

A APPENDIX

A.1 PROOF

A.1.1 THEOREM 1

Theorem 1. *When the ODE pipeline is applied to an initial distribution $x_0 \sim p_0$, the Wasserstein distance between the two marginal distributions at any intermediate times t and $t + \Delta t$ is controlled by the mean kinetic energy at time t :*

$$W_2[p_t, p_{t+\Delta t}] \leq \Delta t \sqrt{\mathbb{E}_{x \sim p_t} \|v_\theta(x_t, t)\|^2}$$

Proof. From the definition of the W_2 distance in Eq.7, we have:

$$W_2[p_t, p_{t+\Delta t}] \leq \sqrt{\mathbb{E}_{x \sim p_t} \|x_{t+\Delta t} - x_t\|^2} \quad (21)$$

For the discrete ODE pipeline using the explicit Euler update:

$$x_{t+\Delta t} = x_t + v_\theta(x_t, t)\Delta t \quad (22)$$

Substituting this into the previous bound yields:

$$W_2[p_t, p_{t+\Delta t}] \leq \sqrt{\mathbb{E}_{x \sim p_t} \|v_\theta(x_t, t)\Delta t\|^2} \quad (23)$$

$$= \Delta t \sqrt{\mathbb{E}_{x \sim p_t} \|v_\theta(x_t, t)\|^2} \quad (24)$$

□

A.1.2 THEOREM 2

Theorem 2. *From time 0 to T , the cumulative one-step W_2 distances over the ODE pipeline is controlled by the total kinetic energy A_{model} :*

$$\begin{aligned} \sum_t W_2^2[p_t, p_{t+\Delta t}] &\leq \delta t \int_0^{T\delta t} \mathbb{E}_{x \sim p_s} \|v_\theta(x_s, s)\|^2 ds \\ A_{model} &:= \int_0^{T\delta t} \mathbb{E}_{x \sim p_s} \|v_\theta(x_s, s)\|^2 ds \end{aligned}$$

Proof. According to Theorem 1, squaring the one-step Wasserstein distance and then summing over the entire time horizon of the ODE pipeline yields:

$$\sum_{t=0}^{T-1} W_2^2[p_t, p_{t+\Delta t}] \leq \sum_{t=0}^{T-1} \Delta t^2 \mathbb{E}_{x \sim p_t} \|v_\theta(x_t, t)\|^2 \quad (25)$$

For any $t \in [0, (T-1)\delta t]$, by applying time rescaling and the Benamou–Brenier inequality, the discrete summation can be mapped to its continuous-time counterpart:

$$\begin{aligned} \sum_{t=0}^{T-1} W_2^2[p_t, p_{t+\Delta t}] &\leq \sum_{t=0}^{T-1} \delta t \int_t^{t+\delta t} \int \|v_\theta(x_s, s)\|^2 dp_s ds \\ &= \delta t \int_0^{T\delta t} \int \|v_\theta(x_s, s)\|^2 dp_s ds \\ &= \delta t \int_0^{T\delta t} \mathbb{E}_{x \sim p_s} \|v_\theta(x_s, s)\|^2 ds \end{aligned} \quad (26)$$

In this formulation, we define the integral term as the total kinetic energy of the velocity field:

$$A_{model} := \int_0^{T\delta t} \mathbb{E}_{x \sim p_s} \|v_\theta(x_s, s)\|^2 ds \quad (27)$$

□

A.1.3 CORRECTION TERM OF THE SECOND-ORDER TTM

For the second-order correction term $\frac{1}{2}h_t^2 \frac{d\epsilon_\theta}{d\gamma_t}$, we first differentiate the numerator and denominator with respect to t :

$$\frac{1}{2}h_t^2 \frac{d\epsilon_\theta}{d\gamma_t} = \frac{1}{2}h_t^2 \frac{d\epsilon_\theta/dt}{d\gamma_t/dt} \quad (28)$$

Then, we apply a simple finite difference scheme to compute the numerator:

$$\frac{d\epsilon_\theta(x_t, t)}{dt} = \frac{\epsilon_\theta(x_{t+\Delta t}, t + \Delta t) - \epsilon_\theta(x_t, t)}{\Delta t} \quad (29)$$

