Maximum Likelihood Training of Implicit Nonlinear Diffusion Models

Dongjun Kim*[1] Byeonghu Na*[1] Se Jung Kwon[2]
Dongsoo Lee[2] Wanmo Kang[1] Il-Chul Moon[1,3]

[1] KAIST [2] CLOVA [3] summary.ai

Contribution of the paper

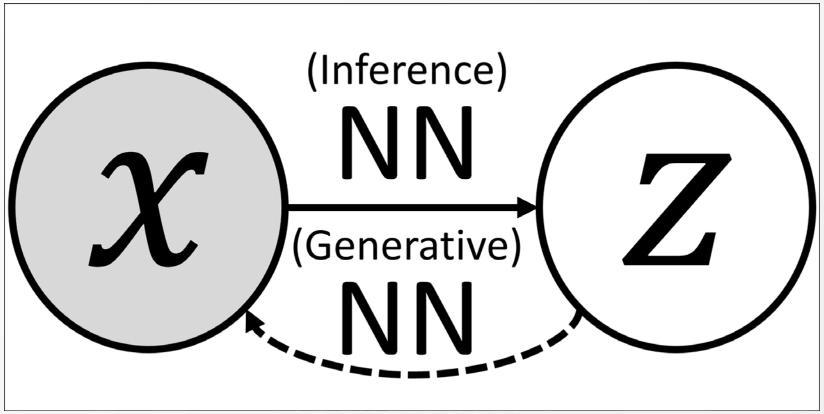
We introduce a Nonlinear Diffusion Model

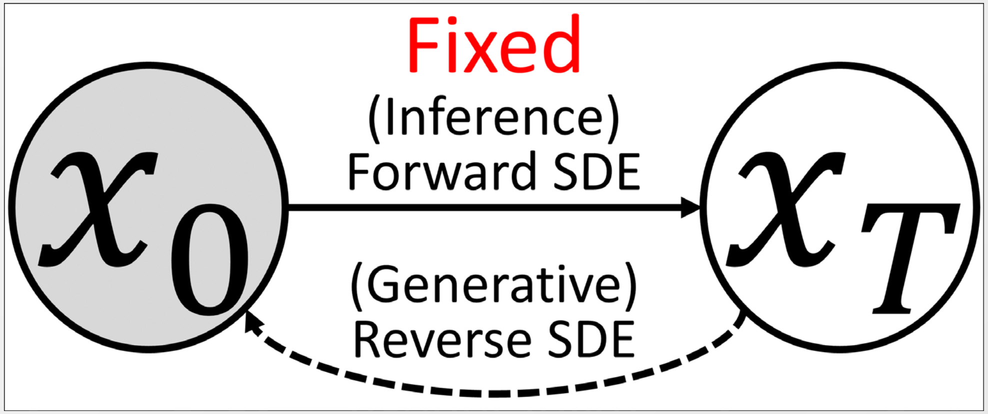
Contribution of the paper

We introduce a Nonlinear Diffusion Model

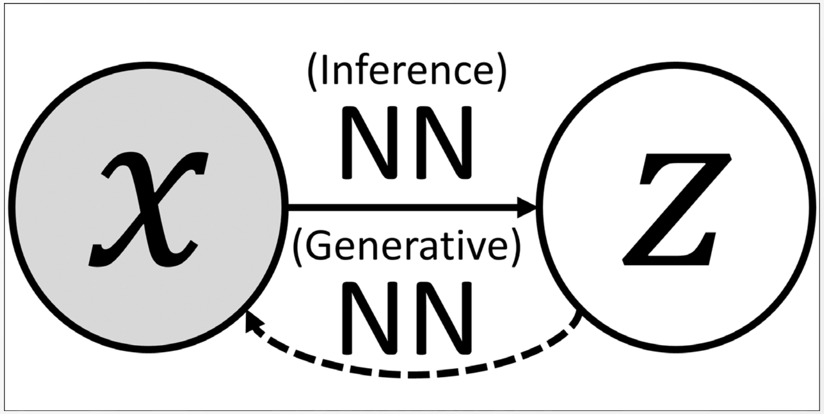
	Discrete Diffusion	Continuous Diffusion
Linear	NCSN/DDPM	NCSN++/DDPM++
Semi-Linear	SBP	-
Fully Nonlinear	DiffFlow	INDM

