Variational Symplectic Accelerated Optimization on Lie Groups

Taeyoung Lee, Molei Tao, and Melvin Leok

Abstract— **There has been significant interest in generalizations of the Nesterov accelerated gradient descent algorithm due to its improved performance guarantee compared to the standard gradient descent algorithm, and its applicability to large scale optimization problems arising in deep learning. A particularly fruitful approach is based on numerical discretizations of differential equations that describe the continuous time limit of the Nesterov algorithm, and a generalization involving time-dependent Bregman Lagrangian and Hamiltonian dynamics that converges at an arbitrarily fast rate to the minimum. We develop a Lie group variational discretization based on an extended path space formulation of the Bregman Lagrangian on Lie groups, and analyze its computational properties with two examples in attitude determination and vision-based localization.**

I. Introduction

Nesterov's accelerated gradient descent algorithm [1] was introduced in 1983, and it exhibits the convergence rate of $\mathcal{O}(1/k^2)$ when applied to a convex objective function, which is faster than the $\mathcal{O}(1/k)$ convergence rate of standard gradient descent methods. It is shown in [2] that this rate of convergence is optimal for the class of first-order gradient methods. This improved rate of convergence over the standard gradient method is referred to as acceleration, and there is a great interest in developing systematic approaches to the construction of efficient accelerated optimization algorithms, driven by potential applications in deep learning.

A continuous time limit of the Nesterov algorithm was studied in [3], whose flow converges to the minimum at $\mathcal{O}(1/t^2)$, and this was generalized in [4] using a time-dependent Bregman Lagrangian and Hamiltonian to obtain higher-order convergence of $\mathcal{O}(1/t^p)$ for arbitrary $p \geq 2$. However, it has been shown that discretizing Bregman dynamics is not trivial as common discretizations fail to achieve the higher convergence rate guaranteed in the continuous time limit. As such, there have been several attempts to construct accelerated optimization algorithms using geometric structure-preserving discretizations of the Bregman dynamics [5].

A natural class[1] of geometric numerical integrators [6] for discretizing such Lagrangian or Hamiltonian systems is variational integrators [9], [10]. They are constructed by a discrete analogue of Hamilton's variational principle, and therefore, their numerical flows are symplectic. They also satisfy a discrete Noether's theorem that relates symmetries with momentum conservation properties, and further exhibit excellent exponentially long-time energy stability. One complication is that such methods are typically developed for autonomous Lagrangian and Hamiltonian systems on the Euclidean space. To address this, variational integrators have been developed on a Lie group [11], and time-adaptive Hamiltonian variational integrators have been proposed [12].

In this paper, we focus on the optimization problem to minimize an objective function defined on an a Lie group. Optimization on a manifold or a Lie group appears in various areas of machine learning, engineering, and applied mathematics [13], [14], and respecting the geometric structure of manifolds yields more accurate and efficient optimization schemes, when compared to methods based on embeddings in a higher-dimensional Euclidean space with algebraic constraints, or using local coordinates.

In particular, we formulate a Bregman Lagrangian system on a Lie group, and we further discretize it using the extended Lie group variational integrator to construct an intrinsic accelerated optimization scheme, which inherits the desirable properties of variational integrators while also preserving the group structure. Compared with [12] where the evolution of the stepsize is prescribed, the proposed scheme adaptively adjusts the stepsize according to the extended variational principle at the cost of increased computational load. The resulting computational properties of the proposed approach are analyzed with two examples in attitude determination and vision-based localization, where it is observed that the scheme exhibits an interesting convergence of the adaptive stepsize, and the variational discretization provides robustness against the choice of stepsize, which is exploited in the numerical experiments to improve computational efficiency. We also present benchmark studies against other discretization schemes applied to the Bregman dynamics, and other accelerated optimization schemes on a Lie group [7].

II. Extended Lagrangian Mechanics

This section presents Lagrangian mechanics for non-autonomous systems on a Lie group. It is referred to as *extended* Lagrangian mechanics as the variational principle is extended to include reparamerization of time [9]. These are developed in both of continuous-time and discrete-time formulations. The latter yields a *Lie group variational integrator* [11], which will be applied to accelerated optimization using the Bregman Lagrangian in the next section.

Taeyoung Lee, Mechanical and Aerospace Engineering, George Washington University, Washington, DC 20052. tylee@gwu.edu

Molei Tao, Mathematics, Georgia Institute of Technology, Atlanta, GA 30332. mtao@gatech.edu

Melvin Leok, Mathematics, University of California–San Diego, La Jolla, CA 92093. mleok@ucsd.edu

The research was partially supported by NSF under the grants CNS-1837382, CMMI-1760928, DMS-1813635, IIP-1747760, DMS-1847802, ECCS-1936776, and by AFOSR FA9550-18-1-0288 and DoD 13106725.

[1] Note that other classes of discretization methods exist, such as those based on splitting (e.g., [6], [7]) and composition (e.g., [8]), and such approaches also arise in variational discretization [9].

Consider an n-dimensional Lie group G. Let $\mathfrak{g}$ be the associated Lie algebra, or the tangent space at the identity, i.e., $\mathfrak{g} = \mathsf{T}_e\mathsf{G}$. Consider a left trivialization of the tangent bundle of the group $\mathsf{TG} \simeq \mathsf{G} \times \mathfrak{g}$, $(g, \dot{g}) \mapsto (g, L_{g^{-1}}\dot{g}) \equiv (g, \xi)$ More specifically, let $\mathsf{L} : \mathsf{G} \times \mathsf{G} \to \mathsf{G}$ be the left action defined such that $\mathsf{L}_g h = gh$ for $g, h \in \mathsf{G}$. Then the left trivialization is a map $(g, \dot{g}) \mapsto (g, L_{g^{-1}}\dot{g}) \equiv (g, \xi)$, where $\xi \in \mathfrak{g}$, and the kinematics equation can be written as

$$\dot{g} = g\xi. \tag{1}$$

Further, suppose $\mathfrak{g}$ is equipped with an inner product $\langle \cdot, \cdot \rangle$, which induces an inner product on $\mathsf{T}_g\mathsf{G}$ via left trivialization. For any $v, w \in \mathsf{T}_g\mathsf{G}$, $\langle w, v \rangle_{\mathsf{T}_g\mathsf{G}} = \langle \mathsf{T}_g\mathsf{L}_{g^{-1}}v, \mathsf{T}_g\mathsf{L}_{g^{-1}}w \rangle_{\mathfrak{g}}$. Given the inner product, we identify $\mathfrak{g} \simeq \mathfrak{g}^*$ and $\mathsf{T}_g\mathsf{G} \simeq \mathsf{T}^*_g\mathsf{G} \simeq G \times \mathfrak{g}^*$ via the Riesz representation. Throughout this paper, the pairing is also denoted by the dot product $\cdot$. Let $\mathbf{J} : \mathfrak{g} \to \mathfrak{g}^*$ be chosen such that $\langle \mathbf{J}(\xi), \zeta \rangle$ is positive-definite and symmetric as a bilinear form of $\xi, \zeta \in \mathfrak{g}$. Define the metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ with $\langle\!\langle \xi, \zeta \rangle\!\rangle = \langle \mathbf{J}(\xi), \zeta \rangle$. This serves as a left-invariant Riemmanian metric on G. Also $\|\xi\|^2 = \langle\!\langle \xi, \xi \rangle\!\rangle$ for any $\xi \in \mathfrak{g}$. The adjoint operator is denoted by $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$, and the ad operator is denoted by $\mathrm{ad}_\xi : \mathfrak{g} \to \mathfrak{g}$. See, for example [15] for detailed preliminaries.

