

837

# Appendix

838

## 7 Proof

839

Below is the explanation and proof of Proposition 4.1:

840

*Proof.* Since  $\epsilon_\theta$  is SE(3)-equivariant by assumption, we have for any  $h \in \text{SE}(3)$ ,

841

$$\epsilon_\theta(h \cdot \mathcal{G}') = h \cdot \epsilon_\theta(\mathcal{G}'), \quad \text{where } \mathcal{G}' = \mathbf{f}(\mathcal{G}_\tau, \mathcal{C}).$$

We consider each component:

842

- The coupling operator  $\mathbf{f}$  augments  $\mathcal{G}_\tau$  with control  $\mathcal{C}$  in a way that respects the SE(3) structure: global controls modify features invariantly; subgraph controls are merged geometrically; frame controls concatenate along the temporal axis. Thus,  $\mathbf{f}$  is SE(3)-equivariant.
- The decoupling operator  $\mathbf{g}$  selects a subset of nodes or frames without altering their coordinates. Therefore, it commutes with SE(3) action:  $\mathbf{g}(h \cdot \mathcal{G}'') = h \cdot \mathbf{g}(\mathcal{G}'')$ .

843

844

845

846

847

Combining the above, we explicitly see that for any  $h \in \text{SE}(3)$  defined as  $h(\mathbf{x}) = R\mathbf{x} + \mathbf{d}$ , we have:

848

$$\begin{aligned} & \mathbf{g} \circ \epsilon_\theta \circ \mathbf{f}(h \cdot (\mathcal{G}_\tau, \mathcal{C})) \\ &= \mathbf{g}(\epsilon_\theta(h \cdot \mathbf{f}(\mathcal{G}_\tau, \mathcal{C}))) \\ &= \mathbf{g}(h \cdot \epsilon_\theta(\mathbf{f}(\mathcal{G}_\tau, \mathcal{C}))) \\ &= h \cdot \mathbf{g}(\epsilon_\theta(\mathbf{f}(\mathcal{G}_\tau, \mathcal{C}))) \\ &= R \cdot (\mathbf{g} \circ \epsilon_\theta \circ \mathbf{f}(\mathcal{G}_\tau, \mathcal{C})) + \mathbf{d} \\ &= h \cdot (\mathbf{g} \circ \epsilon_\theta \circ \mathbf{f}(\mathcal{G}_\tau, \mathcal{C})) \end{aligned}$$

Therefore, the composed function  $\mathbf{g} \circ \epsilon_\theta \circ \mathbf{f}$  is SE(3)-equivariant.

849

850

□

## 8 More Details on Experiments

851

### 8.1 Hyper-parameters

852

We provide the detailed hyper-parameters of GeoAda in Table 8. We adopt Adam optimizer with betas (0.9, 0.999) and  $\epsilon = 10^{-8}$ . For all experiments, we use the linear noise schedule per [11] with  $\beta_{\text{start}} = 0.02$  and  $\beta_{\text{end}} = 0.0001$ .

853

854

855

**Table 8:** Hyper-parameters of GeoAda in the experiments.

	n_layer	hidden	time_emb_dim	$\mathcal{T}$	batch_size	learning_rate
N-body	6	128	32	1000	128	0.0001
MD	6	128	32	1000	128	0.0001
CMU Mocap	6	64	32	100	128	0.0001

### 8.2 Baselines

856

#### Full FT.

857

**Full FT** fully fine-tunes the pre-trained model  $f$  during downstream training. The entire model is updated to fit the target task.

858

859

## PARTIAL- $k$ .

**PARTIAL- $k$**  fine-tunes only the last  $k$  layers of the model  $f$ , while freezing the remaining layers. This method balances adaptability with parameter efficiency by limiting the number of updated layers.

## Graph Prompt Feature (GPF).

In **GPF**, the pre-trained encoder  $f$  is kept frozen, and a learnable prompt vector  $p$  is injected into the input feature space. During training, only the prompt vector  $p$  and the prediction head  $\theta$  are optimized. This method enables task adaptation through a lightweight, task-specific prompt without modifying the backbone model. In our implementation, we replace the original MLP head with a three-layer Equivariant Geometric Trajectory Network (EGTN), which ensures the projection head maintains geometric consistency with the model.

## Graph Prompt Feature-Plus (GPF-plus).

Extending GPF, this variant constructs node-wise prompt vectors using  $k$  learnable basis vectors  $p_1^b, \dots, p_k^b$  and a set of learnable linear weights  $a_1, \dots, a_k$ . These components are used to compute node-specific prompts  $p_i$  via a compositional mechanism. The model  $f$  remains frozen, while prediction head  $\theta$ , learnable basis vectors  $p_i^b$ , and learnable linear weights  $a_i$  are optimized.

## Prompt-Template

We prepend a learnable prompt layer (Equivariant Geometric Trajectory Network) to adapt new inputs to the distribution seen during pretraining, following with prediction head  $\theta$ .

## MLP- $k$ (EGTN- $k$ ).

This baseline freezes the entire pre-trained model  $f$  and replaces the prediction head with a  $k$ -layer multilayer perceptron (MLP). To preserve equivariance in our setting, we replace the MLP with an Equivariant Geometric Trajectory Network (EGTN) block. Only the EGTN-based head is trained during the downstream task.