Motivation: VAE vs. Diffusion Model

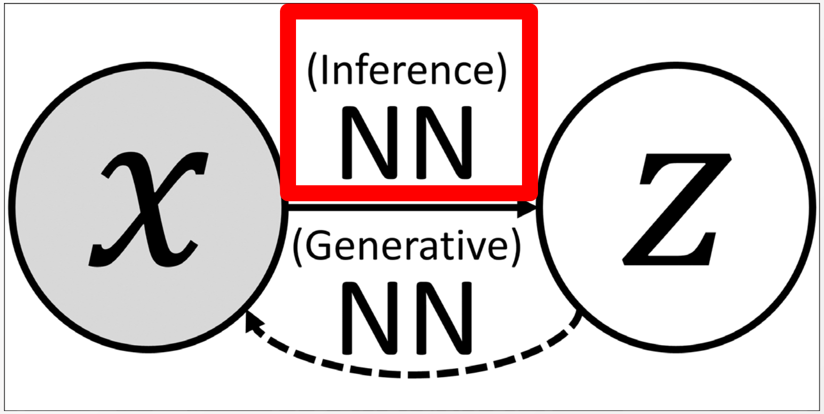




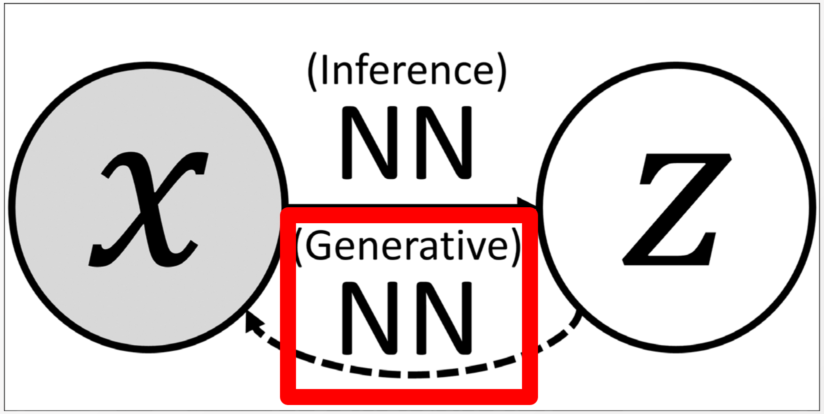
Motivation: VAE vs. Diffusion Model



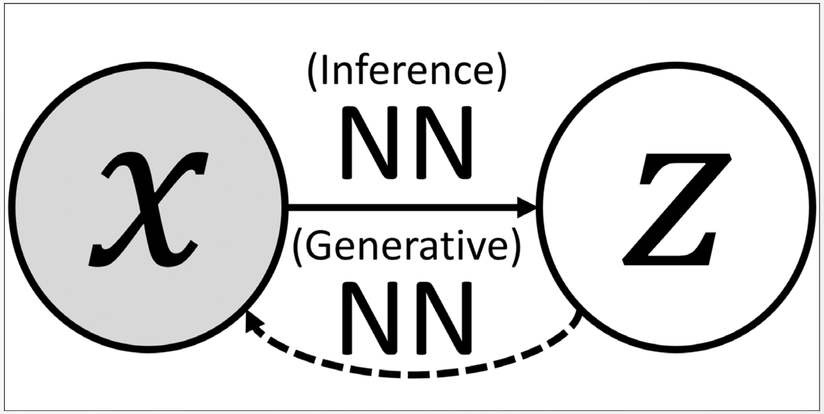
Motivation: VAE vs. Diffusion Model

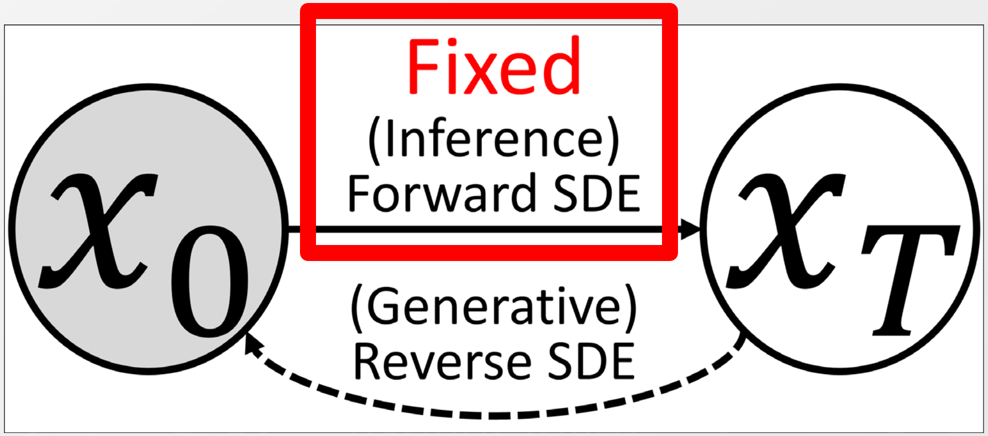


Motivation: VAE vs. Diffusion Model

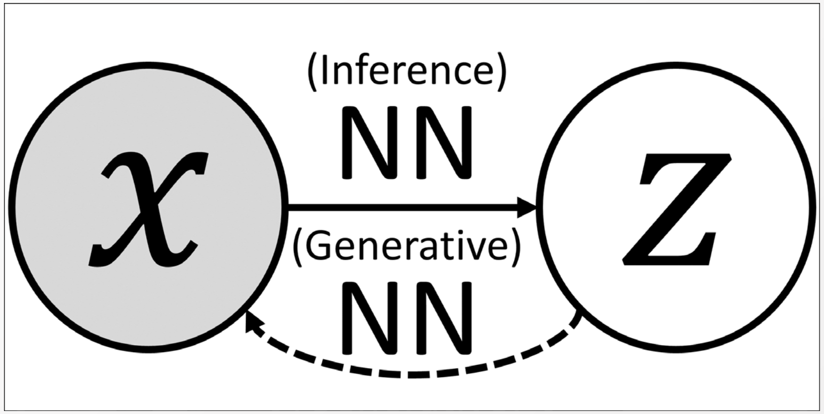


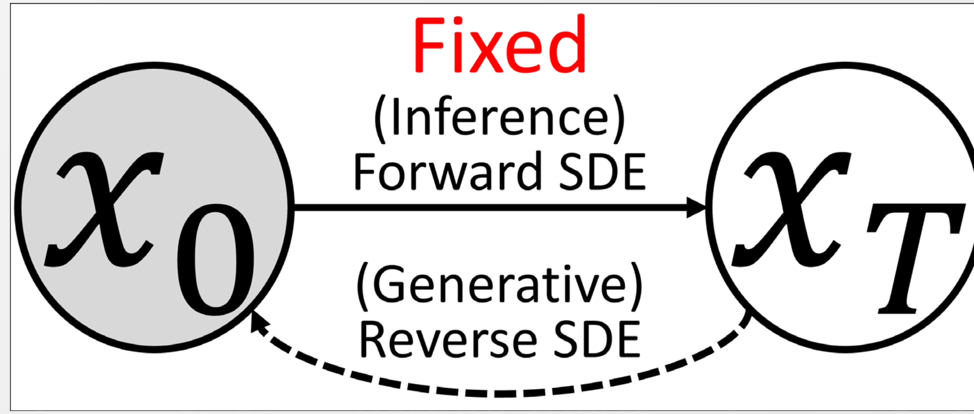
Motivation: VAE vs. Diffusion Model

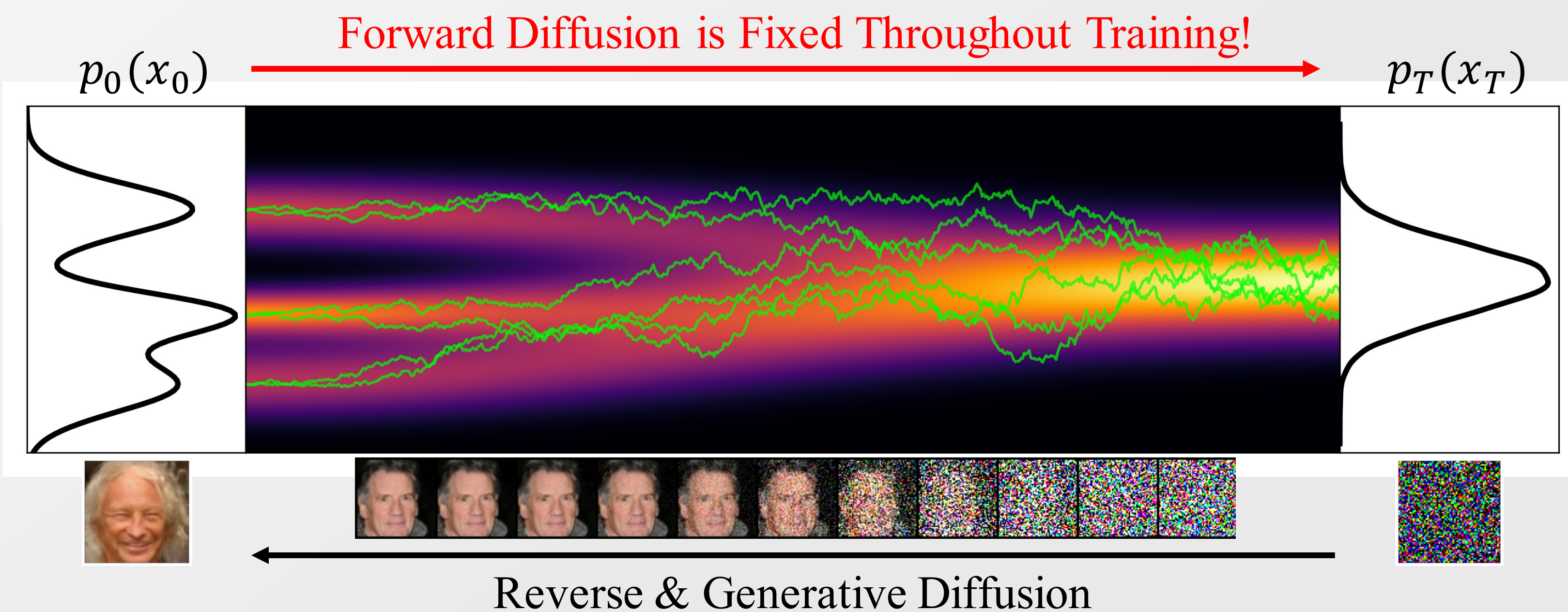




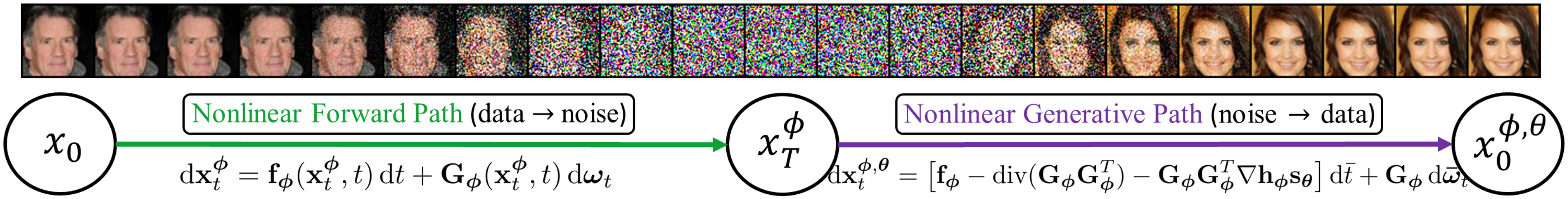
Motivation: VAE vs. Diffusion Model





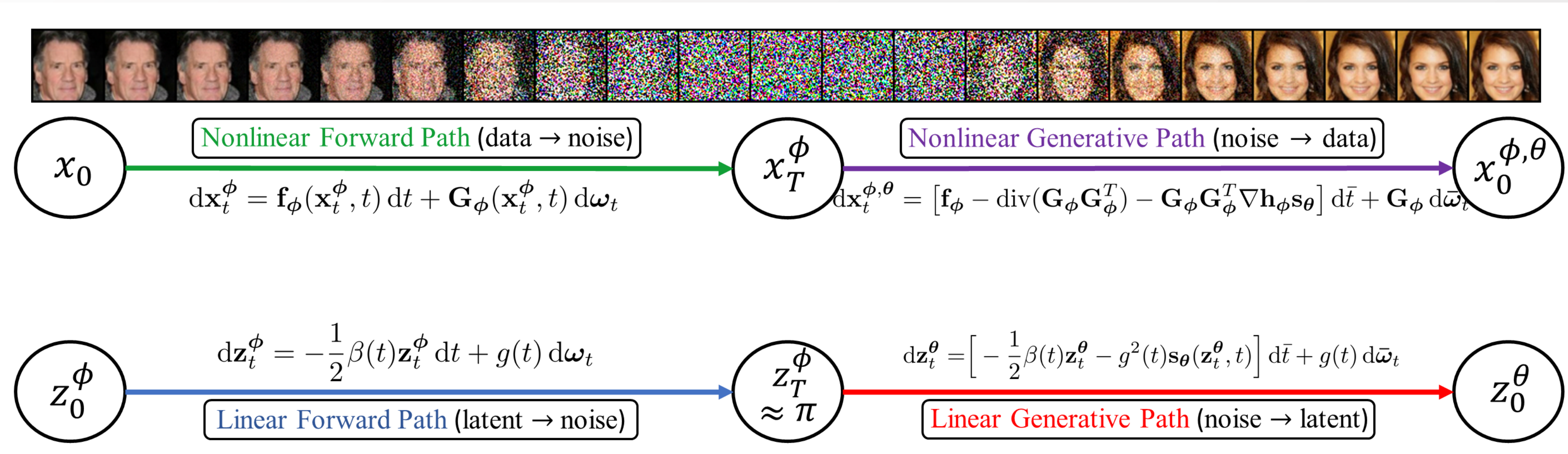


Implicit Nonlinear Diffusion Model



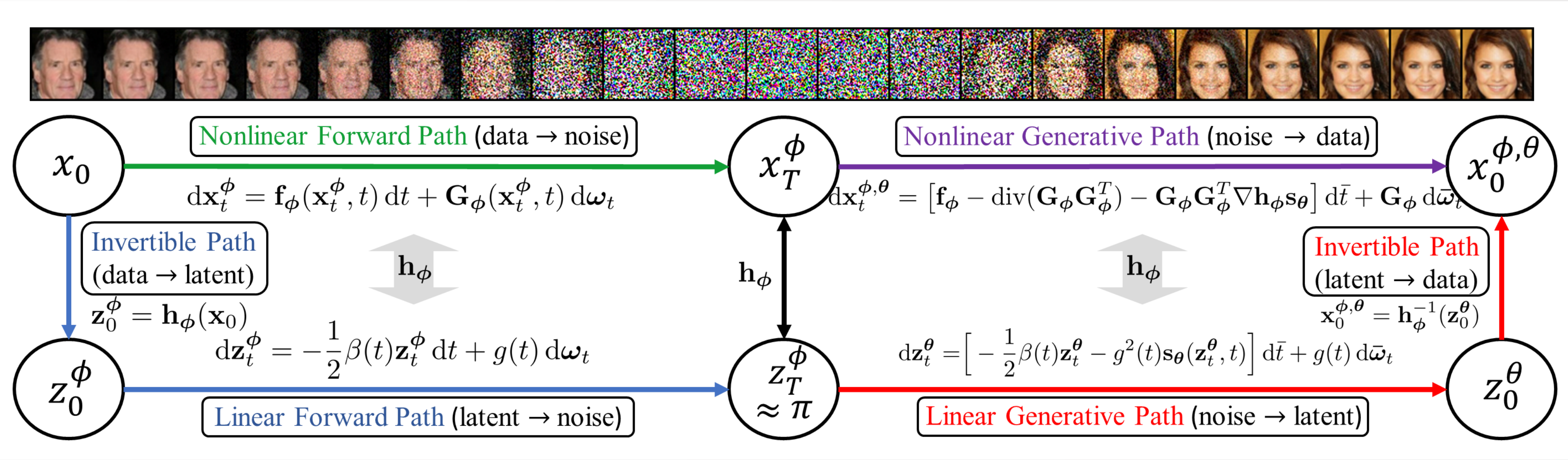
Nonlinear Diffusion on **Data** Space

Implicit Nonlinear Diffusion Model



Nonlinear Diffusion on **Data** Space =
