UC San Diego

Title

High-order symplectic Lie group methods on $ SO(n) $ using the polar decomposition

Permalink

https://escholarship.org/uc/item/7zv987kh

Authors

Shen, Xuefeng
Tran, Khoa
Leok, Melvin

HIGH-ORDER SYMPLECTIC LIE GROUP METHODS ON $SO(n)$ USING THE POLAR DECOMPOSITION

Xuefeng Shen, Khoa Tran and Melvin Leok*

Department of Mathematics, University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0112, USA

(Communicated by Robert McLachlan)

Abstract. A variational integrator of arbitrarily high-order on the special orthogonal group $SO(n)$ is constructed using the polar decomposition and the constrained Galerkin method. It has the advantage of avoiding the second order derivative of the exponential map that arises in traditional Lie group variational methods. In addition, a reduced Lie–Poisson integrator is constructed and the resulting algorithms can naturally be implemented by fixed-point iteration. The proposed methods are validated by numerical simulations on $SO(3)$ which demonstrate that they are comparable to variational Runge–Kutta–Munthe-Kaas methods in terms of computational efficiency. However, the methods we have proposed preserve the Lie group structure much more accurately and and exhibit better near energy preservation.

1. Introduction.

1.1. Overview.
Given a configuration manifold Q, variational integrators provide a useful method of deriving symplectic integrators for Lagrangian mechanics on the tangent bundle TQ in terms of the Lagrangian L, or for Hamiltonian mechanics on the cotangent bundle T^*Q in terms of the Hamiltonian H. It involves discretizing Hamilton's principle or Hamilton's phase space principle rather than the Euler–Lagrange or Hamilton's equations. Discrete Lagrangian variational mechanics is described in terms of a discrete Lagrangian $L_d(q_0, q_1)$, which is an approximation of the exact discrete Lagrangian,

$$L_d^{\text{exact}}(q_0, q_1) = \underset{\substack{q \in C^2([0,h],Q) \\ q(0)=q_0,\ q(h)=q_1}}{\text{ext}} \int_0^h L(q(t), \dot{q}(t))\, dt. \tag{1}$$

The discrete Hamilton's principle states that the discrete action sum is stationary for variations that vanish at the endpoints. This yields the discrete Euler–Lagrange

equations,

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

where D_i denotes the partial derivative with respect to the i-th argument. This defines an update map on $Q \times Q$, where $(q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$. The update map can equivalently be described in terms of the discrete Legendre transforms,

$$p_k = -D_1 L_d(q_k, q_{k+1}), \qquad p_{k+1} = D_2 L_d(q_k, q_{k+1}), \qquad (2)$$

which defines an update map $(q_k, p_k) \mapsto (q_{k+1}, p_{k+1})$ on T^*Q that automatically preserves the canonical symplectic structure on T^*Q.

The order of the variational integrator depends on how accurately $L_d(q_0, q_1)$ approximates $L_d^{\text{exact}}(q_0, q_1)$. To derive a high-order discrete Lagrangian, a typical approach is the Galerkin method [14]. This involves considering the definition of the exact discrete Lagrangian, replacing $C^2([0, h], Q)$ with a finite-dimensional function space, and approximating the integral with a numerical quadrature formula. When the configuration manifold Q is a linear space and the polynomials of degree less than or equal to s are chosen, the classical symplectic partitioned Runge–Kutta methods are recovered. Subsequently, Leok and Shingel [11] introduced the shooting-based discrete Lagrangian, which allows one to construct a symplectic integrator from an arbitrary one-step method.

When the configuration manifold Q is a Lie group G, the construction of the discrete Lagrangian is more complicated than the linear space case. Leok [10] proposed parametrizing curves on the Lie group using the exponential map, namely a curve $g(t)$ connecting g_0 and g_1 that is represented by

$$g(t) = g_0 \cdot \exp(\epsilon(t)),$$

where $\epsilon(t) \in \mathfrak{g}$ is a curve on the Lie algebra of G with fixed endpoints $\epsilon(0) = 0$ and $\epsilon(h) = \log(g_0^{-1} g_1)$. This allows one to replace variations in $g(t)$ by variations in $\epsilon(t)$ on the Lie algebra $\mathfrak{g}$, which is a linear space. This yields the following expression for the exact discrete Lagrangian,

$$L_d^{\text{exact}}(g_0, g_1) = \underset{\substack{\epsilon \in C^2([0,h], \mathfrak{g}) \\ \epsilon(0)=0, \ \epsilon(h)=\log(g_0^{-1} g_1)}}{\text{ext}} \int_0^h L(g_0 \cdot \exp(\epsilon(t)), g_0 \cdot \text{dexp}_{\epsilon(t)}(\dot{\epsilon}(t))) \, dt, \quad (3)$$

where $\text{dexp}_\epsilon(\dot{\epsilon}) = \exp(\epsilon) \cdot \frac{1 - e^{-\text{ad}_\epsilon}}{\text{ad}_\epsilon}(\dot{\epsilon})$ is the tangent lift of the exponential map. If $\epsilon(t)$ is restricted to a finite-dimensional function space and the integral is replaced with a quadrature rule, we obtain the Galerkin Lie group variational integrators. The error analysis and implementation details for such methods can be found in [8, 3]. The above construction can be naturally extended to any retraction [1] on G, which is a diffeomorphism from a neighborhood of $0 \in \mathfrak{g}$ to neighborhood of $e \in G$ that satisfies a rigidity condition. The main disadvantage of Galerkin Lie group variational integrators is the term dexp in (3). This implies that the resulting discrete Euler–Lagrange equations involve $\text{d}^2\exp$, which cannot be calculated exactly in general and requires a truncated series expansion.

1.2. **Contributions.** In this paper, we focus on the Lie group $SO(n)$ as our configuration space. By using the fact that every invertible square matrix can be uniquely decomposed into the product of an orthogonal matrix and a symmetric positive-definite matrix via the polar decomposition, we will circumvent the disadvantage discussed previously: Instead of parametrizing curves on $SO(n)$ by the exponential map or a retraction, $SO(n)$ is embedded naturally in the space

$GL_+(n) = \{A \in \mathbb{R}^{n \times n} \mid \det(A) > 0\}$, an open subset of $\mathbb{R}^{n \times n}$. Given fixed endpoints g_0 and g_1, we will construct interpolating polynomials in $GL_+(n)$ while ensuring that the internal points remain on $SO(n)$ by using the polar decomposition. Furthermore, we do not extend the Lagrangian L to $GL_+(n)$ but instead project the trajectory onto $SO(n)$ in the same way. The variational integrator in Lagrangian form is derived following the usual variational approach for the constrained Galerkin method and the Hamiltonian form is derived using the discrete Legendre transforms.

For a system with rotational symmetry, we obtain a simpler integrator using Lie–Poisson reduction on the Hamiltonian side. Namely, if L is $SO(n)$-invariant, the constructed discrete Lagrangian is also $SO(n)$-invariant and we can construct a reduced symplectic Lie–Poisson integrator. Lastly, we consider the dipole on a stick problem from [3], conduct numerical experiments using our method, and compare these to the variational Runge–Kutta–Munthe-Kaas methods (VRKMK) from the same reference.

2. **Background.**

2.1. **Notation.**
Recall that the Lie algebra of $SO(n)$ is the set $\mathrm{Skew}(n) = \{\Omega \in \mathbb{R}^{n \times n} \mid \Omega^T = -\Omega\}$, with the matrix commutator as the Lie bracket. Here, the inner products on $\mathbb{R}^{n \times n}$ and $\mathrm{Skew}(n)$ are introduced, and we identify these spaces with their dual spaces using the Riesz representations. For any $A, B \in \mathbb{R}^{n \times n}$, the inner product is given by

$$\mathrm{tr}\left(AB^T\right) = \sum_{i,j=1}^n a_{ij} b_{ij}.$$

For any $\Omega, \tilde{\Omega} \in \mathrm{Skew}(n)$, the inner product is defined by

$$\langle \Omega, \tilde{\Omega} \rangle = \sum_{i<j} \Omega_{ij} \tilde{\Omega}_{ij} = \frac{1}{2} \mathrm{tr}(\Omega \tilde{\Omega}^T).$$

In addition, consider the operator $\mathrm{Asym} \colon \mathbb{R}^{n \times n} \to \mathrm{Skew}(n)$, defined by $\mathrm{Asym}(A) = A - A^T$. The following properties can be easily verified:

(a) For any $A, B \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(AB^T) = \mathrm{tr}(A^T B)$.
(b) For any $A, B, P, Q \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(A(PBQ)^T) = \mathrm{tr}((P^T A Q^T)B^T)$.
(c) For any $\Omega \in \mathrm{Skew}(n), A \in \mathbb{R}^{n \times n}$, $\langle \Omega, \mathrm{Asym}(A) \rangle = \sum_{i<j} \Omega_{ij}(A_{ij} - A_{ji}) = \mathrm{tr}(\Omega A^T)$.

In particular, we note that (c) gives a relationship between the two inner products. Lastly, given the choice of inner products, Riesz representation allows us to identify $(\mathbb{R}^{n \times n})^*$ with $\mathbb{R}^{n \times n}$ and $\mathrm{Skew}(n)^*$ with $\mathrm{Skew}(n)$.

2.2. **Polar decomposition.**
We introduce the *polar decomposition* and the retraction on $SO(n)$ described in [4]. Given $A \in GL(n)$, it decomposes uniquely as

$$A \longrightarrow UP, \qquad U \in O(n), \ P \in \mathrm{Sym}^+(n),$$

where $\mathrm{Sym}^+(n)$ is the set of $n \times n$ symmetric positive-definite matrices. This is the polar decomposition of A, and we denote it as a coproduct mapping by $\mathrm{pol} \colon GL(n) \to O(n) \times \mathrm{Sym}^+(n)$. The map of interest is the projection $\mathbb{P} = \pi_1 \circ \mathrm{pol} \colon GL(n) \to O(n)$ defined by

$$\mathbb{P}(A) = U,$$

where π_1 is the projection onto the first coordinate. In particular, when $A \in GL_+(n)$, we have $U \in SO(n)$. A fast and efficient algorithm for calculating the projection of the polar decomposition is by Newton iteration,

$$U_{k+1} = \frac{1}{2}\left(U_k + U_k^{-T}\right), \ U_0 = A. \tag{4}$$

Now, this projection can be used to construct a retraction on $SO(n)$ from its Lie algebra $\mathrm{Skew}(n)$ relative to the identity element I,

$$\mathbb{P}(I + \Omega) = U,$$

where $\Omega \in \mathrm{Skew}(n)$. This provides a diffeomorphism between a neighborhood of $0 \in \mathrm{Skew}(n)$ and a neighborhood of $I \in SO(n)$. To calculate its inverse, suppose that $I + \Omega = UP$ and take the transpose on both sides to obtain $I - \Omega = PU^T$, which implies that $U^T(I + \Omega) = (I - \Omega)U$. Thus, we have that

$$U^T\Omega + \Omega U + U^T - U = 0. \tag{5}$$

This is a *Lyapunov equation*, and it is well-known that matrix equations of the form $AX + XB + C = 0$ have a unique solution if and only if for any $\lambda \in \sigma(A), \mu \in \sigma(B)$, $\lambda + \mu \neq 0$. For U in the neighborhood of identity, its eigenvalues lie in the open right-half plane, which ensures that a unique solution to (5) exists. In principle, this Lyapunov equation can be solved using classical algorithms [2, 7]. For convenience, we denote the solution to the Lyapunov equation as

$$X = \mathrm{Lyap}(A, B, C).$$

Next we introduce the tangent map and its adjoint for $\mathbb{P}$, which are essential for the derivation of the variational integrator.