## 8.3 Details of the datasets

### 8.3.1 Global Type Control

**Pretrain Dataset** The statistics of the pretrained datasets on Global Type Control are presented in Table. 9.

**Table 9:** Pretrain Dataset statistics by Global Type.

Type	Washwindow	Directing Traffic	Basketball Signal	Pretrain
train	12126	9557	7776	29459
val	1342	2346	1920	5588
test	1342	2346	1920	5588

**Downstream datasets** The statistics of the downstream datasets utilized for the models pretrained on Global Type Control are presented in Table. 10.

**Table 10:** Downstream Dataset statistics by Global Type.

Type	Running	Walking	Jumping	Basketball	Soccer
train	245	869	1345	1044	1210
val	47	145	1008	254	264
test	47	145	1008	254	264

## 8.4 Model

### 8.4.1 Geometric Trajectory Diffusion Models

**Unconditional Generation** For unconditional generation, we model the trajectory distribution subject to SE(3)-invariance. The following theorem provides constraints for the prior and transition kernel.

894 **Theorem 8.1.** *If the prior  $p_{\mathcal{T}}(\mathbf{x}_{\mathcal{T}}^{[T]})$  is SE(3)-invariant, and the transition kernels  $p_{\tau-1}(\mathbf{x}_{\tau-1}^{[T]} \mid$   
895  $\mathbf{x}_{\tau}^{[T]}), \forall \tau \in \{1, \dots, \mathcal{T}\}$  are SE(3)-equivariant, then the marginal  $p_{\tau}(\mathbf{x}_{\tau}^{[T]})$  at any step  $\tau \in$   
896  $\{0, \dots, \mathcal{T}\}$  is also SE(3)-invariant.*

897 **Prior in the translation-invariant subspace.** The prior is built on a translation-invariant subspace  
898  $\mathcal{X}_{\mathbf{P}} \subset \mathcal{X}$ , induced by a linear transformation  $\mathbf{P}$ :

$$\mathbf{P} = \mathbf{I}_D \otimes \left( \mathbf{I}_{TN} - \frac{1}{TN} \mathbf{1}_{TN} \mathbf{1}_{TN}^{\top} \right)$$

899 which results in a restricted Gaussian distribution supported only on the subspace, denoted  $\tilde{\mathcal{N}}(\mathbf{0}, \mathbf{I})$ ,  
900 and is isotropic and SO(3)-invariant. To sample, one samples from  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  and projects it onto the  
901 subspace.

902 **Transition kernel.** The transition kernel is parameterized in the subspace  $\mathcal{X}_{\mathbf{P}}$ , given by:

$$p_{\theta}(\tilde{\mathbf{x}}_{\tau-1}^{[T]} \mid \tilde{\mathbf{x}}_{\tau}^{[T]}) = \tilde{\mathcal{N}}(\tilde{\boldsymbol{\mu}}_{\theta}(\tilde{\mathbf{x}}_{\tau}^{[T]}, \tau), \sigma_{\tau}^2 \mathbf{I})$$

903 where the mean function  $\tilde{\boldsymbol{\mu}}_{\theta}$  is SO(3)-equivariant. The function is re-parameterized as:

$$\tilde{\boldsymbol{\mu}}_{\theta}(\tilde{\mathbf{x}}_{\tau}^{[T]}, \tau) = \frac{1}{\sqrt{\alpha_{\tau}}} \left( \tilde{\mathbf{x}}_{\tau}^{[T]} - \frac{\beta_{\tau}}{\sqrt{1 - \alpha_{\tau}}} \tilde{\boldsymbol{\epsilon}}_{\theta}(\mathbf{x}_{\tau}^{[T]}, \tau) \right)$$

904 where  $\tilde{\boldsymbol{\epsilon}}_{\theta} = P \circ \mathbf{f}_{\theta}$  is an SO(3)-equivariant adaptation of the proposed EGTN.

905 **Training and inference.** The VLB is optimized for training, with the objective:

$$\mathcal{L}_{\text{uncond}} := \mathbb{E}_{\mathbf{x}_0^{[T]}, \tilde{\boldsymbol{\epsilon}} \sim \tilde{\mathcal{N}}(\mathbf{0}, \mathbf{I}), \tau \sim \text{Unif}(1, \mathcal{T})} \left[ \|\tilde{\boldsymbol{\epsilon}} - \tilde{\boldsymbol{\epsilon}}_{\theta}(\tilde{\mathbf{x}}_{\tau}^{[T]}, \tau)\|^2 \right]$$

906 The inference process involves projecting intermediate samples onto the subspace  $\mathcal{X}_{\mathbf{P}}$ .

907 **Conditional Generation** In conditional generation, the target distribution is SE(3)-equivariant with  
908 respect to the given frames. The following theorem provides constraints for the prior and transition  
909 kernel.

910 **Theorem 8.2.** *If the prior  $p_{\mathcal{T}}(\mathbf{x}_{\mathcal{T}}^{[T]} \mid \mathbf{x}_c^{[T_c]})$  is SE(3)-equivariant, and the transition kernels  
911  $p_{\tau-1}(\mathbf{x}_{\tau-1}^{[T]} \mid \mathbf{x}_{\tau}^{[T]}, \mathbf{x}_c^{[T_c]}), \forall \tau \in \{1, \dots, \mathcal{T}\}$  are SE(3)-equivariant, the marginal  $p_{\tau}(\mathbf{x}_{\tau}^{[T]} \mid \mathbf{x}_c^{[T_c]}),$   
912  $\forall \tau \in \{0, \dots, \mathcal{T}\}$  is SE(3)-equivariant.*

913 **Flexible equivariant prior.** We provide a guideline for distinguishing feasible prior designs. The  
914 prior  $\mathcal{N}(\boldsymbol{\mu}(\mathbf{x}_c^{[T_c]}), \mathbf{I})$  is SE(3)-equivariant if  $\boldsymbol{\mu}(\mathbf{x}_c^{[T_c]})$  is SE(3)-equivariant. The mean function  
915  $\boldsymbol{\mu}(\mathbf{x}_c^{[T_c]})$  serves as an anchor to transition geometric information from the given frames to the target  
916 distribution. For instance, the anchor can be defined as:

$$\mathbf{x}_r^{[T]} = \mathbf{1}_{T \times N} \otimes \sum_{s \in [T_c]} w^{(s)} \bar{\mathbf{x}}_c^{(s)}$$

917 where the weights satisfy  $\sum_{s \in [T_c]} w^{(s)} = 1$ .