A. Continuous-Time Extended Lagrangian Mechanics

Consider a non-autonomous (left-trivialized) Lagrangian $L(t, g, \xi) : \mathbb{R} \times \mathsf{G} \times \mathfrak{g} \to \mathbb{R}$ on the *extended state space*. The corresponding *extended path space* is composed of the curves $(c_t(a), c_g(a))$ on $\mathbb{R} \times \mathsf{G}$ parameterized by $a > 0$. To ensure that the reparameterized time increases monotonically, we require $c'_t(a) > 0$. For a given time interval $[t_0, t_f]$, the corresponding interval $[a_0, a_f]$ for a is chosen such that $t_0 = c_t(a_0)$ and $t_f = c_t(a_f)$. For any path $(c_t(a), c_g(a))$ over $[a_0, a_f]$ in the extended space, the *associated curve* is

$$g(t) = c_g(c_t^{-1}(t)), \tag{2}$$

on G over the time interval $[t_0, t_f]$. For a given extended path, define the *extended action integral* as

$$\mathfrak{G}(c_t, c_g) = \int_{t_0}^{t_f} L(t, g, \xi)\bigg|_{g(t) = c_g(c_t^{-1}(t))} dt, \tag{3}$$

where the Lagrangian is evaluated on the associated curve (2), and ξ satisfies the kinematics equation (1).

Taking the variation of $\mathfrak{G}$ with respect to the extended path, we obtain the Euler–Lagrange equation according to the variational principle in the extended phase space. As discussed in [9, Sec. 4.2.2], the resulting Euler–Lagrange equations depend only on the associated curve (2), not on the extended path (c_t, c_g) itself, and the variational principle does not dictate how the curve should be reparameterized.

Further, the resulting Euler–Lagrange equation share the exactly same form as (unextended) Lagrangian mechanics for the associated curve. As such, the Euler–Lagrange equation for non-autonomous Lagrangian $L(t, g, \xi) : \mathbb{R} \times \mathsf{G} \times \mathfrak{g} \to \mathbb{R}$ can be written as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \xi}\right) - \mathrm{ad}^*_\xi \frac{\partial L}{\partial \xi} - \mathsf{T}^*_e\mathsf{L}_g(\mathbf{D}_g L) = 0, \tag{4}$$

where $\mathbf{D}_g$ stands for the differential with respect to g (see [16, Sec. 8.6.3] for derivation of the above equation for autonomous Lagrangians).

Introducing the Legendre transform $\mu = \frac{\partial L}{\partial \xi} \in \mathfrak{g}^*$, and assuming that it is invertible, the Euler–Lagrange equation can be rewritten as

$$\dot{\mu} - \mathrm{ad}^*_\xi \mu - \mathsf{T}^*_e\mathsf{L}_g(\mathbf{D}_g L) = 0. \tag{5}$$

B. Extended Lie Group Variational Integrator

Variational integrators are geometric numerical integration schemes that can be viewed as discrete-time mechanics derived from a discretization of the variational principle for Lagrangian mechanics [9]. The discrete-time flows of variational integrators are symplectic and they exhibit a discrete analogue of Noether's theorem. This provides long-term structural stability in the resulting numerical simulations. For Lagrangian mechanics evolving on a Lie group, the corresponding Lie group variational integrators were developed in [11].

Here, we develop extended Lie group variational integrators by discretizing the extended variational principle presented above, following the general framework of [9]. The *extended discrete path space* is composed of the sequence $\{(t_k, g_k)\}_{k=0}^N$ on $\mathbb{R} \times \mathsf{G}$, satisfying $t_{k+1} > t_k$. Next, the discrete kinematics equation is chosen to be

$$g_{k+1} = g_k f_k, \tag{6}$$

for $f_k \in \mathsf{G}$ representing the relative update over a single timestep. The discrete Lagrangian $L_d(t_k, t_{k+1}, g_k, f_k) : \mathbb{R} \times \mathbb{R} \times \mathsf{G} \times \mathsf{G} \to \mathbb{R}$ is chosen such that the following *extended discrete action sum*

$$\mathfrak{G}_d(\{(t_k, g_k)\}_{k=0}^N) = \sum_{k=0}^{N-1} L_d(t_k, t_{k+1}, g_k, f_k), \tag{7}$$

approximates (3).

Proposition 1: The discrete path $\{(g_k, f_k)\}_{k=0}^{N-1}$ that extremizes the discrete action sum (7) subject to fixed endpoints satisfies the following discrete Euler–Lagrange equation,

$$\begin{aligned} \mathsf{T}^*_e\mathsf{L}_{g_k}(\mathbf{D}_{g_k}L_{d_k}) - \mathrm{Ad}^*_{f_k^{-1}}(\mathsf{T}^*_e\mathsf{L}_{f_k}(\mathbf{D}_{f_k}L_{d_k})) \\ + \mathsf{T}^*_e\mathsf{L}_{f_{k-1}}(\mathbf{D}_{f_{k-1}}L_{d_{k-1}}) = 0, \end{aligned} \tag{8}$$

$$\mathbf{D}_{t_k}L_{d_{k-1}} + \mathbf{D}_{t_k}L_{d_k} = 0, \tag{9}$$

which together with the discrete kinematic equation (6) defines an extended Lie group variational integrator.

Proof: From (6), $\delta f_k = -g_k^{-1}(\delta g_k)g_k^{-1}g_{k+1} + g_k^{-1}\delta g_{k+1}$. Since δg_k can be written as $\delta g_k = g_k\eta_k$ for $\eta_k \in \mathfrak{g}$,

$$f_k^{-1}\delta f_k = -\mathrm{Ad}_{f_k^{-1}}\eta_k + \eta_{k+1}. \tag{10}$$

Take the variation of (7) and substitute (10) to obtain

$$\begin{aligned} \delta\mathfrak{G}_d = \sum_{k=0}^{N-1} & \mathsf{T}^*_e\mathsf{L}_{g_k}(\mathbf{D}_{g_k}L_{d_k}) \cdot \eta_k \\ & + \mathsf{T}^*_e\mathsf{L}_{f_k}(\mathbf{D}_{f_k}L_{d_k}) \cdot (-\mathrm{Ad}_{f_k^{-1}}\eta_k + \eta_{k+1}) \\ & + \mathbf{D}_{t_k}L_{d_k} \cdot \delta t_k + \mathbf{D}_{t_{k+1}}\mathbf{D}_{d_k} \cdot \delta t_{k+1}. \end{aligned}$$

Since the endpoints are fixed, we have $\eta_0 = 0$ and $\delta t_0 = 0$. Therefore in the above expression, the range of summation for the terms paired with η_k and δt_k can be reduced to $1 \leq k \leq N-1$. Also, using $\eta_N = 0$ and $\delta t_N = 0$, for the other terms paired with η_{k+1} and δt_{k+1}, the terms can be reindexed by reducing the subscripts by one and summed over the same range. According to the variational principle, $\delta \mathfrak{G}_d = 0$ for any η_k and δt_k, which yields (8) and (9). ■

The most notable difference compared to the continuous-time counterpart is that in addition to the discrete Euler–Lagrange equation (8), we have the additional equation (9) for the evolution of the discrete time. This is because the discrete action sum $\mathfrak{G}_d$ depends on the complete extended path $\{(t_k, g_k)\}_{k=1}^N$. Whereas the continuous-time action $\mathfrak{G}$ is only a function of the associated curve (2).

