A PROOFS

Note that $\boldsymbol{h}$ is identical for both real and learned systems. Unless otherwise specified, we omit $\boldsymbol{h}$ in the following discussions.

A.1 PROOF OF PROPOSITION 3.1

According to the constraints of the initial conditions, the initial position $\boldsymbol{q}_\theta^{(t_0)}$ is $E(3)$-equivariant and the initial velocity $\dot{\boldsymbol{q}}_\theta^{(t_0)}$ is $O(3)$-eqivariant and translation-invariant. Given that $\mathcal{G}_1$ and back-bone GNN f_θ are $O(3)$-equivariant and translation-invariant, for any translation vector $\boldsymbol{b} \in \mathbb{R}^3$, orthogonal matrix $\boldsymbol{A} \in \mathbb{R}^{3\times 3}$ and $t' \in [0, \Delta t]$, we have

$$
\begin{aligned}
\Psi_{t',g_\theta,f_\theta}(\boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{b}) &= g_\theta(\boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{b}) + \mathcal{G}_1(f_\theta(\boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{b}), t') \\
&= \boldsymbol{A}g_\theta(\boldsymbol{q}_\theta^{(t_0)}) + \mathcal{G}_1(\boldsymbol{A}f_\theta(\boldsymbol{q}_\theta^{(t)}), t') \\
&= \boldsymbol{A}\dot{\boldsymbol{q}}_\theta^{(t_0)} + \boldsymbol{A}\mathcal{G}_1(f_\theta(\boldsymbol{q}_\theta^{(t_0)}), t') \\
&= \boldsymbol{A}\dot{\boldsymbol{q}}_\theta^{(t_0+t')} = \boldsymbol{A}\Psi_{\Delta t,g_\theta,f_\theta}(\boldsymbol{q}_\theta^{(t_0)}).
\end{aligned}
\tag{17}
$$

Thus, $\Psi_{t',g_\theta,f_\theta}(\boldsymbol{q}_\theta^{(t_0)})$ is proved to be $O(3)$-equivariant and translation-invariant. Then, for $\boldsymbol{q}_\theta^{(t_0+t')}$, we have

$$
\begin{aligned}
\Phi_{t',g_\theta}(\boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{b}) &= \boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{b} + \mathcal{G}_2(\Psi_{t',g_\theta,f_\theta}(\boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{b}), t') \\
&= \boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \mathcal{G}_2(\boldsymbol{A}\Psi_{t',g_\theta,f_\theta}(\boldsymbol{q}_\theta^{(t_0)}), t') + \boldsymbol{b} \\
&= \boldsymbol{A}\boldsymbol{q}_\theta^{(t_0)} + \boldsymbol{A}\mathcal{G}_2(\Psi_{t',g_\theta,f_\theta}(\boldsymbol{q}_\theta^{(t_0)}), t') + \boldsymbol{b} \\
&= \boldsymbol{A}\boldsymbol{q}_\theta^{(t_0+t')} + \boldsymbol{b}.
\end{aligned}
\tag{18}
$$

Therefore, $\boldsymbol{q}_\theta^{(t_0+t')} = \Phi_{t',g_\theta}(\boldsymbol{q}_\theta^{(t_0)})$ is E(3)-equivariant. Note that the composition of $E(3)$-equivariant functions is still $E(3)$-equivariant. For any time $t \in [t_0, t_1]$, $\boldsymbol{q}_\theta^{(t)}$, which is generated via iteratively calling integrator Φ_{t',g_θ} (Eq. 11) with suitable $t' \in [0, \Delta t]$, is $E(3)$-equivariant. Overall, the approximated trajectory $\boldsymbol{q}_\theta$ is $E(3)$-equivariant. For example, if the backbone is EGNN, the message passing is defined by

$$
\boldsymbol{m}_{ij}^{(t)} = \phi_e\left(\|\boldsymbol{q}_{\theta,i}^{(t)} - \boldsymbol{q}_{\theta,j}^{(t)}\|^2, \boldsymbol{h}_i, \boldsymbol{h}_j, a_{ij}\right), \quad \ddot{\boldsymbol{q}}_{\theta,i}^{(t)} = \frac{1}{N-1}\sum_{j\in\mathcal{N}_i}(\boldsymbol{q}_{\theta,i}^{(t)} - \boldsymbol{q}_{\theta,j}^{(t)})\phi_q(\boldsymbol{m}_{ij}^{(t)}).
$$
$$
\boldsymbol{h}_i^{(t+\Delta t)} = \boldsymbol{h}_i^{(t)} + \phi_h(\boldsymbol{h}_i^{(t)}, \sum_{j\in\mathcal{N}_i}\boldsymbol{m}_{ij}^{(t)}).
\tag{19}
$$

Here ϕ_e, ϕ_x denotes Multi-Layer Perceptrons (MLP) whose output is a scalar and the output of ϕ_h is a vector. The non-geometric features are updated via skip connections. Analogous to neural ODE methods, the model parameters are shared among all iterations.

Specifically, most widely used numerical solvers for motion equations are E(3) equivariant. For example, a symplectic Euler integrator computes

$$
\boldsymbol{q}^{(t+\Delta t)} = \boldsymbol{q}^{(t)} + \dot{\boldsymbol{q}}^{(t+\Delta t)}\Delta t, \; \dot{\boldsymbol{q}}^{(t+\Delta t)} = \dot{\boldsymbol{q}}^{(t)} + \ddot{\boldsymbol{q}}^{(t)}\Delta t.
\tag{20}
$$

It is straightforward to show that

$$
\boldsymbol{A}\dot{\boldsymbol{q}}^{(t)} + \boldsymbol{A}\ddot{\boldsymbol{q}}^{(t+\Delta t)}\Delta t = \boldsymbol{A}\dot{\boldsymbol{q}}^{(t+\Delta t)}, \; \boldsymbol{A}\boldsymbol{q}^{(t)} + \boldsymbol{b} + \boldsymbol{A}\dot{\boldsymbol{q}}^{(t+\Delta t)}\Delta t = \boldsymbol{A}\boldsymbol{q}^{(t+\Delta t)} + \boldsymbol{b}.
\tag{21}
$$

