 L_d(q_{k-1}, q_k) - \langle \lambda_k, \nabla \mathcal{C}(q_k) \rangle \right] \delta q_k - \sum_{k=0}^{N-1} \langle \delta \lambda_{k+1}, \mathcal{C}(q_{k+1}) \rangle.
\end{aligned}$$

If the discrete constrained Euler–Lagrange equations (A.12) are satisfied, then each term vanishes and $\delta \mathfrak{S}_d^{\pm} = 0$. Conversely, if $\delta \mathfrak{S}_d^{\pm} = 0$, then a discrete fundamental theorem of the calculus of variations yields the discrete constrained Euler–Lagrange equations (A.12). □

A.8. **Proof of Theorem 3.5.**

Theorem A.8. *The Type II discrete Hamilton's phase space variational principle*

$$\delta \mathfrak{S}_d^{+} \left(\{(q_k, p_k, \lambda_k)\}_{k=0}^N \right) = 0 \tag{A.13}$$

is equivalent to the discrete constrained right Hamilton's equations

$$q_{k+1} = D_2 H_d^+(q_k, p_{k+1}), \qquad p_k = D_1 H_d^+(q_k, p_{k+1}) + \langle \lambda_k, \nabla \mathcal{C}(q_k) \rangle, \qquad \mathcal{C}(q_k) = 0, \tag{A.14}$$

where $H_d^+(q_k, p_{k+1})$ is defined via equation (3.16).

Proof. Using the fact that $\delta q_0 = 0$ and $\delta p_N = 0$ since (q_0, p_N) is fixed, we obtain the following expression for the variations of $\mathfrak{S}_d^+$:

$$\begin{aligned}
\delta\mathfrak{S}_d^+ &= \delta\left(p_N q_N - \sum_{k=0}^{N-1}\left[p_{k+1}q_{k+1} - H_d^+(q_k,p_{k+1}) - \langle\lambda_k, \mathcal{C}(q_k)\rangle\right]\right) = \delta\left(-\sum_{k=0}^{N-2} p_{k+1}q_{k+1} + \sum_{k=0}^{N-1}\left[H_d^+(q_k,p_{k+1}) + \langle\lambda_k,\mathcal{C}(q_k)\rangle\right]\right)\\
&= -\sum_{k=0}^{N-2}\left(q_{k+1}\delta p_{k+1} + p_{k+1}\delta q_{k+1}\right) + \sum_{k=0}^{N-1}\left(D_1H_d^+(q_k,p_{k+1})\delta q_k + D_2H_d^+(q_k,p_{k+1})\delta p_{k+1}\right) + \sum_{k=0}^{N-1}\left(\langle\lambda_k, \nabla\mathcal{C}(q_k)\delta q_k\rangle + \langle\delta\lambda_k, \mathcal{C}(q_k)\rangle\right)\\
&= -\sum_{k=1}^{N-1}\left(q_k\delta p_k + p_k\delta q_k\right) + \sum_{k=1}^{N-1} D_1H_d^+(q_k,p_{k+1})\delta q_k + \sum_{k=0}^{N-2} D_2H_d^+(q_k,p_{k+1})\delta p_{k+1} + \sum_{k=0}^{N-1}\left(\langle\lambda_k, \nabla\mathcal{C}(q_k)\delta q_k\rangle + \langle\delta\lambda_k, \mathcal{C}(q_k)\rangle\right)\\
&= -\sum_{k=1}^{N-1}\left(q_k\delta p_k + p_k\delta q_k\right) + \sum_{k=1}^{N-1} D_1H_d^+(q_k,p_{k+1})\delta q_k + \sum_{k=1}^{N-1} D_2H_d^+(q_{k-1},p_k)\delta p_k + \sum_{k=0}^{N-1}\left(\langle\lambda_k, \nabla\mathcal{C}(q_k)\delta q_k\rangle + \langle\delta\lambda_k, \mathcal{C}(q_k)\rangle\right)\\
&= \sum_{k=1}^{N-1}\left[-q_k + D_2H_d^+(q_{k-1},p_k)\right]\delta p_k + \sum_{k=0}^{N-1}\langle\delta\lambda_k, \mathcal{C}(q_k)\rangle + \sum_{k=1}^{N-1}\left[-p_k + D_1H_d^+(q_k,p_{k+1}) + \langle\lambda_k, \nabla\mathcal{C}(q_k)\rangle\right]\delta q_k.
\end{aligned}$$