The discrete Euler–Lagrange equation for the discrete time (9) is associated with an energy. Define the discrete energy to be

$$E_k^+ = -\mathbf{D}_{t_{k+1}} L_{d_k}, \tag{11}$$

$$E_k^- = \mathbf{D}_{t_k} L_{d_k}. \tag{12}$$

Then, (9) can be rewritten as

$$E_{k-1}^+ = E_k^-, \tag{13}$$

which reflects the evolution of the discrete energy. When the discrete Lagrangian is autonomous, (13) implies the conservation of discrete energy, thereby yielding a symplectic-energy-momentum integrator [17].

To implement (8) and (9) as a numerical integrator, it is more convenient to introduce the *extended discrete Legendre transforms*, $\mathbb{F}^{\pm} L_{d_k} : \mathbb{R} \times \mathbb{R} \times \mathsf{G} \times \mathsf{G} \to \mathbb{R} \times \mathbb{R} \times \mathsf{G} \times \mathfrak{g}^*$ as

$$\mathbb{F}^+ L_{d_k}(t_k, t_{k+1}, g_k, f_k) = (t_{k+1}, E_{k+1}, g_{k+1}, \mu_{k+1}), \tag{14}$$

$$\mathbb{F}^- L_{d_k}(t_k, t_{k+1}, g_k, f_k) = (t_k, E_k, g_k, \mu_k). \tag{15}$$

where

$$\mu_k = -\mathsf{T}_e^* \mathsf{L}_{g_k}(\mathbf{D}_{g_k} L_{d_k}) + \mathrm{Ad}^*_{f_k^{-1}}(\mathsf{T}_e^* \mathsf{L}_{f_k}(\mathbf{D}_{f_k} L_{d_k})), \tag{16}$$

$$\mu_{k+1} = \mathsf{T}_e^* \mathsf{L}_{f_k}(\mathbf{D}_{f_k} L_{d_k}), \tag{17}$$

and E_{k+1} and E_k are given by (11) and (12), respectively.

The resulting discrete flow map is defined by $\mathbb{F}^+ L_{d_k} \circ (\mathbb{F} L_{d_k})^{-1}$. More specifically, for given (t_k, E_k, g_k, μ_k), (12) and (16) are solved together for t_{k+1}, f_k with the constraint $t_{k+1} > t_k$. Then, $(E_{k+1}, g_{k+1}, \mu_{k+1})$ are computed by (11), (6), and (17), respectively. This yields the discrete flow map $(t_k, E_k, g_k, \mu_k) \to (t_{k+1}, E_{k+1}, g_{k+1}, \mu_{k+1})$ consistent with (8) and (9). While the flow map is expressed in terms of E for convenience, the initial value of E_0 is often selected by choosing the initial timestep h_0 and calculating the corresponding value of E_0 through (12). This inherits the desirable properties of variational integrators, and the group structure is also preserved through (6).

III. Bregman Lagrangian Systems on G

Let $\mathsf{f} : \mathsf{G} \to \mathbb{R}$ be a real-valued smooth function on G. We focus on the optimization problem:

$$\min_{g \in \mathsf{G}} \mathsf{f}(g). \tag{18}$$

A variational accelerated optimization scheme for the above problem was developed in [7], where the Nesterov accelerated gradient (NAG) descent on a finite-dimensional vector space was intrinsically generalized to a Lie group. In this section, we introduce an intrinsic formulation of Bregman Lagrangian dynamics [4], which encompasses a larger class of accelerated optimization scheme, including NAG. More importantly, the continuous dynamics guarantees polynomial convergence rates up to an arbitrary order.

A. Continuous-Time Bregman Dynamics

The Bregman Lagrangian $L(t, g, \xi) : \mathbb{R} \times \mathsf{G} \times \mathfrak{g} \to$ is

$$L(t, g, \xi) = \frac{t^{\lambda p + 1}}{2p} \|\xi\|^2 - Cpt^{(\lambda+1)p-1} \mathsf{f}(g), \tag{19}$$

where $\|\xi\|^2 = \langle\!\langle \xi, \xi \rangle\!\rangle = \langle \mathbf{J}(\xi), \xi \rangle$, for $p, C > 0$, and $\lambda \geq 1$. When $\mathsf{G} = \mathbb{R}^n$ and $\lambda = 1$, this recovers the Bregman Lagrangian for vector spaces [4], and it yields the continuous-time limit of Nesterov's accelerated gradient descent for $p = 2$ [18]. Also, in case $p = 3$, it corresponds to the continuous-time limit of Nesterov's accelerated cubic-regularized Newton's method [19]. When G is considered as a Riemannian manifold, this corresponds to the p-Bregman Lagrangian in [20]. The additional term λ accounts for the sectional curvature and diameter of the manifold [21].

The left-trivialized derivative of the objective function is

$$\nabla_{\mathsf{L}} \mathsf{f}(g) = \mathsf{T}_e^* \mathsf{L}_g(\mathbf{D}_g \mathsf{f}(g)). \tag{20}$$

Applying (4) to (19), the corresponding Euler–Lagrange equations are given below.

Proposition 2: The Euler–Lagrange equations corresponding to the Bregman Lagrangian (19) are

$$\frac{d\mathbf{J}(\xi)}{dt} + \frac{\lambda p + 1}{t} \mathbf{J}(\xi) - \mathrm{ad}_\xi^* \mathbf{J}(\xi) + Cp^2 t^{p-2} \nabla_{\mathsf{L}} \mathsf{f}(g) = 0, \tag{21}$$

and (1). Further, the corresponding continuous flow locally converges to the minimizer g^* of f with the rate given by

$$\mathsf{f}(g(t)) - \mathsf{f}(g^*) \in \mathcal{O}(t^{-p}), \tag{22}$$

when f is geodesically convex.