This property also holds for Velocity Verlet

$$
\boldsymbol{q}^{(t+\Delta t)} = \boldsymbol{q}^{(t)} + \dot{\boldsymbol{q}}^{(t)}\Delta t + \frac{1}{2}\ddot{\boldsymbol{q}}^{(t)}\Delta t^2, \; \dot{\boldsymbol{q}}^{(t+\Delta t)} = \dot{\boldsymbol{q}}^{(t)} + \frac{1}{2}(\ddot{\boldsymbol{q}}^{(t)} + \ddot{\boldsymbol{q}}^{(t+\Delta t)})\Delta t,
\tag{22}
$$

and Leapfrog

$$
\boldsymbol{q}^{(t+\Delta t)} = \boldsymbol{q}^{(t)} + \dot{\boldsymbol{q}}^{(t)}\Delta t + \frac{1}{2}\ddot{\boldsymbol{q}}^{(t)}\Delta t^2, \; \dot{\boldsymbol{q}}^{(t+\Delta t)} = \dot{\boldsymbol{q}}^{(t)} + \ddot{\boldsymbol{q}}^{(t+\Delta t)}\Delta t.
\tag{23}
$$

The proof is the same as symplectic Euler.

A.2 PROOF OF THEOREM 4.1

We first prove the uniqueness of the solution in our dynamic system. The key to this proof is to convert the high-order non-linear ODE to a first-order non-linear ODE. Let's define a new variable $\boldsymbol{Q}^{(t)} = (\dot{\boldsymbol{q}}^{(t)}, \boldsymbol{q}^{(t)})$ and $\dot{\boldsymbol{Q}}^{(t)}$ as its first-order derivative with respect t. Then in terms of this new variable, the second-order non-linear ODE becomes

$$\dot{\boldsymbol{Q}}^{(t)} = (F(\boldsymbol{Q}^{(t)}), G(\boldsymbol{Q}^{(t)})), \tag{24}$$

where $F(\boldsymbol{Q}^{(t)})) = f(\boldsymbol{q}^{(t)})$ and $G(\boldsymbol{Q}^{(t)}) = \dot{\boldsymbol{q}}^{(t)}$. Furthermore, let's define $H(\boldsymbol{Q}^{(t)}) = (F(\boldsymbol{Q}^{(t)}), G(\boldsymbol{Q}^{(t)}))$, and Eq. 24 is reformatted as

$$\dot{\boldsymbol{Q}}^{(t)} = H(\boldsymbol{Q}^{(t)}), \quad \boldsymbol{Q}^{(t_0)} = (\dot{\boldsymbol{q}}^{(t_0)}, \boldsymbol{q}^{(t_0)}), \tag{25}$$

which is exactly a first-order non-linear ODE with a known initial condition. Note that G and f are both Lipschitz continuous, thus H is Lipschitz continuous. Then, based on Picard's existence theorem (Coddington et al., 1956), the aforementioned non-linear ODE has a unique solution $\boldsymbol{Q}^{(t)}$, $\forall t \in [t_0, t_1]$. Subsequently, our system has a unique solution $\boldsymbol{q}^{(t)}$, $\forall t \in [t_0, t_1]$.

A.3 PROOF OF PROPOSITION 4.2

Due to the uniqueness of the solution, if the realistic measurement $\boldsymbol{q}^{(t_1)}$ is given, then the trajectory $\boldsymbol{q}^{(t)}$ is fully determined over the time interval $[t_0, t_1]$. Under SEGNO framework, without loss of generality, we define the discrepancy between $\boldsymbol{q}_\theta^{(t_1)}$ and $\boldsymbol{q}^{(t_1)}$ as

$$d(\boldsymbol{q}_\theta^{(t_1)}, \boldsymbol{q}^{(t_1)}) = ||\boldsymbol{q}_\theta^{(t_1)} - \boldsymbol{q}^{(t_1)}||_p = ||(\Phi_{\Delta t, g_\theta})^\tau (\boldsymbol{q}^{(t_0)}) - \phi_{T,g}(\boldsymbol{q}^{(t_0)})||_p, \tag{26}$$

where $|| \cdot ||_p$ represents the p-norm. According to Eq. 9, $\Phi_{\Delta t, g_\theta}$ is fully determined by f_θ. To minimize $d(\boldsymbol{q}_\theta^{(t_1)}, \boldsymbol{q}^{(t_1)})$, our neural ODE update scheme requires $f_\theta(\boldsymbol{q}_\theta^{(t_0+k\Delta t)})$ accurately approximate $f(\boldsymbol{q}^{(t_0+k\Delta t)})$ for $k = 0, \cdots, \tau - 1$. Then, under the assumption of universal approximation of neural networks (Hornik et al., 1989) and with sufficiently small Δt, there exists a $f_{\theta*}$ such that $f_{\theta*}(\boldsymbol{q}_{\theta*}^{(t)}, \boldsymbol{h}) = f(\boldsymbol{q}^{(t)}, \boldsymbol{h})$, $\forall t \in [t_0, t_1]$.

A.4 PROOF OF THEOREM 4.3

The proof sketch is as follows: We first revisit the core definitions pertaining to neural ODE in SEGNO and introduce its variant with Euler integrator, then derive the bound for first-order approximation error $||g_\theta - g||_\infty$, and finally extend the results to the second-order case $||f_\theta - f||_\infty$ to finish the proof.

