GeodesicNVS: Probability Density Geodesic Flow Matching for Novel View Synthesis

Supplementary Material

7. Derivations

This section provides detailed derivations of the Euler–Lagrange equation and functional derivative for the density-based geodesic formulation.

We recall the weighted length functional

$$S[\gamma] = \int_0^1 L\big(t, \gamma(t), \dot{\gamma}(t)\big)\, dt, \qquad \gamma : [0,1] \to \mathbb{R}^d \tag{18}$$

with the Lagrangian written in three equivalent forms:

$$\begin{aligned} L(t, \gamma, \dot{\gamma}) &= \|\dot{\gamma}\|_{G(\gamma)} && (19)\\ &= p(\gamma)^{-1}\, \|\dot{\gamma}\| && (20)\\ &= \sqrt{\dot{\gamma}^\top G(\gamma)\, \dot{\gamma}}, && (21) \end{aligned}$$

where the metric tensor is defined as

$$G(\gamma) = p(\gamma)^{-2}\, I. \tag{22}$$

7.1. Euler-Lagrange Equation

The Euler-Lagrange equation is given by:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\gamma}}\right) - \frac{\partial L}{\partial \gamma} = 0. \tag{23}$$

Compute partial derivatives using $L = p(\gamma)^{-1}\|\dot{\gamma}\|$:

$$\begin{aligned} \frac{\partial L}{\partial \dot{\gamma}} &= p(\gamma)^{-1}\widehat{\dot{\gamma}}, && (24)\\ \frac{\partial L}{\partial \gamma} &= -p(\gamma)^{-1}\nabla \log p(\gamma)\|\dot{\gamma}\|. && (25) \end{aligned}$$

The time derivative yields:

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\gamma}}\right) = &-p(\gamma)^{-1}(\nabla \log p(\gamma)\cdot \dot{\gamma})\widehat{\dot{\gamma}} \\ &+ p(\gamma)^{-1}\frac{(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma}}{\|\dot{\gamma}\|}. \end{aligned} \tag{26}$$

Substituting into Euler–Lagrange:

$$\begin{aligned} -p(\gamma)^{-1}(\nabla \log p(\gamma)\cdot \dot{\gamma})\widehat{\dot{\gamma}} + p(\gamma)^{-1}\frac{(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma}}{\|\dot{\gamma}\|} \\ + p(\gamma)^{-1}\nabla \log p(\gamma)\|\dot{\gamma}\| = 0. \end{aligned} \tag{27}$$

Multiply by $p(\gamma)$ and simplify using $\widehat{\dot{\gamma}} = \frac{\dot{\gamma}}{\|\dot{\gamma}\|}$:

$$\frac{(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma}}{\|\dot{\gamma}\|} - (\nabla \log p(\gamma)\cdot \dot{\gamma})\widehat{\dot{\gamma}} + \nabla \log p(\gamma)\|\dot{\gamma}\| = 0. \tag{28}$$

Note the identity:

$$\nabla \log p(\gamma)\|\dot{\gamma}\| - (\nabla \log p(\gamma)\cdot \dot{\gamma})\widehat{\dot{\gamma}} = \|\dot{\gamma}\|(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\nabla \log p(\gamma). \tag{29}$$

Thus:

$$\frac{(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma}}{\|\dot{\gamma}\|} + \|\dot{\gamma}\|(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\nabla \log p(\gamma) = 0. \tag{30}$$

Multiplying by $\|\dot{\gamma}\|$ and using the fact that for constant-speed paths, $(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma} = \ddot{\gamma}$, recovers the equivalent formulas in the main text (4):

$$\boxed{\ddot{\gamma} + \|\dot{\gamma}\|^2(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\nabla \log p(\gamma) = 0.} \tag{31}$$

7.2. Functional Derivative

The functional derivative is:

$$\frac{\delta S}{\delta \gamma} = \frac{\partial L}{\partial \gamma} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\gamma}}\right). \tag{32}$$