918 The weights  $\mathbf{w}^{(t,s)}$  are derived as:

$$\mathbf{W}_{t,s} = [\boldsymbol{\gamma} \otimes \hat{\mathbf{h}}_c^{[T_c]}]_{t,s} \in \mathbb{R}^N, \quad \mathbf{w}^{(t,s)} = \begin{cases} \mathbf{W}_{t,s} & \text{if } s < T_c - 1, \\ \mathbf{1}_N - \sum_{s=0}^{T_c-2} \mathbf{W}_{t,s} & \text{if } s = T_c - 1. \end{cases}$$

919 where  $\boldsymbol{\gamma} \in \mathbb{R}^T$  are learnable parameters, ensuring the constraint for translation equivariance.

920 **Transition kernel.** To match the proposed prior, we modify both the forward and reverse processes.  
921 The forward process is defined as:

$$q(\mathbf{x}_{\tau}^{[T]} \mid \mathbf{x}_{\tau-1}^{[T]}, \mathbf{x}_c^{[T_c]}) := \mathcal{N}(\mathbf{x}_{\tau}^{[T]}; \mathbf{x}_r + \sqrt{1 - \beta_{\tau}}(\mathbf{x}_{\tau-1}^{[T]} - \mathbf{x}_r), \beta_{\tau} \mathbf{I}),$$

922 which ensures that  $q(\mathbf{x}_{\tau}^{[T]} \mid \mathbf{x}_c^{[T_c]})$  matches the equivariant prior  $\mathbf{x}_r$  (proof in App. ??). The reverse  
923 transition kernel is:

$$p_{\tau-1}(\mathbf{x}_{\tau-1}^{[T]} \mid \mathbf{x}_{\tau}^{[T]}, \mathbf{x}_c^{[T_c]}) = \mathcal{N}(\boldsymbol{\mu}_{\theta}(\mathbf{x}_{\tau}^{[T]}, \tau, \mathbf{x}_c^{[T_c]}), \sigma_{\tau}^2 \mathbf{I}).$$

We adopt the noise prediction objective for the reverse process, rewriting  $\mu_\theta$  as:

$$\mu_\theta(\mathbf{x}_\tau^{[T]}, \mathbf{x}_c^{[T_c]}, \tau) = \mathbf{x}_\tau^{[T]} + \frac{1}{\sqrt{\alpha_\tau}} \left( \mathbf{x}_\tau^{[T]} - \mathbf{x}_\tau^{[T]} - \frac{\beta_\tau}{\sqrt{1-\alpha_\tau}} \epsilon_\theta(\mathbf{x}_\tau^{[T]}, \mathbf{x}_c^{[T_c]}, \tau) \right),$$

where the denoising network  $\epsilon_\theta$  is implemented as an EGTN with translation invariance, ensuring the translation equivariance of  $\mu_\theta$ .

**Training and inference.** Optimizing the VLB of our diffusion model leads to the following objective:

$$\mathcal{L}_{\text{cond}} := \mathbb{E}_{\mathbf{x}_0^{[T]}, \mathbf{x}_c^{[T_c]}, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tau \sim \text{Unif}(1, T)} \left[ \|\epsilon - \epsilon_\theta(\mathbf{x}_\tau^{[T]}, \mathbf{x}_c^{[T_c]}, \tau)\|^2 \right].$$

## 8.5 Evaluation Metrics in the Unconditional Case

All these metrics are evaluated on a set of model samples with the same size as the testing set.

**Marginal score** is computed as the absolute difference of two empirical probability density functions. Practically, we collect the  $x, y, z$  coordinates at each time step marginalized over all nodes in all systems in the predictions and the ground truth (testing set). Then we split the collection into 50 bins and compute the MAE in each bin, finally averaged across all time steps to obtain the score. Note that on MD17, instead of computing the pdf on coordinates, we compute the pdf on the length of the chemical bonds, which is a clearer signal that correlates to the validity of the generated MD trajectory, since during MD simulation the bond lengths are usually stable with very small vibrations. Marginal score gives a broad statistical measurement how each dimension of the generated samples align with the original data.

**Classification score** is computed as the cross-entropy loss of a sequence classification model that aims to distinguish whether the trajectory is generated by the model or from the testing set. To be specific, we construct a dataset mixed by the generated samples and the testing set, and randomly split it into 80% and 20% subsets for training and testing. Then the model is trained on the training set and the classification score is computed as the cross-entropy on the testing set. We use a 1-layer EqMotion with a classification head as the model. The classification score provided intuition on how difficult it is to distinguish the generated samples and the original data.