A.4.1 EULER INTEGRATOR IN SEGNO

As mentioned in Eq. 10, the Euler integrator $\Phi_{\Delta t, g_\theta}$, as mentioned in Eq. 20, approaches $\phi_{\Delta t, g_\theta}$ via

$$\begin{aligned} g_\theta(\boldsymbol{q}_\theta^{(t+\Delta t)}) &= g_\theta(\boldsymbol{q}_\theta^{(t)}) + f_\theta(\boldsymbol{q}_\theta^{(t)})\Delta t, \\ \Phi_{\Delta t, g_\theta}(\boldsymbol{q}_\theta^{(t)}) &= \boldsymbol{q}_\theta^{(t)} + g_\theta(\boldsymbol{q}_\theta^{(t+\Delta t)})\Delta t. \end{aligned} \tag{27}$$

Considering Eq. 11 and 12, it is straightforward to reframe our learning objective in the context of a neural ODE:

$$\mathcal{L}_{\text{train}} = \sum_{s \in \mathcal{D}_{\text{train}}} ||\boldsymbol{q}_{\theta,s}^{(t_1)} - \boldsymbol{q}_s^{(t_1)}||^2 = \sum_{s \in \mathcal{D}_{\text{train}}} ||(\Phi_{\Delta t, g_\theta})^\tau (\boldsymbol{q}_s^{(t_0)}) - \phi_{T,g}(\boldsymbol{q}_s^{(t_0)})||^2. \tag{28}$$

A.4.2 APPROXIMATION ERROR OF f_θ

In our dynamical system, g_θ and g are entirely determined by f_θ and f respectively. Thus, we first establish the boundedness of $||g_\theta - g||_\infty$, then demonstrate the approximation error of $||f_\theta - f||_\infty$.

Lemma A.1. *For $\boldsymbol{q}^{(t_0)} \in \mathbb{R}^3$ as the initial position of a trajectory, $r_a, r_b, T, \tau > 0$ and $k \in \mathbb{Z}^+$, a given ODE solver that is τ compositions of an Euler Integrator $\Phi_{\Delta t, g_\theta}$ with $\Delta t = T/\tau$, a set of points on exact trajectory associated with time increment interval $[k\Delta t, T + k\Delta t]$ as*

$$V_{k\Delta t}^{T+k\Delta t} = \{\phi_{t',g}(\boldsymbol{q}^{(t_0)})|k\Delta t \le t' \le T + k\Delta t\}, \tag{29}$$

, and denote

$$\mathcal{L}_{k\Delta t}^{T+k\Delta t} = \frac{1}{T}||(\Phi_{\Delta t,g_\theta})^\tau - \phi_{T,g}||_{\mathcal{B}(V_{k\Delta t}^{T+k\Delta t}, r_a)}, \tag{30}$$

and suppose that g and g_θ are analytic and bounded by m within $\mathcal{B}(V_{k\Delta t}^{T+k\Delta t}, r_a + r_b)$, the union of complex balls centered at $\boldsymbol{q} \in V_{k\Delta t}^{T+k\Delta t}$ with radius $r_a + r_b$. Then, there exist constants T_0 and C that depends on r_a/m, r_b/m, τ, and $\Phi_{\Delta t, g_\theta}$, such that, if $0 < T < T_0$, $\forall t \in [t_0 + k\Delta t, t_0 + T + k\Delta t]$,

$$||g_\theta(\boldsymbol{q}^{(t)}) - g(\boldsymbol{q}^{(t)})||_\infty \le Cm\Delta t + \frac{e}{e-1}\mathcal{L}_{k\Delta t}^{T+k\Delta t}, \tag{31}$$

where e is the base of the natural logarithm.

Proof. This result can be directly derived from Theorem 3.1, 3.2 and Corollary 3.3 in (Zhu et al., 2022) via replacing the integrator from Runge-Kutta integrator with Euler integrator since Euler integrator has been proven to satisfy the prerequisites of theorems in Appendix B.2 in (Zhu et al., 2022). □

Since $\mathcal{B}(V_0^T, r_1) \subset \mathcal{B}(V_0^{2T}, r_1)$, per our assumption, then g and g_θ are both analytic and bounded by m_1 in $\mathcal{B}(V_0^T, r_1)$. To utlize Lemma A.1, we set $k = 0$ and $T = t_1 - t_0$. With suitable r_1 and m_1, we have $T < T_0$ and, $\forall t \in [t_0, t_1]$,

$$||g_\theta(\boldsymbol{q}^{(t)}) - g(\boldsymbol{q}^{(t)})||_\infty \le C_1 m_1 \Delta t + \frac{e}{e-1}\mathcal{L}_0^T, \tag{32}$$

where $\mathcal{L}_0^T = \frac{1}{T}||(\Phi_{\Delta t,g_\theta})^\tau - \phi_{T,g}||_{\mathcal{B}(V_0^T, r_1)}$ and C_1 is a control constant.

To establish the connection between g and f, we only focus on the first time step Δt instead of the entire time interval T. Recall that g is obtained by integrating f, we have

$$\psi_{\Delta t,g,f}(\boldsymbol{q}^{(t_0)}) = g(\boldsymbol{q}^{(t_0+\Delta t)}) = g(\boldsymbol{q}^{(t_0)}) + \int_{t_0}^{t_0+\Delta t} f(\boldsymbol{q}^{(t)})\, dt, \tag{33}$$

as the flow map for the first-order derivative of the exact trajectory in a single step. As defined in Eq. 27, its corresponding Euler integrator has the form

$$g_\theta(\boldsymbol{q}_\theta^{(t_0+\Delta t)}) = \Psi_{\Delta t,g_\theta,f_\theta}(\boldsymbol{q}^{(t_0)}) = g_\theta(\boldsymbol{q}_\theta^{(t_0)}) + f_\theta(\boldsymbol{q}_\theta^{(t_0)})\Delta t. \tag{34}$$