Proof: We have

$$\frac{\partial L}{\partial \xi} = \frac{t^{\lambda p+1}}{p} \mathbf{J}(\xi)$$

Substituting this into (4) and using (20),

$$\frac{t^{\lambda p+1}}{p} \frac{d\mathbf{J}(\xi)}{dt} + \frac{(\lambda p + 1) t^{\lambda p}}{p} \mathbf{J}(\xi) - \frac{t^{\lambda p+1}}{p} \mathrm{ad}_\xi^* \mathbf{J}(\xi)$$
$$+ Cpt^{(\lambda+1)p-1} \nabla_{\mathsf{L}} \mathsf{f}(g) = 0.$$

Dividing both sides by $\frac{t^{\lambda p+1}}{p}$ yields (21). The convergence property is established by [20, Theorem 3.2]. ■

Therefore, the optimization problem on G can be addressed by numerically integrating (21) from an initial guess. However, it has been observed that a naïve discretization is not able to match the polynomial convergence rate established in [4]. Further, we need a guarantee that the discrete trajectory evolves on the Lie group.

These two challenges can be addressed by applying a Lie group variational integrator, as their structure-preserving properties provides long-term numerical stability, and preservation of the group structure. In the subsequent section, we derive Lie group variational integrators for the Bregman Lagrangian system.

B. Lie Group Variational Integrator for Bregman Dynamics

Let $h_k = t_{k+1} - t_k$ and $t_{k,k+1} = t_k + h_k/2$. We consider the following form of the discrete Lagrangian

$$\begin{aligned} L_d(t_k, t_{k+1}, g_k, f_k) = \frac{\phi(t_{k,k+1})}{h_k} T_d(f_k) - \frac{h_k}{2}\theta(t_k)\mathsf{f}(g_k) \\ - \frac{h_k}{2}\theta(t_{k+1})\mathsf{f}(g_k f_k), \end{aligned} \tag{23}$$

where $T_d(f_k) : \mathsf{G} \to \mathbb{R}$ is chosen such that it approximates $T(f_k) \approx h_k^2 \|\xi_k\|^2/2$, and $\phi, \theta : \mathbb{R} \to \mathbb{R}$ are

$$\phi(t) = \frac{t^{\lambda p+1}}{p}, \tag{24}$$

$$\theta(t) = Cpt^{(\lambda+1)p-1}. \tag{25}$$

The corresponding variational integrators are presented as follows.

Proposition 3: The discrete-time Euler–Lagrange equations, or the Lie group variational integrator for the discrete Lagrangian (23) corresponding to the Bregman Lagrangian (19) are given by

$$\mu_k = \frac{\phi_{k,k+1}}{h_k} \mathrm{Ad}^*_{f_k^{-1}}(\mathsf{T}^*_e \mathsf{L}_{f_k}(\mathbf{D}_{f_k} T_{d_k})) + \frac{h_k \theta_k}{2} \nabla_L \mathsf{f}_k, \tag{26}$$

$$\mu_{k+1} = \mathrm{Ad}^*_{f_k}(\mu_k - \frac{h_k\theta_k}{2}\nabla_L \mathsf{f}_k) - \frac{h_k \theta_{k+1}}{2}\nabla_{\mathsf{L}} \mathsf{f}_{k+1}, \tag{27}$$

$$\begin{aligned} E_k = \frac{\phi'_{k,k+1}}{2h_k} T_{d_k} - \frac{h_k \theta'_k}{2}\mathsf{f}_k \\ + \frac{\phi_{k,k+1}}{h_k^2} T_{d_k} + \frac{\theta_k}{2}\mathsf{f}_k + \frac{\theta_{k+1}}{2}\mathsf{f}_{k+1}, \end{aligned} \tag{28}$$

$$\begin{aligned} E_{k+1} = -\frac{\phi'_{k,k+1}}{2h_k} T_{d_k} + \frac{h_k \theta'_{k+1}}{2}\mathsf{f}_{k+1} \\ + \frac{\phi_{k,k+1}}{h_k^2} T_{d_k} + \frac{\theta_k}{2}\mathsf{f}_k + \frac{\theta_{k+1}}{2}\mathsf{f}_{k+1}, \end{aligned} \tag{29}$$

together with (6).

Proof: These can be derived by substituting (23) into (16), (17), (12), and (11), respectively. ■

As discussed at the end of Section III, these provide symplectic and momentum-preserving discrete time flow maps. Since these correspond to a discretization of the Bregman Lagrangian system, they can be considered as a geometric numerical integrator for (21), or utilized as an optimization algorithm on G. If $T_d(f_k) = T_d(f_k^{-1})$, then the discrete Lagrangian is self-adjoint, and the above integrator is symmetric and therefore at least second-order accurate.

IV. Optimization on G

In this section, we present both of the continuous Bregman Lagrangian system and the Lie group variational integrator for several Lie groups.

A. Euclidean Space $\mathbb{R}^n$

Suppose $\mathsf{G} = \mathbb{R}^n$, with the additive group action, and the inner product is chosen to be $\langle x, y \rangle = x^T y$ for any $x, y \in \mathbb{R}^n$. Let $\mathbf{J}(\dot{x}) = I_{n\times n}\dot{x}$, and $\lambda = 1$.

From (21), the continuous Euler–Lagrange equation is given by

$$\ddot{x} + \frac{p+1}{t}\dot{x} + Cp^2 t^{p-2}\nabla \mathsf{f}(x) = 0, \tag{30}$$

which recovers the differential equation derived in [4].

Next, we develop variational integrators. The discrete kinematics equation (6) is rewritten as $x_{k+1} = x_k + \Delta x_k$ for $\Delta x_k \in \mathbb{R}^n$. The kinetic energy term in (23) is chosen as

$$T_d = \frac{1}{2}\|\Delta x_k\|^2. \tag{31}$$

According to Proposition 3, we obtain the discrete Euler–Lagrange equations as follows.

Proposition 4: When $G = \mathbb{R}^n$, the variational integrator for the discrete Bregman Lagrangian (23) is given by

$$v_k = \frac{\phi_{k,k+1}}{h_k}\Delta x_k + \frac{h_k\theta_k}{2}\nabla \mathsf{f}_k, \tag{32}$$

$$v_{k+1} = v_k - \frac{h_k \theta_k}{2}\nabla \mathsf{f}_k - \frac{h_k\theta_{k+1}}{2}\nabla \mathsf{f}_{k+1}, \tag{33}$$

and (28), (29) with (31).

These are implicit as (32) and (28) should be solved together for Δx_k and h_k. One straightforward approach is fixed-point iteration. For a given h_k, (32) can be solved explicitly for Δ_k, which yields x_{k+1}. Then, (28) can be solved for h_k. These procedures are iterated until h_k converges.