Linear Diffusion on **Latent** Space +

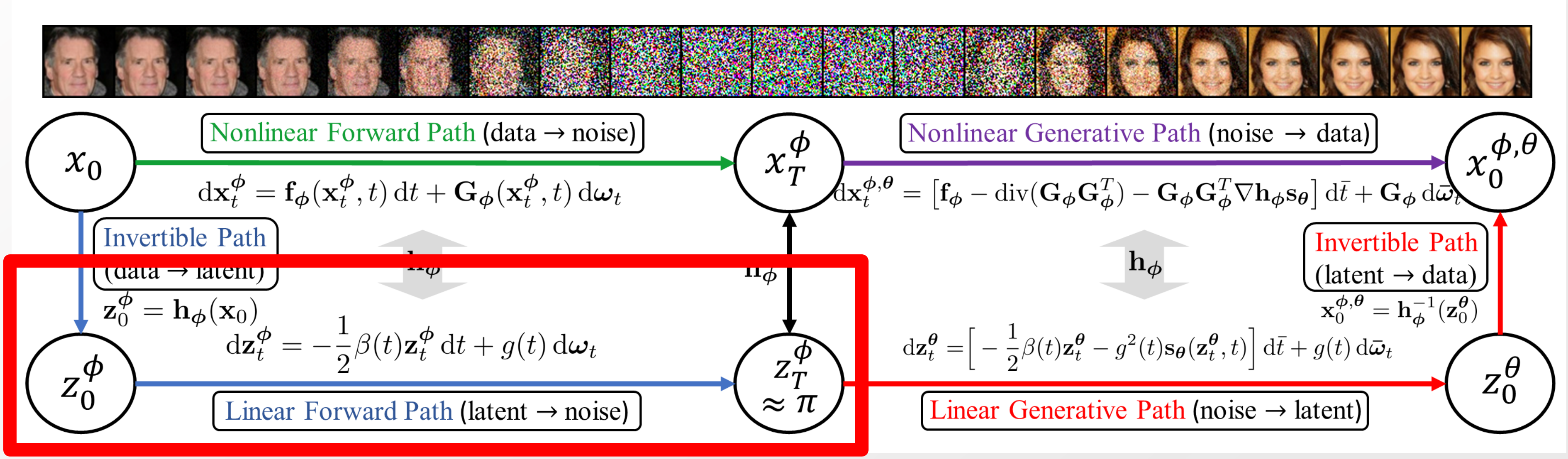
Implicit Nonlinear Diffusion Model



Nonlinear Diffusion on **Data** Space =

Linear Diffusion on **Latent** Space + *Invertible* **Transformation**

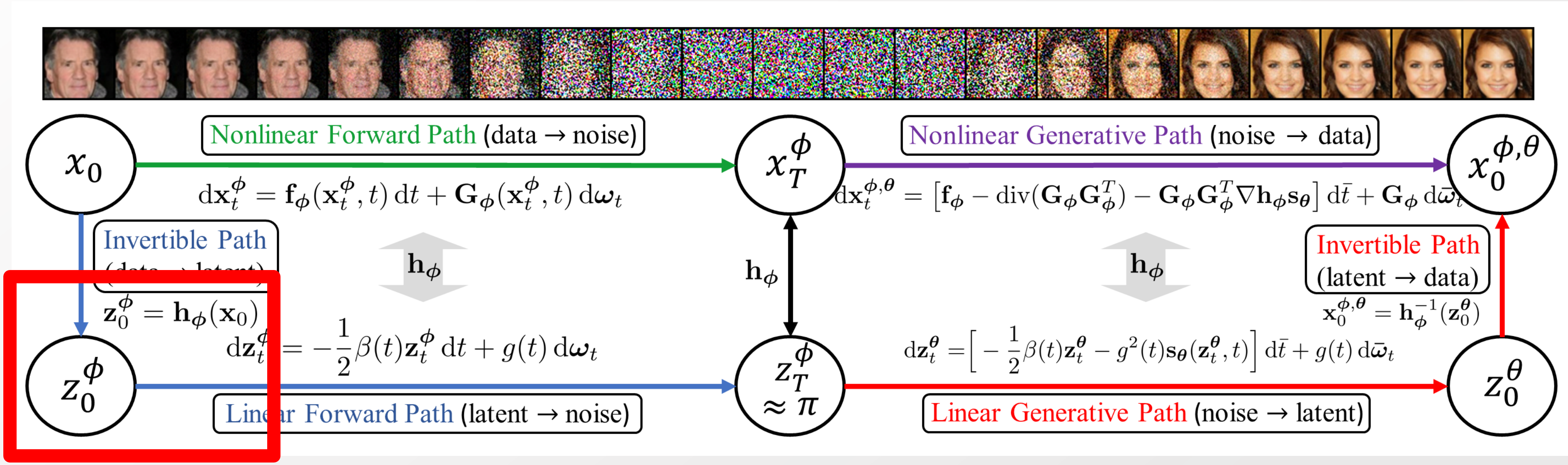
Implicit Nonlinear Diffusion Model



Linear Forward **Latent** Diffusion: $\left\{z_t^{\phi}\right\}_{t=0}^{T}$

ϕ: flow parameter

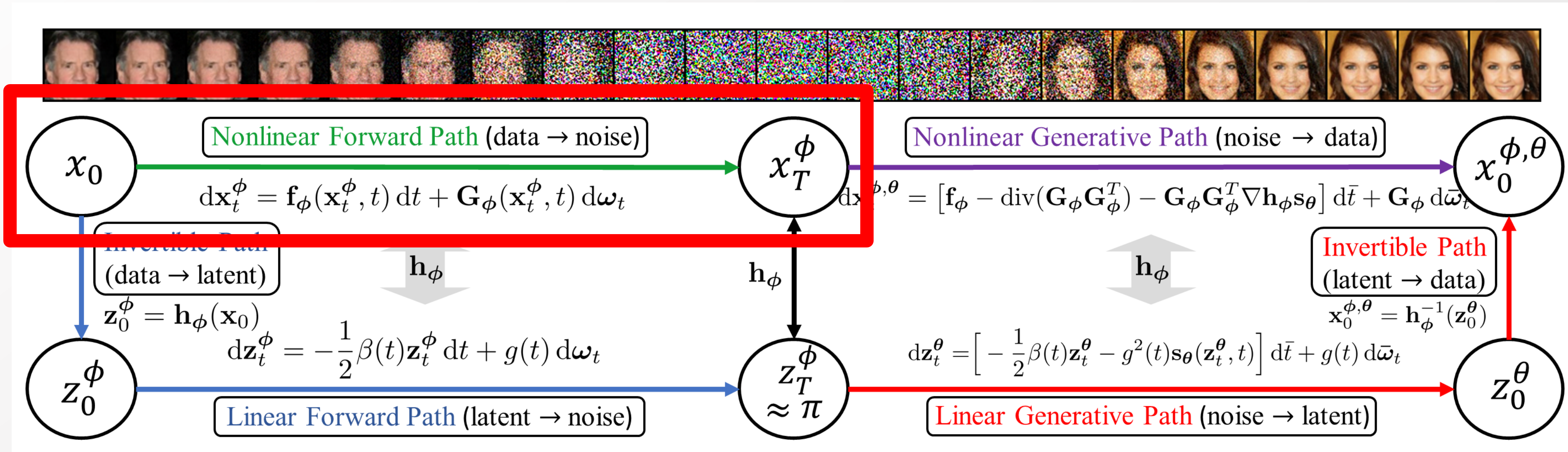
Implicit Nonlinear Diffusion Model



Linear Forward **Latent** Diffusion: $\left\{z_t^\phi\right\}_{t=0}^{T}$

ϕ: flow parameter

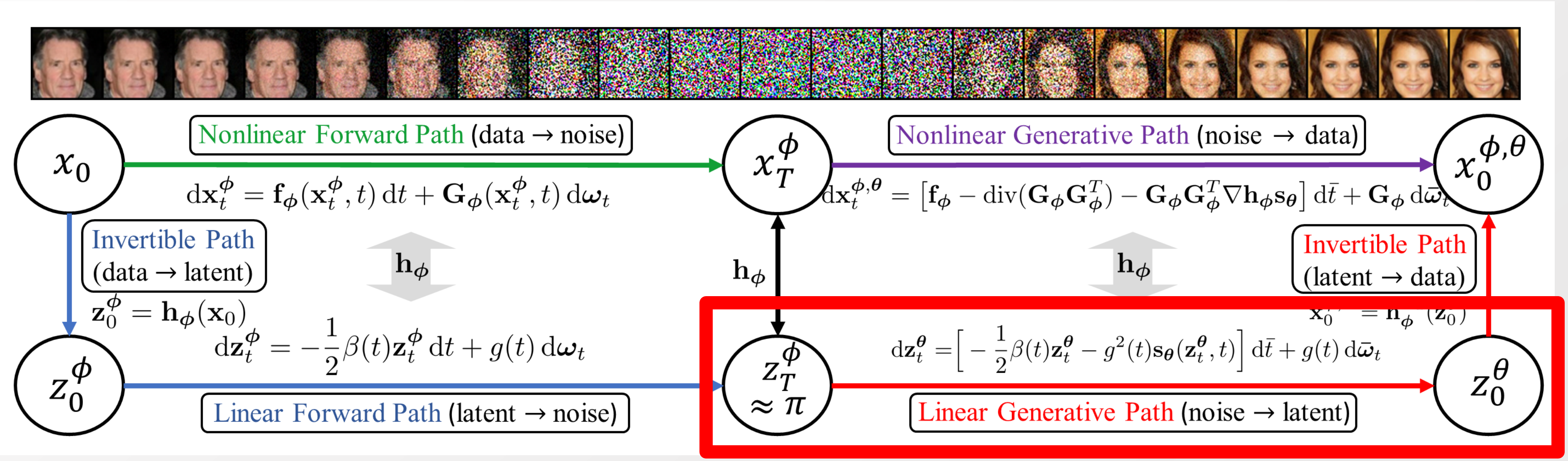
APPLIED ARTIFICIAL INTELLIGENCE LAB



Nonlinear Forward **Data** Diffusion: $\left\{x_t^\phi \coloneqq h_\phi^{-1}\left(z_t^\phi\right)\right\}_{t=0}^{T}$