Note that $\mathcal{B}(V_0^T, r_2) \subset \mathcal{B}(V_0^{2T}, r_2)$, f and f_θ are both analytic and bounded by m_2 in $\mathcal{B}(V_0^T, r_2)$. In lemma A.1, we substitute $\phi_{\Delta t,g}$ by $\Psi_{\Delta t,g_\theta,f_\theta}$ and set $T = \Delta t$ with $\tau = 1$ and $k = 0$, then the corresponding loss becomes

$$
\begin{aligned}
\mathcal{L}_{0,1}^T &= \frac{1}{\Delta t}||\Psi_{\Delta t,g_\theta,f_\theta} - \psi_{\Delta t,g,f}||_{\mathcal{B}(V_0^{\Delta t}, r_2)} \\
&= \sup_{\boldsymbol{q} \in \mathcal{B}(V_0^{\Delta t}, r_2)} \frac{1}{\Delta t}||\Psi_{\Delta t,g_\theta,f_\theta}(\boldsymbol{q}) - \psi_{\Delta t,g,f}(\boldsymbol{q})||_\infty \\
&= \frac{1}{\Delta t}||\Psi_{\Delta t,g_\theta,f_\theta}(\boldsymbol{q}^*) - \psi_{\Delta t,g,f}(\boldsymbol{q}^*)||_\infty \\
&= \frac{1}{\Delta t}||g_\theta(\tilde{\boldsymbol{q}}_\theta^{(t_0+\Delta t)}) - g(\tilde{\boldsymbol{q}}^{(t_0+\Delta t)})||_\infty,
\end{aligned}
\tag{35}
$$

where $\boldsymbol{q}^*$ denote the point that maximizes the Eq. 35 and $\tilde{\boldsymbol{q}}^{(t)}$ is another trajectory with $\tilde{\boldsymbol{q}}^{(t_0)} = \boldsymbol{q}^*$ because $\boldsymbol{q}^* \in V_0^{\Delta t}$ may not holds. For both $\tilde{\boldsymbol{q}}^{(t)}$ and $\boldsymbol{q}^t$, the actual position and estimation from SEGNO at $t_0 + \Delta t$ have the form

$$
\begin{aligned}
\tilde{\boldsymbol{q}}^{(t_0+\Delta t)} &= \tilde{\boldsymbol{q}}^{(t_0)} + \int_{t_0}^{t_0+\Delta t} g(\tilde{\boldsymbol{q}}^{(t)})\, dt, \\
\tilde{\boldsymbol{q}}_\theta^{(t_0+\Delta t)} &= \tilde{\boldsymbol{q}}^{(t_0)} + g_\theta(\tilde{\boldsymbol{q}}^{(t_0)})\Delta t + f_\theta(\tilde{\boldsymbol{q}}^{(t_0)})\Delta t^2, \\
\boldsymbol{q}_\theta^{(t_0+\Delta t)} &= \boldsymbol{q}^{(t_0)} + g_\theta(\boldsymbol{q}^{(t_0)})\Delta t + f_\theta(\boldsymbol{q}^{(t_0)})\Delta t^2.
\end{aligned}
\tag{36}
$$

Then, with suitable Δt and r_1, $g(\tilde{\boldsymbol{q}}^{(t_0+\Delta t)})$, $g_\theta(\tilde{\boldsymbol{q}}_\theta^{(t_0+\Delta t)})$ and $g_\theta(\boldsymbol{q}_\theta^{(t_0+\Delta t)})$ are bounded by m_1 since $\tilde{\boldsymbol{q}}^{(t_0+\Delta t)}, \tilde{\boldsymbol{q}}_\theta^{(t_0+\Delta t)}, \boldsymbol{q}_\theta^{(t_0+\Delta t)} \in \mathcal{B}(V_T, r_1)$. Given such, the aforementioned loss is transformed into

$$
\begin{aligned}
\mathcal{L}_{0,1}^T &\leq \frac{1}{\Delta t} \big[||g_\theta(\tilde{\boldsymbol{q}}_\theta^{(t_0+\Delta t)}) - g_\theta(\tilde{\boldsymbol{q}}^{(t_0+\Delta t)})||_\infty + ||g_\theta(\tilde{\boldsymbol{q}}^{(t_0+\Delta t)}) - g_\theta(\boldsymbol{q}^{(t_0+\Delta t)})||_\infty \\
&\quad + ||g_\theta(\boldsymbol{q}^{(t_0+\Delta t)}) - g(\boldsymbol{q}^{(t_0+\Delta t)})||_\infty + ||g(\boldsymbol{q}^{(t_0+\Delta t)}) - g(\tilde{\boldsymbol{q}}^{(t_0+\Delta t)})||_\infty \big] \\
&\leq \frac{1}{\Delta t}(6m_1 + C_1 m_1 \Delta t + \frac{e}{e-1}\mathcal{L}_0^T).
\end{aligned}
\tag{37}
$$

Subsequently, given that $\Delta t < T < T_0$, we have, $\forall t \in [t_0, t_0 + \Delta t]$,

$$
\begin{aligned}
||f_\theta(\boldsymbol{q}^{(t)}) - f(\boldsymbol{q}^{(t)})||_\infty &\leq C_2 m_2 \Delta t + \frac{e}{e-1}\mathcal{L}_{0,1}^T, \\
&\leq C_2 m_2 \Delta t + \frac{e}{e-1}\big[\frac{1}{\Delta t}(6m_1 + C_1 m_1 \Delta t + \frac{e}{e-1}\mathcal{L}_0^T)\big], \\
&\leq C_2 m_2 \Delta t + (\frac{e}{e-1})^2 \frac{\mathcal{L}_0^T}{\Delta t} + \frac{e}{e-1}(\frac{6m_1}{\Delta t} + C_1 m_1), \\
&\leq O(\Delta t + \frac{\mathcal{L}_0^T}{\Delta t}).
\end{aligned}
\tag{38}
$$