If the discrete constrained right Hamilton's equations (A.14) are satisfied, then each term vanishes and $\delta\mathfrak{S}_d^+ = 0$. Conversely, if $\delta\mathfrak{S}_d^+ = 0$, then a discrete fundamental theorem of the calculus of variations yields the discrete constrained right Hamilton's equations (A.14). □

A.9. **Proof of Theorem 3.6.**

Theorem A.9. *The Type III discrete Hamilton's phase space variational principle*

$$\delta\mathfrak{S}_d^-\left(\{(q_k,p_k,\lambda_k)\}_{k=0}^N\right) = 0 \tag{A.15}$$

is equivalent to the discrete constrained left Hamilton's equations

$$q_k = -D_2H_d^-(q_{k+1},p_k), \qquad p_{k+1} = -D_1H_d^-(q_{k+1},p_k) - \langle\lambda_{k+1}, \nabla\mathcal{C}(q_{k+1})\rangle, \qquad \mathcal{C}(q_k) = 0, \tag{A.16}$$

where $H_d^-(q_{k+1},p_k)$ is defined via equation (3.17).

Proof. Using the fact that $\delta q_N = 0$ and $\delta p_0 = 0$ since (q_N, p_0) is fixed, we obtain the following expression for the variations of $\mathfrak{S}_d^-$:

$$\begin{aligned}
\delta\mathfrak{S}_d^- &= \delta\left(-p_0q_0 - \sum_{k=0}^{N-1}\left[-p_kq_k - H_d^-(q_{k+1},p_k) - \langle\lambda_{k+1},\mathcal{C}(q_{k+1})\rangle\right]\right) = \delta\left(\sum_{k=1}^{N-1} p_kq_k + \sum_{k=0}^{N-1}\left[H_d^-(q_{k+1},p_k) + \langle\lambda_{k+1},\mathcal{C}(q_{k+1})\rangle\right]\right)\\
&= \sum_{k=1}^{N-1}\left(q_k\delta p_k + p_k\delta q_k\right) + \sum_{k=0}^{N-1}\left(D_1H_d^-(q_{k+1},p_k)\delta q_{k+1} + D_2H_d^-(q_{k+1},p_k)\delta p_k\right) + \sum_{k=0}^{N-1}\left(\langle\lambda_{k+1}, \nabla\mathcal{C}(q_{k+1})\delta q_{k+1}\rangle + \langle\delta\lambda_{k+1}, \mathcal{C}(q_{k+1})\rangle\right)\\
&= \sum_{k=0}^{N}\left(q_k\delta p_k + p_k\delta q_k\right) + \sum_{k=0}^{N-2} D_1H_d^-(q_{k+1},p_k)\delta q_{k+1} + \sum_{k=1}^{N-1} D_2H_d^-(q_{k+1},p_k)\delta p_k + \sum_{k=0}^{N-1}\left(\langle\lambda_{k+1}, \nabla\mathcal{C}(q_{k+1})\delta q_{k+1}\rangle + \langle\delta\lambda_{k+1}, \mathcal{C}(q_{k+1})\rangle\right)\\
&= \sum_{k=1}^{N-1}\left(q_k\delta p_k + p_k\delta q_k\right) + \sum_{k=1}^{N-1} D_1H_d^-(q_k,p_{k-1})\delta q_k + \sum_{k=1}^{N-1} D_2H_d^-(q_{k+1},p_k)\delta p_k + \sum_{k=1}^{N}\langle\lambda_k, \nabla\mathcal{C}(q_k)\delta q_k\rangle + \sum_{k=0}^{N-1}\langle\delta\lambda_{k+1}, \mathcal{C}(q_{k+1})\rangle\\
&= \sum_{k=1}^{N-1}\left[q_k + D_2H_d^-(q_{k+1},p_k)\right]\delta p_k + \sum_{k=0}^{N-1}\langle\delta\lambda_{k+1}, \mathcal{C}(q_{k+1})\rangle + \sum_{k=1}^{N-1}\left[p_k + D_1H_d^-(q_k,p_{k-1}) + \langle\lambda_k, \nabla\mathcal{C}(q_k)\rangle\right]\delta q_k.
\end{aligned}$$

If the discrete constrained left Hamilton's equations (A.16) are satisfied, then each term vanishes and $\delta\mathfrak{S}_d^- = 0$. Conversely, if $\delta\mathfrak{S}_d^- = 0$, then a discrete fundamental theorem of the calculus of variations yields the discrete constrained left Hamilton's equations (A.16). □

References

[1] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications.*, volume 75 of *Applied Mathematical Sciences*. Springer, New York, second edition, 1988.