ϕ: flow

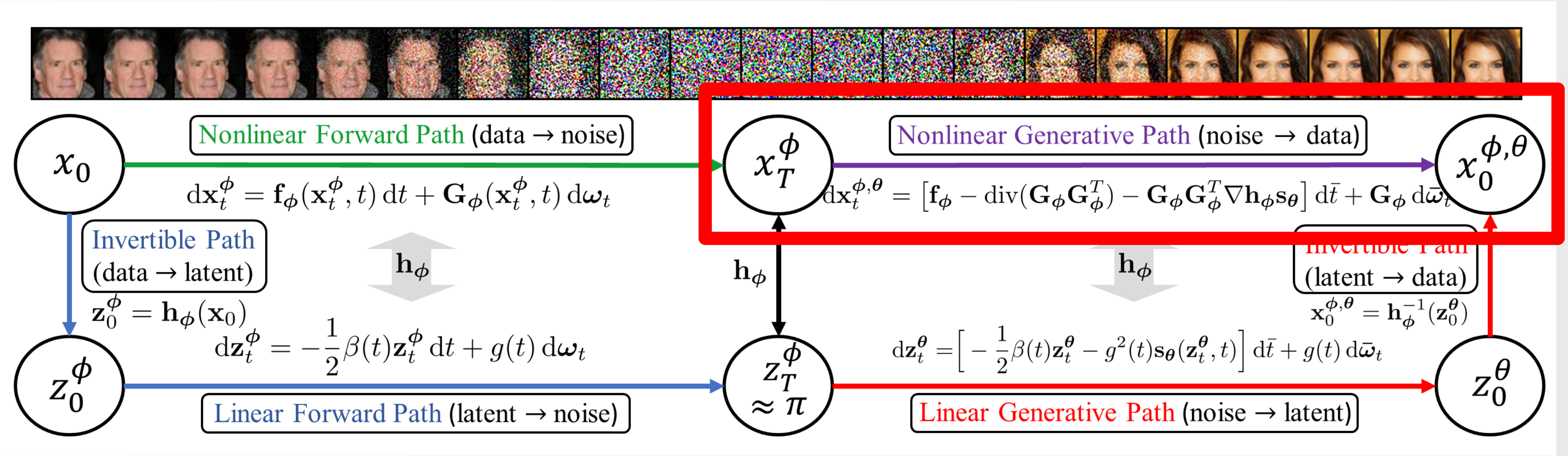
Implicit Nonlinear Diffusion Model



Linear Reverse **Latent** Diffusion: $\{z_t^\theta\}_{t=0}^T$

θ: score parameter

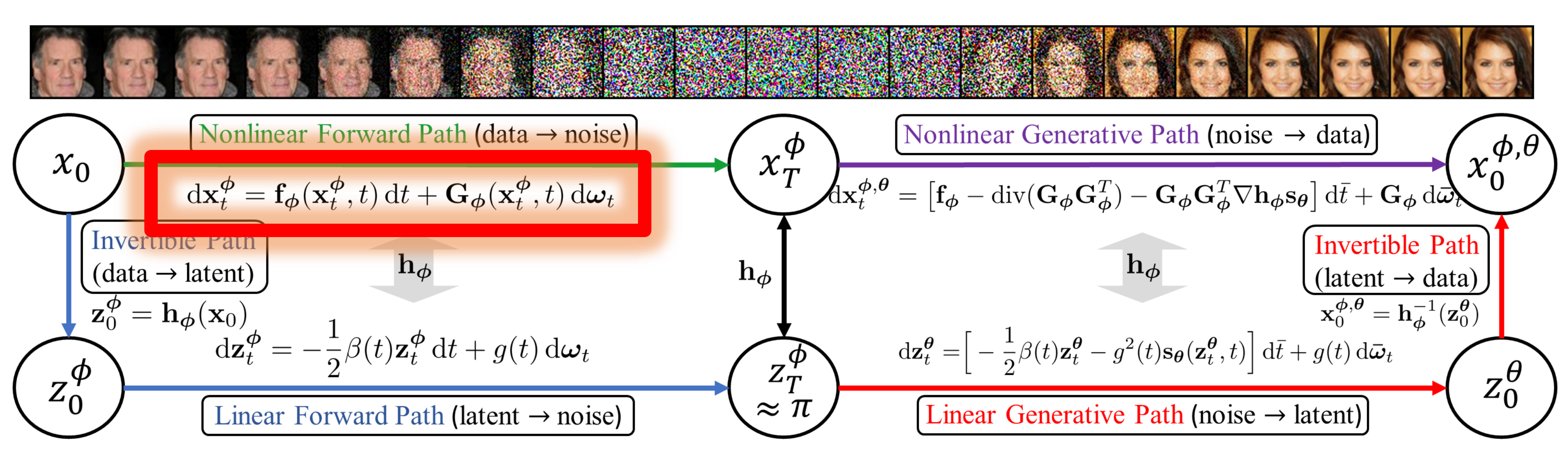
Implicit Nonlinear Diffusion Model



Nonlinear Reverse **Data** Diffusion: $\left\{x_t^{\phi,\theta} := h_{\phi}^{-1}(z_t^{\theta})\right\}_{t=0}^{T}$

ϕ: flow
θ: score

Implicit Nonlinear Diffusion Model

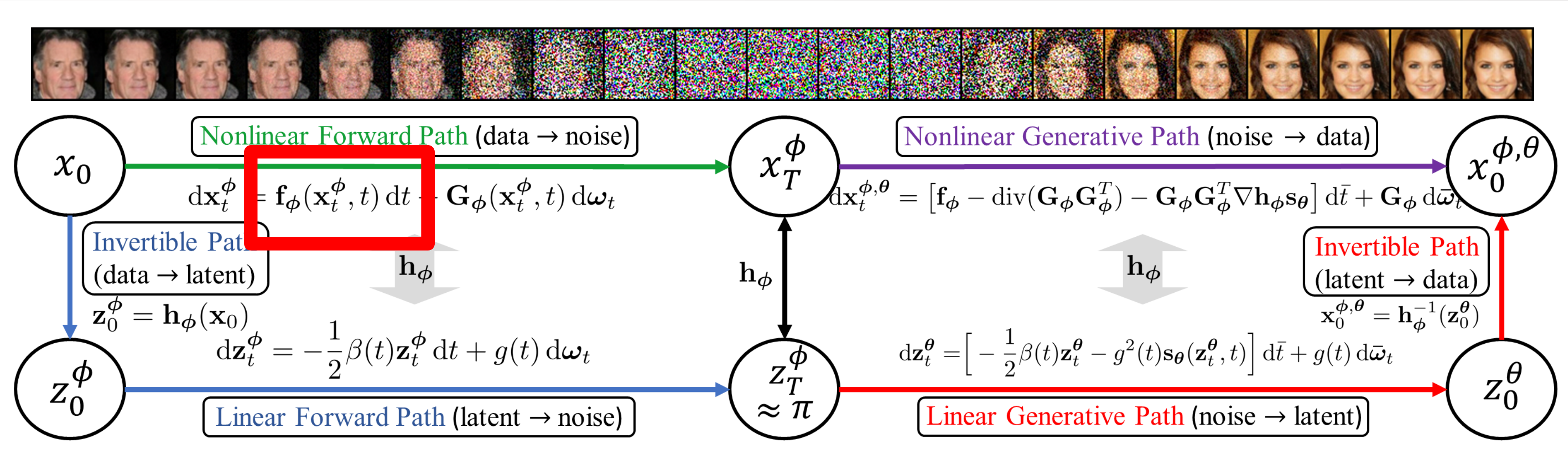


Nonlinear Forward **Data** Diffusion: $\left\{x_t^\phi := h_\phi^{-1}\left(z_t^\phi\right)\right\}_{t=0}^T$

Nonlinear Reverse **Data** Diffusion: $\left\{x_t^{\phi,\theta} := h_\phi^{-1}\left(z_t^\theta\right)\right\}_{t=0}^T$

ϕ: flow
θ: score

Implicit Nonlinear Diffusion Model

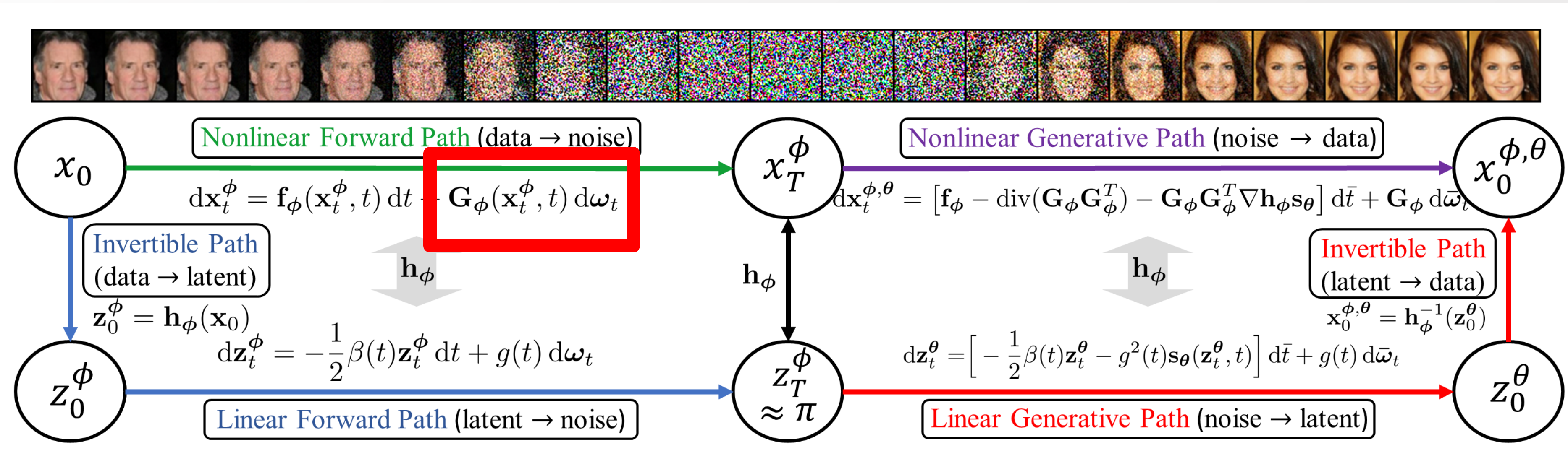


Nonlinear Forward **Data** Diffusion: $\left\{x_t^\phi := h_\phi^{-1}\left(z_t^\phi\right)\right\}_{t=0}^{T}$

Nonlinear Reverse **Data** Diffusion: $\left\{x_t^{\phi,\theta} := h_\phi^{-1}\left(z_t^\theta\right)\right\}_{t=0}^{T}$

ϕ: flow
θ: score

Implicit Nonlinear Diffusion Model

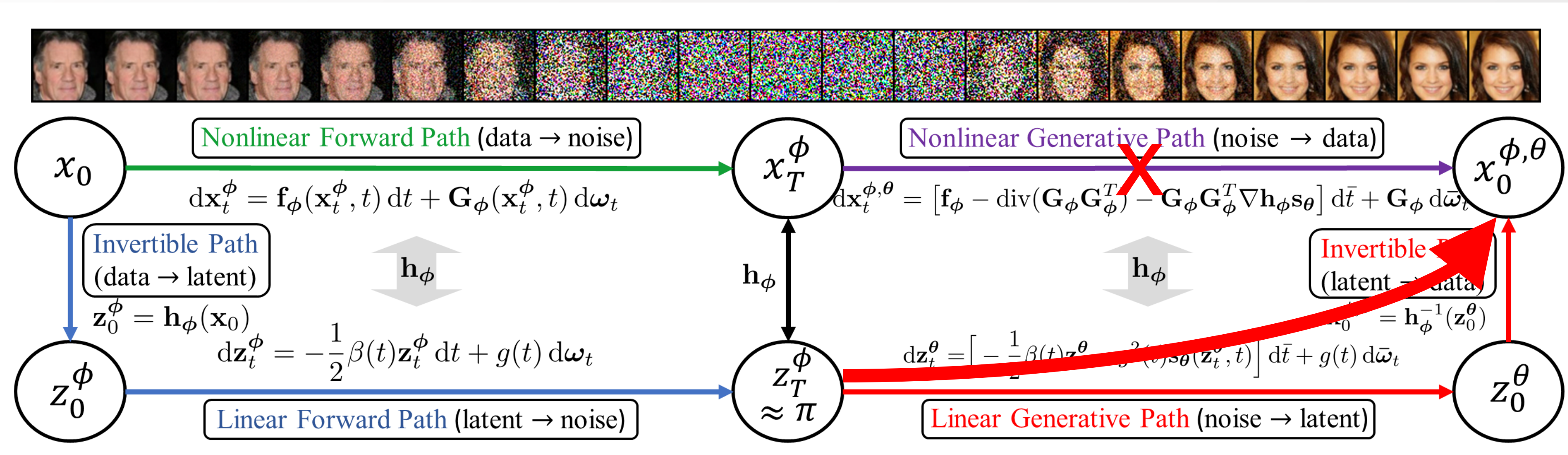


Nonlinear Forward **Data** Diffusion: $\left\{x_t^{\phi} := h_\phi^{-1}\left(z_t^{\phi}\right)\right\}_{t=0}^{T}$