2.2.1. *The tangent map.* Consider the polar decomposition $A(t) = U(t)P(t)$ and differentiate both sides to yield $\dot{A} = \dot{U}P + U\dot{P}$. By left-trivialization on $SO(n)$, we can write $\dot{U} = U\Omega$, where $\Omega \in \mathrm{Skew}(n)$. Rearranging gives $\dot{P} = U^T\dot{A} - \Omega P$, and since $\dot{P} \in \mathrm{Sym}^+(n)$, we get that $U^T\dot{A} - \Omega P = \dot{A}^T U + P\Omega$. Consequently, we may write it in the form of a Lyapunov equation,

$$P\Omega + \Omega P + \dot{A}^T U - U^T\dot{A} = 0. \tag{6}$$

We see that the tangent map of the polar decomposition $d\mathbb{P}_A : \mathbb{R}^{n \times n} \to \mathrm{Skew}(n)$ is given by $d\mathbb{P}_A(\dot{A}) = \Omega$, where we solve the Lyapunov equation (6) to obtain Ω.

2.2.2. *The adjoint of the tangent map.* The adjoint of $d\mathbb{P}_A$ can be defined as

$$\mathrm{tr}(d\mathbb{P}_A^*(\Omega)B^T) = \langle \Omega, d\mathbb{P}_A(B) \rangle, \qquad \forall \Omega \in \mathrm{Skew}(n), B \in \mathbb{R}^{n \times n}.$$

Recall that $d\mathbb{P}_A(B)$ involves solving the Lyapunov equation (6). We aim to compute $d\mathbb{P}_A^*$, so we define the following two maps,

$$\phi \colon \mathrm{Skew}(n) \longrightarrow \mathrm{Skew}(n), \qquad \Omega \longmapsto \Omega P + P\Omega,$$

$$\psi \colon \mathbb{R}^{n \times n} \longrightarrow \mathrm{Skew}(n), \qquad B \longmapsto U^T B - B^T U,$$

where $A = UP$ is fixed. Therefore, $d\mathbb{P}_A$ can be viewed as the composition of ψ and ϕ^{-1},

$$d\mathbb{P}_A = (\phi^{-1} \circ \psi) \colon \mathbb{R}^{n \times n} \longrightarrow \mathrm{Skew}(n),$$

and $d\mathbb{P}_A^*(\Omega) = (\phi^{-1} \circ \psi)^*(\Omega) = \psi^* \circ (\phi^*)^{-1}(\Omega)$. We shall derive the expressions for ϕ^* and ψ^* by considering the Riesz representations for our domains and codomains. For the adjoint of ϕ, let $\Omega, \tilde{\Omega} \in \mathrm{Skew}(n)$, then

$$
\begin{aligned}
\langle \phi^*(\Omega), \tilde{\Omega} \rangle &= \langle \Omega, \phi(\tilde{\Omega}) \rangle = \langle \Omega, \tilde{\Omega}P + P\tilde{\Omega} \rangle \\
&= \langle \Omega, \mathrm{Asym}(\tilde{\Omega}P) \rangle = \mathrm{tr}(\Omega(\tilde{\Omega}P)^T) \\
&= \mathrm{tr}((\Omega P)\tilde{\Omega}^T) \\
&= \langle \Omega P + P\Omega, \tilde{\Omega} \rangle.
\end{aligned}
$$

Thus, $\phi^* = \phi$, and ϕ is Hermitian. For ψ^*, let $\Omega \in \mathrm{Skew}(n)$ and $B \in \mathbb{R}^{n \times n}$, then

$$
\begin{aligned}
\mathrm{tr}(\psi^*(\Omega)B^T) &= \langle \Omega, \psi(B) \rangle = \langle \Omega, U^T B - B^T U \rangle \\
&= \langle \Omega, \mathrm{Asym}(U^T B) \rangle = \mathrm{tr}(\Omega(U^T B)^T) \\
&= \mathrm{tr}((U\Omega)B^T).
\end{aligned}
$$

Therefore, $\psi^*(\Omega) = U\Omega$, and we obtain

$$
d\mathbb{P}_A^*(\Omega) = (\psi^* \circ (\phi^*)^{-1})(\Omega) = (\psi^* \circ \phi^{-1})(\Omega) = U\,\mathrm{Lyap}(P, \Omega^T),
$$

where $\mathrm{Lyap}(P, \Omega^T) = \mathrm{Lyap}(P, P, \Omega^T)$. Finally, we state a lemma that will be used later:

Lemma 2.1. $\mathbb{P}(I + S) = I$ *if and only if* $S \in \mathrm{Sym}(n)$ *and* $\lambda > -1$ *for all* $\lambda \in \sigma(S)$.

3. **Lagrangian variational integrators on the rotation group** $SO(n)$**.** Let the configuration manifold be the rotation group $G = SO(n)$, and $L \colon G \times \mathfrak{g} \to \mathbb{R}$ be a left-trivialized Lagrangian. We shall construct a discrete Lagrangian following the approach of constrained Galerkin methods on $GL_+(n)$ (see Appendix A). Denote the internal points by $\{U_i\}_{i=1}^s \subset G$ and the left-trivialized internal tangent vectors by $\{\Omega_i\}_{i=1}^s \subset \mathfrak{g}$. Fixing the endpoints g_0 and g_1, we have

$$
L_d(g_0, g_1) = h \sum_{i=1}^s b_i L(U_i, \Omega_i),
$$

subject to the constraint

$$
g_1 = \mathbb{P}\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i \right), \tag{7}
$$

where the internal points U_i are represented by

$$
U_i = \mathbb{P}\left(g_0 + h \sum_{j=1}^s a_{ij} U_j \Omega_j \right). \tag{8}
$$

The expressions inside the parentheses in (7) and (8) correspond to a Runge–Kutta method in the embedding space. But, since these points may not lie on the Lie group G, they are projected to G using the polar decomposition.

Observe that (7) is equivalent to the condition that $\mathbb{P}\left(g_1^T(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i) \right) = I$. Suppose that h is small, and g_0 and g_1 are close enough to each other, then $g_1^T\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i \right)$ is in the neighborhood of I. By Lemma 2.1, (7) holds if and only if $g_1^T\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i \right) \in \mathrm{Sym}(n)$, and so it is equivalent to

$$
\mathrm{Asym}\left(g_1^T\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i \right) \right) = 0.
$$

Now, we can construct a discrete Lagrangian with the constraint using a Lagrange multiplier $\Lambda \in \text{Skew}(n)$,

$$\tilde{F}(g_0, g_1, \{\Omega_i\}_{i=1}^s, \Lambda) = h \sum_{i=1}^s b_i L(U_i, \Omega_i) + \left\langle \Lambda, \text{Asym}\left(g_1^T\left(g_0 + h\sum_{i=1}^s b_i U_i \Omega_i\right)\right)\right\rangle.$$

The corresponding variational integrator is given by

$$\begin{cases} 0 = \dfrac{\partial \tilde{F}}{\partial \Omega_i}, \qquad i = 1, 2 \ldots s, & (9a) \\[2mm] 0 = \dfrac{\partial \tilde{F}}{\partial \Lambda}, & (9b) \\[2mm] U_i = \mathbb{P}\left(g_0 + h\sum_{j=1}^s a_{ij} U_j \Omega_j\right), & (9c) \\[2mm] -p_0 = \dfrac{\partial \tilde{F}}{\partial g_0}, & (9d) \\[2mm] p_1 = \dfrac{\partial \tilde{F}}{\partial g_1}. & (9e) \end{cases}$$

We shall compute the above equations more explicitly. It is easy to see that (9b) is equivalent to the constraint (7). Let us turn to (9a), where the main difficulty is the implicit dependence of $\{U_i\}_{i=1}^s$ on $\{\Omega_i\}_{i=1}^s$ that involves solving a nonlinear system (8). Suppose that $k \in \{1, 2, \ldots, s\}$ is fixed, and we vary Ω_k such that $\Omega_k \to \Omega_k(t)$ with $\Omega_k(0) = \Omega_k$ and $\dot{\Omega}_k(0) = \delta\Omega_k$, while $\{\Omega_i\}_{i \neq k}$ remain fixed. Then,

$$U_i(t) = \mathbb{P}\left(g_0 + h\sum_{j \neq k} a_{ij} U_j(t)\Omega_j + ha_{ik} U_k(t)\Omega_k(t)\right).$$

Differentiating both sides and letting $\dot{U}_i = U_i X_{ik}$, we have that

$$X_{ik} = d\mathbb{P}_{A_i}\left(h\sum_{j=1}^s a_{ij} U_j X_{jk}\Omega_j + ha_{ik} U_k \delta\Omega_k\right), \qquad (10)$$

where $A_i = g_0 + h\sum_{j=1}^s a_{ij} U_j \Omega_j$. Then, (10) can be rewritten as

$$X_{ik} - d\mathbb{P}_{A_i}\left(h\sum_{j=1}^s a_{ij} U_j X_{jk}\Omega_j\right) = hd\mathbb{P}_{A_i}(a_{ik} U_k \delta\Omega_k). \qquad (11)$$

In order to represent $\{X_{ik}\}_{i=1}^s$ in terms of $\delta\Omega_k$, we define three maps,

$$\tilde{\psi}_k : \text{Skew}(n) \to \text{Skew}(n)^s, \qquad \delta\Omega_k \mapsto \{d\mathbb{P}_{A_i}(a_{ik} U_k \delta\Omega_k)\}_{i=1}^s,$$

$$\tilde{\phi} : \text{Skew}(n)^s \to \text{Skew}(n)^s, \ \{X_{ik}\}_{i=1}^s \mapsto \left\{X_{ik} - d\mathbb{P}_{A_i}\left(h\sum_{j=1}^s a_{ij} U_j X_{jk}\Omega_j\right)\right\}_{i=1}^s,$$

$$\tilde{\pi}_i : \text{Skew}(n)^s \to \text{Skew}(n), \quad \{\Omega_i\}_{i=1}^s \mapsto \Omega_i.$$

Then,

$$X_{ik} = (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ h\tilde{\psi}_k)(\delta\Omega_k) = h(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)(\delta\Omega_k). \qquad (12)$$

Now, we compute $\frac{\partial \tilde{F}}{\partial \Omega_k}$ by evaluating $\frac{d}{dt}\big|_{t=0} \tilde{F}(g_0, g_1, \ldots, \Omega_k(t), \ldots, \Lambda)$ and expressing $\frac{\partial L}{\partial U} : G \times \mathfrak{g} \to \mathfrak{g}^*$ as a left-trivialized cotangent vector. Since this is a straightforward calculation of the variation, we present the result here for equation (9a)

(see section C.1 for details),

$$0 = h \sum_{i=1}^{s} b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* \left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T) \right)$$
$$+ b_k \left(\frac{\partial L}{\partial \Omega}(U_k, \Omega_k) + \mathrm{Asym}(U_k^T g_1 \Lambda) \right), \quad (13)$$

for any $k = 1, 2, \ldots, s$.