To extend the boundness up to t_1, we can easily utilize Lemma A.1 with different $k = 1, \cdots, \tau-1$ (Eq. 32) to repeat the above derivation, and obtain, $\forall t \in [t_0 + k\Delta t, t_0 + (k+1)\Delta t]$,

$$
||f_\theta(\boldsymbol{q}^{(t)}) - f(\boldsymbol{q}^{(t)})||_\infty \leq O(\Delta t + \frac{\mathcal{L}_{k\Delta t}^{T+k\Delta t}}{\Delta t}),
\tag{39}
$$

where $\mathcal{L}_{k\Delta t}^{T+k\Delta t} = \frac{1}{T}||(\Phi_{\Delta t, g_\theta})^\tau - \phi_{T,g}||_{\mathcal{B}(V_{k\Delta t}^{T+k\Delta t}, r_1)}$. Therefore, $\forall t \in [t_0, t_1]$,

$$
\begin{aligned}
||f_\theta(\boldsymbol{q}^{(t)}) - f(\boldsymbol{q}^{(t)})||_\infty &\leq \sup_{k=0,\cdots,\tau-1} O(\Delta t + \frac{\mathcal{L}_{k\Delta t}^{T+k\Delta t}}{\Delta t}) \\
&\leq O(\Delta t + \frac{\mathcal{L}_0^{2T}}{\Delta t}),
\end{aligned}
\tag{40}
$$

where $\mathcal{L}_0^{2T} = \frac{1}{T}||(\Phi_{\Delta t, g_\theta})^\tau - \phi_{T,g}||_{\mathcal{B}(V_0^{2T}, r_1)}$ and it concludes the proof.

A.5 PROOF OF COROLLARY 4.4

A.5.1 LOCAL TRUNCATION ERROR

We use Taylor series expansion to approximate the trajectory at time $t + \Delta t$ as

$$
\boldsymbol{q}^{(t+\Delta t)} = \boldsymbol{q}^{(t)} + g(\boldsymbol{q}^{(t)})\Delta t + \frac{1}{2}f(\boldsymbol{q}^{(t)})\Delta t^2 + O(\Delta t^3).
\tag{41}
$$

Then the local truncation error $\epsilon_{t+\Delta t}$, $\forall t \in [t_0, t_1]$, equals to

$$
\begin{aligned}
\epsilon_{t+\Delta t} &= ||\boldsymbol{q}^{(t+\Delta t)} - \boldsymbol{q}^{(t)} - g(\boldsymbol{q}^{(t)})\Delta t - f_\theta(\boldsymbol{q}^{(t)})\Delta t^2||_2 \\
&= ||\boldsymbol{q}^{(t)} + g(\boldsymbol{q}^{(t)})\Delta t + \frac{1}{2}f(\boldsymbol{q}^{(t)})\Delta t^2 + O(\Delta t^3) \\
&\quad - \boldsymbol{q}^{(t)} - g(\boldsymbol{q}^{(t)})\Delta t - f_\theta(\boldsymbol{q}^{(t)})\Delta t^2||_2 \\
&= ||\frac{1}{2}f(\boldsymbol{q}^{(t)})\Delta t^2 - f_\theta(\boldsymbol{q}^{(t)})\Delta t^2 + O(\Delta t^3)||_2 \\
&= ||\frac{1}{2}(f(\boldsymbol{q}^{(t)}) - f_\theta(\boldsymbol{q}^{(t)}))\Delta t^2 - \frac{1}{2}f_\theta(\boldsymbol{q}^{(t)})\Delta t^2 + O(\Delta t^3)||_2 \\
&\leq ||\frac{1}{2}(f(\boldsymbol{q}^{(t)}) - f_\theta(\boldsymbol{q}^{(t)}))\Delta t^2||_2 + ||\frac{1}{2}f_\theta(\boldsymbol{q}^{(t)})\Delta t^2||_2 + O(\Delta t^3) \\
&\leq \frac{\sqrt{3}\Delta t^2}{2}||f(\boldsymbol{q}^{(t)}) - f_\theta(\boldsymbol{q}^{(t)}))||_\infty + ||\frac{1}{2}m_2\Delta t^2||_2 + O(\Delta t^3) \\
&\leq \frac{\sqrt{3}\Delta t^2}{2}O(\Delta t + \frac{2\mathcal{L}_0^{2T}}{\Delta t}) + O(\Delta t^2).
\end{aligned}
\tag{42}
$$

16

Here the last inequality is due to Theorem 4.3. If the learned model achieves a near-perfect approximation of the true trajectory, resulting in a significantly diminished loss $\mathcal{L}_0^{2T}$. Then, $\forall t \in [t_0, t_1]$,

$$\epsilon_{t+\Delta t} \leq O(\Delta t^2). \tag{43}$$

A.5.2 GLOBAL TRUNCATION ERROR

Considering $t \in [t_0, t_1]$ and $k = 1, \cdots, \tau - 1$, the boundness of the global truncation error can be derived in a recursive way via