Nonlinear Reverse **Data** Diffusion: $\left\{x_t^{\phi,\theta} := h_\phi^{-1}\left(z_t^{\theta}\right)\right\}_{t=0}^{T}$

ϕ: flow
θ: score

Implicit Nonlinear Diffusion Model



Nonlinear Forward **Data** Diffusion: $\left\{x_t^{\phi} := h_{\phi}^{-1}\left(z_t^{\phi}\right)\right\}_{t=0}^{T}$

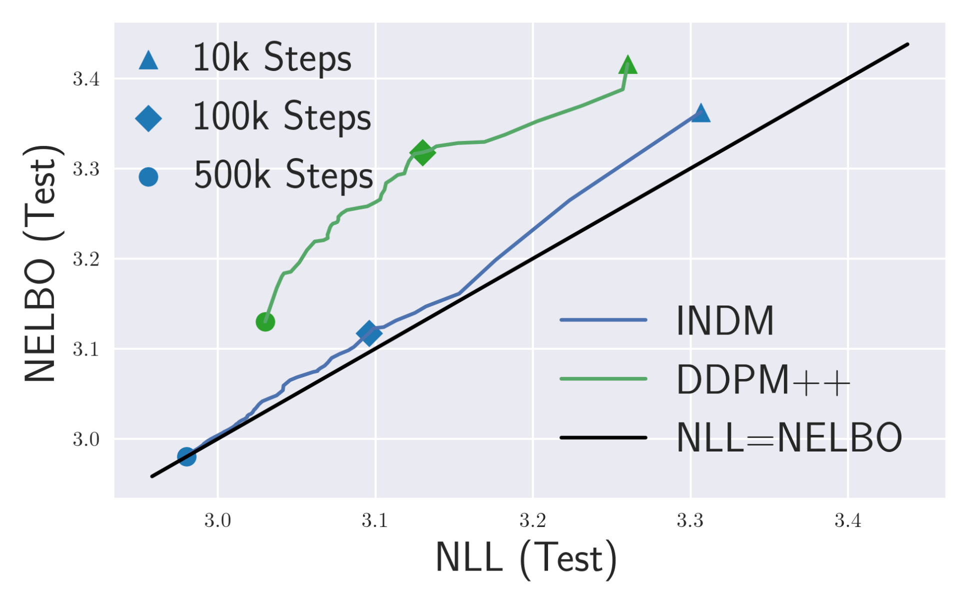
Nonlinear Reverse **Data** Diffusion: $\left\{x_t^{\phi,\theta} := h_{\phi}^{-1}(z_t^{\theta})\right\}_{t=0}^{T}$

ϕ: flow
θ: score

Characteristics of INDM

- INDM is the first continuous fully nonlinear diffusion model
  1. INDM training is fast
  2. INDM training is MLE
  3. INDM sampling is robust
  4. INDM enables image-to-image translation

Characteristics of INDM

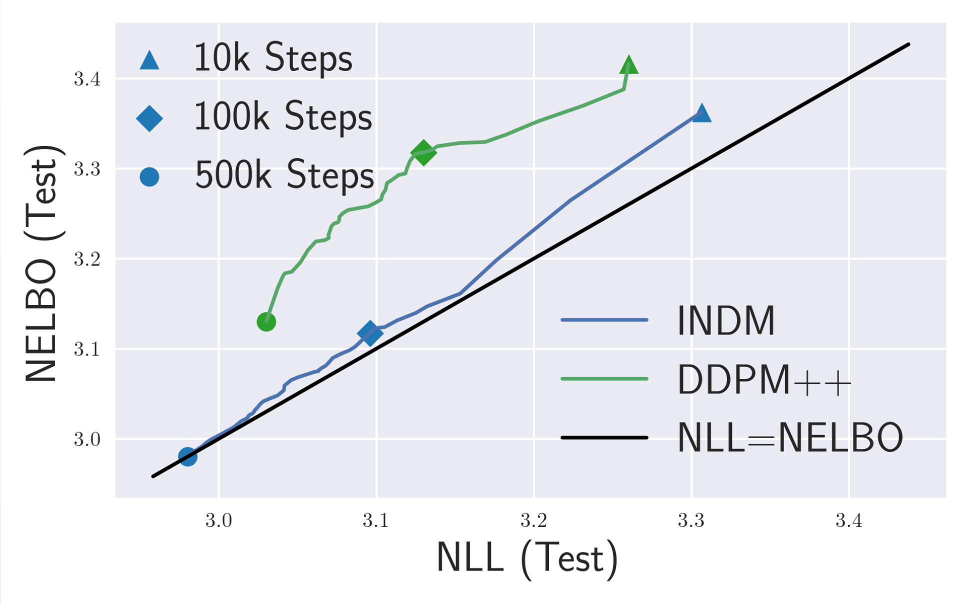


The learning curve of INDM is strictly under that of DDPM

Fast Training

Characteristics of INDM

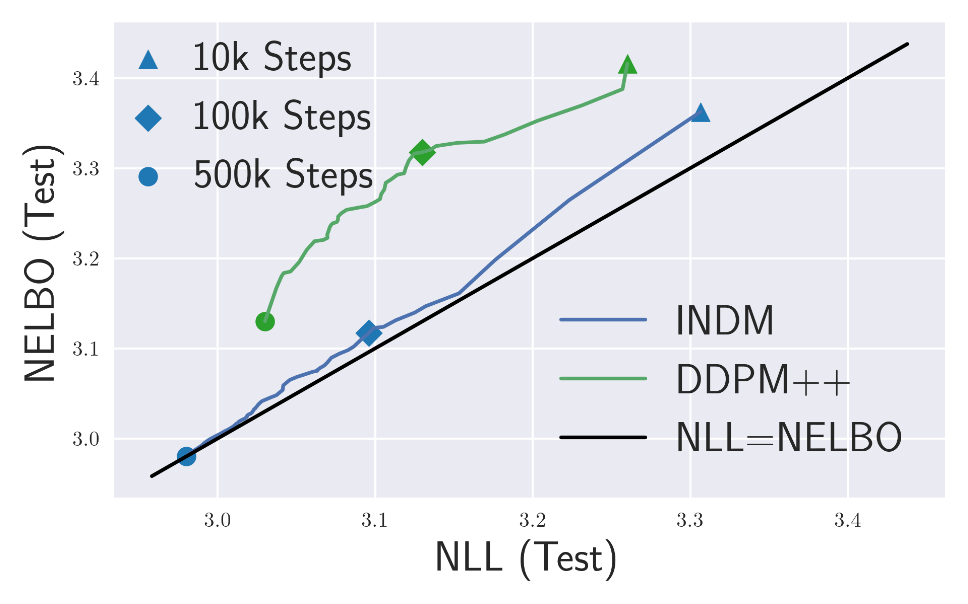
Fast Training | MLE Training | Robust Sampling | Image-to-Image Translation



Original Loss (linear diffusion)

Characteristics of INDM

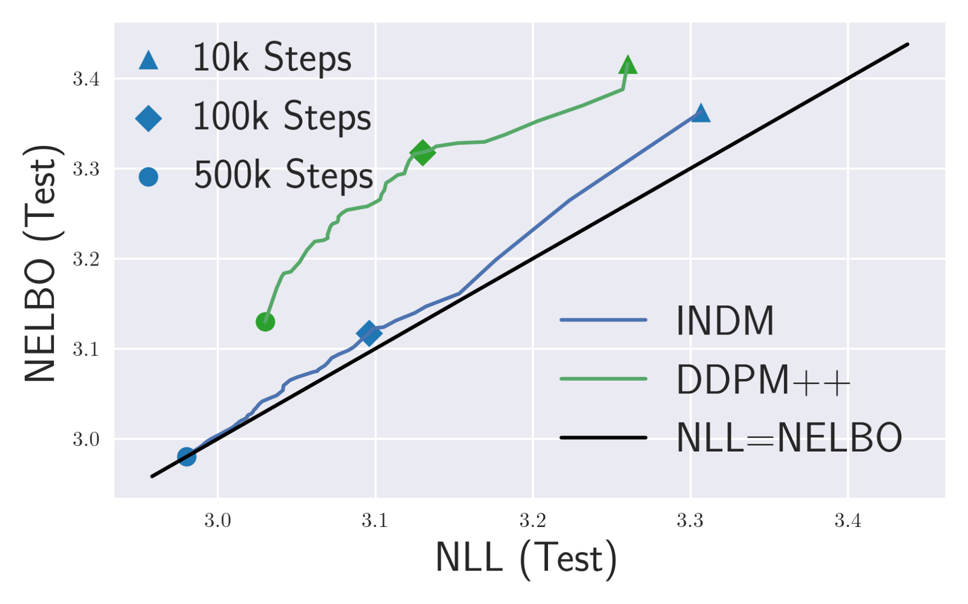
Fast Training | MLE Training | Robust Sampling | Image-to-Image Translation



Original Loss (linear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Characteristics of INDM

