S^3 Forced Variational Integrator Network

Hu Hanyang
Instructor: Melvin Leok

University of California San Diego
MATH 273A Course Project

December 9, 2025

Outline

Introduction

What are **Variational Integrator Networks (VINs)**[1]?

Motivation. Encode prior knowledge of the underlying physical laws that govern the dynamical systems into the model design.

Example. Consider the Lagrangian $L_\theta(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}\dot{\mathbf{q}}^\top \mathbf{M}_\theta \dot{\mathbf{q}} - U_\theta(\mathbf{q})$, then the Velocity-Verlet method

$$\mathbf{q}_{k+1} = \mathbf{q}_k + h\dot{\mathbf{q}}_k - \frac{h^2}{2}\mathbf{M}_\theta^{-1}\nabla U_\theta(\mathbf{q}_k)$$
$$\dot{\mathbf{q}}_{k+1} = \dot{\mathbf{q}}_k - h\mathbf{M}_\theta^{-1}\left(\frac{\nabla U_\theta(\mathbf{q}_k) + \nabla U_\theta(\mathbf{q}_{k+1})}{2}\right)$$

can serve as the feed-forward architecture of the VIN.

[1] Saemundsson et al., *Variational Integrator Networks for Physically Structured Embeddings*.

Advantages of VINs

Compared to **Hamiltonian neural networks** that learns a parameterized Hamiltonian by minimizing the loss function

$$\mathcal{L}_{\mathrm{HNN}} = \left\| \frac{\partial \mathcal{H}_\theta}{\partial \mathbf{p}} - \dot{\mathbf{q}} \right\| + \left\| \frac{\partial \mathcal{H}_\theta}{\partial \mathbf{q}} + \dot{\mathbf{p}} \right\|$$

VINs have the following advantages:

1. Automatically enforce symplecticity, momentum preservation, and approximate energy conservation.
2. Do not need data to sufficiently cover the configuration space.

Motivation

In robotics and control applications (e.g., model-predictive control), we want to model $(\mathbf{q}_k, \dot{\mathbf{q}}_k, \mathbf{u}_k) \mapsto (\mathbf{q}_{k+1}, \dot{\mathbf{q}}_{k+1})$.

We need to consider external forcing (e.g., control, damping, contact).

The **Forced Variational Integrator Networks (FVINs)**[2] was presented for this purpose, following with the **Lie Group Forced Variational Integrator Networks (LieFVINs)**[3] on $\mathbf{SE}(3)$.

Goal. Understand forced variational integrators and extend LieFVINs on the unit quaternion group S^3.

[2] Havens and Chowdhary, *Forced Variational Integrator Networks for Prediction and Control of Mechanical Systems*.

[3] Duruisseaux et al., *Lie Group Forced Variational Integrator Networks for Learning and Control of Robot Systems*.

Outline

Continuous Forced System

Notations. Configuration manifold Q, control manifold $\mathcal{U}$, Lagrangian $L : TQ \to \mathbb{R}$, external force $f_L : TQ \times \mathcal{U} \to T^*Q$.

Lagrange-d'Alembert principle.

$$\delta \underbrace{\int_0^T L(\mathbf{q}(t), \dot{\mathbf{q}}(t))\, dt}_{\text{action integral}} + \underbrace{\int_0^T f_L(\mathbf{q}(t), \dot{\mathbf{q}}(t), \mathbf{u}(t)) \cdot \delta \mathbf{q}(t)\, dt}_{\text{virtual work}} = 0$$

subject to $\delta \mathbf{q}(0) = \delta \mathbf{q}(T) = 0$.

Forced Euler-Lagrange equations.

$$\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{q}}} = f_L(\mathbf{q}(t), \dot{\mathbf{q}}(t), \mathbf{u}(t)).$$

Discrete Forced System

Notations. Discrete Lagrangian $L_d : Q \times Q \to \mathbb{R}$, discrete controlled forces $f_d^{\pm} : Q \times Q \times \mathcal{U} \to T^*Q$.