Using our previous results:

$$\begin{aligned} \frac{\delta S}{\delta \gamma} = &-p(\gamma)^{-1}\nabla \log p(\gamma)\|\dot{\gamma}\| + p(\gamma)^{-1}(\nabla \log p(\gamma)\cdot \dot{\gamma})\widehat{\dot{\gamma}} \\ &- p(\gamma)^{-1}\frac{(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma}}{\|\dot{\gamma}\|}. \end{aligned} \tag{33}$$

Factor out the projection identity $(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)$:

$$\frac{\delta S}{\delta \gamma} = -\frac{1}{p(\gamma)\|\dot{\gamma}\|}\left[(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\left(\nabla \log p(\gamma)\|\dot{\gamma}\|^2 + \ddot{\gamma}\right)\right]. \tag{34}$$

For constant-speed paths, $(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\ddot{\gamma} = \ddot{\gamma}$, yielding:

$$\boxed{\frac{\delta S}{\delta \gamma} = \frac{-1}{p(\gamma)\|\dot{\gamma}\|}\left[(I - \widehat{\dot{\gamma}}\widehat{\dot{\gamma}}^\top)\nabla \log p(\gamma) + \frac{\ddot{\gamma}}{\|\dot{\gamma}\|^2}\right].} \tag{35}$$

This represents the normal component of the path variation, as noted in the main text (5), consistent with the geometric decomposition in the Euler–Lagrange equation.

7.3. Probability density along the path

To compute the functional derivative norm, we estimate the probability density along paths. Following [7, 52], we compute relative probabilities $p(\gamma(t))/p(\gamma(a))$ since absolute density estimation is challenging.

For the conservative vector field $\nabla \log p$, the relative log-probability between $\gamma(a)$ and $\gamma(t)$ is path-independent:

$$\log \widetilde{p}_a(\gamma(b)) := \log p(\gamma(b)) - \log p(\gamma(a)) \tag{36}$$

$$= \int_a^b \dot{\gamma}(t)^\mathsf{T} \nabla \log p(\gamma(t))\, \mathrm{d}t. \tag{37}$$

We approximate this integral using the trapezoidal rule. Let $f(\tau) = \dot{\gamma}(\tau)^\mathsf{T} \nabla \log p(\gamma(\tau))$ and sample the path at $n+1$ equally spaced points $t_i = a + i/n$ for $i = 0, \ldots, n$. The relative log-probability at $\gamma(t_i)$ is:

$$\log \widetilde{p}_a(\gamma(t_i)) \approx \frac{1}{n} \left(\frac{f(a)}{2} + \sum_{k=1}^{i-1} f(t_k) + \frac{f(t_i)}{2} \right). \tag{38}$$

This enables computation of $\widetilde{p}_a(\gamma(t_i)) = p(\gamma(t_i))/p(\gamma(a))$ along the curve, which is used in evaluating (5) for comparison of geodesic gradient norm in Fig. 7.

8. Further Implementation Details

This section provides comprehensive implementation details for the key components of our Probability Density Geodesic Flow Matching (PDG-FM) framework, complementing the methodological descriptions in Sec. 4.

8.1. Data-to-Data Flow Matching Implementation

Training Configuration Our Linear-D2D-FM, as well as Free3D and Naive FM, are trained using the linear interpolant formulation from Eq. 6. All models inherit the publicly released checkpoint from Zero-1-to-3 [30] to ensure fair comparison.

Training is conducted in the latent space at 32×32 resolution, obtained by encoding 256×256 input images through the VAE from Stable Diffusion [36], which provides an $8\times$ downscaling factor. We train all models for 20,000 steps with a global batch size of 256, using the AdamW optimizer with learning rate 1×10^{-5} and standard weight decay.

8.2. Variational Distillation of Geodesics

The geodesic distillation follows Algorithm 1 using the LVIS-annotated Objaverse subset with 12 random views per object and category annotations.