**Prediction score** is computed as the MSE loss of a train-on-synthetic-test-on-real sequence to sequence model. In detail, we train a 1-layer EqMotion on the sampled dataset with the task of predicting the second half of the trajectory given the first half. We then evaluate the model on the testing set and report the MSE as the prediction score. Prediction score provides intuition on the capability of the generative model on generating synthetic data that well aligns with the ground truth.

## 9 More Experiments and Discussions

### 9.1 Molecular

Additional experimental results on the Malonaldehyde and Naphthalene are shown below:

**Table 11:** Comparisons for Molecular Dynamics prediction on MD17 dataset (all results reported by  $\times 10^{-1}$ ). The best results are highlighted in bold. Results averaged over 5 runs

Scenarios	Malonaldehyde				Naphthalene			
	FT		Pretrain		FT		Pretrain	
	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
<b>Pretrain</b>	3.235 $\pm$ 0.012	5.189 $\pm$ 0.023	<b>0.962</b> $\pm$ 0.007	1.584 $\pm$ 0.021	1.416 $\pm$ 0.003	2.268 $\pm$ 0.005	0.714 $\pm$ 0.002	0.972 $\pm$ 0.006
<b>FT</b>	<b>0.897</b> $\pm$ 0.002	<b>1.511</b> $\pm$ 0.009	1.405 $\pm$ 0.006	2.237 $\pm$ 0.023	<b>0.555</b> $\pm$ 0.001	<b>0.867</b> $\pm$ 0.010	nan	nan
<b>PARTIAL-k</b> [10]	0.981 $\pm$ 0.004	1.675 $\pm$ 0.015	1.230 $\pm$ 0.003	2.110 $\pm$ 0.006	0.653 $\pm$ 0.002	0.903 $\pm$ 0.003	2.083 $\pm$ 0.009	1.629 $\pm$ 0.007
<b>MLP-k</b>	0.997 $\pm$ 0.005	1.694 $\pm$ 0.010	1.291 $\pm$ 0.004	2.051 $\pm$ 0.015	0.718 $\pm$ 0.001	0.969 $\pm$ 0.005	nan	nan
<b>Prompt-Tem</b>	1.092 $\pm$ 0.019	2.003 $\pm$ 0.056	2.323 $\pm$ 0.024	3.351 $\pm$ 0.081	0.972 $\pm$ 0.006	1.593 $\pm$ 0.021	nan	nan
<b>GPF</b> [5]	1.176 $\pm$ 0.012	1.931 $\pm$ 0.030	282.1 $\pm$ 157.9/inf	24.97 $\pm$ 24.43/inf	0.758 $\pm$ 0.002	1.005 $\pm$ 0.005	nan	nan
<b>GPF-plus</b> [5]	1.018 $\pm$ 0.003	1.793 $\pm$ 0.010	3.527 $\pm$ 0.501	4.719 $\pm$ 1.397	0.717 $\pm$ 0.003	0.873 $\pm$ 0.006	1.891 $\pm$ 0.056	2.674 $\pm$ 0.156
<b>GeoAda</b>	<b>0.862</b> $\pm$ 0.002	<b>1.414</b> $\pm$ 0.014	<b>0.963</b> $\pm$ 0.007	<b>1.573</b> $\pm$ 0.018	<b>0.581</b> $\pm$ 0.002	<b>0.822</b> $\pm$ 0.004	<b>0.714</b> $\pm$ 0.001	<b>0.969</b> $\pm$ 0.007

### 9.2 Human Motion

Additional experimental results on the *jumping* and *soccer* scenarios are presented below. We also report the standard deviations across all experiments.

**Table 12: Short-term prediction on running from the CMU Mocap dataset.**

scenarios	running				pretrain			
	80	160	320	400	80	160	320	400
Pretrain	28.74 $\pm$ 0.34	57.99 $\pm$ 0.33	126.20 $\pm$ 0.93	159.37 $\pm$ 1.15	<b>7.941</b> $\pm$ 0.02	<b>16.84</b> $\pm$ 0.04	<b>39.91</b> $\pm$ 0.43	<b>52.45</b> $\pm$ 0.07
Full FT	20.34 $\pm$ 0.32	35.26 $\pm$ 0.19	60.58 $\pm$ 1.14	70.20 $\pm$ 1.36	nan	nan	nan	nan
PARTIAL- <i>k</i>	20.47 $\pm$ 0.32	36.04 $\pm$ 0.40	68.15 $\pm$ 0.70	82.59 $\pm$ 0.86	nan	nan	nan	nan
MLP- <i>k</i>	23.35 $\pm$ 0.39	44.44 $\pm$ 0.71	84.27 $\pm$ 1.58	101.42 $\pm$ 2.11	nan	nan	nan	nan
GPF	19.17 $\pm$ 0.28	32.85 $\pm$ 0.66	52.83 $\pm$ 1.03	60.90 $\pm$ 1.54	21.38 $\pm$ 0.97	35.44 $\pm$ 1.05	65.23 $\pm$ 1.76	79.55 $\pm$ 1.11
GPF-plus	19.11 $\pm$ 0.54	31.93 $\pm$ 0.81	<b>49.85</b> $\pm$ 1.21	<b>56.67</b> $\pm$ 1.39	nan	nan	nan	nan
<b>GeoAda</b>	<b>18.70</b> $\pm$ 0.37	<b>33.56</b> $\pm$ 0.25	<b>50.26</b> $\pm$ 0.42	<b>55.54</b> $\pm$ 0.36	<b>7.972</b> $\pm$ 0.02	<b>16.91</b> $\pm$ 0.05	<b>40.06</b> $\pm$ 0.42	<b>52.61</b> $\pm$ 0.07