[2] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[3] K. Ahn and S. Sra. From Nesterov's estimate sequence to Riemannian acceleration. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 84–118. PMLR, 09–12 Jul 2020.

[4] F. Alimisis, A. Orvieto, G. Bécigneul, and A. Lucchi. Practical accelerated optimization on Riemannian manifolds, 2020.

[5] 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.

[6] V. I. Arnol′d. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1989. Translated from the Russian by K. Vogtmann and A. Weinstein.

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

[8] V. Duruisseaux and M. Leok. A variational formulation of accelerated optimization on Riemannian manifolds. 2021.
[9] V. Duruisseaux and M. Leok. Accelerated optimization on Riemannian manifolds via projected variational integrators. in preparation, 2021.
[10] V. Duruisseaux, J. Schmitt, and M. Leok. Adaptive Hamiltonian variational integrators and applications to symplectic accelerated optimization. 2020.
[11] L. Eldén and H. Park. A Procrustes problem on the Stiefel manifold. *Numerische Mathematik*, 82(4): 599–619, 1999. doi: 10.1007/s002110050432.
[12] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013. ISBN 9781421407944.
[13] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.
[14] J. Hall and M. Leok. Spectral Variational Integrators. *Numer. Math.*, 130(4):681–740, 2015.
[15] D. D. Holm, T. Schmah, and C. Stoica. *Geometric mechanics and symmetry*, volume 12 of *Oxford Texts in Applied and Engineering Mathematics*. Oxford University Press, Oxford, 2009. From finite to infinite dimensions.
[16] J. Jost. *Riemannian geometry and geometric analysis*. Universitext. Springer, Cham, 7th edition, 2017.
[17] S. Lang. *Fundamentals of Differential Geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer -Verlag, New York, 1999. ISBN 9780387985930.
[18] J. M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2018.
[19] M. Leok and T. Ohsawa. Variational and geometric structures of discrete Dirac mechanics. *Found. Comput. Math.*, 11(5):529–562, 2011.
[20] M. Leok and J. Zhang. Discrete Hamiltonian variational integrators. *IMA Journal of Numerical Analysis*, 31(4):1497–1532, 2011.
[21] Y. Liu, F. Shang, J. Cheng, H. Cheng, and L. Jiao. Accelerated first-order methods for geodesically convex optimization on Riemannian manifolds. In *Advances in Neural Information Processing Systems*, volume 30, pages 4868–4877. Curran Associates, Inc., 2017.
[22] 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.
[23] J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numer.*, 10:357–514, 2001.
[24] J. Nash. The imbedding problem for Riemannian manifolds. *Annals of Mathematics*, 63(1):20–63, 1956. ISSN 0003486X.
[25] 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.
[26] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004.
[27] S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36:1549–1570, 1999.
[28] J. M. Schmitt and M. Leok. Properties of Hamiltonian variational integrators. *IMA Journal of Numerical Analysis*, 38(1):377–398, 03 2017.
[29] 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.
[30] 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.
[31] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages 1139–1147, 2013.
[32] H. Whitney. The singularities of a smooth n-manifold in $(2n-1)$-space. *Annals of Mathematics*, 45(2): 247–293, 1944. ISSN 0003486X.
[33] H. Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *Annals of Mathematics*, 45(2): 220–246, 1944. ISSN 0003486X.

[34] 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.

[35] H. Zhang and S. Sra. First-order methods for geodesically convex optimization. In *29th Annual Conference on Learning Theory*, pages 1617–1638, 2016.

[36] H. Zhang and S. Sra. An estimate sequence for geodesically convex optimization. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1703–1723, 06–09 Jul 2018.