Recall that $\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k \colon \mathrm{Skew}(n) \to \mathrm{Skew}(n)$ and its dual is given by

$$(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* = \tilde{\psi}_k^* \circ (\tilde{\phi}^*)^{-1} \circ \tilde{\pi}_i^*.$$

Therefore, let us derive the explicit expressions for the adjoints of our three proposed maps, so we may write $(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^*$ explicitly. The adjoint of $\tilde{\pi}_i$ is easy to compute, and for any $S \in \mathrm{Skew}(n)$,

$$\tilde{\pi}_i^*(S) = (0, \ldots, S, \ldots 0), \quad (14)$$

where S is in the i-th position. For the adjoint of $\tilde{\phi}$, we consider the identity

$$\langle \tilde{\phi}^*(S_1, \ldots, S_s), (\tilde{S}_1, \ldots, \tilde{S}_s) \rangle = \langle (S_1, \ldots, S_s), \tilde{\phi}(\tilde{S}_1, \ldots, \tilde{S}_s) \rangle,$$

for any $(S_1, \ldots, S_s), (\tilde{S}_1, \ldots, \tilde{S}_s) \in \mathrm{Skew}(n)^s$. Using the properties of the inner products again, we obtain the explicit expression (see section C.2 for details),

$$\tilde{\phi}^*(S_1, \ldots, S_s) = \left\{ S_j - \mathrm{Asym}\left(hU_j^T \sum_{i=1}^{s} a_{ij} d\mathbb{P}_{A_i}^*(S_i)\Omega_j^T \right) \right\}_{j=1}^{s}. \quad (15)$$

Similarly, we consider the same identity and techniques to obtain the explicit expression for $\tilde{\psi}_k^*$ (see section C.3 for details),

$$\tilde{\psi}_k^*(S_1, \ldots, S_s) = \mathrm{Asym}\left(U_k^T \sum_{i=1}^{s} a_{ik} d\mathbb{P}_{A_i}^*(S_i) \right). \quad (16)$$

Combining (14), (15), and (16), $(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^*(S)$ for $S \in \mathrm{Skew}(n)$ can be computed as follows,

$$\begin{cases} S_j = S \cdot \delta_{ij} - \mathrm{Asym}\left(hU_j^T \sum_{l=1}^{s} a_{lj} d\mathbb{P}_{A_l}^*(S_l)\Omega_j^T \right), & j = 1, 2 \ldots s, \quad (17\mathrm{a}) \\ (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^*(S) = \mathrm{Asym}\left(U_k^T \sum_{l=1}^{s} a_{lk} d\mathbb{P}_{A_l}^*(S_l) \right). & (17\mathrm{b}) \end{cases}$$

We can first calculate $\{S_l\}_{l=1}^{s}$ from (17a) by using fixed-point iteration, and then substitute the result into (17b) to obtain $(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^*(S)$. This combined with (13) gives an explicit formula for (9a).

We now derive an explicit formula for (9d). Notice that U_i depends on g_0 by the nonlinear system in (8), so we can use the method of variations again. If we vary $g_0 \to g_0(t)$ such that $g_0(0) = g_0$ and $\dot{g}_0(0) = g_0 \delta g_0$, we obtain

$$U_i(t) = \mathbb{P}\left(g_0(t) + h \sum_{j=1}^{s} a_{ij} U_j(t)\Omega_j \right).$$

Differentiating on both sides and letting $\dot{U}_i = U_i Y_i$, where $Y_i \in \mathrm{Skew}(n)$ is a left-trivialized tangent vector, we obtain

$$Y_i = d\mathbb{P}_{A_i}\left(g_0 \delta g_0 + h \sum_{j=1}^{s} a_{ij} U_j Y_j \Omega_j \right),$$

which can be rewritten as

$$Y_i - d\mathbb{P}_{A_i}\left(h \sum_{j=1}^{s} a_{ij} U_j Y_j \Omega_j \right) = d\mathbb{P}_{A_i}(g_0 \delta g_0).$$

Similar to the approach used for X_{ik}, we introduce a new map,

$$\tilde{\varphi} \colon \mathrm{Skew}(n) \longrightarrow \mathrm{Skew}(n)^s, \quad \delta g_0 \longmapsto \{d\mathbb{P}_{A_i}(g_0 \delta g_0)\}_{i=1}^s \,,$$

then $Y_i = (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})(\delta g_0)$. The explicit expression for $\tilde{\varphi}^*$ can be written as (see section C.4 for details),

$$\tilde{\varphi}^*(S_1, \ldots, S_s) = \mathrm{Asym}\left(g_0^T \sum_{i=1}^s d\mathbb{P}^*_{A_i}(S_i)\right). \tag{18}$$

As such, $(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^*(S)$ can be computed as follows,

$$\begin{cases} S_j - \mathrm{Asym}\left(h U_j^T \sum_{l=1}^s a_{lj} d\mathbb{P}^*_{A_l}(S_l) \Omega_j^T\right) = S \cdot \delta_{ij}, \qquad j = 1, 2, \ldots s, \tag{19a} \\ (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^*(S) = \mathrm{Asym}\left(g_0^T \sum_{l=1}^s d\mathbb{P}^*_{A_l}(S_l)\right). \tag{19b} \end{cases}$$

Now, we compute $\frac{\partial \tilde{F}}{\partial g_0}$, which gives us (9d) explicitly (see section C.5 for details),

$$-p_0 = h \sum_{i=1}^s b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^* \left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T)\right)$$
$$+ \mathrm{Asym}(g_0^T g_1 \Lambda). \tag{20}$$

For equation (9e), it is easy to show that

$$p_1 = \mathrm{Asym}\left(g_1^T \left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i\right) \Lambda^T\right). \tag{21}$$

Combining (13), (7), (8), (20), and (21), we obtain a Lagrangian variational integrator on $SO(n)$,

$$\begin{cases} 0 = h \sum_{i=1}^s b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* \left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T)\right) \\ \qquad + b_k \left(\frac{\partial L}{\partial \Omega}(U_k, \Omega_k) + \mathrm{Asym}(U_k^T g_1 \Lambda)\right) \tag{22a} \\ g_1 = \mathbb{P}\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i\right), \tag{22b} \\ U_i = \mathbb{P}\left(g_0 + h \sum_{j=1}^s a_{ij} U_j \Omega_j\right), \tag{22c} \\ -p_0 = h \sum_{i=1}^s b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^* \left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T)\right) \\ \qquad + \mathrm{Asym}(g_0^T g_1 \Lambda), \tag{22d} \\ p_1 = \mathrm{Asym}\left(g_1^T \left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i\right) \Lambda^T\right). \tag{22e} \end{cases}$$

The integrator gives an update map $(g_0, p_0) \mapsto (g_1, p_1)$ on the cotangent bundle. In particular, one may solve for $(\{\Omega\}_{k=1}^s, g_1, \{U_i\}_{i=1}^s, \Lambda)$ simultaneously using equations (22a)–(22d). Unfortunately, while (22b)–(22d) can be written in fixed-point form for the variables g_1, $\{U_i\}_{i=1}^s$, and Λ, (22b) is implicit for $\{\Omega\}_{k=1}^s$. However, we can arrive at a fixed-point form for (22a) on the Hamiltonian side if L is hyper-regular.

4. **Hamiltonian variational integrators on the rotation group** $SO(n)$. It is often desirable to transform a numerical method from the Lagrangian side to the Hamiltonian side, which is possible if L is hyperregular. The same mechanical system can be represented by either a Lagrangian or a Hamiltonian, and they are related by the Legendre transform. In Euclidean space, this gives

$$(TQ, L) \underset{\mathbb{F}H}{\overset{\mathbb{F}L}{\rightleftarrows}} (T^*Q, H),$$

and we have the following relationships,

$$\frac{\partial L}{\partial \dot{q}}(q, \dot{q}) = p, \quad \frac{\partial H}{\partial p}(q, p) = \dot{q}, \quad \frac{\partial L}{\partial q}(q, \dot{q}) = -\frac{\partial H}{\partial q}(q, p).$$

Given a Lie group G, a left-trivialized Lagrangian $L \colon G \times \mathfrak{g} \to \mathbb{R}$, and its corresponding Hamiltonian $H \colon G \times \mathfrak{g}^* \to \mathbb{R}$, it is easy to verify that similar relationships hold for these trivializations,

$$\frac{\partial L}{\partial \epsilon}(g, \epsilon) = \mu, \quad \frac{\partial H}{\partial \mu}(g, \mu) = \epsilon, \quad \frac{\partial L}{\partial g}(g, \epsilon) = -\frac{\partial H}{\partial g}(g, \mu). \tag{23}$$

Using (23) and denoting the corresponding internal cotangent vectors by $\{\mu_k\}_{k=1}^s$, (22) can be transformed to the Hamiltonian form as

$$
\begin{cases}
\mu_k = -\operatorname{Asym}(U_k^T g_1 \Lambda) \\
\qquad + h \sum_{i=1}^s \frac{b_i}{b_k} (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* \\
\qquad\qquad \left(\frac{\partial H}{\partial U}(U_i, \mu_i) - \operatorname{Asym}(U_i^T g_1 \Lambda \Omega_i^T) \right), & \text{(24a)} \\[2mm]
g_1 = \mathbb{P}\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i \right), & \text{(24b)} \\[2mm]
U_i = \mathbb{P}\left(g_0 + h \sum_{j=1}^s a_{ij} U_j \Omega_j \right), & \text{(24c)} \\[2mm]
\operatorname{Asym}(g_0^T g_1 \Lambda) = -p_0 + h \sum_{i=1}^s b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^* \\
\qquad\qquad \left(\frac{\partial H}{\partial U}(U_i, \mu_i) - \operatorname{Asym}(U_i^T g_1 \Lambda \Omega_i^T) \right), & \text{(24d)} \\[2mm]
p_1 = \operatorname{Asym}\left(g_1^T \left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i \right) \Lambda^T \right), & \text{(24e)} \\[2mm]
\Omega_i = \frac{\partial H}{\partial \mu}(U_i, \mu_i). & \text{(24f)}
\end{cases}
$$

In the above algorithm, Ω_i is given explicitly by (24f) and only serves to reduce the redundancy in the computations because they are used numerous times in the other expressions. Similarly, $(\tilde{\pi}_i \circ \tilde{\phi}^{-1})^*$ shows up in both (24a) and (24d), so one can save computational effort by reusing the shared solution to (17a) and (19a). Then, the first four equations can be solved simultaneously by fixed-point iterations, meaning the variables $(\{\mu_k\}_{k=1}^s, g_1, \{U_i\}_{i=1}^s, \Lambda)$ are updated concurrently in each iteration. Observe that the fixed-point form for Λ in (24d) requires solving a Lyapunov equation. Finally, p_1 is solved explicitly in (24e) after solving for $(\{\mu_k\}_{k=1}^s, g_1, \{U_i\}_{i=1}^s, \Lambda)$. We shall call the integrators defined by (24) the *variational polar decomposition* method or VPD for short.