**Table 13: Short-term prediction on walking from the CMU Mocap dataset.**

scenarios	walking				pretrain			
	80	160	320	400	80	160	320	400
Pretrain	15.49 $\pm$ 0.07	30.66 $\pm$ 0.16	68.71 $\pm$ 0.33	89.23 $\pm$ 0.43	<b>7.941</b> $\pm$ 0.02	<b>16.84</b> $\pm$ 0.04	<b>39.91</b> $\pm$ 0.43	<b>52.45</b> $\pm$ 0.07
Full FT	<b>10.01</b> $\pm$ 0.17	<b>15.12</b> $\pm$ 0.14	<b>23.98</b> $\pm$ 0.13	<b>28.49</b> $\pm$ 0.19	nan	nan	nan	nan
PARTIAL- <i>k</i>	10.47 $\pm$ 0.81	16.97 $\pm$ 0.43	30.97 $\pm$ 0.60	37.69 $\pm$ 0.60	17.26 $\pm$ 0.23	31.75 $\pm$ 0.36	66.29 $\pm$ 0.44	84.27 $\pm$ 0.79
MLP- <i>k</i>	10.86 $\pm$ 1.07	18.74 $\pm$ 1.37	35.78 $\pm$ 1.92	44.86 $\pm$ 1.35	nan	nan	nan	nan
GPF	10.79 $\pm$ 0.14	16.79 $\pm$ 0.16	28.32 $\pm$ 0.21	33.48 $\pm$ 0.22	22.04 $\pm$ 0.38	37.10 $\pm$ 0.40	71.24 $\pm$ 0.39	88.23 $\pm$ 0.97
GPF-plus	10.17 $\pm$ 0.16	16.23 $\pm$ 0.12	26.29 $\pm$ 0.22	31.54 $\pm$ 0.19	nan	nan	nan	nan
<b>GeoAda</b>	<b>8.92</b> $\pm$ 1.02	<b>13.82</b> $\pm$ 1.26	<b>22.99</b> $\pm$ 1.30	<b>26.68</b> $\pm$ 1.31	<b>7.932</b> $\pm$ 0.03	<b>16.90</b> $\pm$ 0.04	<b>38.96</b> $\pm$ 0.47	<b>52.54</b> $\pm$ 0.10

**Table 14: Short-term prediction comparison on the basketball action from the CMU Mocap dataset.**

scenarios	basketball				pretrain			
	80	160	320	400	80	160	320	400
Pretrain	17.92 $\pm$ 0.06	35.89 $\pm$ 0.12	78.48 $\pm$ 0.47	101.12 $\pm$ 0.69	<b>7.941</b> $\pm$ 0.02	<b>16.84</b> $\pm$ 0.04	<b>39.91</b> $\pm$ 0.43	<b>52.45</b> $\pm$ 0.07
FT	16.95 $\pm$ 0.11	30.33 $\pm$ 0.17	58.28 $\pm$ 0.41	71.93 $\pm$ 0.54	-	-	-	-
PARTIAL- <i>k</i>	17.54 $\pm$ 0.12	32.03 $\pm$ 0.43	62.96 $\pm$ 0.86	78.95 $\pm$ 0.98	-	-	-	-
MLP- <i>k</i>	17.60 $\pm$ 0.50	34.76 $\pm$ 1.26	72.90 $\pm$ 2.36	86.34 $\pm$ 1.53	-	-	-	-
GPF	18.48 $\pm$ 0.09	31.94 $\pm$ 0.14	58.80 $\pm$ 0.28	73.56 $\pm$ 0.29	21.38 $\pm$ 0.11	35.44 $\pm$ 0.27	65.23 $\pm$ 0.30	79.55 $\pm$ 0.51
GPF-plus	17.17 $\pm$ 0.07	31.86 $\pm$ 0.15	58.48 $\pm$ 0.32	72.71 $\pm$ 0.24	-	-	-	-
<b>GeoAda</b>	<b>16.85</b> $\pm$ 0.24	<b>29.71</b> $\pm$ 0.41	<b>57.59</b> $\pm$ 0.39	<b>71.19</b> $\pm$ 0.48	<b>7.898</b> $\pm$ 0.05	<b>16.79</b> $\pm$ 0.05	<b>39.70</b> $\pm$ 0.04	<b>52.50</b> $\pm$ 0.07

**Table 15: Short-term prediction comparison on the jumping action from the CMU Mocap dataset.**

scenarios	jumping				pretrain			
	80	160	320	400	80	160	320	400
Pretrain	-	-	-	-	<b>7.941</b> $\pm$ 0.02	<b>16.84</b> $\pm$ 0.04	<b>39.91</b> $\pm$ 0.43	<b>52.45</b> $\pm$ 0.07
FT	-	-	-	-	26.67 $\pm$ 0.10	50.83 $\pm$ 0.29	94.13 $\pm$ 0.58	112.66 $\pm$ 0.71
PARTIAL- <i>k</i>	26.01 $\pm$ 0.08	49.19 $\pm$ 0.09	95.84 $\pm$ 0.13	116.24 $\pm$ 0.23	19.08 $\pm$ 0.17	38.53 $\pm$ 0.34	79.77 $\pm$ 0.52	99.85 $\pm$ 1.07
MLP- <i>k</i>	22.63/nan	44.68/nan	88.93/nan	108.43/nan	15.32 $\pm$ 0.30	31.92 $\pm$ 0.26	66.50 $\pm$ 0.42	82.55 $\pm$ 0.93
GPF	28.74 $\pm$ 0.19	51.97 $\pm$ 0.19	97.80 $\pm$ 0.34	117.94 $\pm$ 0.37	$\pm$ 0.08 -	-	-	-
GPF-plus	-	-	-	-	-	-	-	-
<b>GeoAda</b>	<b>25.91</b> $\pm$ 0.09	<b>48.83</b> $\pm$ 0.83	<b>91.51</b> $\pm$ 0.07	<b>109.24</b> $\pm$ 0.60	<b>7.956</b> $\pm$ 0.03	<b>16.82</b> $\pm$ 0.04	<b>39.55</b> $\pm$ 0.47	<b>52.57</b> $\pm$ 0.09