Discrete Lagrange-d'Alembert principle.

$$\delta \sum_{k=0}^{N-1} L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) + \sum_{k=0}^{N-1} \underbrace{[f_d^-(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k) \cdot \delta \mathbf{q}_k + f_d^+(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k) \cdot \delta \mathbf{q}_{k+1}]}_{\approx \int_{t_k}^{t_{k+1}} f_L(\mathbf{q}(t), \dot{\mathbf{q}}(t), \mathbf{u}(t)) \cdot \delta \mathbf{q}(t) dt} = 0$$

subject to $\delta \mathbf{q}_0 = \delta \mathbf{q}_N = 0$.

Forced Discrete Euler-Lagrange Equation.

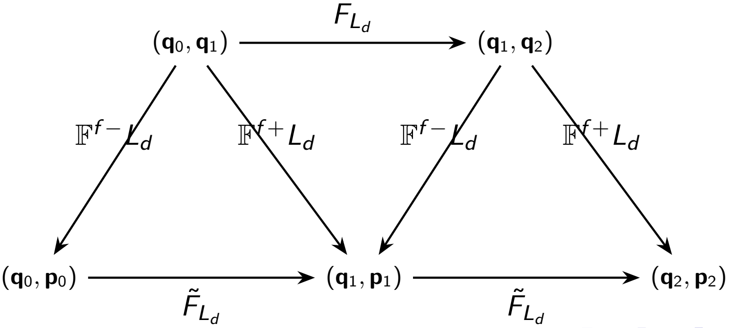
$$\mathrm{D}_1 L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) + D_2 L_d(\mathbf{q}_{k-1}, \mathbf{q}_k) + f_d^-(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k) + f_d^+(\mathbf{q}_{k-1}, \mathbf{q}_k, \mathbf{u}_{k-1}) = 0.$$

Forced Discrete Legendre Transform

Define $\mathbb{F}^{f\pm}L_d : Q \times Q \to T^*Q$ by

$$\mathbb{F}^{f+}L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) = (\mathbf{q}_{k+1}, D_2L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) + f_d^+(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k))$$
$$\mathbb{F}^{f-}L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) = (\mathbf{q}_k, -D_1L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) - f_d^-(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k))$$

so that $\mathbb{F}^{f\pm}L_d$ is consistent with the continuous Legendre transform $\mathbb{F}L$ when the discrete Lagrangian L_d and discrete forces $f_d^\pm$ are exact.



Symplecticity and Forced Discrete Noether's Theorem

In general, we have $\mathbf{dq}_k^i \wedge \mathbf{dp}_k^i \neq \mathbf{dq}_{k+1}^j \wedge \mathbf{dp}_{k+1}^j$ since

$$
\begin{aligned}
0 &= \mathbf{d}^2 L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) \\
&= \mathbf{dq}_k^i \wedge \mathbf{dp}_k^i - \mathbf{dq}_{k+1}^j \wedge \mathbf{dp}_{k+1}^j \\
&\quad - \underbrace{\left[\mathbf{d}f_d^{-i}(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k) \wedge \mathbf{dq}_k^i + \mathbf{d}f_d^{+j}(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k)) \wedge \mathbf{dq}_{k+1}^j\right]}_{\text{extra terms from forcing}}.
\end{aligned}
$$

Forced Discrete Noether's Theorem.

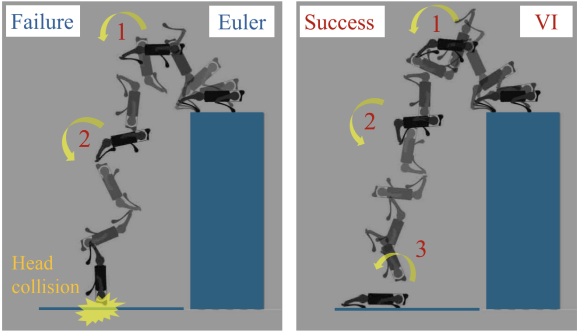
For a G-invariant discrete Lagrangian, if the discrete forces are orthogonal to the group action in the sense that

$$\langle f_d^-(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k), \xi_Q(\mathbf{q}_k)\rangle + \langle f_d^+(\mathbf{q}_k, \mathbf{q}_{k+1}, \mathbf{u}_k), \xi_Q(\mathbf{q}_{k+1})\rangle = 0$$

then the discrete momentum map $J_d : Q \times Q \to \mathfrak{g}^*$ is preserved.