Preprocessing Pipeline Image pairs at 512×512 resolution are encoded to 64×64 ambient latents x_0, x_1 using the VAE encoder. For each object pair with shared category text prompt, we apply Text Inversion [20] to fine-tune CLIP text embeddings, yielding dedicated embeddings c_0, c_1 for respective views. The inversion runs for 500 steps with learning rate 0.005 using AdamW optimizer.

The ambient latents x_0, x_1 and their corresponding text embeddings c_0, c_1 are then processed through DDIM-F(x, c, τ) with $\tau = 0.6$, classifier-free guidance scale cfg $=$ 1, and 30 NFE (using a 50-step DDIM scheduler) to obtain the final latent pairs z_0, z_1 in the DDIM-F latent space.

GeodesicNet Optimization The teacher network ϕ_ξ operates in the DDIM-F latent space and is optimized using AdamW with learning rate 1×10^{-6} and uniform timestep sampling. The optimization uses score gradients computed via Eq. 12, with conditioning employing $c_t - c_{\text{neg}}$ where $c_t = (1-t)c_0 + tc_1$ represents time-linear text conditioning. The negative prompt targets unrealistic image artifacts: "A doubling image, unrealistic, artifacts, distortions, unnatural blending, ghosting effects, overlapping edges, harsh transitions, motion blur, poor resolution, low detail".

The student network ϕ_η operates in ambient latent space and is optimized using AdamW with learning rate 1×10^{-3}, minimizing the MSE loss between predicted x_t and DDIM-B(z_t, c_t, τ) reconstructions. Both networks use identical UNet architectures with 128 channels and 2 residual blocks.

8.3. Probability Density Geodesic Flow Matching

We train two Data-to-Data Flow Matching variants: Linear-D2D-FM using linear interpolants; and PDG-FM (geodesic) using geodesic interpolants from the trained GeodesicNet ϕ_η. Both models initialize from the Zero-1-to-3 checkpoint and train at 64×64 latent space for 40,000 steps with global batch size 32.

8.4. Computational Efficiency

All models are trained on NVIDIA RTX A6000 GPUs with 48GB memory.

Our two-phase approach (geodesic distillation followed by flow matching) provides significant efficiency advantages: GeodesicNet training requires fewer samples than end-to-end metric flow matching; flow model training is detached from score function evaluation, avoiding expensive density gradient computations during FM training; and the modular design enables independent improvements to geodesic estimation and flow matching components.

9. Further Experiment Results

CFM Ablation We analyze two key design choices in Data-to-Data Flow Matching: noise augmentation and training timestep sampling. Table 5 presents the quantitative evaluation.

For noise augmentation, we compare our default setting (eps=400) with a lower-noise variant (eps=50). Both configurations are evaluated using 100 inference steps. Results demonstrate that moderate noise augmentation (eps=400) consistently outperforms the low-noise alternative across all metrics, indicating that appropriate noise regularization prevents overfitting while preserving semantic consistency.

For timestep sampling, we compare our default lognormal sampling (std=1, mean=0) against discrete uniform sampling strategies optimized for few-step inference. The discrete distributions U_{10} and U_4 sample training timesteps that align with specific inference steps (10 and 4 NFE respectively). Our default lognormal sampling is evaluated using 10 inference steps. Notably, these discrete strategies outperform continuous uniform sampling when evaluated at their target inference steps, with U_{10} achieving the best overall balance across evaluation metrics.