**Table 16: Short-term prediction comparison on the soccer action from the CMU Mocap dataset.**

scenarios	soccer				pretrain			
	80	160	320	400	80	160	320	400
Pretrain	-	-	-	-	<b>7.941</b> $\pm$ 0.02	<b>16.84</b> $\pm$ 0.04	<b>39.91</b> $\pm$ 0.43	<b>52.45</b> $\pm$ 0.07
FT	17.65 $\pm$ 0.17	31.43 $\pm$ 0.35	59.76 $\pm$ 0.44	74.30 $\pm$ 0.54	-	-	-	-
PARTIAL- <i>k</i>	18.83 $\pm$ 0.10	32.86 $\pm$ 0.07	64.58 $\pm$ 0.37	81.53 $\pm$ 0.50	14.14 $\pm$ 0.41	24.95 $\pm$ 0.38	50.89 $\pm$ 0.40	64.24 $\pm$ 0.51
MLP- <i>k</i>	-	-	-	-	-	-	-	-
GPF	19.28 $\pm$ 0.10	32.18 $\pm$ 0.14	69.58 $\pm$ 0.42	73.63 $\pm$ 0.63	15.70 $\pm$ 0.30	28.31 $\pm$ 0.28	58.57 $\pm$ 0.42	74.28 $\pm$ 0.61
GPF-plus	19.11 $\pm$ 0.18	32.03 $\pm$ 0.31	59.67 $\pm$ 0.46	74.25 $\pm$ 0.65	15.22 $\pm$ 0.23	26.22 $\pm$ 0.29	52.02 $\pm$ 0.47	65.17 $\pm$ 0.53
<b>GeoAda</b>	<b>17.04</b> $\pm$ 0.14	<b>30.03</b> $\pm$ 0.12	<b>53.51</b> $\pm$ 0.25	<b>64.78</b> $\pm$ 0.42	<b>7.961</b> $\pm$ 0.02	<b>16.90</b> $\pm$ 0.03	<b>39.46</b> $\pm$ 0.41	<b>52.43</b> $\pm$ 0.09

**Table 17: Long-term prediction on CMU Mocap: Running and Walking.**

scenarios	running		pretrain		walking		pretrain	
	560	1000	560	1000	560	1000	560	1000
Pretrain	219.16 $\pm$ 2.18	314.85 $\pm$ 3.03	<b>77.06</b> $\pm$ 0.47	<b>130.51</b> $\pm$ 0.27	129.43 $\pm$ 0.77	212.94 $\pm$ 1.90	<b>77.06</b> $\pm$ 0.47	<b>130.51</b> $\pm$ 0.27
Full FT	85.14 $\pm$ 2.36	97.02 $\pm$ 1.45	nan	nan	36.92 $\pm$ 1.37	52.58 $\pm$ 0.62	nan	nan
PARTIAL- <i>k</i> [10]	102.85 $\pm$ 1.68	108.47 $\pm$ 2.03	nan	nan	51.36 $\pm$ 1.93	84.72 $\pm$ 0.67	118.82 $\pm$ 0.58	182.88 $\pm$ 0.79
MLP- <i>k</i>	127.67 $\pm$ 2.46	131.59 $\pm$ 2.90	nan	nan	62.97 $\pm$ 1.45	102.34 $\pm$ 0.86	nan	nan
GPF [5]	61.92 $\pm$ 1.02	71.42 $\pm$ 1.27	nan	nan	42.37 $\pm$ 0.31	52.24 $\pm$ 0.38	119.43 $\pm$ 0.60	171.74 $\pm$ 0.55
GPF-plus [5]	63.56 $\pm$ 1.54	71.60 $\pm$ 0.95	nan	nan	41.31 $\pm$ 0.35	56.47 $\pm$ 0.4	nan	nan
<b>GeoAda</b>	<b>60.88</b> $\pm$ 0.82	<b>70.22</b> $\pm$ 0.20	<b>77.22</b> $\pm$ 0.45	<b>130.17</b> $\pm$ 0.26	<b>34.52</b> $\pm$ 2.26	<b>50.49</b> $\pm$ 0.33	<b>78.12</b> $\pm$ 0.49	<b>129.97</b> $\pm$ 0.30

**Table 18:** Long-term prediction on CMU Mocap: **Jumping and Soccer.**

scenarios	jumping		pretrain		soccer		pretrain	
	560	1000	560	1000	560	1000	560	1000
Pretrain	—	—	77.06±0.47	130.51±0.27	—	—	77.06±0.47	130.51±0.27
Full FT	—	—	145.76±1.23	199.34±2.01	101.44 ±0.51	157.11±0.96	—	—
PARTIAL- <i>k</i> [10]	149.06±0.37	181.52±0.81	135.25±0.92	186.16±1.58	113.88 ±0.75	170.50±1.98	89.63 ±1.62	142.41±1.84
MLP- <i>k</i>	140.11/nan	184.77/nan	111.71±0.74	158.55±1.37	—	—	—	—
GPF [5]	151.50±0.65	194.94±0.34	—	—	100.23±0.88	151.84±1.23	104.03±0.81	161.56±1.31
GPF-plus [5]	—	—	—	—	101.15 ±0.86	153.43 ±0.79	90.12 ±1.10	142.30±1.12
<b>GeoAda</b>	139.46 ±0.29	184.01±0.60	76.98±0.51	130.72±0.26	84.91±0.46	125.91 ±0.75	77.19 ±0.48	130.06±0.31