Example Application of Forced VI

A recent application in robotics for modeling aerial maneuvers.[4]



[4] Beck et al., "High Accuracy Aerial Maneuvers on Legged Robots using Variational Integrator Discretized Trajectory Optimization".

Outline

Lagrangian on $\mathbb{R}^3 \times S^3$

Consider the Lagrangian $L : T\mathbf{SE}(3) \to \mathbb{R}$

$$L_\theta(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{R}, \dot{\mathbf{R}}) = \frac{1}{2}\dot{\mathbf{x}}^\top \mathbf{M}_\theta \dot{\mathbf{x}} + \frac{1}{2}\omega^\top \mathbf{J}_\theta \omega - U_\theta(\mathbf{R})$$

where $\omega = (\mathbf{R}^\top \dot{\mathbf{R}})^\vee \in \mathbb{R}^3 \cong \mathfrak{so}(3)$ is the angular velocity.

We can lift it to $\hat{L} : T(\mathbf{R}^3 \times S^3) \to R$ using the Lie group homomorphism $\Phi : \mathbf{q} \mapsto (2q_s^2 - 1)\mathbb{I} + 2\mathbf{q}_v\mathbf{q}_v^\top + 2q_s[\mathbf{q}_v]_\times \in \mathbf{SO}(3)$, such that

$$\hat{L}_\theta(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}\dot{\mathbf{x}}^\top \mathbf{M}_\theta \dot{\mathbf{x}} + 2\xi^\top \mathbf{J}_\theta \xi - U_\theta(\Phi(\mathbf{q}))$$

where $\xi = \mathrm{Im}(\mathbf{q}^* \dot{\mathbf{q}})$ since $(\mathbf{q}, \dot{\mathbf{q}}) = (\mathbf{q}, \mathbf{q}\xi) \xrightarrow{T_\mathbf{q}\Phi} (\Phi(\mathbf{q}), 2\Phi(\mathbf{q})[\xi]_\times)$.

Forced Variational Integrator Network on $\mathbb{R}^3 \times S^3$

A straightforward extension to the result of Shen et al.[5]

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h\mathbf{M}_\theta^{-1}\mathbf{p}_k - \frac{h^2}{2}\mathbf{M}_\theta^{-1}\nabla_\mathbf{x} U_\theta(\mathbf{x}_k, \mathbf{q}_k) + h\mathbf{M}_\theta^{-1} f_d^{\mathbf{x}-}$$

$$\mathbf{p}_{k+1} = \mathbf{p}_k - h\frac{\nabla_\mathbf{x} U_\theta(\mathbf{x}_k, \mathbf{q}_k) + \nabla_\mathbf{x} U_\theta(\mathbf{x}_{k+1}, \mathbf{q}_{k+1})}{2} + f_d^{\mathbf{x}+} + f_d^{\mathbf{x}-}$$

$$\pi_k = -\frac{4}{h} G(\mathbf{q}_{k+1}^*\mathbf{q}_k)\mathbf{J}_\theta \mathrm{Im}(\mathbf{q}_{k+1}^*\mathbf{q}_k) + \frac{h}{2} H(\mathbf{q}_k)\nabla_\mathbf{q} U_\theta(\mathbf{x}_k, \mathbf{q}_k) - f_d^{\mathbf{q}-}$$