	Objaverse			
Method	FID ↓	CLIP-S ↑	SSIM ↑	PSNR ↑
Linear-D2D-FM	5.4324	88.9855	0.8634	20.8447
Linear-D2D-FM(eps=50)	5.8050	88.3582	0.8532	20.0291
Linear-D2D-FM	5.8223	88.9749	0.8688	21.3045
Linear-D2D-FM, U_{10}	5.5146	88.9185	0.8706	21.5952
Linear-D2D-FM, U_4	7.4414	88.1954	0.8784	22.2092

Table 5. **Ablation study of noise augmentation (top) and timestep sampling (bottom) in Linear-D2D-FM.**

Extended Linear-D2D-FM Qualitative Results Figures 9 and 10 provide additional qualitative comparisons of our Data-to-Data FM against Noise-to-Data FM and Free3D baselines. Our method maintains superior structural consistency and detail preservation, particularly in the challenging 10 NFE setting. For Free3D, we use cfg=1 at 10NFE as conditional diffusion models exhibit significant performance degradation with classifier-free guidance during accelerated inference. The deterministic coupling in Linear-D2D-FM enables more faithful geometry reconstruction and reduces artifacts observed in diffusion-based approaches under limited inference budgets.

Extended PDG-FM Qualitative Results Fig. 8 presents additional comparisons between our Probability Density Geodesic FM (PDG-FM) and linear interpolation baselines. Visual analysis shows that our geodesic Data-to-Data FM (PDG-FM) preserves finer details across viewpoint changes, as geodesic interpolants maintain manifold consistency throughout transformations. These observations align with our quantitative findings in the main text, demonstrating the benefits of manifold-aware interpolation.

Figure 8. **Extended Qualitative Results** of PDG-FM (geodesic) and Linear-D2D-FM on Objaverse.

Figure 9. **Extended Qualitative Results** on Objaverse.

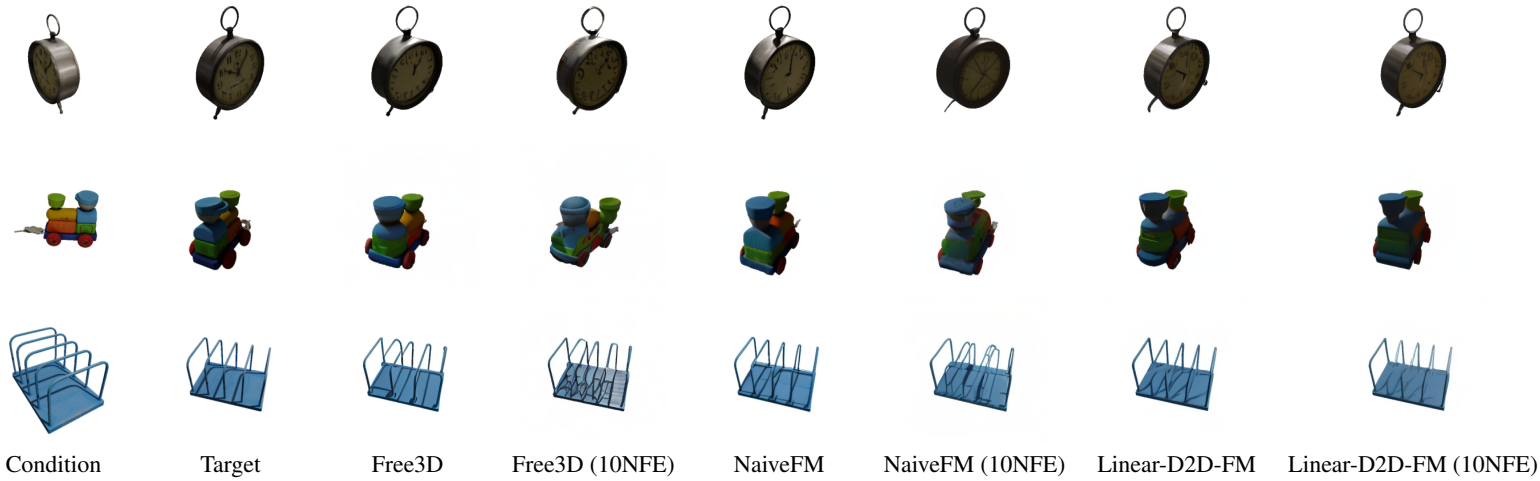

Figure 10. **Extended Qualitative Results** on GSO30.