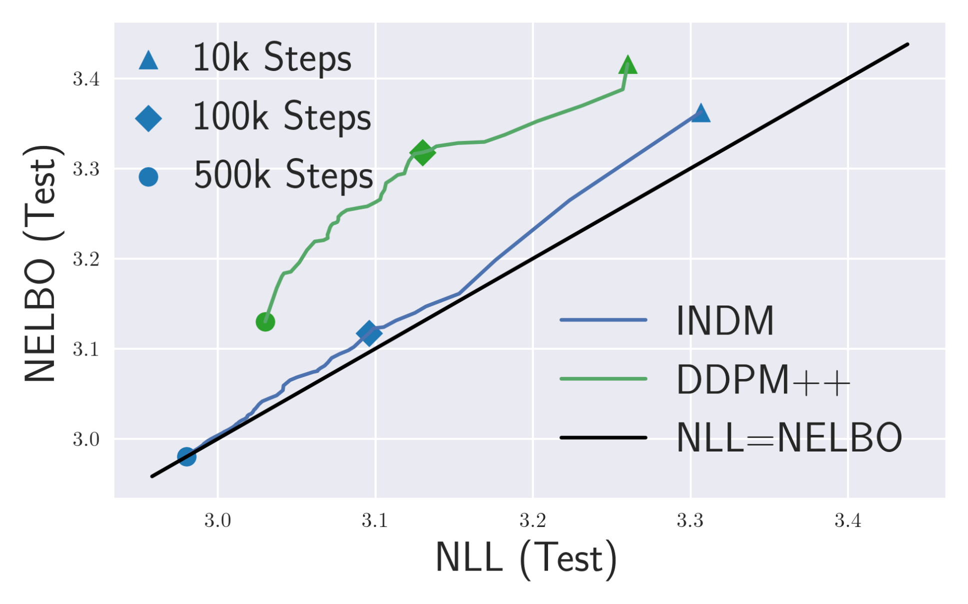


Original Loss (linear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Our Loss (nonlinear diffusion)

Characteristics of INDM



Original Loss (linear diffusion)

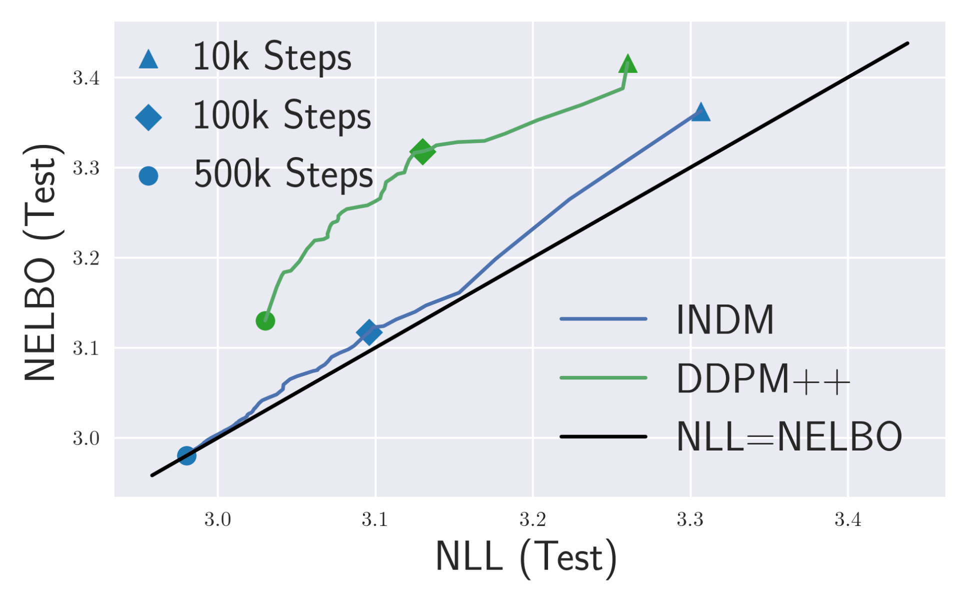
$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Our Loss (nonlinear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t^{\phi} - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Characteristics of INDM

Fast Training | MLE Training | Robust Sampling | Image-to-Image Translation



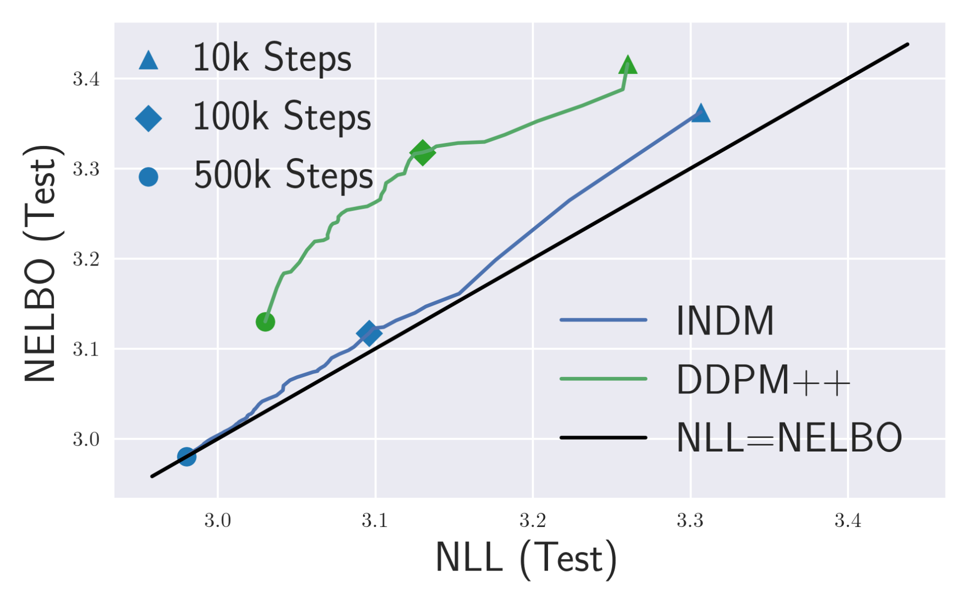
Original Loss (linear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Our Loss (nonlinear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t^{\phi} - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Characteristics of INDM



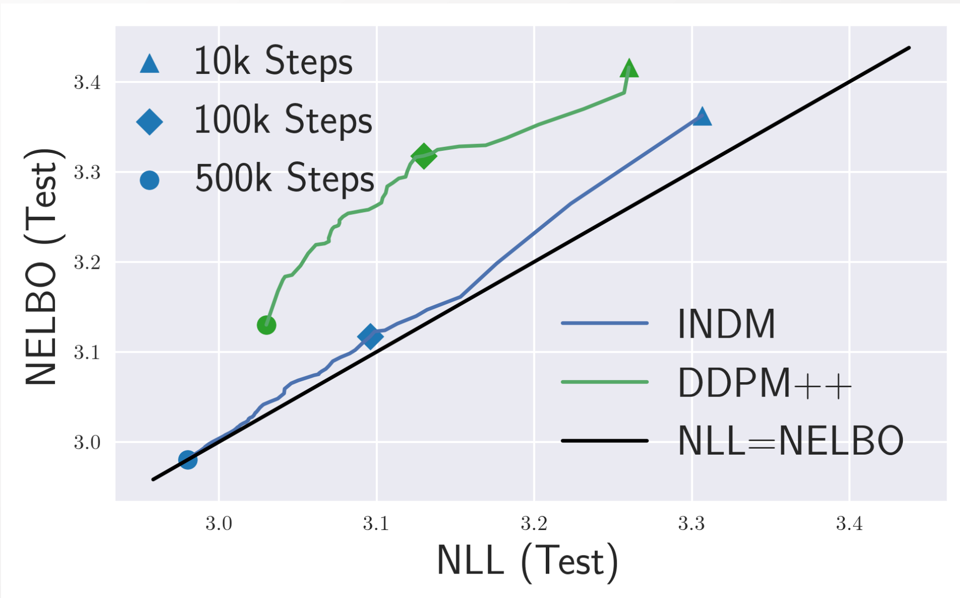
Original Loss (linear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Our Loss (nonlinear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t^{\phi} - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Characteristics of INDM



Original Loss (linear diffusion)

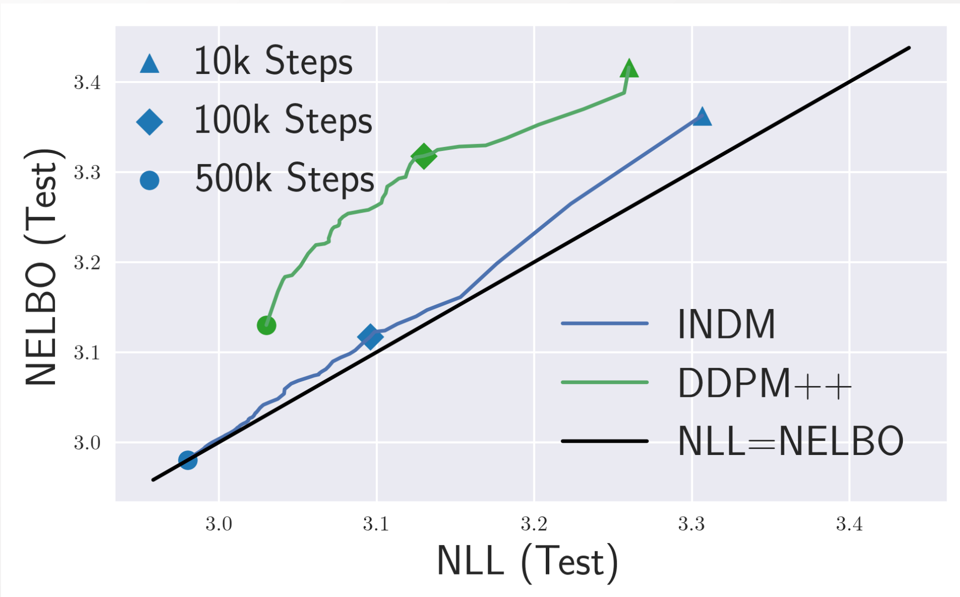
$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Our Loss (nonlinear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t^{\phi} - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Bidirectional

Characteristics of INDM



Original Loss (linear diffusion)

$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

Our Loss (nonlinear diffusion)

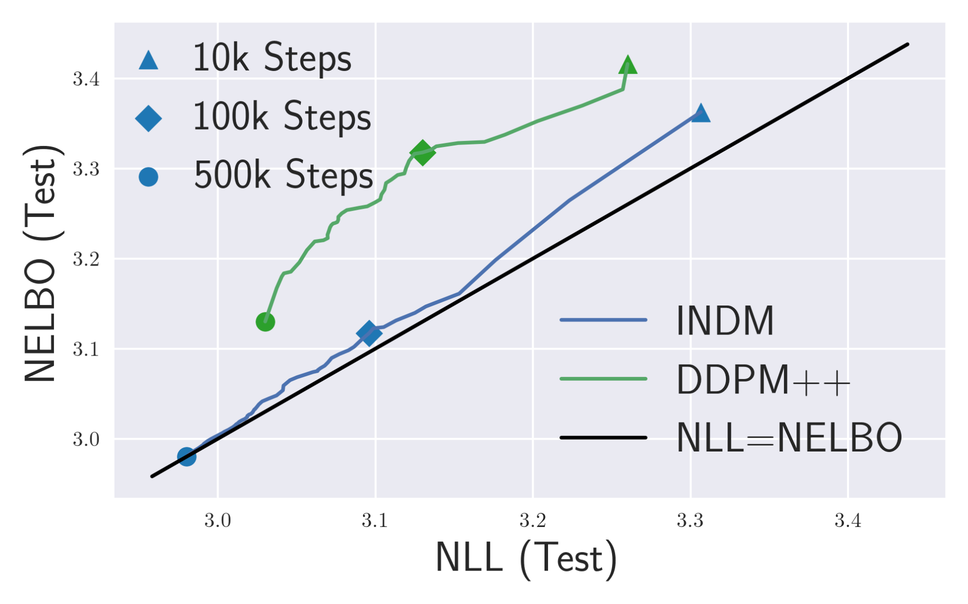
$$\int_0^T \lambda(t)\mathbb{E}[\|\nabla \log p_t^{\phi} - \mathbf{s}_{\boldsymbol{\theta}}\|_2^2]\,\mathrm{d}t$$