**Table 19:** Long-term prediction on CMU Mocap: **Basketball.**

scenarios	basketball		pretrain	
	560	1000	560	1000
Pretrain	143.49±1.19	223.99±2.23	77.06±0.47	130.51±0.27
Full FT	94.59±0.58	132.34±1.30	nan	nan
PARTIAL- <i>k</i> [10]	106.84 ±1.00	146.27 ±1.24	nan	nan
MLP- <i>k</i>	107.30±2.76	149.58±2.24	nan	nan
GPF [5]	97.16±0.35	128.29±0.43	nan	nan
GPF-plus [5]	104.54 ±0.26	130.76±1.28	104.51±0.57	155.02 ±0.51
<b>GeoAda</b>	91.03±0.33	120.35±0.44	76.94±0.45	129.81±0.30

### 9.3 Ablations

#### 9.3.1 Parameter efficiency analysis

As shown in Table 20, we explore the impact of varying the number of equivariant adapter blocks. Increasing the number of trainable copy layers generally improves performance, but introduces more parameters and computational cost, revealing a trade-off between performance and efficiency. We have computed the number of tunable parameters for all baselines and GeoAda. The statistics are presented in Table 21. Except for Full Fine-Tuning, all methods have a comparable number of tunable parameters, ensuring a fair comparison.

**Table 20:** Different numbers of adapter blocks

Number	ADE	FDE
1	1.321	3.088
2	1.291	2.968
3	<b>1.108</b>	<b>2.621</b>
4	1.104	2.588
5	1.106	2.686

**Table 21:** The number of tunable parameters for different tuning strategies.

Dataset	Tuning Strategy	Total Parameters	Tunable Parameters
Charged Particle	Full FT	1418252 ~5.41 MB	1418252 ~5.41 MB
	PARTIAL- <i>k</i>	1418252 ~5.41 MB	711302 ~2.71 MB
	MLP- <i>k</i>	2125202 ~8.11 MB	711302 ~2.71 MB
	Prompt-Tem	1418252 ~5.41 MB	711302 ~2.71 MB
	GPF	2125266 ~8.11 MB	711366 ~2.71 MB
	GPF-plus	2125847 ~8.11 MB	711947 ~2.72 MB
	<b>GeoAda</b>	2125691 ~8.11 MB	711791 ~2.72 MB
MD17	Full FT	1424268 ~5.43 MB	1424268 ~5.43 MB
	PARTIAL- <i>k</i>	1424268 ~5.43 MB	716166 ~2.73 MB
	MLP- <i>k</i>	2132370 ~8.13 MB	716166 ~2.73 MB
	Prompt-Tem	1424268 ~5.43 MB	716166 ~2.73 MB
	GPF	2132434 ~8.13 MB	716230 ~2.73 MB
	GPF-plus	2135079 ~8.14 MB	718875 ~2.74 MB
	<b>GeoAda</b>	2132844 ~8.14 MB	716640 ~2.73 MB
CMU Mocap	Full FT	368012 ~1.40 MB	368012 ~1.40 MB
	PARTIAL- <i>k</i>	368012 ~1.40 MB	185990 ~0.71 MB
	MLP- <i>k</i>	550034 ~2.10 MB	185990 ~0.71 MB
	GPF	550098 ~2.10 MB	186054 ~0.71 MB
	GPF-plus	553259 ~2.11 MB	189215 ~0.72 MB
	<b>GeoAda</b>	550075 ~2.10 MB	186031 ~0.71 MB

### 9.3.2 Ablative Architectures

We study the following ablative architectures as shown in Figure 6, Figure 7, Figure 8:

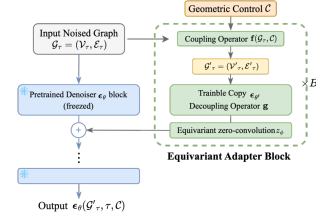


Figure 6: GeoAda

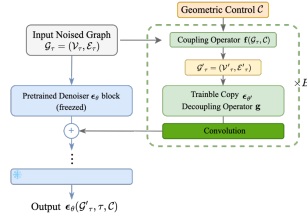


Figure 7: w/o zero convolution

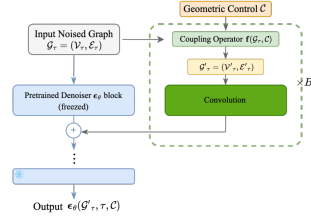


Figure 8: w/o trainable copy

**Proposed GeoAda.** The proposed architecture in the main paper.

**Without Zero Convolution.** Replacing the zero convolutions with standard convolution layers initialized with Gaussian weights.

**Lightweight Layers.** This architecture does not use a trainable copy, and directly initializes single convolution layers.