B. Three-Dimensional Special Orthogonal Group $\mathsf{SO}(3)$

Next, consider $\mathsf{SO}(3) = \{R \in \mathbb{R}^{3\times3} \,|\, R^T R = I_{3\times3}, \det(R)] = 1\}$. Its Lie algebra is $\mathfrak{so}(3) = \{S \in \mathbb{R}^{3\times3} \,|\, S^T = -S\}$ with the matrix commutator as the Lie bracket. This is identified with $\mathbb{R}^3$ through the *hat* map $\hat{\cdot} : \mathbb{R}^3 \to \mathfrak{so}(3)$ defined such that $\hat{x} \in \mathfrak{so}(3)$ and $\hat{x}y = x \times y$ for any $x, y \in \mathbb{R}^3$. The inverse of the hat map is denoted by the *vee* map $\vee : \mathfrak{so}(3) \to \mathbb{R}^3$. The inner product is given by

$$\langle \hat{\eta}, \hat{\xi} \rangle_{\mathfrak{so}(3)} = \frac{1}{2}\mathrm{tr}\Big[\hat{\eta}^T \hat{\xi}\Big] = \eta^T \xi = \langle \eta, \xi \rangle_{\mathbb{R}^3}.$$

The metric is chosen as

$$\langle \mathbf{J}(\hat{\eta}), \hat{\xi} \rangle_{\mathfrak{so}(3)} = \mathrm{tr}\Big[\hat{\eta}^T J_d \hat{\xi}\Big] = \eta^T J \xi = \langle J\eta, \xi \rangle_{\mathbb{R}^3}, \tag{34}$$

where $J \in \mathbb{R}^{3\times3}$ is a symmetric, positive-definite matrix, and $J_d = \frac{1}{2}\mathrm{tr}[J]\, I_{3\times3} - J \in \mathbb{R}^{3\times3}$. Further,

$$\begin{aligned} \mathrm{ad}_\eta \xi = \eta \times \xi, \qquad \mathrm{ad}^*_\eta \xi = \xi \times \eta, \\ \mathrm{Ad}_F \eta = F\eta, \qquad \mathrm{Ad}^*_F \eta = F^T \eta. \end{aligned}$$

Consider

$$L(t,R,\Omega)=\frac{t^{p+1}}{2p}\Omega\cdot J\Omega-Cpt^{2p-1}\mathsf{f}(R).$$

From (21), the Euler–Lagrange equations are given by

$$J\dot{\Omega}+\frac{p+1}{t}J\Omega+\hat{\Omega}J\Omega+Cp^2t^{p-2}\nabla_{\mathsf{L}}\mathsf{f}(R)=0, \qquad (35)$$

$$\dot{R}=R\hat{\Omega}. \qquad (36)$$

Next, we derive variational integrators. The kinematics equation is written as

$$R_{k+1}=R_kF_k, \qquad (37)$$

for $F_k\in\mathsf{SO}(3)$. Similar with [11], the angular velocity is approximated with $\hat{\Omega}_k\approx\frac{1}{h_k}R_k^T(R_{k+1}-R_k)=\frac{1}{h_k}(F_k-I_{3\times3})$. Substituting this into (34),

$$T_d(F_k)=\mathrm{tr}[(I_{3\times3}-F_k)J_d], \qquad (38)$$

which satisfies $T_d(F_k)=T_d(F_k^T)$.

Proposition 5: When $\mathsf{G}=\mathsf{SO}(3)$, the Lie group variational integrator for the discrete Bregman Lagrangian (23) with (38) is given by

$$\mu_k=\frac{\phi_{k,k+1}}{h_k}(F_kJ_d-J_dF_k^T)^\vee+\frac{h_k\theta_k}{2}\nabla_{\mathsf{L}}\mathsf{f}_k, \qquad (39)$$

$$\mu_{k+1}=F_k^T\mu_k-\frac{h_k\theta_k}{2}\nabla_{\mathsf{L}}\mathsf{f}_k-\frac{h_k\theta_k}{2}\nabla_{\mathsf{L}}\mathsf{f}_{k+1}, \qquad (40)$$

together with (29), (37), (28), and (38).

Proof: Let $\delta F_k=F_k\hat{\chi}_k$. The derivative of (38) is

$$\mathbf{D}_{F_k}T_{d_k}\cdot\delta F_k=\mathrm{tr}[-F_k\hat{\chi}_kJ_d]=(J_dF_k-F_k^TJ_d)^\vee\cdot\chi,$$

where the last equality is from the identity, $\mathrm{tr}[-\hat{x}A]=x\cdot(A-A^T)^\vee$ for any $x\in\mathbb{R}^3$ and $A\in\mathbb{R}^{3\times3}$. Thus, $\mathsf{T}_I^*\mathsf{L}_{F_k}(\mathbf{D}_{F_k}T_{d_k})=(J_dF_k-F_k^TJ_d)^\vee$. Substituting this into (26) and (27) yields (39) and (40), respectively. ■

To implement these, (40) and (28) should be solved together for h_k and F_k. For a given h_k, computational approaches to solve (39) for F_k are presented in [22, Sec 3.3.8]. When $J=I_{3\times3}$, or equivalently when $J_d=\frac{1}{2}I_{3\times3}$, (39) can be solved explicitly to obtain

$$F_k=\exp\left(\frac{\sin^{-1}\|a\|}{\|a\|}\hat{a}\right), \qquad (41)$$

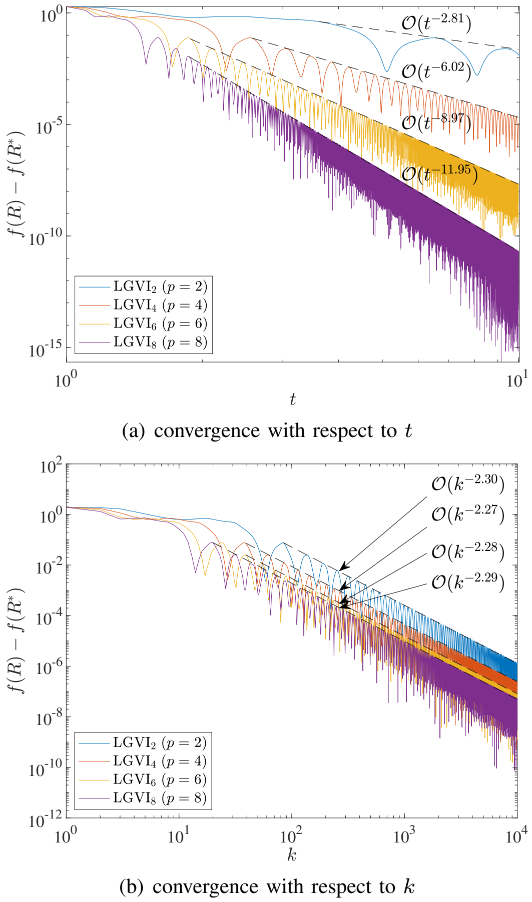
where $a=\frac{h_k}{\phi_{k,k+1}}(\mu_k-\frac{h_k\theta_k}{2}\nabla_{\mathsf{L}}\mathsf{f}_k)\in\mathbb{R}^3$. This can replace (39).

C. Product of $\mathbb{R}^n$ and $\mathsf{SO}(3)$

Suppose $\mathsf{G}=\mathsf{SO}(3)\times\mathbb{R}^n$. As it is the direct product of $\mathsf{SO}(3)$ and $\mathbb{R}^n$, the variation of the action sum is decomposed into two parts of $\mathsf{SO}(3)$ and $\mathbb{R}^n$. Therefore, the continuous Euler–Lagrange equations on $\mathsf{SO}(3)\times\mathbb{R}^n$ are given by (30) and (35), after replacing $\nabla\mathsf{f}(x)$ of (30) with $\nabla_x\mathsf{f}(R,x)$, and replacing $\nabla_{\mathsf{L}}f(R)$ of (35) with $\mathsf{T}_I^*\mathsf{L}_R(\mathbf{D}_R\mathsf{f}(R,x))$.