4.1. **Lie–Poisson integrator by reduction.** We consider a G-invariant Hamiltonian system given by H on the cotangent bundle T^*G. In this case, Hamilton's equations can be reduced to a Lie–Poisson system on $\mathfrak{g}^*$. If the Hamiltonian is hyperregular, then both the Lagrangian and the corresponding constrained Galerkin discrete Lagrangian $L_d(g_0, g_1)$ will be G-invariant. As such, (2) naturally reduces to yield a Lie–Poisson integrator (see Appendix B). We only consider the reduction on the Hamiltonian side due to the nature of the constrained Galerkin methods, which give an update map on the cotangent bundle.

The discrete Lagrangian we have constructed becomes

$$L_d(g_0, g_1) = \underset{\{\Omega_i\}_{i=1}^s}{\text{ext}} \sum_{i=1}^s b_i \boldsymbol{l}(\Omega_i),$$

where

$$U_i = \mathbb{P}\left(g_0 + h \sum_{j=1}^s a_{ij} U_j \Omega_j\right),$$

$$g_1 = \mathbb{P}\left(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i\right),$$

and $\boldsymbol{l} \colon \mathfrak{g} \to \mathbb{R}$ is the reduced Lagrangian. It is easy to verify that our system is G-invariant, meaning

$$L_d(g \cdot g_0, g \cdot g_1) = L_d(g_0, g_1) = \underset{\{\Omega_i\}_{i=1}^s}{\text{ext}} \sum_{i=1}^s b_i \boldsymbol{l}(\Omega_i),$$

where

$$g \cdot U_i = \mathbb{P}\left((g \cdot g_0) + h \sum_{j=1}^s a_{ij}(g \cdot U_j \Omega_j)\right),$$

$$g \cdot g_1 = \mathbb{P}\left((g \cdot g_0) + h \sum_{i=1}^s b_i(g \cdot U_i \Omega_i)\right).$$

Therefore, the variational integrator (24) can theoretically be reduced to a Lie–Poisson integrator. By letting $g_0^T g_1 = f_0$ and $U_i^T g_1 = \Theta_i$, (24) simplifies to

$$
\left\{
\begin{aligned}
\mu_k &= -\operatorname{Asym}(\Theta_k \Lambda) - h \sum_{i=1}^s \frac{b_i}{b_k} (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* (\operatorname{Asym}(\Theta_i \Lambda \Omega_i^T)), &\text{(25a)} \\
0 &= \operatorname{Asym}\left(f_0^T + h \sum_{i=1}^s b_i \Theta_i^T \Omega_i\right), &\text{(25b)} \\
\Theta_i^T &= \mathbb{P}\left(f_0^T + h \sum_{j=1}^s a_{ij} \Theta_j^T \Omega_j\right), &\text{(25c)} \\
\operatorname{Asym}(f_0 \Lambda) &= -p_0 - h \sum_{i=1}^s b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^* (\operatorname{Asym}(\Theta_i \Lambda \Omega_i^T)), &\text{(25d)} \\
p_1 &= \operatorname{Asym}\left(\left(f_0^T + h \sum_{i=1}^s b_i \Theta_i^T \Omega_i\right) \Lambda^T\right), &\text{(25e)} \\
\Omega_i &= \frac{\partial \boldsymbol{h}}{\partial \mu}(\mu_i), &\text{(25f)}
\end{aligned}
\right.
$$

where $\boldsymbol{h}$ is the reduced Hamiltonian. Multiplying by g_1^T on both sides of $g_1 = \mathbb{P}(g_0 + h \sum_{i=1}^s b_i U_i \Omega_i)$ yields

$$I = \mathbb{P}\left(f_0^T + h \sum_{i=1}^s b_i \Theta_i^T \Omega_i\right).$$

Suppose that h is small and g_0 and g_1 are close, then $\left(f_0^T + h \sum_{i=1}^s b_i \Theta_i^T \Omega_i\right)$ is in the neighborhood of I. By Lemma 2.1, this is equivalent to

$$\mathrm{Asym}\left(f_0^T + h \sum_{i=1}^s b_i \Theta_i^T \Omega_i\right) = 0,$$

which can be regarded as the fixed-point equation for f_0. The first four equations can be solved using fixed-point iteration for the variables $(\{\mu_k\}_{k=1}^s, f_0, \{\Theta_i\}_{i=1}^s, \Lambda)$ as in our previous discussions. Then, p_1 can be calculated explicitly.

In the above algorithm, we also need to figure out the reduced version of $(\tilde\pi_i \circ \tilde\phi^{-1} \circ \tilde\psi_k)^*$ and $(\tilde\pi_i \circ \tilde\phi^{-1} \circ \tilde\varphi)^*$. Note that (17a) and (17b) involve $U_j^T d\mathbb{P}_{A_i}^*$; in addition, g_0, g_1, U_i, and $A_i = g_0 + h \sum_{j=1}^s a_{ij} U_j \Omega_j = U_i P_i$ are reduced, so we need a reduced version of $U_j^T d\mathbb{P}_{A_i}^*$ as well. Multiplying A_i on the left by g_1^T, we obtain

$$g_1^T A_i = f_0^T + h \sum_{j=1}^s a_{ij} \Theta_j^T \Omega_j = (g_1^T U_i) P_i.$$

Then, $(g_1^T U_i) P_i$ is the polar decomposition of $f_0^T + h \sum_{j=1}^s a_{ij} \Theta_j^T \Omega_j$, and for $S \in \mathrm{Skew}(n)$,

$$U_j^T d\mathbb{P}_{A_i}^*(S) = U_j^T \cdot U_i \, \mathrm{Lyap}(P_i, S^T) = \Theta_j g_1^T U_i \, \mathrm{Lyap}(P_i, S^T),$$
$$= \Theta_j \mathbb{P}\left(f_0^T + h \sum_{j=1}^s a_{ij} \Theta_j^T \Omega_j\right) \mathrm{Lyap}(P_i, S^T).$$

This is the reduced version of $U_j^T d\mathbb{P}_{A_i}^*(S)$, and so $(\tilde\pi_i \circ \tilde\phi^{-1} \circ \tilde\psi_k)^*(S)$ can be computed as follows,

$$\begin{cases} S_j - \mathrm{Asym}\left(h \Theta_j \sum_{l=1}^s a_{lj} d\mathbb{P}_{A_l}^*(S_l) \Omega_j^T\right) = S \cdot \delta_{ij}, & j = 1, 2 \ldots s, \qquad \text{(26a)} \\ (\tilde\pi_i \circ \tilde\phi^{-1} \circ \tilde\psi_k)^*(S) = \mathrm{Asym}\left(\Theta_k \sum_{l=1}^s a_{lk} d\mathbb{P}_{A_l}^*(S_l)\right), & \text{(26b)} \end{cases}$$

where $\{A_i\}$ are redefined to be $A_i = f_0^T + h \sum_{j=1}^s a_{ij} \Theta_j^T \Omega_j$. For $(\tilde\pi_i \circ \tilde\phi^{-1} \circ \tilde\varphi)^*$, we need to compute $g_0^T d\mathbb{P}_{A_i}^*(S)$ for some $S \in \mathrm{Skew}(n)$, which is given by

$$g_0^T d\mathbb{P}_{A_i}^*(S) = g_0^T \cdot U_i \, \mathrm{Lyap}(P_i, S^T) = f_0 g_1^T U_i \, \mathrm{Lyap}(P_i, S^T)$$
$$= f_0 \mathbb{P}\left(f_0^T + h \sum_{j=1}^s a_{ij} \Theta_j^T \Omega_j\right) \mathrm{Lyap}(P_i, S^T).$$

Hence, we have

$$\begin{cases} S_j - \mathrm{Asym}\left(h \Theta_j \sum_{l=1}^s a_{lj} d\mathbb{P}_{A_l}^*(S_l) \Omega_j^T\right) = S \cdot \delta_{ij}, & j = 1, 2 \ldots s, \qquad \text{(27a)} \\ (\tilde\pi_i \circ \tilde\phi^{-1} \circ \tilde\varphi)^*(S) = \mathrm{Asym}\left(f_0 \sum_{l=1}^s a_{lk} d\mathbb{P}_{A_l}^*(S_l)\right). & \text{(27b)} \end{cases}$$

5. **Numerical experiments.** We test our methods and compare them to the variational Runge–Kutta–Munthe-Kaas (VRKMK) methods from Bogfjellmo and Marthinsen [3] for the dipole on a stick problem that they considered. In particular, our configuration space is $SO(3)$, and its Lie algebra is identified with $\mathbb{R}^3$. We shall only recall the mathematical expressions here, so for a thorough description of the system one should refer to the reference above. The right-trivialized and left-trivialized Hamiltonians, $H^R, H^L \colon SO(3) \times \mathbb{R}^3 \to \mathbb{R}$, can be written as

$$H^R(g, p) = \frac{1}{2} p^T g J^{-1} g^T p + U(g), \qquad \text{(28)}$$

$$H^L(g, \tilde p) = \frac{1}{2} \tilde p^T J^{-1} \tilde p + U(g), \qquad \text{(29)}$$

where $U(g) = me_3^T g e_3 + q\beta \left(\|gy_+^0 - z\|_2^{-1} - \|gy_-^0 - z\|_2^{-1} \right)$. Note that $J = m \, \text{diag}(1+\alpha^2, 1, \alpha^2)$, with $m = 1$ and $\alpha = 0.1$. The constant vectors are $y_\pm^0 = (0, \pm\alpha, -1)^T$ and $z = (0, 0, -3/2)^T$. Lastly, $q = \beta = 1$, and $\| \cdot \|_2$ is our usual Euclidean norm.