Given that $\gamma_t = \sqrt{\frac{1-\bar{\alpha}_t^2}{\bar{\alpha}_t}} = \sqrt{\frac{1}{\bar{\alpha}_t} - \bar{\alpha}_t}$, differentiating with respect to t yields:

$$\begin{aligned} \frac{d\gamma_t}{dt} &= \frac{1}{2} \left(\frac{1}{\bar{\alpha}_t} - \bar{\alpha}_t \right)^{-\frac{1}{2}} \cdot \frac{d}{dt} \left(\frac{1}{\bar{\alpha}_t} - \bar{\alpha}_t \right) \\ &= \frac{1}{2} \left(\frac{1}{\bar{\alpha}_t} - \bar{\alpha}_t \right)^{-\frac{1}{2}} \left(-\frac{1}{\bar{\alpha}_t^2} \frac{d\bar{\alpha}_t}{dt} - \frac{d\bar{\alpha}_t}{dt} \right) \\ &= -\frac{1}{2} \left(\frac{1}{\bar{\alpha}_t} - \bar{\alpha}_t \right)^{-\frac{1}{2}} \left(\frac{1 + \bar{\alpha}_t^2}{\bar{\alpha}_t^2} \frac{d\bar{\alpha}_t}{dt} \right) \\ &= -\frac{1 + \bar{\alpha}_t^2}{2\bar{\alpha}_t^2 \gamma_t} \frac{d\bar{\alpha}_t}{dt} \end{aligned} \quad (30)$$

Since $\bar{\alpha}_t$ is a predefined set of hyperparameters, its derivative with respect to t can also be approximated using finite differences. Then, substituting the above into Eq. 28, we obtain:

$$\frac{1}{2}h_t^2 \frac{d\epsilon_\theta}{d\gamma_t} \approx -\frac{\Delta \gamma_t^2 \bar{\alpha}_t^2 \gamma_t \Delta \epsilon_{\theta_t}}{(1 + \bar{\alpha}_t^2) \Delta \bar{\alpha}_t} \quad (31)$$

A.2 COMPUTE E_t

In Eq. 10 we define the non-optimality term E_t . This section details the algorithm for computing its cumulative sum $\sum_t E_t$.

For clarity of notation, we denote the root-mean-square norm of the model-predicted velocity field under the current distribution p_t in Eq. 10 as K_t :

$$\sqrt{\mathbb{E}_{x \sim p_t} \|v_\theta(x_t, t)\|^2} \approx \sqrt{\frac{1}{N} \sum_{i=1}^N \|v_\theta(x_t^i, t)\|^2} := K_t \quad (32)$$

Accordingly, Eq. 10 can be rewritten in a simplified form as:

$$E_t = \Delta t \cdot K_t - W_2[p_t, p_{t+\Delta t}] \quad (33)$$

Based on this, the cumulative sum is computed as

$$\mathcal{E}_{0:T-1} = \sum_{t=0}^{T-1} E_t \quad (34)$$

This metric characterizes the discrepancy between the model-provided upper bound on kinetic energy and the true Wasserstein distance, reflecting the degree of non-optimality of the overall velocity field relative to the optimal transport.

In our experiments, we approximate the W_2 distance between consecutive time-step distributions using an entropy-regularized Sinkhorn (Cuturi, 2013) optimal transport algorithm. The implementation employs the GeomLoss (SamplesLoss) interface with the cost function exponent set to $p = 2$, corresponding to the W_2 distance. To maintain numerical stability in high-dimensional sample spaces, we set the blur parameter to 0.01 and the multi-scale scaling factor to 0.9. When iterating on the GPU, these parameters strike a balance between accuracy and computational efficiency. We provide pseudocode in Algorithm 1.

Algorithm 1 Cumulative Non-Optimality Computation**Require:**

Samples $\{x_t^i\}$ for times $t = 0, \dots, T$
 Model velocity field $v_\theta(x, t)$
 Time step Δt
 Sinkhorn parameters: $p = 2$, blur = 0.01, scaling = 0.9

Ensure: $\{W2_t\}, \{E_t\}$, cumulative error $\mathcal{E}$

```

1: Initialize  $\mathcal{E} \leftarrow 0$ 
2: for  $t = 0$  to  $T - 1$  do
3:    $K_t \leftarrow \sqrt{\frac{1}{N} \sum_i \|v_\theta(x_t^i, t)\|^2}$ 
4:    $W2_{sq_t} \leftarrow \text{Sinkhorn}(x_t, x_{t+1}; p, \text{blur}, \text{scaling}), W2_t \leftarrow \sqrt{W2_{sq_t}}$ 
5:   Compute non-optimality term:  $E_t \leftarrow \Delta t \cdot K_t - W2_t$ 
6:   Update cumulative error:  $\mathcal{E} \leftarrow \mathcal{E} + E_t$ 
7: end for
8: return  $\mathcal{E}$ 