→ ←

Bidirectional

Fast Training

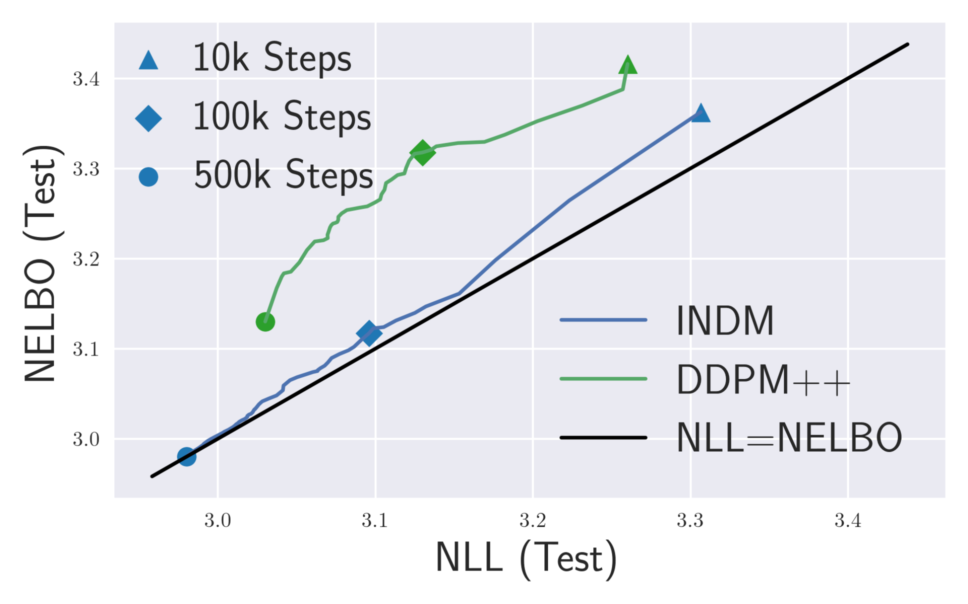
Characteristics of INDM



The learning curve of INDM is close to the line of MLE training

MLE Training (!)