Both forms of the Hamiltonian are written here because while our method was developed for left-trivialized systems using Hamilton's principle, their method was developed for right-trivialized systems using the Hamilton–Pontryagin principle. As a result, both discretizations yield symplectic variational integrators for the Hamiltonian with their corresponding choice of trivialization. In particular, we have

$$\tilde{p} = g^T p \tag{30}$$

as a relationship between the dual elements of the corresponding trivialization: p is the dual representation of $\xi \in \mathfrak{g}$ for the right-trivialization $\dot{g} = \xi g$, and $\tilde{p}$ is the dual representation of $\tilde{\xi} \in \mathfrak{g}$ for the left-trivialization $\dot{g} = g\tilde{\xi}$. Consequently, we note that the left-trivialized cotangent vector $\frac{\partial H}{\partial g}$ in VPD (24) is computed as

$$\frac{\partial H}{\partial g}(g, \tilde{p}) = \text{Asym}\left(g^{-1} \nabla_g H^L(g, \tilde{p}) \right),$$

where $\nabla_g H^L(g, \tilde{p})$ is the matrix derivative of $H^L(g, \tilde{p})$ with respect to g. On the other hand, the right-trivialized cotangent vector is computed as

$$\frac{\partial H}{\partial g}(g, p) = \text{Asym}\left(\nabla_g H^R(g, p) g^{-1} \right),$$

when implementing the VRKMK method.

For our tests, we also have the same initial data from [3],

$$g(0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$p(0) = g(0) J g(0)^T e_2,$$

for the VRKMK methods, and so (30) gives $\tilde{p}(0)$ to complete the initial data for the VPD methods. Since both methods involve fixed-point iteration, we terminate the processes when the norm $\| \cdot \|_2$ between the current and previous iteration is less than 10^{-15} for each variable. In particular, the norm for vectors is the Euclidean norm and for matrices it is the induced matrix norm. Lastly, we ran these implementations in Wolfram Mathematica 12 on a personal computer with the following specifications: Operating system: Windows 10; CPU: 3.7GHz AMD Ryzen 5 5600X; RAM: G.Skill Trident Z Neo Series 32Gb DDR4 3600 C16; Storage: 1TB Rocket NVMe PCIe 4.0 writing/reading up to 5000/4400 MB/s.

5.1. **Order tests.** We run both methods based on the one-, two-, and three-stage Gauss–Legendre methods and a third order Runge–Kutta method for comparisons, and these methods are shown as Butcher tableaux in Table 1.

We compute the errors in $(g(0.5), p(0.5))$ for VRKMK and errors in $(g(0.5), \tilde{p}(0.5))$ for VPD with respect to a reference solution and these errors are shown in Figure 1. The reference solution was calculated using the sixth order VPD method with step size $h = 0.001$. The errors for VRKMK is computed as $\|g(0.5)^T p(0.5) - \tilde{p}_{\text{ref}}\|_2 + \|g(0.5) - g_{\text{ref}}\|_2$ while the errors for VPD is computed as $\|\tilde{p}(0.5) - \tilde{p}_{\text{ref}}\|_2 + \|g(0.5) - g_{\text{ref}}\|_2$. The black-dashed lines are reference lines for the appropriate orders: In the sixth order comparison plot, the errors from the fixed-point iteration are dominating for smaller step sizes, so the theoretical order is not observed in those regimes.

$$
\begin{array}{c|cc}
\frac{1}{2} & \frac{1}{2} \\
\hline
 & 1
\end{array}
$$

(A) second order Gauss–Legendre method

$$
\begin{array}{c|ccc}
0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
1 & -1 & 2 & 0 \\
\hline
 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6}
\end{array}
$$

(B) third order Runge–Kutta method

$$
\begin{array}{c|cc}
\frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\
\frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\
\hline
 & \frac{1}{2} & \frac{1}{2}
\end{array}
$$

(C) fourth order Gauss–Legendre method

$$
\begin{array}{c|ccc}
\frac{1}{2} - \frac{\sqrt{15}}{10} & \frac{5}{36} & \frac{2}{9} - \frac{\sqrt{15}}{15} & \frac{5}{36} - \frac{\sqrt{15}}{30} \\
\frac{1}{2} & \frac{5}{36} + \frac{\sqrt{15}}{24} & \frac{2}{9} & \frac{5}{36} - \frac{\sqrt{15}}{24} \\
\frac{1}{2} + \frac{\sqrt{15}}{10} & \frac{5}{36} + \frac{\sqrt{15}}{30} & \frac{2}{9} + \frac{\sqrt{15}}{15} & \frac{5}{36} \\
\hline
 & \frac{5}{18} & \frac{4}{9} & \frac{5}{18}
\end{array}
$$

(D) sixth order Gauss–Legendre method

TABLE 1. Butcher tableaux for the comparison tests

Otherwise, both methods achieve their theoretical orders, and they are quite comparable, noting that VPD exhibits a noticeable, smaller error constant in the second order method.

5.2. **Long-term behaviors.** The long-term energy behaviors for both methods are presented in Figures 2 – 5. For second, third, and fourth order, we fixed the step size $h = 0.01$ and ran 10^5 integration steps to observe the energy errors. For second order, the energy errors have magnitude orders of 10^{-4} for VRKMK and 10^{-5} for VPD; for third order, they have magnitude orders of 10^{-7} for VRKMK and 10^{-6} for VPD; for fourth order, they have magnitude orders of 10^{-9} for both methods.

For the sixth order, we consider step size $h = 1/26$ to avoid the regime where the errors in the fixed-point iteration are dominating. This step size corresponds to the third point from the right in the order comparison plots. We also run 10^5 integration steps and observe the energy errors in Figure 5. The energy error for the VPD method is stable with an order of magnitude of 10^{-10}. On the other
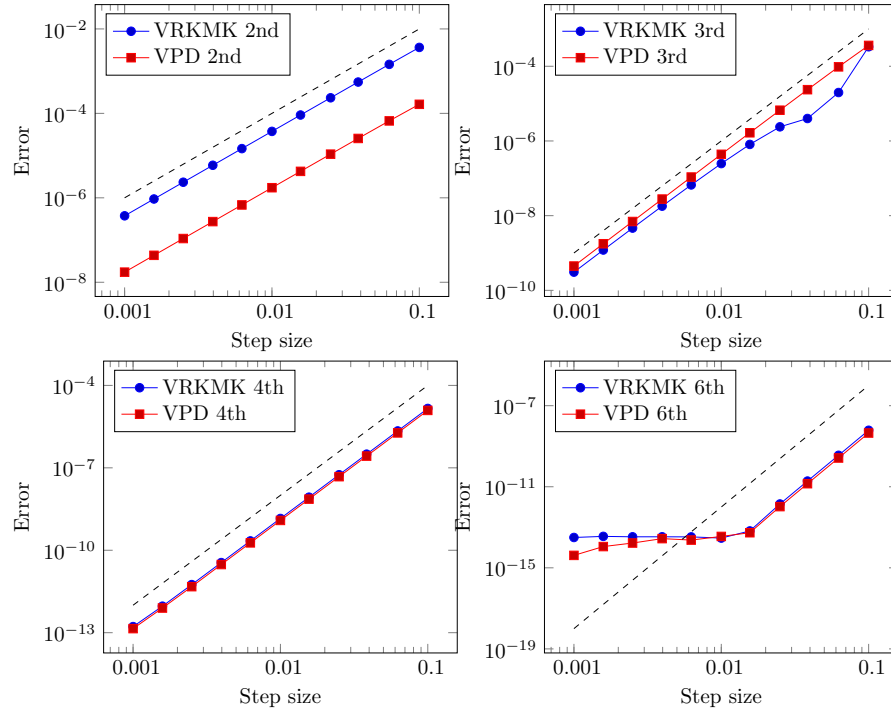
FIGURE 1. Order comparison plots between VRKMK and VPD methods: The black-dashed lines are references for the corresponding orders.

hand, the energy error in VRKMK exhibits a slow increase over the integration time span. We investigated this phenomenon and attribute this to the framework of each method. In VPD, g_1 in (24b) and the internal points U_i in (24c) are updated in each integration step via polar decomposition. As a result, the Lie group structure is always preserved up to machine precision for both the configuration space elements and internal points. However, in VRKMK, both g_1 and the internal points U_i are updated by the left action of $SO(n)$. This left action is a matrix multiplication which is performed to machine precision, but with a sufficient number of multiplications, the round-off error will still accumulate. Consequently, the Lie group structure is not as well-preserved in comparison, and this is illustrated in the comparison of orthogonality errors $\|gg^T - I_3\|_2$ in Figure 6.

We also observed that projecting the sixth order VRKMK method onto the rotation group using the polar decomposition at each timestep does not recover the near energy preservation typical of symplectic methods. Presumably, this is because the projection subtly compromises the symplecticity of the method. This is consistent with the observation made in [9], that both the Lie group structure and symplecticity need to be preserved for the methods to exhibit near energy preservation.