$$
\begin{aligned}
\mathcal{E}_{t+(k+1)\Delta t} &= ||\boldsymbol{q}^{(t+(k+1)\Delta t)} - \boldsymbol{q}_\theta^{(t+(k+1)\Delta t)}||_2 \\
&= ||\boldsymbol{q}^{(t+(k+1)\Delta t)} - \boldsymbol{q}_\theta^{(t+k\Delta t)} - g_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})\Delta t - f_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})\Delta t^2||_2 \\
&= ||\boldsymbol{q}^{(t+k\Delta t)} - \boldsymbol{q}_\theta^{(t+k\Delta t)} + \boldsymbol{q}^{(t+(k+1)\Delta t)} - \boldsymbol{q}^{(t+k\Delta t)} - g(\boldsymbol{q}^{(t+k\Delta t)})\Delta t \\
&\quad - f_\theta(\boldsymbol{q}^{(t+k\Delta t)})\Delta t^2 + f_\theta(\boldsymbol{q}^{(t+k\Delta t)})\Delta t^2 - f_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})\Delta t^2 \\
&\quad + g(\boldsymbol{q}^{(t+k\Delta t)})\Delta t - g_\theta(\boldsymbol{q}^{(t+k\Delta t)})\Delta t \\
&\quad + g_\theta(\boldsymbol{q}^{(t+k\Delta t)})\Delta t - g_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})\Delta t||_2 \\
&\leq \mathcal{E}_{t+k\Delta t} + \epsilon_{t+(k+1)\Delta t} + ||f_\theta(\boldsymbol{q}^{(t+k\Delta t)}) - f_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})||_2\Delta t^2 \\
&\quad + \left[||g_\theta(\boldsymbol{q}^{(t+k\Delta t)}) - g_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})||_2 + ||g(\boldsymbol{q}^{(t+k\Delta t)}) - g_\theta(\boldsymbol{q}^{(t+k\Delta t)})||_2\right]\Delta t.
\end{aligned} \tag{44}
$$

Note that g_θ and f_θ satisfy the Lipschitz continuity, we have

$$
\begin{aligned}
||g_\theta(\boldsymbol{q}^{(t+k\Delta t)}) - g_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})||_2 &\leq L_g||\boldsymbol{q}^{(t+k\Delta t)} - \boldsymbol{q}_\theta^{(t+k\Delta t)}||_2 = L_g||\mathcal{E}_{t+k\Delta t}||_2, \\
||f_\theta(\boldsymbol{q}^{(t+k\Delta t)}) - f_\theta(\boldsymbol{q}_\theta^{(t+k\Delta t)})||_2 &\leq L_f||\boldsymbol{q}^{(t+k\Delta t)} - \boldsymbol{q}_\theta^{(t+k\Delta t)}||_2 = L_f||\mathcal{E}_{t+k\Delta t}||_2,
\end{aligned} \tag{45}
$$

where L_g and L_f denote Lipschitz constant for g_θ and f_θ respectively. Given that the learned model achieves a near-perfect approximation of the true trajectory, by Lemma A.1 and Eq. 32, we obtain

$$||g(\boldsymbol{q}^{(t+k\Delta t)}) - g_\theta(\boldsymbol{q}^{(t+k\Delta t)})||_2 \leq \sqrt{3}\left[C_1 m_1 \Delta t + \frac{e}{e-1}\mathcal{L}_0^{2T}\right] \leq O(\Delta t). \tag{46}$$

Thus, the global truncation error in Eq. 44 is transformed into

$$\mathcal{E}_{t+(k+1)\Delta t} \leq (1 + L_g\Delta t + L_f\Delta t^2)\mathcal{E}_{t+k\Delta t} + O(\Delta t^2). \tag{47}$$

Note that the global truncation error for the first step $\mathcal{E}_{t+\Delta t}$, where $k = 1$, is exactly the local truncation error $\epsilon_{t+\Delta t}$ mentioned in Eq. 42. Given that $k \leq \tau$, we can derive, $\forall t \in [t_0, t_1]$ and $k = 1, \cdots, \tau$,

$$
\begin{aligned}
\mathcal{E}_{t+k\Delta t} &\leq (1 + L_g\Delta t + L_f\Delta t^2)^{k-1}O(\Delta t^2) + \cdots + (1 + L_g\Delta t + L_f\Delta t^2)O(\Delta t^2) \\
&\leq O(\tau\Delta t^2) = O(\frac{T}{\Delta t}\Delta t^2) = O(\Delta t).
\end{aligned} \tag{48}
$$

B IMPLEMENTATION DETAILS

B.1 MORE DETAILS ON SIMULATED N-BODY SYSTEMS

N-body charged system We use the same N-body charged system code[2] with previous work (Satorras et al., 2021; Brandstetter et al., 2021). They inherit the 2D implementation of (Kipf et al., 2018) and extend it to 3 dimensions. System trajectories are generated in 0.001 timestep and unbounded with virtual boxes. The initial location is sampled from a Gaussian distribution (mean $\mu = 0$, standard deviation $\sigma = 0.5$), and the initial velocity is a random vector of norm 0.5.

N-body gravity system The code[3] of gravitational N-body systems is provided by (Brandstetter et al., 2021). They implement it under the same framework as the above charged N-body systems. System trajectories are generated in 0.001 timestep, using gravitational interaction and no boundary conditions. Particle positions are initialized from a unit Gaussian, particle velocities are initialized with a norm equal to one, random direction, and particle mass is set to one.

---

[2]https://github.com/vgsatorras/egnn
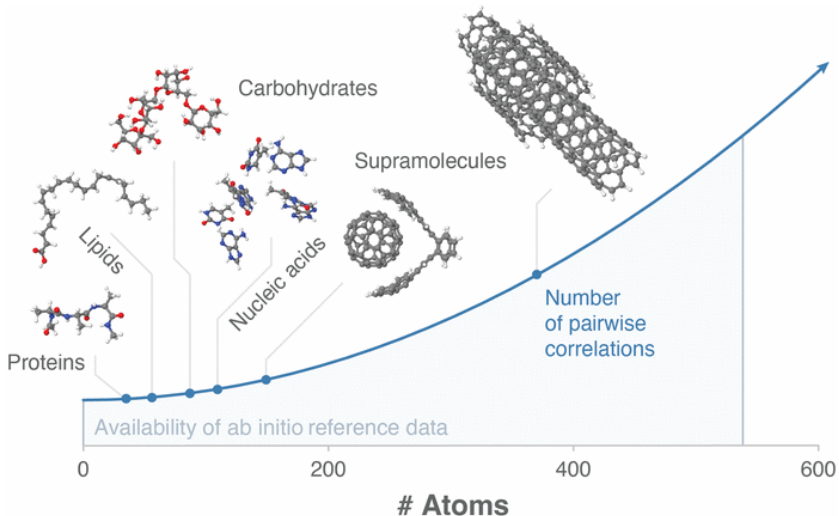[3]https://github.com/RobDHess/Steerable-E3-GNN

Figure 5: Molecular structures of MD22 dataset, which is borrowed from Fig.1 of original paper (Chmiela et al., 2023).