Similarly, the corresponding Lie group variational integrators are also given by (32), (33), (39), and (40), in addition to the energy equations (28) and (29) with

$$T_{d_k}(F_k,\Delta x_k)=\frac{1}{2}\|\Delta x_k\|^2+\mathrm{tr}[(I_{3\times3}-F_k)J_d].$$



(a) convergence with respect to t

(b) convergence with respect to k

Fig. 1. Convergence rate of LGVI in Proposition 5 for varying p

V. Numerical Examples

A. Optimization on $\mathsf{SO}(3)$

Consider the objective function given by

$$\mathsf{f}(R)=\frac{1}{2}\|A-R\|_{\mathcal{F}}^2=\frac{1}{2}(\|A\|_{\mathcal{F}}^2+3)-\mathrm{tr}[A^TR], \qquad (42)$$

where $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm, and $A\in\mathbb{R}^{3\times3}$. Optimization of the above function appears in the least-squares estimation of attitude, referred to as Wahba's problem [23]. Let the singular value decomposition of $A=USV^T$ for a diagonal $S\in\mathbb{R}^{3\times3}$ and $U,V\in\mathsf{O}(3)$. The optimal attitude is explicitly given by $R^*=U\mathrm{diag}[1,1,\det(UV)]V^T$. The left-trivialized gradient is $\nabla_{\mathsf{L}}\mathsf{f}(R)=(A^TR-R^TA)^\vee$.

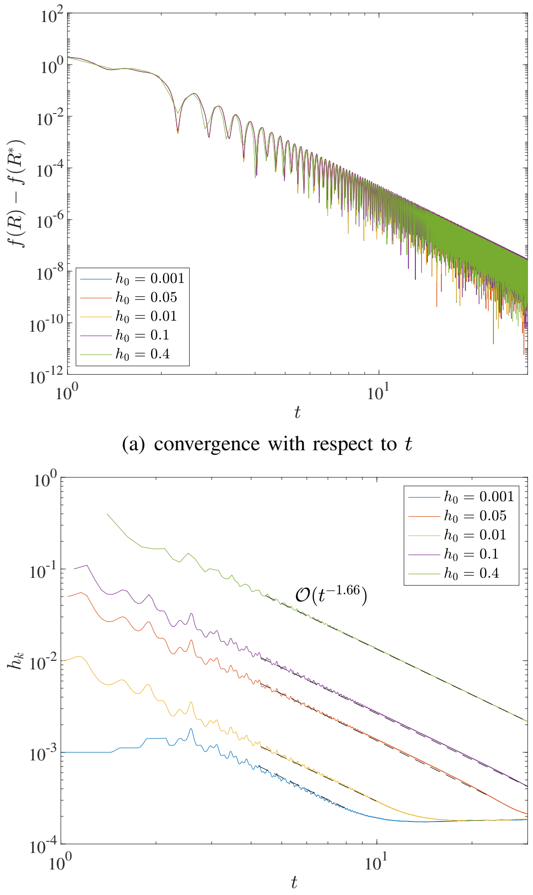
1) Order of Convergence: First, we check if the theoretical order of convergence guaranteed by Proposition 2 is achieved by the discrete Euler–Lagrange equations presented in Proposition 3. The elements of the matrix A in (42) are randomly chosen from the uniform distribution on $[0,1]$. The initial guess of R_0 is chosen such that the initial error is 0.9π in terms of the Euler-axis rotation. Lie group variational integrators (LGVI) in Proposition 5 are simulated with fixed $J=I_{3\times3}$, $C=1$, and $h_0=0.1$ for varying $p\in\{2,4,6,8\}$. Since $J=I_{3\times3}$, (39) is replaced by (41). The remaining implicit equation (28) is solved for h_k via the Matlab equation solver, `lsqnonlin` with the tolerance of 10^{-4}. The initial guess for h_k is provided by h_{k-1}.

The resulting convergence rate represented by $\mathsf{f}-\mathsf{f}^*$ over t_k is illustrated in Figure 1.(a), where the empirical convergence rate computed by manual fitting are also marked. It is shown that LGVI empirically achieved the order of convergence greater than the theoretical guarantee of $\mathcal{O}(t^{-p})$. It has been reported that naïve discretizations of Bregman Lagrangian systems are not able to match the theoretical convergence rate, or it might cause numerical instability [4], [5]. These results suggest that LGVIs do not suffer from these discretization issues, and their performance are consistent with the continuous-time analysis.

Next, given that the step size h_k is adjusted adaptively according to (28) and (29), it is likely that numerical simulation with higher p requires a smaller step size. In fact, the average step sizes are given by 6.15×10^{-2}, 6.50×10^{-3}, 4.89×10^{-4} and 1.21×10^{-5}, respectively for $p\in\{2,4,6,8\}$. To examine the effects of the step size variations, the convergence with respect to the discrete time step is illustrated in Figure 1.(b). It turns out that all of four cases of p exhibit the similar order of long-term convergence, approximately $\mathcal{O}(k^{-2.3})$. This is not surprising, as Nesterov [2] showed that for every smooth first-order method, there exists a convex, L-smooth objective function, such that the rate of convergence is bounded from below by $\mathcal{O}(k^{-2})$, but it does not preclude the possibility of faster rates of convergence for strongly convex functions.

However, the case of higher p benefits from faster initial convergence, and as a result, the terminal error for $p=4$ is more than 400 times smaller than that of $p=2$.

2) Effects of Initial Step Size: As discussed at the end of Section III, the extended LGVI requires choosing the initial step size h_0. Here, we study the effects of h_0 in the convergence. More specifically, the order is fixed to $p=4$, and the initial step size is varied as $h_0\in\{0.001, 0.05, 0.01, 0.1, 0.4\}$. The corresponding results are illustrated at Figure 2. Interestingly, in Figure 2.(a), the convergence with respect to t is not much affected by the initial step size h_0. Next, Figure 2.(b) presents the time-evolution of the step size, and it is shown that the step size computed by (28) decreases at the approximate order of $\mathcal{O}(t^{-1.6})$ for all cases. This might have been caused by the fact that the forcing term in (35) increases over time. Another notable feature is that after a certain period, the step sizes tend to converge. More specifically, the step size initialized by $h_0=0.001$ converges to 1.8×10^{-4} when $t>10$, which is joined by the case of $h_0=0.005$ later. It is expected that the next case for $h_0=0.01$ would follow the similar trend if the simulation time is increased. This implies a certain stability property of the extended LGVI in the step size. Furthermore, observe that for the wide range of variations of step sizes presented in Figure 2.(b), the convergence in Figure 2.(a) is fairly consistent, which suggests that the LGVI is robust to the choice of the step size.