5.3. **Runtime comparison.** We also have data for runtime comparison of VPD and VRKMK for the two-stage and three-stage Gauss–Legendre methods in Figure 7. For each step size h, we recorded the runtime of each method in seconds and
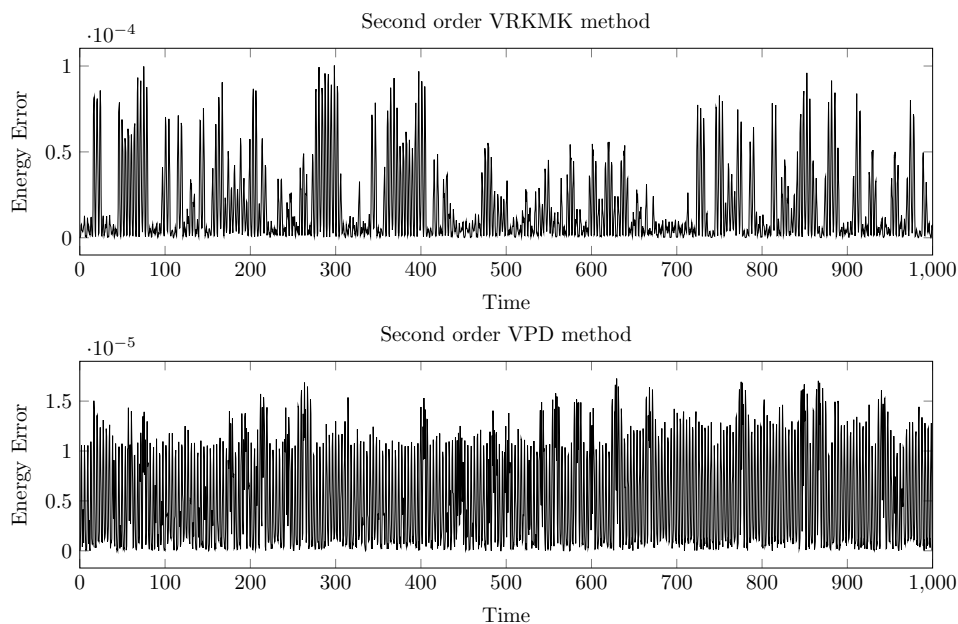
FIGURE 2. Long-term energy error for second order VRKMK/VPD methods.
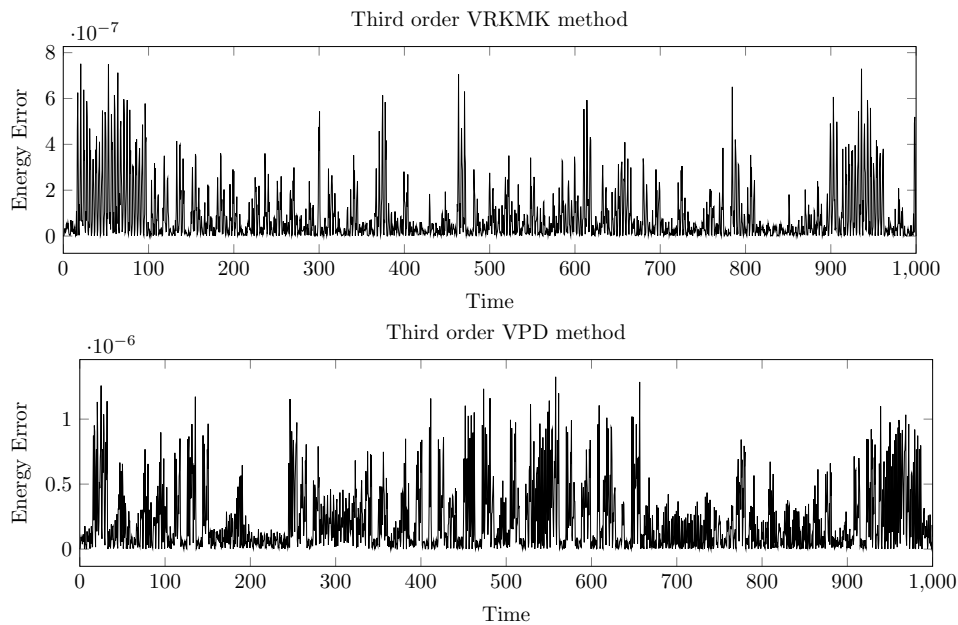

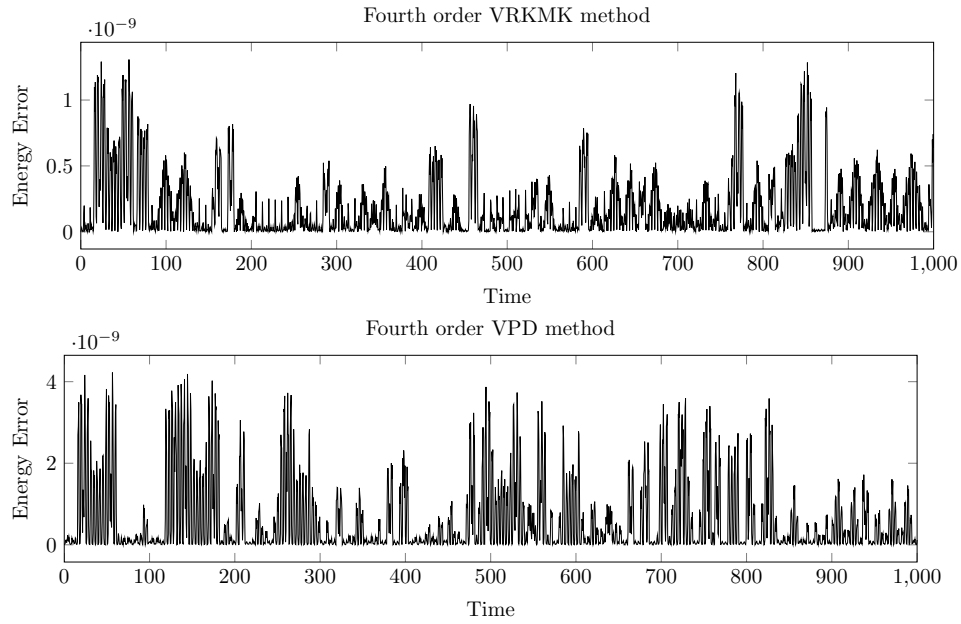FIGURE 3. Long-term energy error for third order VRKMK/VPD methods.

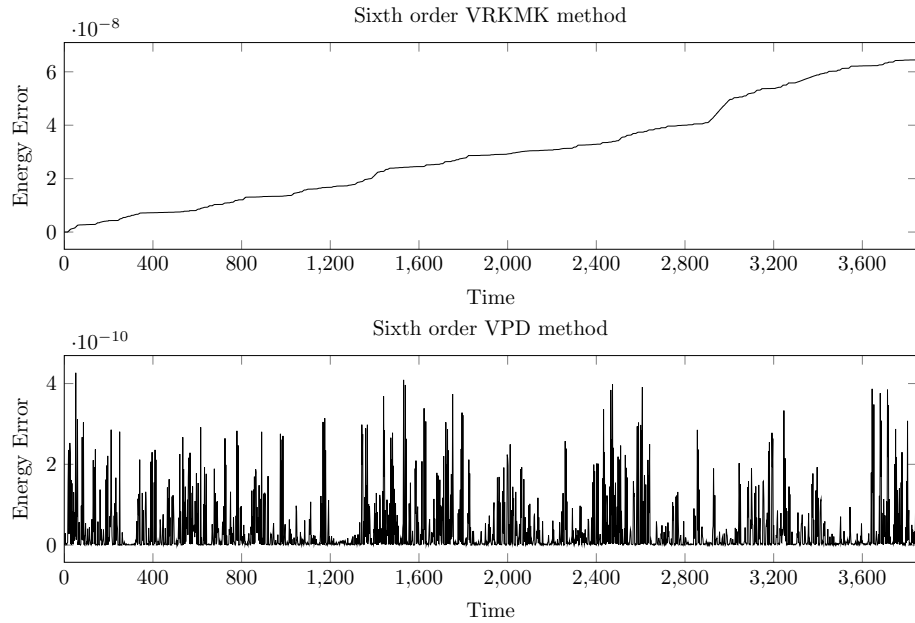FIGURE 4. Long-term energy error for fourth order VRKMK/VPD methods.

FIGURE 5. Long-term energy error for sixth order VRKMK/VPD methods.

Sixth order VRKMK method
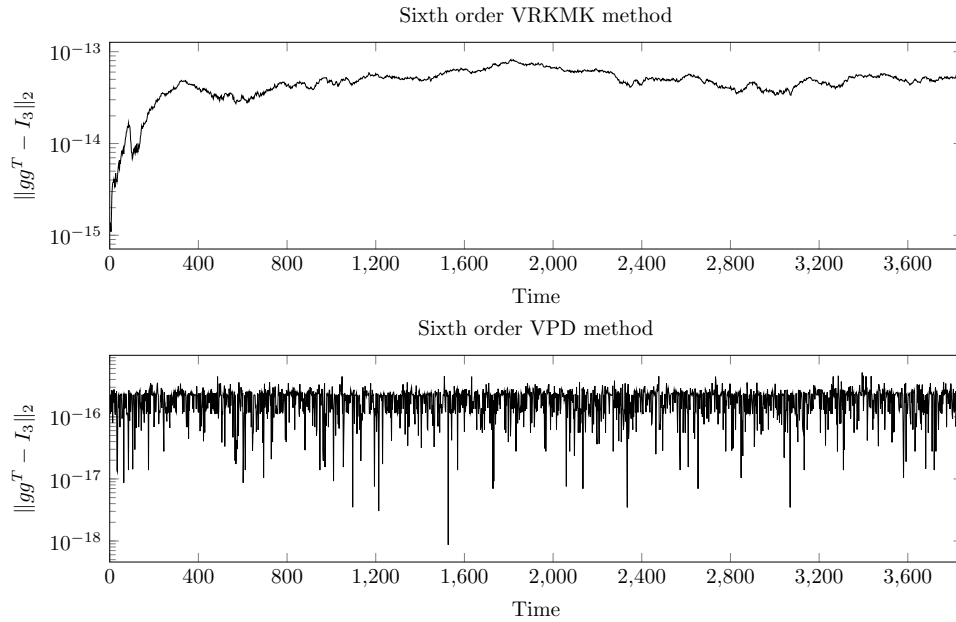

Sixth order VPD method

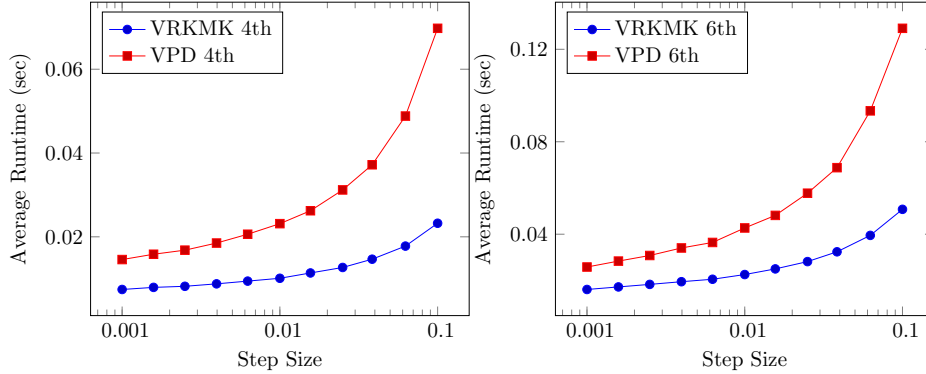FIGURE 6. Orthogonality error for sixth order VRKMK/VPD methods.

FIGURE 7. Run-time comparison for fourth and sixth order VRKMK/VPD methods.

repeated the execution 256 times to compute the average runtime. The runtime for VPD methods is on average longer than for VRKMK methods. This may be due in parts to the Lyapunov solutions and multiple layers of fixed-point iterations required in the polar decomposition in (4) and $(\tilde{\pi}_i \circ \tilde{\phi}^{-1})^*$ in the same equations (17a) and (19a).

6. **Conclusion.** By applying the polar decomposition in a constrained Galerkin method, we constructed high-order Lie group symplectic integrators on $SO(n)$, which we refer to as the variational polar decomposition method. Furthermore, the

integrator can be reduced to a Lie–Poisson integrator when the underlying mechanical system is group-invariant. We performed extensive numerical experiments comparing our proposed method to the variational Runge–Kutta–Munthe-Kaas method from [3]. These comparisons provide insights into each method and highlight an advantage of our proposed method, which is the preservation of Lie group structure up to machine precision for both the configurations and internal points. This appears to be important for high-order methods to exhibit the near energy preservation that one expects for symplectic integrators when applied to dynamics on Lie groups.