$$\pi_{k+1} = \frac{4}{h} G(\mathbf{q}_k^*\mathbf{q}_{k+1})\mathbf{J}_\theta \mathrm{Im}(\mathbf{q}_k^*\mathbf{q}_{k+1}) - \frac{h}{2} H(\mathbf{q}_{k+1})\nabla_\mathbf{q} U_\theta(\mathbf{x}_{k+1}, \mathbf{q}_{k+1}) + f_d^{\mathbf{q}+}$$

where $G(\mathbf{q}) = \mathbf{q}_s\mathbb{I} - [\mathbf{q}_v]_\times$, $H(\mathbf{q}) = (-\mathbf{q}_v, G(\mathbf{q}))$ and $\mathbf{q}_{k+1} = \mathbf{q}_k \exp(\xi_k)$.

Left-trivialized momenta. Compute momenta in $\mathbf{R}^3 \cong (\mathfrak{s}^3)^*$

$$\pi_k = -T_\mathbf{e}^* L_{\mathbf{q}_k} D_1 L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) - f_d^-$$
$$\pi_{k+1} = T_\mathbf{e}^* L_{\mathbf{q}_{k+1}} D_2 L_d(\mathbf{q}_k, \mathbf{q}_{k+1}) + f_d^+$$

[5]Shen and Leok, *Lie group variational integrators for rigid body problems using quaternions*.

Overcoming Double-Covering Issue

Observation. For any $\mathbf{R} \in \mathbf{SO}(3)$, $\Phi^{-1}(\mathbf{R}) = \{\pm\mathbf{q}\} \subseteq S^3$.

How to fix this? To ensure consistency under this double covering, all black-box components should satisfy the symmetry condition

$$\mathbf{J}_\theta(\mathbf{q}) = \mathbf{J}_\theta(-\mathbf{q}), \quad U_\theta(\mathbf{x}, \mathbf{q}) = U_\theta(\mathbf{x}, -\mathbf{q}), \quad F_\theta^\pm(\mathbf{x}, \mathbf{q}, \mathbf{u}) = F_\theta^\pm(\mathbf{x}, -\mathbf{q}, \mathbf{u}).$$

where $F_\theta^\pm(\mathbf{x}, \mathbf{q}, \mathbf{u})$ are the models used to approximate the discrete forces.

$$\hat{\mathbf{J}}_\theta(\mathbf{q}) = \frac{1}{2}(\mathbf{J}_\theta(\mathbf{q}) + \mathbf{J}_\theta(-\mathbf{q}))$$
$$\hat{U}_\theta(\mathbf{x}, \mathbf{q}) = \frac{1}{2}(U_\theta(\mathbf{x}, \mathbf{q}) + U_\theta(\mathbf{x}, -\mathbf{q}))$$
$$\hat{F}_\theta^\pm(\mathbf{x}, \mathbf{q}, \mathbf{u}) = \frac{1}{2}(F_\theta^\pm(\mathbf{x}, \mathbf{q}, \mathbf{u}) + F_\theta^\pm(\mathbf{x}, -\mathbf{q}, \mathbf{u}))$$

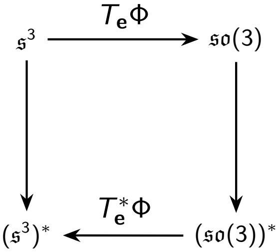
Relationship with FVIN on $\mathbf{SE}(3)$

Use the following mapping

$$(\mathbf{x}, \mathbf{R}, \dot{\mathbf{x}}, \dot{\mathbf{R}}) \mapsto (\mathbf{x}, \mathbf{q}, \mathbf{p}, \pi) = (\mathbf{x}, \mathbf{q}, \mathbf{M}_\theta \dot{\mathbf{x}}, 2\mathbf{J}_\theta (\mathbf{R}^\top \dot{\mathbf{R}})^\vee) \in T^*(\mathbb{R}^3 \times S^3)$$

since $\mathbf{p} = \mathbf{M}_\theta \dot{\mathbf{x}}$, $\pi = 4\mathbf{J}_\theta \xi$ and $2\xi = \omega = (\mathbf{R}^\top \dot{\mathbf{R}})^\vee$.