Hyperparameter We empirically find that the following hyperparameters generally work well, and use them across most experimental evaluations: Adam optimizer with learning rate 0.001, the number of epochs 500, hidden dim 64, weight decay 1×10^{-12}, and layer number 4. We set the iteration time of SEGNO to 8. The representation degrees of SE(3)-transformer and TFN are set to 3 and 2. The number of training, validation, and testing sets are 3000, 2000, and 2000, respectively.

B.2 MORE DETAILS ON MD22

Molecules and Hyperparameter The molecular structures of MD22 are displayed in Figure 5, which is borrowed from their paper (Chmiela et al., 2023). We use the following hyperparameters across all experimental evaluations: Adam optimizer with learning rate 0.0005, the number of epochs 500, hidden dim 64, weight decay 1×10^{-12}, and layer number 4. The iteration time of SEGNO is searched from 4 to 8. The representation degrees of SE(3)-transformer and TFN are set to 2. Due to the limited memory, the batch size of TFN is searched in 50, 30, 10 according to the size of molecules. The number of training, validation, and testing sets are 500, 2000, and 2000, respectively. The threshold for graph construction is set to 2.5 for all molecules.

B.3 MORE DETAILS ON MOTION CAPTURE

Hyperparameter We use the following hyperparameters across all experimental evaluations: Adam optimizer with learning rate 0.001, the number of epochs 3000, hidden dim 64, weight decay 1×10^{-12}, and layer number 4. The iteration time of SEGNO is set to 4. We adopt a random split strategy introduced by Huang et al. (2022) where train/validation/test data contains 200/600/600 frame pairs.

C ADDITIONAL EXPERIMENTS

C.1 ACCURACY OF LEARNED LATENT TRAJECTORY

It is interesting to see how models learn the latent trajectory between the input and output states. Accordingly, we train models on 1000ts on two datasets and make the test on shorter time steps by performing SEGNO on the smaller τ step with the same ratio. For the baselines, we treat the forward timestep of each hidden layer as the same and extract their object position information as the prediction. Table 5 reports the mean and standard deviation of each setting. From Table 5 we can observe that:

Table 5: The generalization from long-term to short-term. All models are trained on 1000ts and test on 250/500/750/1000 ts. Mean squared error ($\times 10^{-2}$) and the standard deviation are reported. Results averaged across 5 runs.

Method	Charged					Gravity				
	250 ts	500 ts	750 ts	1000 ts	Avg	250 ts	500 ts	750 ts	1000 ts	Avg
GNN	$73.40_{\pm 9.60}$	$31.79_{\pm 5.28}$	$12.86_{\pm 2.81}$	$0.826_{\pm 0.08}$	29.72	$181.9_{\pm 26.1}$	$90.33_{\pm 15.5}$	$30.66_{\pm 12.3}$	$0.746_{\pm 0.05}$	75.93
GDE	$92.65_{\pm 25.0}$	$43.94_{\pm 10.9}$	$12.20_{\pm 23.2}$	$0.652_{\pm 0.05}$	37.36	$136.0_{\pm 135}$	$56.80_{\pm 60.6}$	$12.21_{\pm 14.8}$	$0.588_{\pm 0.58}$	51.39
EGNN	$6.756_{\pm 3.05}$	$3.816_{\pm 4.68}$	$3.668_{\pm 0.74}$	$0.568_{\pm 0.09}$	3.702	$7.146_{\pm 7.06}$	$29.70_{\pm 20.4}$	$9.712_{\pm 3.60}$	$0.382_{\pm 0.11}$	11.89
GMN	$10.44_{\pm 7.43}$	$10.92_{\pm 4.67}$	$4.518_{\pm 1.36}$	$0.512_{\pm 0.16}$	6.598	$7.430_{\pm 8.19}$	$9.540_{\pm 12.1}$	$5.730_{\pm 6.69}$	$0.349_{\pm 0.48}$	5.762
SEGNN	$21.78_{\pm 9.69}$	$52.74_{\pm 15.6}$	$34.13_{\pm 14.9}$	$0.342_{\pm 0.04}$	27.25	$10.58_{\pm 6.28}$	$49.63_{\pm 37.7}$	$25.82_{\pm 27.0}$	$0.448_{\pm 0.02}$	21.62
SEGNO	$0.188_{\pm 0.03}$	$0.312_{\pm 0.06}$	$0.360_{\pm 0.06}$	$0.309_{\pm 0.11}$	0.292	$0.064_{\pm 0.02}$	$0.128_{\pm 0.03}$	$0.176_{\pm 0.04}$	$0.210_{\pm 0.07}$	0.145

Table 6: Comparsion ($\times 10^{-2}$) between SEGNO and GNS on simulated N-body systems. Results averaged across 3 runs.

Method	Charged			Gravity		
	1000 ts	1500 ts	2000 ts	1000 ts	1500 ts	2000 ts
GNS	$3.245_{\pm 0.068}$	$11.689_{\pm 0.330}$	$31.632_{\pm 0.206}$	$4.204_{\pm 0.081}$	$17.095_{\pm 0.136}$	$50.275_{\pm 0.201}$
SEGNO-avg.	$2.146_{\pm 0.079}$	$10.145_{\pm 0.034}$	$24.244_{\pm 0.212}$	$1.431_{\pm 0.047}$	$19.488_{\pm 0.978}$	$54.370_{\pm 1.385}$
SEGNO	$\mathbf{0.433}_{\pm 0.013}$	$\mathbf{1.858}_{\pm 0.029}$	$\mathbf{4.285}_{\pm 0.049}$	$\mathbf{0.338}_{\pm 0.027}$	$\mathbf{1.362}_{\pm 0.077}$	$\mathbf{4.017}_{\pm 0.087}$

- Clearly, SEGNO outperforms all other baselines across all settings by a large margin. Notably, when there is a lack of supervised signals at 250/500/750ts, the performance of all other baselines decreases significantly. By contrast, SEGNO achieves similar results as in 1000ts, demonstrating its robust generalization to short-term prediction.