**Results** We present the results of this ablative study in Table 22, Table 23, Table 24 and Table 25.

**Table 22:** Comparisons for Molecular Dynamics prediction on MD17 dataset (all results reported by  $\times 10^{-1}$ ). The best results are highlighted in bold. Results averaged over 5 runs

Scenarios	Aspirin				Benzene			
	FT		Pretrain		FT		Pretrain	
	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
w/o trainable copy	2.232 $\pm$ 0.008	3.501 $\pm$ 0.022	2.197 $\pm$ 0.007	3.449 $\pm$ 0.010	0.607 $\pm$ 0.002	0.952 $\pm$ 0.010	0.584 $\pm$ 0.002	0.948 $\pm$ 0.006
w/o zero conv	0.929 $\pm$ 0.002	1.619 $\pm$ 0.009	1.203 $\pm$ 0.009	1.975 $\pm$ 0.018	0.214 $\pm$ 0.001	0.359 $\pm$ 0.004	0.291 $\pm$ 0.001	0.469 $\pm$ 0.007
GeoAda	<b>0.891</b> $\pm$ 0.003	<b>1.533</b> $\pm$ 0.008	<b>1.060</b> $\pm$ 0.003	<b>1.852</b> $\pm$ 0.012	<b>0.191</b> $\pm$ 0.000	<b>0.319</b> $\pm$ 0.002	<b>0.240</b> $\pm$ 0.002	<b>0.394</b> $\pm$ 0.005

**Table 23:** Ablation study of Short-term prediction on **running** from the CMU Mocap dataset.

scenarios	running				pretrain			
	80	160	320	400	80	160	320	400
millisecond (ms)								
w/o trainable copy	44.52 $\pm$ 0.48	76.98 $\pm$ 1.20	134.91 $\pm$ 2.04	159.75 $\pm$ 2.45	198.16 $\pm$ 2.97	139.24 $\pm$ 0.23	270.04 $\pm$ 0.56	314.89 $\pm$ 0.68
w/o zero conv	19.07 $\pm$ 0.37	34.25 $\pm$ 0.82	51.75 $\pm$ 1.39	55.74 $\pm$ 1.53	nan	nan	nan	nan
GeoAda	<b>18.70</b> $\pm$ 0.37	<b>33.56</b> $\pm$ 0.25	<b>50.26</b> $\pm$ 0.42	<b>55.54</b> $\pm$ 0.36	<b>7.972</b> $\pm$ 0.02	<b>16.91</b> $\pm$ 0.05	<b>40.06</b> $\pm$ 0.42	<b>52.61</b> $\pm$ 0.07

**Table 24:** Ablation study of Short-term prediction on **walking** from the CMU Mocap dataset.

scenarios	walking				pretrain			
	80	160	320	400	80	160	320	400
millisecond (ms)								
w/o trainable copy	23.77 $\pm$ 0.15	43.43 $\pm$ 0.31	84.79 $\pm$ 0.77	105.08 $\pm$ 1.28	nan	nan	nan	nan
w/o zero conv	12.62 $\pm$ 0.07	20.67 $\pm$ 0.20	36.75 $\pm$ 0.52	44.69 $\pm$ 0.45	36.18 $\pm$ 0.12	58.18 $\pm$ 0.08	102.06 $\pm$ 0.05	123.82 $\pm$ 0.04
GeoAda	<b>8.92</b> $\pm$ 0.02	<b>13.82</b> $\pm$ 0.12	<b>22.99</b> $\pm$ 0.30	<b>26.68</b> $\pm$ 0.13	<b>7.932</b> $\pm$ 0.03	<b>16.90</b> $\pm$ 0.04	<b>38.96</b> $\pm$ 0.47	<b>52.54</b> $\pm$ 0.10

**Table 25:** Ablation study of long-term prediction on **running, walking** from the CMU Mocap dataset.

scenarios	running		pretrain		walking		pretrain	
	560	1000	560	1000	560	1000	560	1000
millisecond (ms)								
w/o trainable copy	198.16 $\pm$ 2.97	228.41 $\pm$ 2.96	365.41 $\pm$ 0.63	374.74 $\pm$ 0.48	143.35 $\pm$ 2.26	214.78 $\pm$ 2.82	nan	nan
w/o zero conv	64.69 $\pm$ 0.97	74.64 $\pm$ 0.66	nan	nan	60.28 $\pm$ 1.07	93.99 $\pm$ 1.41	165.86 $\pm$ 0.15	249.90 $\pm$ 0.38
GeoAda	<b>60.88</b> $\pm$ 0.82	<b>70.22</b> $\pm$ 0.02	<b>77.22</b> $\pm$ 0.45	<b>130.17</b> $\pm$ 0.26	<b>34.52</b> $\pm$ 2.26	<b>50.49</b> $\pm$ 0.33	<b>78.12</b> $\pm$ 0.49	<b>129.97</b> $\pm$ 0.30

## 977 9.4 Standard Deviations

978 We have already provided the standard deviations in App. 9.2.

## 979 10 Discussion

980 **Limitation** While GeoAda demonstrates strong empirical performance and theoretical grounding  
981 in preserving  $SE(3)$ -equivariance during fine-tuning, several limitations remain: The effectiveness  
982 of GeoAda hinges on the design of coupling and decoupling operators for control injection. While  
983 theoretically sound, these handcrafted designs may not generalize well to control signals with high-  
984 dimensional or structured semantics. Moreover, although GeoAda is validated across multiple  
985 domains (e.g., particles, molecules, human motion), the evaluations are limited to medium-scale  
986 datasets and relatively small models. Assessing its scalability to larger systems—such as full proteins  
987 or macromolecular assemblies—remains an important direction for future work.

988