```

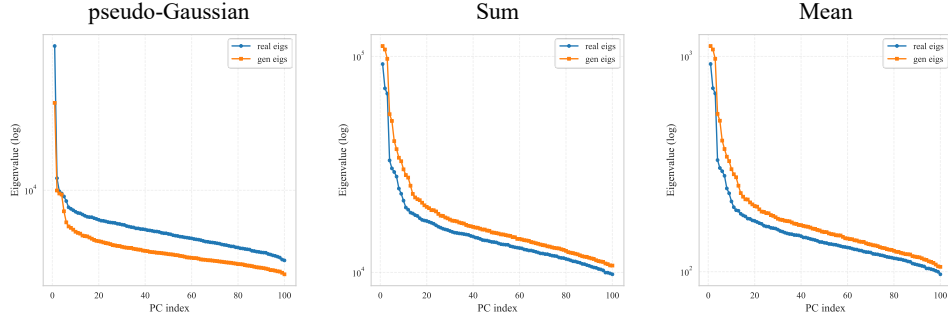


Figure 6: Eigenvalue curves from randomized SVD under different temporal compression methods.

A.3 COMPRESSING CURVATURE FEATURES ALONG TIME

In Fig. 2, we present histograms of curvature features obtained using three different methods for compressing the temporal dimension. From the histograms, one can already discern which method is more discriminative. Here, we further provide principal component analysis (PCA) of these different methods.

We perform randomized SVD on the curvature features computed via temporal summation, mean, and pseudo-Gaussian methods for both real and synthetic images, and obtain the top eigenvalues for each principal component.

As shown in Fig. 6, we compare the top eigenvalue curves obtained from randomized SVD for the three curvature feature extraction methods. The pseudo-Gaussian method exhibits pronounced information concentration: the first principal component is the largest, followed by an almost linear decline until the sixth component, after which the curve flattens, indicating that the major variations are captured by a small number of components. In contrast, the summation and mean methods show more dispersed information, with the first 15 components retaining relatively large eigenvalues and noticeable fluctuations, suggesting that the main information is not well concentrated. This demonstrates that the pseudo-Gaussian approach outperforms the other two in terms of feature concentration, dimensionality reduction efficiency, and potential discriminative power.

Furthermore, regardless of the method, the modes of variation are largely consistent between real and synthetic images. However, the differences between principal components of pseudo-Gaussian features for real versus synthetic images are significantly larger than those of the other methods.

A closer examination of the pseudo-Gaussian method reveals that real images generally have higher eigenvalues than synthetic ones, indicating that their curvature features exhibit greater variability and richer structure in the principal component space. Synthetic images, on the other hand, show

lower eigenvalues, suggesting more uniform or smoother curvature patterns. This difference serves as a potential discriminative indicator for distinguishing real and synthetic images, and corroborates our earlier observation that synthetic images exhibit stronger velocity field consistency and more concentrated distributions.

We also note that, for the summation and mean methods, the eigenvalue curves of synthetic images are slightly higher. We attribute this to the fact that summation/mean operations tend to smooth out local details, which suppresses extreme values in real images while amplifying the more uniform overall curvature of synthetic images, resulting in slightly higher values in the principal component space. In contrast, the pseudo-Gaussian method better preserves local curvature information, which we believe underlies the artifact-focused capability demonstrated in Section. 6.1.

A.4 IMPLEMENTATION DETAILS

Our experiments are conducted on an A800 GPU for a total of 15 epochs. All images are center-cropped to 64×64 before being fed into the model. Random horizontal flipping and rotation are applied as data augmentation. We use the AdamW optimizer, setting the learning rate of the second-order feature pipeline to 1e-5 with a weight decay of 0.05, the zeroth-order feature pipeline to 1e-4 with a weight decay of 0.01, and the shared projection layer and classifier to 1e-3 with zero weight decay. The learning rates are decoupled to account for the differing sensitivities of features at different orders to input variations.

It is worth noting that the selected ODE pipeline operates in the RGB domain, and all conclusions in this work are based on RGB images. However, we anticipate that our approach could be extended to ODE pipelines in latent spaces, potentially supporting larger receptive fields, which remains an avenue for future research.