For future work, it would be interesting to explore the generalization of the proposed method to symmetric spaces, by applying the generalized polar decomposition [15]. This may be of particular interest in the construction of accelerated optimization algorithms on symmetric spaces, following the use of time-adaptive variational integrators for symplectic optimization [6] based on the discretization of Bregman Lagrangian and Hamiltonian flows [16, 5]. Examples of symmetric spaces include the space of Lorentzian metrics, the space of symmetric positive-definite matrices, and the Grassmannian.

Appendix A. **Constrained Galerkin methods.** Our Galerkin variational integrator will involve a discrete Lagrangian that differs from the classical construction in [14]. Traditionally in the linear space setting, (1) is approximated with a quadrature rule

$$L_d(q_0, q_1) = h \sum_{i=1}^{s} b_i L(q(c_i h), \dot{q}(c_i h)) = h \sum_{i=1}^{s} L(Q_i, \dot{Q}_i),$$

and $q(t)$ is approximated by polynomials with degree less than or equal to s with fixed endpoints q_0 and q_1. By choosing interpolation nodes $\{d_0^\nu\}_{\nu=0}^s$ with $d_0^0 = 0, d_0^s = 1$ and interpolation values $\{q_0^\nu\}_{\nu=0}^s$ with $q_0^0 = q_0$ and $q_0^s = q_1$, $q(t)$ can be expressed as $q(t) = \sum_{\nu=0}^s q_0^\nu \phi_\nu \left(\frac{t}{h}\right)$ on $[0, h]$, where $\phi_\nu(t)$ are Lagrange polynomials corresponding to the nodes $\{d_0^\nu\}_{\nu=0}^s$. By taking variations with respect to the interpolation values $\{q_0^\nu\}_{\nu=1}^{s-1}$, $q(t)$ is varied over the finite-dimensional function space,

$$M^s = \{q(t) \mid q(t) \in P_s[0, h], q(0) = q_0, q(h) = q_1\}.$$

Consider the quadrature approximation of the action integral, viewing it as a function of the endpoint and interpolation values,

$$F(q_0, q_1, \{q_0^\nu\}_{\nu=1}^{s-1}) = h \sum_{i=1}^{s} b_i L(q(c_i h), \dot{q}(c_i h)),$$

where $q(t) = \sum_{\nu=0}^s q_0^\nu \phi_\nu(\frac{t}{h})$. Then, a variational integrator (2) can be obtained as follows,

$$\begin{cases} 0 = \dfrac{\partial F}{\partial q_0^\nu}, & \nu = 1, 2 \ldots s - 1, \\[2mm] -p_0 = \dfrac{\partial F}{\partial q_0}, \\[2mm] p_1 = \dfrac{\partial F}{\partial q_1}. \end{cases} \tag{31}$$

However, (31) is often impractical due to the complexity of evaluating $q(c_i h)$ and $\dot{q}(c_i h)$, which involve the Lagrange interpolating polynomials. In addition, computing the root of a system of nonlinear equations in (31) can be challenging because the formulation of a fixed-point problem could be complicated, and convergence issues could arise. In contrast, Runge-Kutta methods are already in fixed-point form and are convergent as long as consistency is satisfied.

Now, note that the finite-dimensional function space M^s does not depend on the choice of nodes $\{d_0^\nu\}_{\nu=1}^{s-1}$, and by a tricky elimination of $\phi_\nu(t)$, (31) can be simplified to yield

$$q_1 = q_0 + h \sum_{i=1}^{s} b_i \dot{Q}_i, \qquad\qquad p_1 = p_0 + h \sum_{i=1}^{s} b_i \dot{P}_i, \qquad (32\text{a})$$

$$Q_i = q_0 + h \sum_{j=1}^{s} a_{ij} \dot{Q}_j, \qquad\qquad P_i = p_0 + h \sum_{j=1}^{s} \tilde{a}_{ij} \dot{P}_j, \qquad (32\text{b})$$

$$P_i = \frac{\partial L}{\partial \dot{q}}(Q_i, \dot{Q}_i), \qquad\qquad \dot{P}_i = \frac{\partial L}{\partial q}(Q_i, \dot{Q}_i), \qquad (32\text{c})$$

where $\tilde{a}_{ij} = b_j(1 - \frac{a_{ji}}{b_i})$. When transformed to the Hamiltonian side, (32) simply recovers the symplectic partitioned Runge–Kutta method.

The same variational integrator can be derived in a much simpler way: Instead of performing variations on internal points $\{q_0^\nu\}_{\nu=1}^{s-1}$, we will perform variations on the internal derivatives $\{\dot{Q}\}_{i=1}^{s}$, subject to the constraint $q_1 = q_0 + h \sum_{i=1}^{s} b_i \dot{Q}_i$. Then, the internal points are reconstructed using the fundamental theorem of calculus and the degree of precision of the quadrature rule to obtain $Q_i = q_0 + h \sum_{j=1}^{s} a_{ij} \dot{Q}_j$. Now, consider the quadrature approximation of the action integral, viewed as a function of the endpoint values and the internal velocities,

$$\tilde{F}(q_0, q_1, \{\dot{Q}_i\}_{i=1}^{s}, \lambda) = h \sum_{i=1}^{s} b_i L(Q_i, \dot{Q}_i) + \lambda^T \left(q_1 - q_0 - h \sum_{i=1}^{s} b_i \dot{Q}_i \right),$$

where λ is a Lagrange multiplier that enforces the constraint. Then, a variational integrator (2) can be obtained as follows,

$$\begin{cases} 0 = \dfrac{\partial \tilde{F}}{\partial \dot{Q}_i}, \qquad i = 1, 2 \ldots s, \\[2mm] 0 = \dfrac{\partial \tilde{F}}{\partial \lambda}, \\[2mm] Q_i = q_0 + h \displaystyle\sum_{j=1}^{s} a_{ij} \dot{Q}_j, \\[2mm] -p_0 = \dfrac{\partial \tilde{F}}{\partial q_0}, \\[2mm] p_1 = \dfrac{\partial \tilde{F}}{\partial q_1}. \end{cases} \qquad (33)$$

Explicitly expanding (33) and eliminating the Lagrange multiplier yields (32) in a much more straightforward manner. This same technique, known as the *constrained Galerkin method*, is adopted on the rotation group $SO(n)$ in order to directly obtain a variational integrator in fixed-point form.

Appendix B. **Euler–Poincaré and Lie–Poisson reductions.** When the Lagrangian L or Hamiltonian H is left-invariant, the mechanical system can be symmetry reduced to evolve on the Lie algebra $\mathfrak{g}$ or its dual $\mathfrak{g}^*$, respectively, assuming some regularity. On the Lagrangian side, this corresponds to Euler–Poincaré reduction [13], which is described by the Euler–Poincaré equations,

$$\frac{d}{dt}\left(\frac{\partial \boldsymbol{l}}{\partial \epsilon}\right) = \mathrm{ad}^*_\epsilon\left(\frac{\partial \boldsymbol{l}}{\partial \epsilon}\right).$$

The above is expressed in terms of the reduced Lagrangian $\boldsymbol{l}(\epsilon) = \boldsymbol{l}(g^{-1}\dot{g}) = L(g,\dot{g})$. As a result, this can be described in terms of a reduced variational principle $\delta \int_a^b \boldsymbol{l}(\epsilon(t))\,dt = 0$ with respect to constrained variations of form $\delta\epsilon = \dot{\eta} + [\epsilon, \eta]$, where $\eta(t)$ is an arbitrary path in the Lie algebra $\mathfrak{g}$ that vanishes at the endpoints, namely $\eta(a) = \eta(b) = 0$. The constraints on the variations $\delta\epsilon$ are induced by the condition that $\epsilon = g^{-1}\dot{g}$ together with arbitrary variations δg that vanish at the endpoints.

On the Hamiltonian side, this corresponds to Lie–Poisson reduction [13]. Recall that the Lie–Poisson structure on $\mathfrak{g}^*$ is given by

$$\{F, G\}(\mu) = \left\langle \mu, \left[\frac{\partial F}{\partial \mu}, \frac{\partial G}{\partial \mu}\right]\right\rangle,$$

and together with the reduced Hamiltonian $\boldsymbol{h}(\mu)$, they gives the Lie–Poisson equations on $\mathfrak{g}^*$,

$$\frac{d\mu}{dt} = \mathrm{ad}^*_{\frac{\partial \boldsymbol{h}}{\partial \mu}}(\mu).$$

If the discrete Lagrangian $L_d(g_0, g_1)$ is also G-invariant, meaning $L_d(g \cdot g_0, g \cdot g_1) = L_d(g_0, g_1)$ for some $g \in G$, then (2) can be reduced to a Lie–Poisson integrator [12],

$$\begin{cases} \mu_0 = \boldsymbol{l}'_d(f_0)f_0^{-1}, \\ \mu_1 = f_0^{-1} \cdot \mu_0 \cdot f_0, \end{cases} \tag{34}$$

where $\boldsymbol{l}_d(f_0) = L_d(e, f_0)$. This algorithm preserves the coadjoint orbits and, hence, the Poisson structure on $\mathfrak{g}^*$.

Appendix C. **Detailed derivations for Lagrangian variational integrators.**

C.1. **Derivations of** $\frac{\partial \tilde{F}}{\partial \Omega_k}$**.** Initially, we have

$$\frac{d}{dt}\bigg|_{t=0} \tilde{F}(g_0, g_1, \ldots, \Omega_k(t), \ldots, \Lambda) =$$

$$h\sum_{i=1}^{s} b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i), X_{ik}\right\rangle + hb_k \left\langle \frac{\partial L}{\partial \Omega}(U_k, \Omega_k), \delta\Omega_k\right\rangle$$

$$+ \left\langle \Lambda, \mathrm{Asym}\left(hg_1^T \sum_{i=1}^{s} b_i U_i X_{ik} \Omega_i + hg_1^T b_k U_k \delta\Omega_k\right)\right\rangle,$$

where we can use the properties of the inner products to express the last term as follows,

$$h\sum_{i=1}^{s} b_i\langle \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T), X_{ik}\rangle + hb_k\langle \mathrm{Asym}(U_k^T g_1 \Lambda), \delta\Omega_k\rangle.$$

Then, we continue using equation (12) to obtain,

$$
\begin{aligned}
\frac{d}{dt}\bigg|_{t=0} &\tilde{F}(g_0, g_1, \ldots, \Omega_k(t), \ldots, \Lambda) \\
&= h \sum_{i=1}^{s} b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T), X_{ik} \right\rangle \\
&\quad + hb_k \left\langle \frac{\partial L}{\partial \Omega}(U_k, \Omega_k) + \mathrm{Asym}(U_k^T g_1 \Lambda), \delta\Omega_k \right\rangle \\
&= h \sum_{i=1}^{s} b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T), h(\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)(\delta\Omega_k) \right\rangle \\
&\quad + hb_k \left\langle \frac{\partial L}{\partial \Omega}(U_k, \Omega_k) + \mathrm{Asym}(U_k^T g_1 \Lambda), \delta\Omega_k \right\rangle \\
&= h^2 \sum_{i=1}^{s} b_i \left\langle (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* \left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T) \right), \delta\Omega_k \right\rangle \\
&\quad + hb_k \left\langle \frac{\partial L}{\partial \Omega}(U_k, \Omega_k) + \mathrm{Asym}(U_k^T g_1 \Lambda), \delta\Omega_k \right\rangle.
\end{aligned}
$$