(a) convergence with respect to t

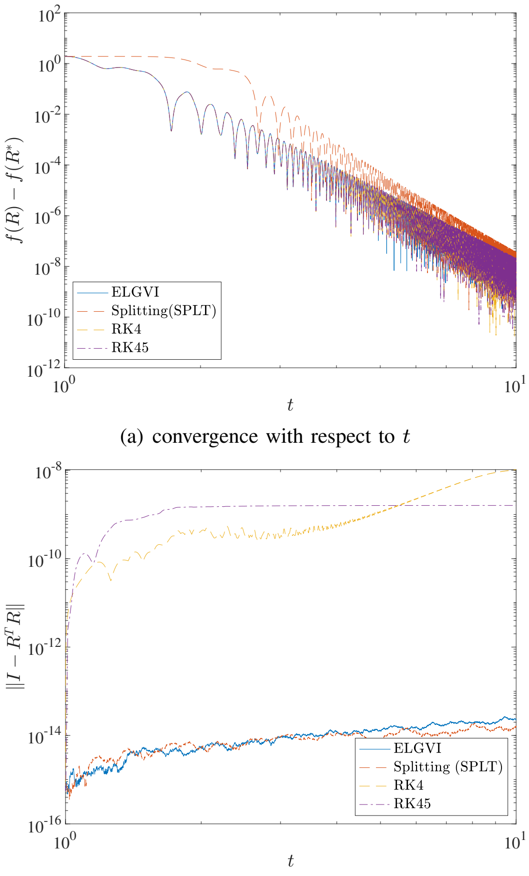
(b) evolution of step size h_k

Fig. 2. Convergence rate of LGVI in Proposition 5 for varying h_0.

3) Comparison with Other Discretizations of Bregman Euler–Lagrange Equation: Next, we compare LGVI with other discretization schemes applied to (35) and (36). Three methods are considered, namely the splitting approach introduced in [7] applied to the proposed continuous dynamics (abbreviated as SPLT), a 4-th order fixed-step Runge–Kutta method (RK4), and a variable stepsize Runge–Kutta method (RK45) implemented by the Matlab `ode45` function with the tolerance of 10^{-8}. More precisely, the evolution of SPLT over step size h is written as $\phi_{h/2}\circ\psi_h\circ\phi_{h/2}$, where ϕ_t is the exact flow map of (36) with fixed Ω, and ψ_t is the exact t-time flow map of (35) with fixed R and $J=I_{3\times3}$.

The goal of this comparison is not to claim that a certain method is superior to the other methods. Rather, it is to identify the numerical properties of LGVI compared with others. Having stated that, LGVI is implicit, and (28) is solved by a general purpose nonlinear solver, instead of a numerical solver tailored for (28). As a consequence, LGVI is substantially slower than the three explicit methods, to the extent that the comparison is not meaningful.

Instead, for a more interesting comparison, we exploit the property of LGVI providing consistent results for a wide range of step sizes, and we only utilize (39) and (40) with a fixed prescribed step size. The resulting scheme, denoted by ELGVI, is explicit as shown in (41). Overall ELGVI is quite comparable with SPLT, but it benefits from a bit faster initial convergence, especially when p is larger and h is smaller. One particular case for $p=6$ and $h=0.001$ is illustrated in Figure 3.(a). With regard to RK4 and RK45,



(a) convergence with respect to t

(b) orthogonality error of R_k

Fig. 3. Comparison with other discretization schemes for Bregman Euler–Lagrange equation

their convergence is almost identical to ELGVI, but as presented in Figure 3.(b), those methods do not preserve the orthogonality of the rotation matrix, which is problematic. Whereas, both of LGVI and SPLT conserve the structure of rotation matrices. Next, the computation time with Intel Core i7 3.2GHz, averaged for 10 executions, are 0.0727, 0.0258, 0.3847, and 1.1476 seconds for ELGVI, SPLT, RK4, and RK45, respectively. It is expected that RK4 requires more computation time as the gradient should be evaluated four times per a step, and it seems that the time-adaptive RK45 algorithm requires more frequent evaluations of the gradient.

4) Comparison with Other Optimization Schemes on Lie Groups: Finally, we compare ELGVI with other optimization schemes on Lie groups. In particular, we consider variationally accelerated Lie-group methods based on the NAG variational principle and operating splitting [7], referred to as Lie-NAG-SC and Lie-NAG-C, which are conformally symplectic and group-structure preserving. Note that Lie-NAG-C corresponds to SPLT with $p = 2$.

Four cases are considered, as labeled in Figure 4 for varying p and h. Compared with Lie-NAG-C, ELGVI exhibits faster convergence at a higher order. This does not contradict Nesterov's oracle lower bound: the continuous Bregman dynamics with $p > 2$ should be discretized by smaller steps as t increases, and therefore, the asymptotic order of convergence is still $\mathcal{O}(1/k^2)$ as illustrated above. However, since ELGVI uses a fixed stepsize, the initial error can decay faster than inverse quadratic, and depending on the level of accuracy required, we can take advantage of it by employing early stopping. On the other hand, Lie-NAG-SC demonstrates exponential convergence asymptotically when applied to strongly convex functions. Overall, if moderate stopping criteria are employed, ELGVI may be preferred, as they exhibit the fastest initial decay of the cost function.

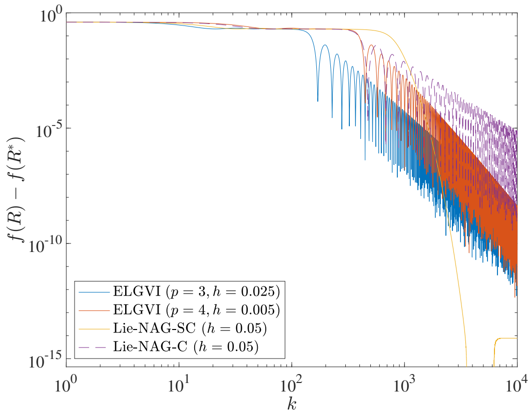


Fig. 4. Comparison with other accelerated optimization schemes on Lie groups

B. *Optimization on* $\mathsf{SO}(3) \times \mathbb{R}^3$

Next, we present an optimization problem on $\mathsf{SO}(3) \times \mathbb{R}^3$ to estimate the position and the attitude of a camera using the KITTI vision benchmark dataset [24]. This is to verify the performance of ELGVI for a non-convex function in a higher-dimensional Lie group, with more relevance to engineering practice. More specifically, we consider $N = 516$ distinct features on a single image frame, where their 2D pixel coordinates in the image plane, and the actual 3D location in the world coordinates are given by $p^i \in \mathbb{R}^3$ and $P^i \in \mathbb{R}^4$, respectively as homogeneous coordinates. Assuming that the camera calibration matrix $K \in \mathbb{R}^{3\times 3}$ is also known, we wish to estimate the pose $(R, x) \in \mathsf{SO}(3) \times \mathbb{R}^3$ of the camera.