For the **left-trivialized discrete forces** $f_d^{\mathbf{q}\pm} \in \mathbb{R}^3 \cong (\mathfrak{s}^3)^*$, we also have $f_d^{\mathbf{R}\pm} = \frac{1}{2} f_d^{\mathbf{q}\pm} \in \mathbb{R}^3 \cong (\mathfrak{so}(3))^*$.



Outline

Ground-Truth Dynamics

$$\ddot{\theta} = -15 \sin\theta + 3u$$

$$\tau = g(\theta)u, \quad g(\theta) = 1, \quad m = \frac{1}{3}, \quad U(\theta) = 5(1 - \cos\theta)$$

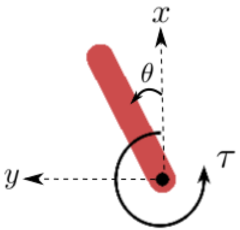
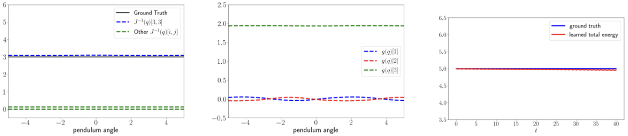
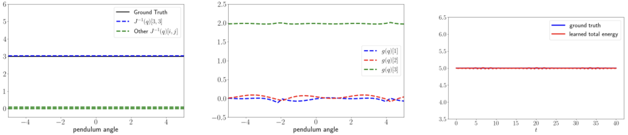


Figure: The inverted pendulum swingup problem in the OpenAI Gymnasium.

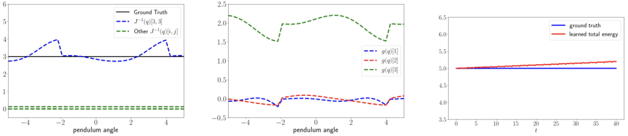
Experiment Results



(a) Inertia matrix of (1) (b) Control gain of (1) (c) Energy conservation of (1)



(d) Inertia matrix of (2) (e) Control gain of (2) (f) Energy conservation of (2)



(g) Inertia matrix of (3) (h) Control gain of (3) (i) Energy conservation of (3)

Figure: (1) Sign-invariance; (2) Fixed inertia matrix $\mathbf{J}_\theta$; (3) Plain.

Training/Testing Loss

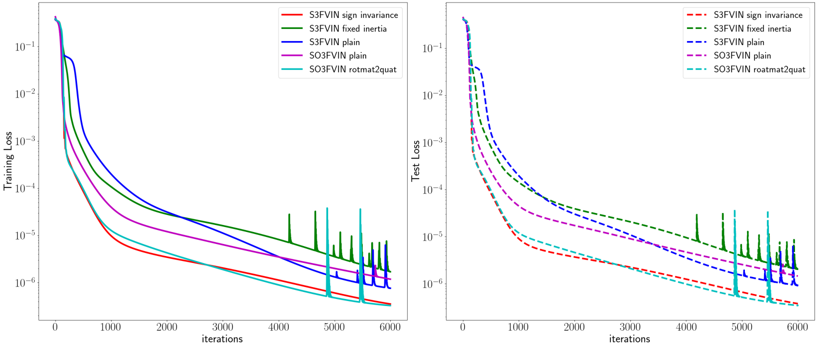


Figure: Loss curves for different variants of **S3FVIN** and **SO3FVIN**. The sign-invariant **S3FVIN** converges fastest and most stably. **SO3FVIN** can achieve comparable performance by parameterizing each black-box component using the transformation $\mathbf{R} \mapsto \mathbf{q}$, but exhibits less stable training.

Outline

Conclusion

Key takeaways from this project:

1. Explicitly enforcing **sign invariance** is a simple yet essential inductive bias for obtaining physically plausible results for **S3FVIN**;
2. Working directly with unit quaternions generally leads to **faster and more stable training** than using 3×3 rotation matrices, largely due to the **more compact representation**.