We finally have

$$
\begin{aligned}
\frac{\partial \tilde{F}}{\partial \Omega_k} = h^2 \sum_{i=1}^{s} b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\psi}_k)^* &\left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T) \right) \\
&+ hb_k \left(\frac{\partial L}{\partial \Omega}(U_k, \Omega_k) + \mathrm{Asym}(U_k^T g_1 \Lambda) \right).
\end{aligned}
$$

C.2. **Explicit expression for $\tilde{\phi}^*$: A derivation.** Let us consider $(S_1, \ldots, S_s)$, $(\tilde{S}_1, \ldots, \tilde{S}_s) \in \mathrm{Skew}(n)^s$, and so

$$
\begin{aligned}
\langle \tilde{\phi}^*&(S_1, \ldots, S_s), (\tilde{S}_1, \ldots, \tilde{S}_s) \rangle \\
&= \langle (S_1, \ldots, S_s), \tilde{\phi}(\tilde{S}_1, \ldots, \tilde{S}_s) \rangle \\
&= \left\langle (S_1, \ldots, S_s), \left\{ \tilde{S}_i - d\mathbb{P}_{A_i} \left(h \sum\nolimits_{j=1}^{s} a_{ij} U_j \tilde{S}_j \Omega_j \right) \right\}_{i=1}^{s} \right\rangle \\
&= \sum_{i=1}^{s} \left\langle S_i, \tilde{S}_i - d\mathbb{P}_{A_i} \left(h \sum\nolimits_{j=1}^{s} a_{ij} U_j \tilde{S}_j \Omega_j \right) \right\rangle \\
&= \sum_{i=1}^{s} \langle S_i, \tilde{S}_i \rangle - \sum_{i=1}^{s} \mathrm{tr} \left(d\mathbb{P}_{A_i}^*(S_i) \left(h \sum\nolimits_{j=1}^{s} a_{ij} U_j \tilde{S}_j \Omega_j \right)^T \right) \\
&= \sum_{i=1}^{s} \langle S_i, \tilde{S}_i \rangle - \sum_{i,j=1}^{s} \mathrm{tr} \left(\left(h a_{ij} U_j^T d\mathbb{P}_{A_i}^*(S_i) \Omega_j^T \right) \tilde{S}_j^T \right) \\
&= \sum_{i=1}^{s} \langle S_i, \tilde{S}_i \rangle - \sum_{j=1}^{s} \mathrm{tr} \left(\left(h U_j^T \sum\nolimits_{i=1}^{s} a_{ij} d\mathbb{P}_{A_i}^*(S_i) \Omega_j^T \right) \tilde{S}_j^T \right) \\
&= \sum_{j=1}^{s} \left\langle S_j - \mathrm{Asym} \left(h U_j^T \sum\nolimits_{i=1}^{s} a_{ij} d\mathbb{P}_{A_i}^*(S_i) \Omega_j^T \right), \tilde{S}_j \right\rangle.
\end{aligned}
$$

This gives us equation (15).

C.3. **Explicit expression for** $\tilde{\psi}_k^*$: **A derivation.** Consider $(S_1, \ldots, S_s) \in$ $\mathrm{Skew}(n)^s$ and $\tilde{S} \in \mathrm{Skew}(n)$, then

$$
\begin{aligned}
\langle \tilde{\psi}_k^*(S_1, \ldots, S_s), \tilde{S} \rangle &= \langle (S_1, \ldots, S_s), \tilde{\psi}_k(\tilde{S}) \rangle \\
&= \left\langle (S_1, \ldots, S_s), \left\{ d\mathbb{P}_{A_i}(a_{ik} U_k \tilde{S}) \right\}_{i=1}^s \right\rangle \\
&= \sum_{i=1}^s \mathrm{tr} \left(d\mathbb{P}_{A_i}^*(S_i)(a_{ik} U_k \tilde{S})^T \right) \\
&= \sum_{i=1}^s \mathrm{tr} \left((a_{ik} U_k^T d\mathbb{P}_{A_i}^*(S_i)) \tilde{S}^T \right) \\
&= \left\langle \mathrm{Asym} \left(U_k^T \sum_{i=1}^s a_{ik} d\mathbb{P}_{A_i}^*(S_i) \right), \tilde{S} \right\rangle.
\end{aligned}
$$

This gives us equation (16).

C.4. **Explicit expression for** $\tilde{\varphi}^*$: **A derivation.** Let us derive $\tilde{\varphi}^*$ by considering $(S_1, \ldots, S_s) \in \mathrm{Skew}(n)^s$ and $\tilde{S} \in \mathrm{Skew}(n)$, and so

$$
\begin{aligned}
\langle \tilde{\varphi}^*(S_1, \ldots, S_s), \tilde{S} \rangle &= \left\langle (S_1, \ldots, S_s), \{ d\mathbb{P}_{A_i}(g_0 \tilde{S}) \}_{i=1}^s \right\rangle \\
&= \sum_{i=1}^s \mathrm{tr} \left(d\mathbb{P}_{A_i}^*(S_i)(g_0 \tilde{S})^T \right) \\
&= \sum_{i=1}^s \mathrm{tr} \left((g_0^T d\mathbb{P}_{A_i}^*(S_i)) \tilde{S}^T \right) \\
&= \left\langle \mathrm{Asym} \left(g_0^T \sum_{i=1}^s d\mathbb{P}_{A_i}^*(S_i) \right), \tilde{S} \right\rangle.
\end{aligned}
$$

This gives us equation (18).

C.5. **Derivation of** $\frac{\partial \tilde{F}}{\delta g_0}$. We compute

$$
\begin{aligned}
\frac{d}{dt} &\bigg|_{t=0} \tilde{F}(g_0(t), g_1, \{\Omega_i\}_{i=1}^s, \Lambda) \\
&= h \sum_{i=1}^s b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i), Y_i \right\rangle + \left\langle \Lambda, \mathrm{Asym} \left(g_1^T \left(g_0 \delta g_0 + h \sum_{i=1}^s b_i U_i Y_i \Omega_i \right) \right) \right\rangle \\
&= h \sum_{i=1}^s b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i), Y_i \right\rangle + \mathrm{tr} \left(\Lambda(g_1^T g_0 \delta g_0)^T \right) + h \sum_{i=1}^s \mathrm{tr} \left(\Lambda(b_i g_1^T U_i Y_i \Omega_i)^T \right) \\
&= h \sum_{i=1}^s b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i), Y_i \right\rangle + \mathrm{tr} \left((g_0^T g_1 \Lambda) \delta g_0^T \right) + h \sum_{i=1}^s \mathrm{tr} \left(b_i(U_i^T g_1 \Lambda \Omega_i^T) Y_i \right) \\
&= h \sum_{i=1}^s b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T), Y_i \right\rangle + \langle \mathrm{Asym}(g_0^T g_1 \Lambda), \delta g_0 \rangle \\
&= h \sum_{i=1}^s b_i \left\langle \frac{\partial L}{\partial U}(U_i, \Omega_i) + \mathrm{Asym}(U_i^T g_1 \Lambda \Omega_i^T), (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})(\delta g_0) \right\rangle \\
&\quad + \langle \mathrm{Asym}(g_0^T g_1 \Lambda), \delta g_0 \rangle
\end{aligned}
$$

Thus, we have

$$\frac{\partial \tilde{F}}{\partial g_0} = h \sum_{i=1}^{s} b_i (\tilde{\pi}_i \circ \tilde{\phi}^{-1} \circ \tilde{\varphi})^* \left(\frac{\partial L}{\partial U}(U_i, \Omega_i) + \text{Asym}(U_i^T g_1 \Lambda \Omega_i^T) \right) + \text{Asym}(g_0^T g_1 \Lambda).$$

REFERENCES

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

[2] R. H. Bartels and G. W. Stewart, Solution of the matrix equation $ax + xb = c$ [f4], *Communications of the ACM*, **15** (1972), 820–826.

[3] G. Bogfjellmo and H. Marthinsen, High-order symplectic partitioned Lie group methods, *Found. Comput. Math.*, **16** (2016), 493–530.

[4] E. Celledoni and B. Owren, A class of intrinsic schemes for orthogonal integration, *SIAM J. Numer. Anal.*, **40** (2002), 2069–2084.

[5] V. Duruisseaux and M. Leok, A variational formulation of accelerated optimization on Riemannian manifolds, *SIAM Journal on Mathematics of Data Science*, Accepted.

[6] V. Duruisseaux, J. Schmitt and M. Leok, Adaptive Hamiltonian variational integrators and applications to symplectic accelerated optimization, *SIAM J. Sci. Comput.*, **43** (2021), A2949–A2980.

[7] G. H. Golub, S. Nash and C. Van Loan, A Hessenberg-Schur method for the problem $ax + xb = c$, *IEEE Trans. Automat. Control*, **24** (1979), 909–913.

[8] J. Hall and M. Leok, Lie group spectral variational integrators, *Found. Comput. Math.*, **17** (2017), 199–257.

[9] T. Lee, M. Leok and N. McClamroch, Lie group variational integrators for the full body problem in orbital mechanics, *Celestial Mech. Dynam. Astronom.*, **98** (2007), 121–144.

[10] M. Leok, Generalized Galerkin variational integrators, arXiv:math/0508360.

[11] M. Leok and T. Shingel, General techniques for constructing variational integrators, *Front. Math. China*, **7** (2012), 273–303.

[12] J. E. Marsden, S. Pekarsky and S. Shkoller, Discrete Euler–Poincaré and Lie-Poisson equations, *Nonlinearity*, **12** (1999), 1647–1662.

[13] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, vol. 17 of Texts in Applied Mathematics, 2nd edition, Springer-Verlag, 1999.

[14] J. E. Marsden and M. West, Discrete mechanics and variational integrators, *Acta Numer.*, **10** (2001), 357–514.

[15] H. Z. Munthe-Kaas, G. R. W. Quispel and A. Zanna, Generalized polar decompositions on Lie groups with involutive automorphisms, *Found. Comput. Math.*, **1** (2001), 297–324.

[16] A. Wibisono, A. C. Wilson and M. I. Jordan, A variational perspective on accelerated methods in optimization, *Proc. Natl. Acad. Sci. USA*, **113** (2016), E7351–E7358.

E-mail address: xus009@ucsd.edu
E-mail address: k2tran@ucsd.edu
E-mail address: mleok@ucsd.edu