This is formulated as an optimization problem to minimize the reprojection error, which is the discrepancy between the actual pixel location of the features and the features projected to the image plane by the current estimate of (R, x) [25]. For example, let $\tilde{p}^i \in \mathbb{R}^3$ be the homogeneous coordinates for the feature corresponding to P^i projected to the image plane by (R, x). From the perspective camera model,

$$\lambda \tilde{p}^i = K[R, x]P^i,$$

for $\lambda > 0$. The corresponding reprojected pixel is determined by the dehomogenization of $\tilde{p}^i$, namely $H^{-1}(\tilde{p}^i) \in \mathbb{R}^2$ corresponding to the first two elements of $\tilde{p}^i$ divided by the last element. The objective function is the sum of the

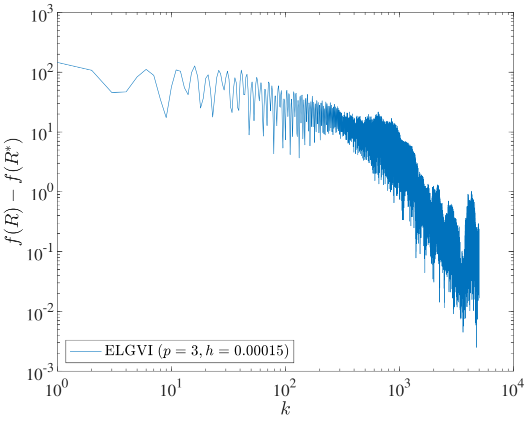
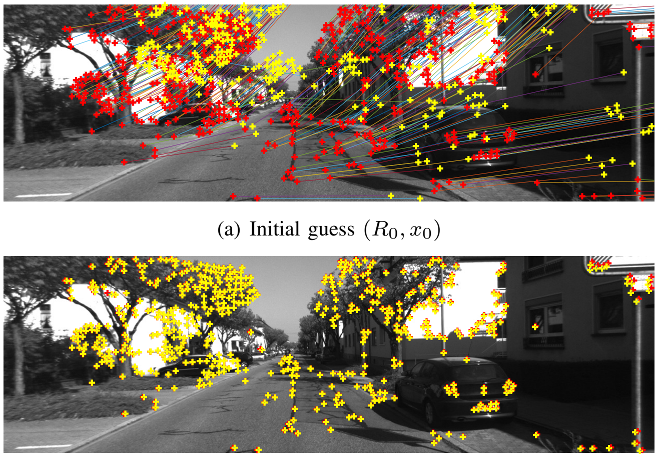


Fig. 5. Optimization on $\mathsf{SO}(3) \times \mathbb{R}^3$: convergence with respect to k



(a) Initial guess (R_0, x_0)

(b) Optimized (R^*, x^*)

Fig. 6. Reprojection error: the red + markers denote the key points detected, and the yellow + markers represent the key points projected by the estimated pose. The paired features are connected by solid lines.

reprojection error given by

$$\mathsf{f}(R, x) = \sum_{i=1}^{N} \|H^{-1}(p^i) - H^{-1}(\tilde{p}^i)\|^2. \tag{43}$$

Figure 5 presents the optimization results by ELGVI, which are comparable to the benchmark examples presented for $\mathsf{SO}(3)$. However, the terminal phase is relatively noisy, partially because the gradients of (43) are evaluated numerically with a finite-difference rule. Figure 6 illustrates the reprojected features before and after the optimization.

VI. Conclusions

In this paper, we proposed a Lie group variational integrator for the Bregman Lagrangian dynamics on Lie groups, to construct an accelerated optimization scheme. The variable stepsize prescribed by the extended variational principle exhibits an interesting convergence property, and the variational discretization is robust to the initial stepsize. It would be interesting to explore the role of variable time-stepping in geometric discretizations of the Bregman dynamics especially compared with Hamiltonian variational integrators.

References

[1] Y. Nesterov, “A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$,” *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[2] ——, *Introductory lectures on convex optimization: A basic course*, 2004.

[3] W. Su, S. Boyd, and E. J. Candes, “A differential equation for modeling nesterov’s accelerated gradient method: Theory and insights,” *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 5312–5354, 2016.

[4] A. Wibisono, A. C. Wilson, and M. I. Jordan, “A variational perspective on accelerated methods in optimization,” *proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. E7351–E7358, 2016.

[5] M. Betancourt, M. I. Jordan, and A. C. Wilson, “On symplectic optimization,” *arXiv preprint arXiv:1802.03653*, 2018.

[6] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-preserving algorithms for ordinary differential equations*, 2nd ed. Berlin: Springer-Verlag, 2006.

[7] M. Tao and T. Ohsawa, “Variational optimization on Lie groups, with examples of leading (generalized) eigenvalue problems,” in *International Conference on Artificial Intelligence and Statistics*, 2020, pp. 4269–4280.

[8] M. Tao, H. Owhadi, and J. E. Marsden, “Nonintrusive and structure preserving multiscale integration of stiff odes, sdes, and hamiltonian systems with hidden slow dynamics via flow averaging,” *Multiscale Modeling & Simulation*, vol. 8, no. 4, pp. 1269–1324, 2010.

[9] J. Marsden and M. West, “Discrete mechanics and variational integrators,” in *Acta Numerica*. Cambridge University Press, 2001, vol. 10, pp. 317–514.

[10] M. Leok and J. Zhang, “Discrete Hamiltonian variational integrators,” *IMA J. Numer. Anal.*, vol. 31, no. 4, pp. 1497–1532, 2011.

[11] T. Lee, M. Leok, and N. McClamroch, “Lie group variational integrators for the full body problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 196, pp. 2907–2924, May 2007.

[12] V. Duruisseaux, J. Schmitt, and M. Leok, “Adaptive Hamiltonian variational integrators and symplectic accelerated optimization,” *arXiv preprint arXiv:1709.01975*, 2021.

[13] J. Hu, X. Liu, Z.-W. Wen, and Y.-X. Yuan, “A brief introduction to manifold optimization,” *Journal of the Operations Research Society of China*, vol. 8, no. 2, pp. 199–248, 2020.

[14] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[15] J. Marsden and T. Ratiu, *Introduction to Mechanics and Symmetry*, 2nd ed., ser. Texts in Applied Mathematics. Springer-Verlag, 1999, vol. 17.

[16] T. Lee, M. Leok, and N. McClamroch, *Global Formulation of Lagrangian and Hamiltonian Dynamics on Manifolds*. Springer, 2018.

[17] C. Kane, J. Marsden, and M. Ortiz, “Symplectic-energy-momentum preserving variational integrators,” *Journal of Mathematical Physics*, vol. 40, no. 7, pp. 3353–3371, 1999.

[18] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.

[19] ——, “Accelerating the cubic regularization of Newton’s method on convex problems,” *Mathematical Programming*, vol. 112, no. 1, pp. 159–181, 2008.

[20] V. Duruisseaux and M. Leok, “A variational formulation of accelerated optimization on Riemannian manifolds,” *arXiv preprint arXiv:2101.06552*, 2021.

[21] F. Alimisis, A. Orvieto, G. Bécigneul, and A. Lucchi, “A continuous-time perspective for modeling acceleration in Riemannian optimization,” in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 1297–1307.

[22] T. Lee, “Computational geometric mechanics and control of rigid bodies,” Ph.D. dissertation, University of Michigan, 2008.

[23] G. Wahba, “A least squares estimate of satellite attitude, Problem 65-1,” *SIAM Review*, vol. 7, no. 5, p. 409, 1965.

[24] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *International Journal of Robotics Research (IJRR)*, 2013.

[25] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-D vision: from images to geometric models*. Springer Science & Business Media, 2012, vol. 26.