- Another interesting point is that SEGNO's error exhibits a distinct trend compared to other baselines. While the errors of other baselines significantly increase with decreasing time steps, SEGNO achieves even smaller errors with shorter time steps. This observation justifies our theoretical results that the error is bound by the chosen timestep.

- Additionally, the standard deviation of SEGNO is much smaller than that of other baselines, indicating the numerical stability of SEGNO. This result further confirms our theoretical finding that SEGNO can obtain a better latent trajectory between two discrete states.

C.2   COMPARISON WITH GNS

We conducted additional evaluations of GNS and PINGO-avg., which are learned by minimizing average acceleration, on two N-body systems. The results of these evaluations are presented in Table 6. We can observe that SEGNO outperforms both GNS and PINGO-avg. In all cases, showing that training second-order neural odes on position loss outperforms training models on average acceleration.

C.3   ROLLOUT COMPARISON ON N-BODY SYSTEMS

We evaluate the generalizability of models for rollout simulation. Specifically, we train all models on 1000ts and use rollout to make the prediction for the longer time step (over 40 rollout steps, indicating over 40000ts.). Figure 6 depicts the mean squared error of all methods on two datasets. We can observe that all baselines experience numerical explosion due to error accumulation during the rollout process, leading to a quick drop in prediction performance. In contrast, SEGNO demonstrates an order-of-magnitude error improvement over other baselines. This numerical stability can be attributed to the Neural ODE framework for modeling position and velocity.

C.4   MORE RESULTS ON CMU MOTION CAPTURE

This section illustrates more visualizations of GMN and SEGNO on modeling object motions. From Figure 7, we can observe that PINGO is able to track the ground-truth trajectories accurately, which is consistent with the performance in Table 4.
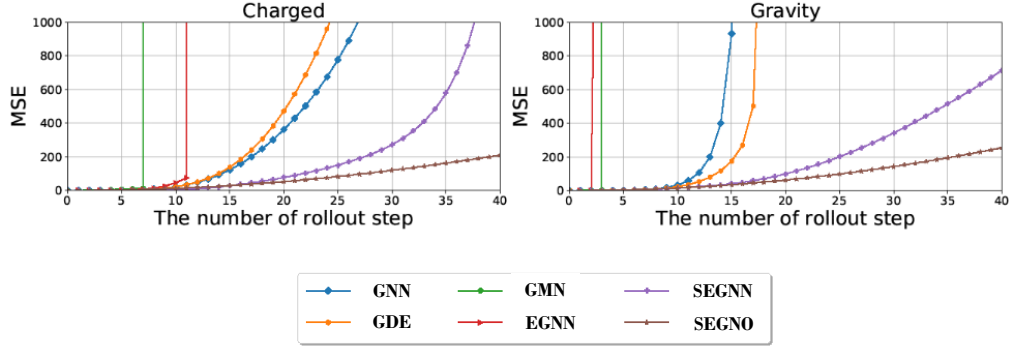
Figure 6: Mean squared error of rollout. Each rollout step is equal to 1000 ts. All models are trained on 1000ts.
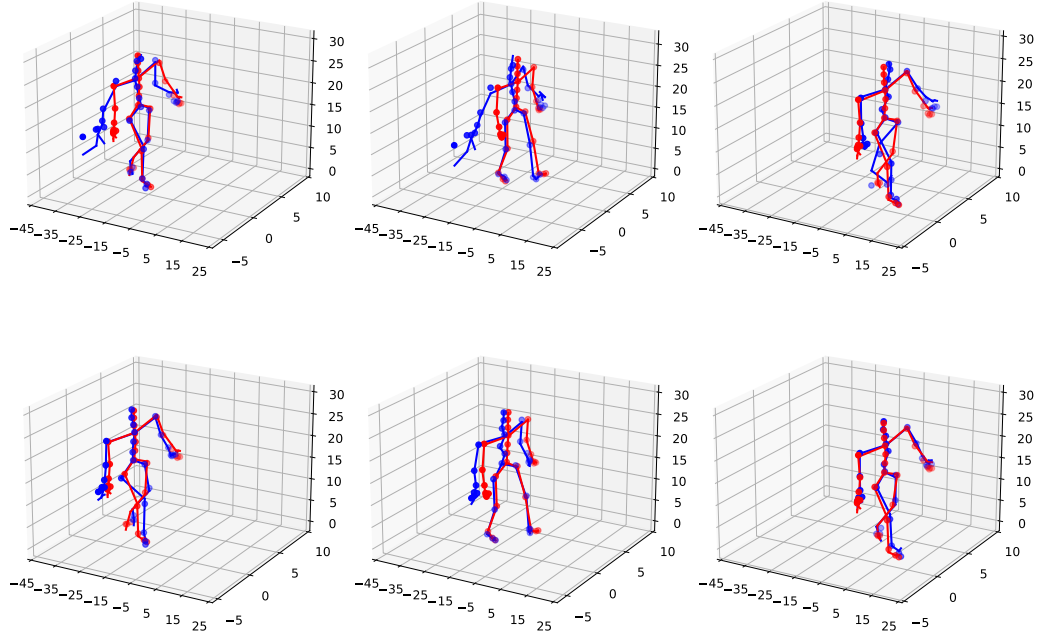

Figure 7: Example snapshots of Motion Capture with 50 time steps. *Top*: Prediction of GMN. *Bottom*: Prediction of SEGNO. Ground truth in red, and prediction in blue.