Characteristics of INDM

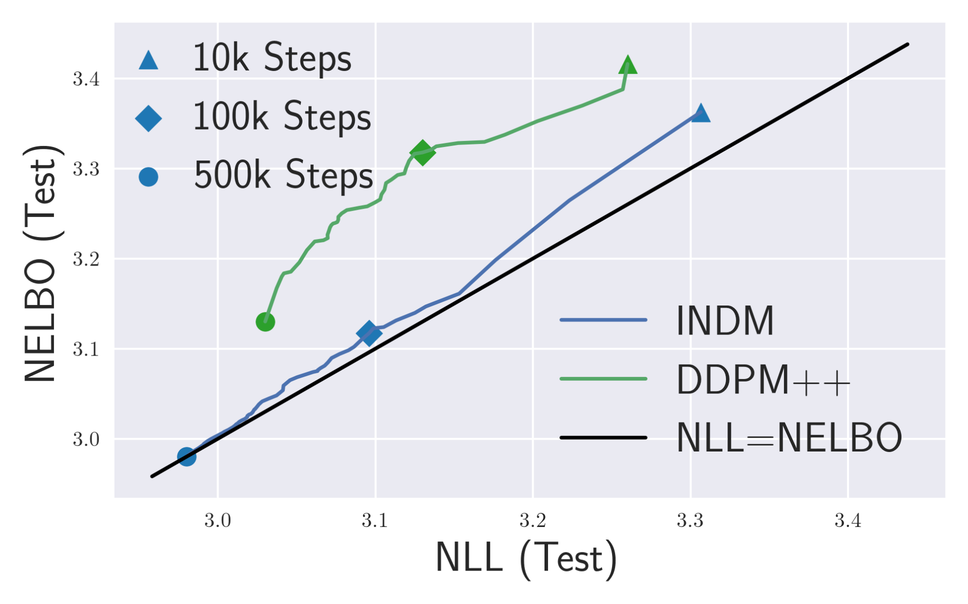


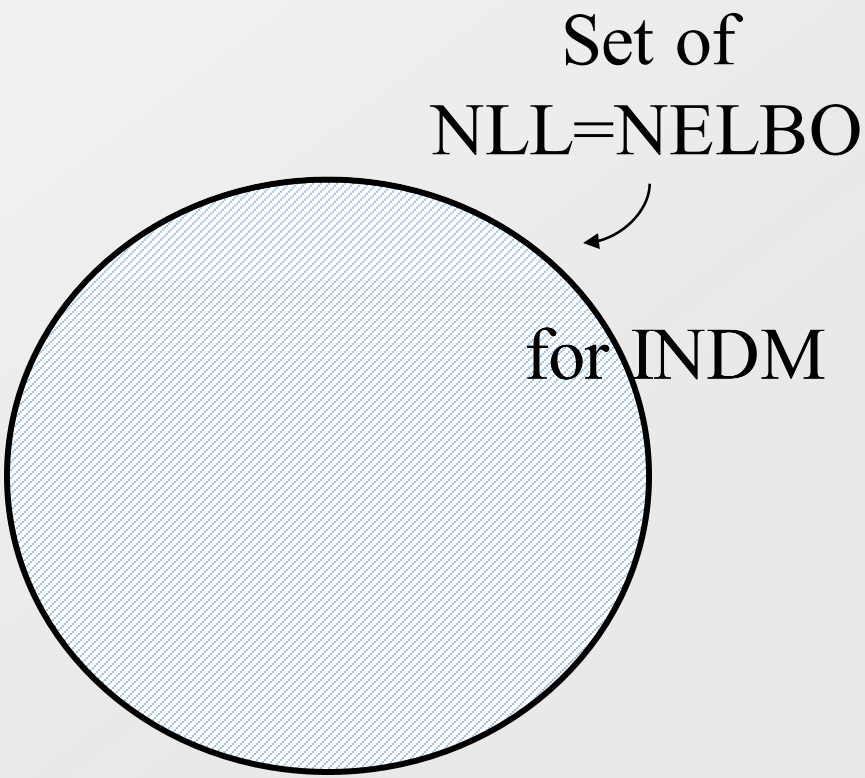
NELBO

The closer the better

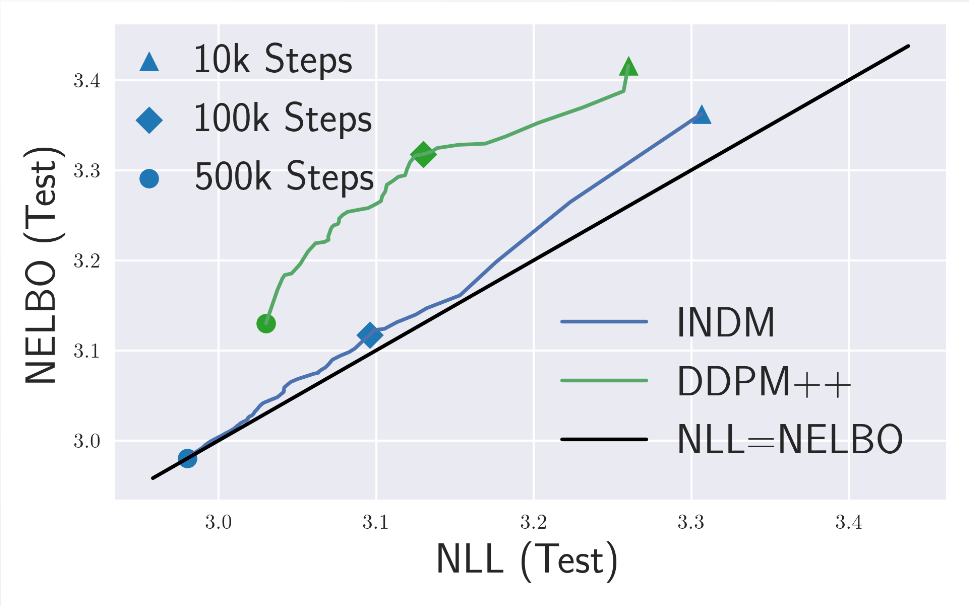
NLL

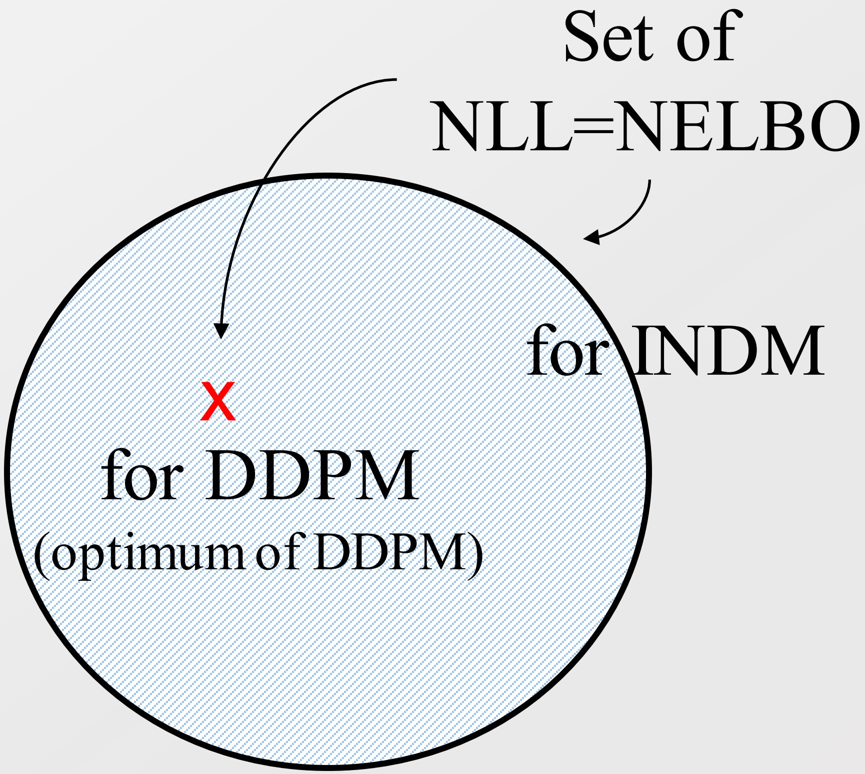
Characteristics of INDM



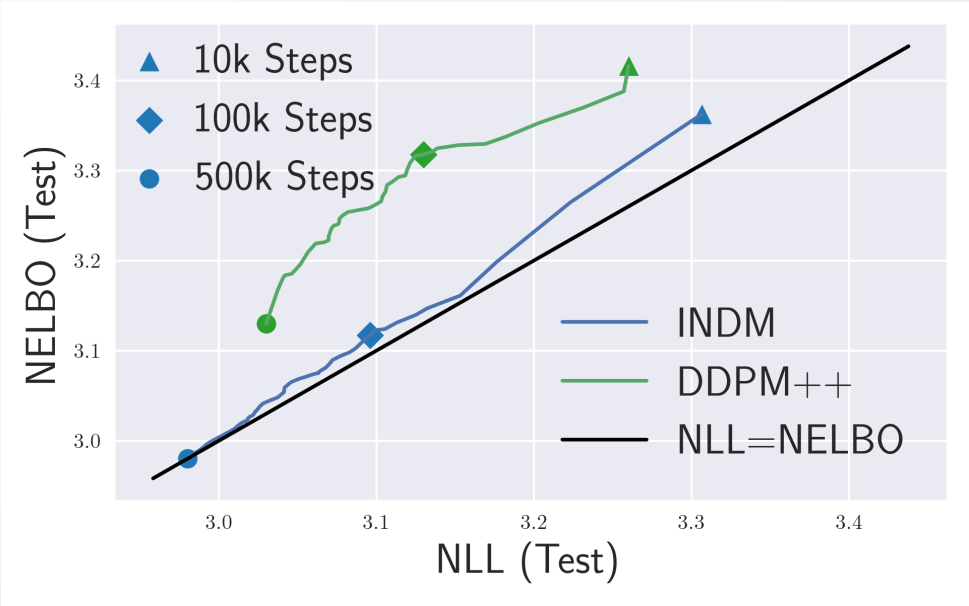


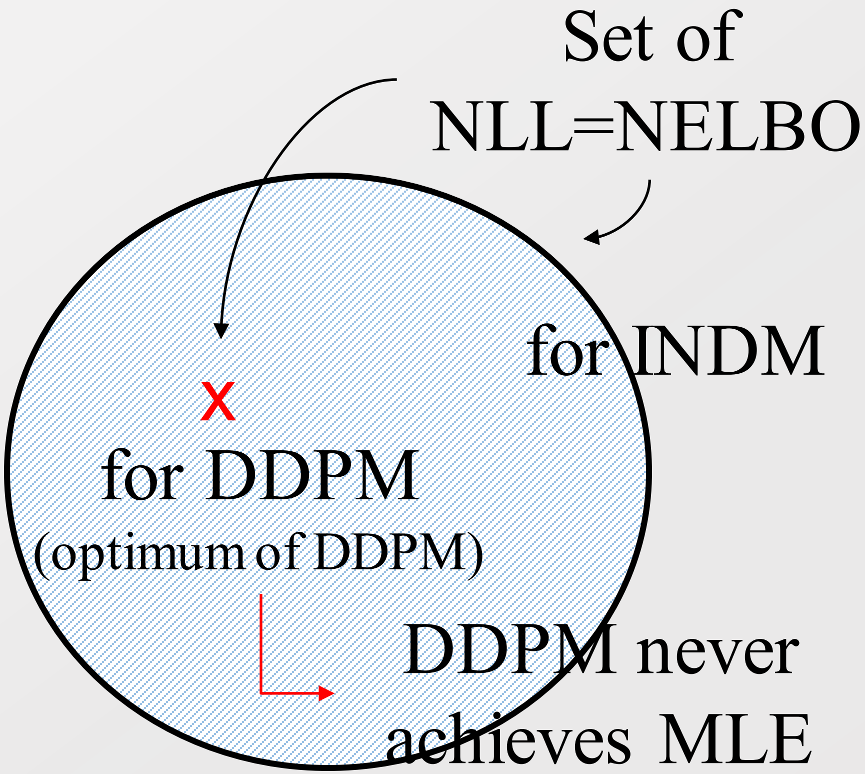
Characteristics of INDM





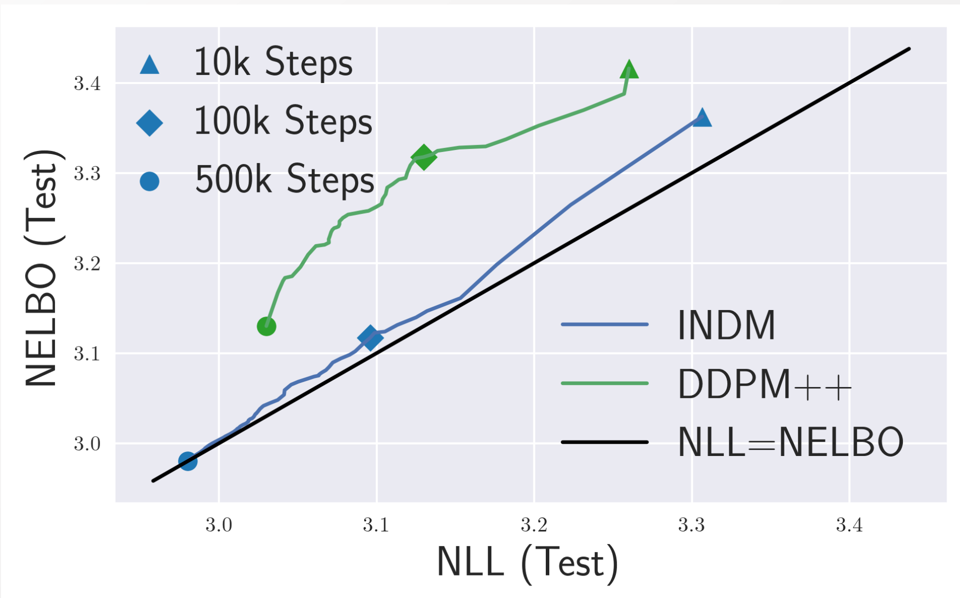
Characteristics of INDM

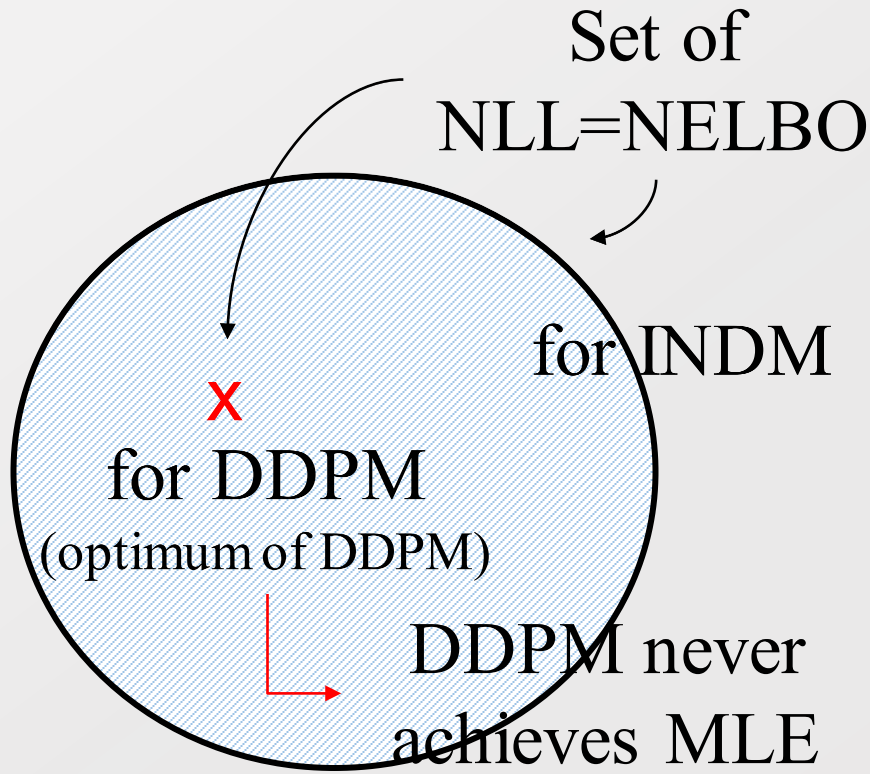




Characteristics of INDM

Fast Training | **MLE Training** | Robust Sampling | Image-to-Image Translation

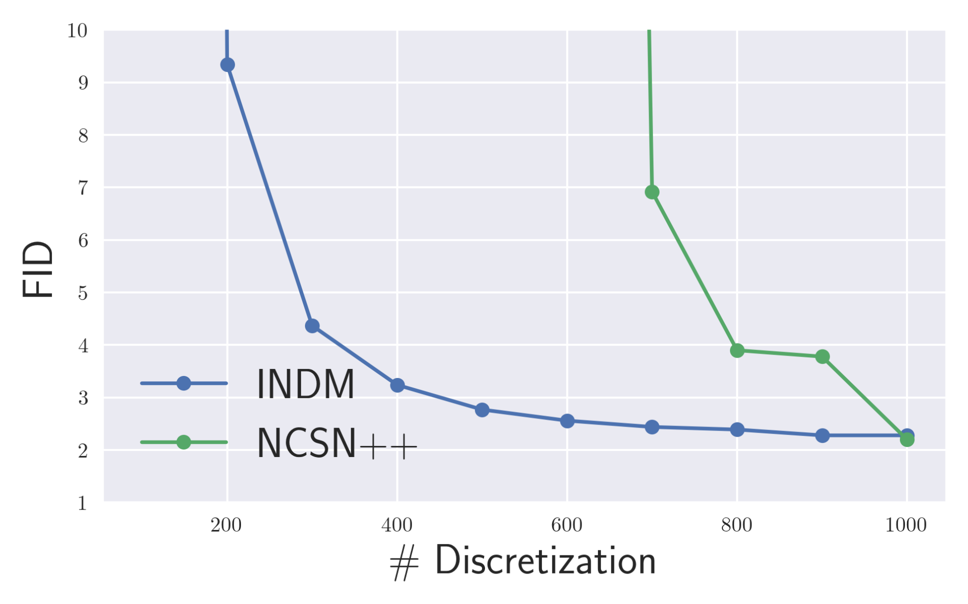




MLE Training (!)

Characteristics of INDM

Fast Training | MLE Training | **Robust Sampling** | Image-to-Image Translation

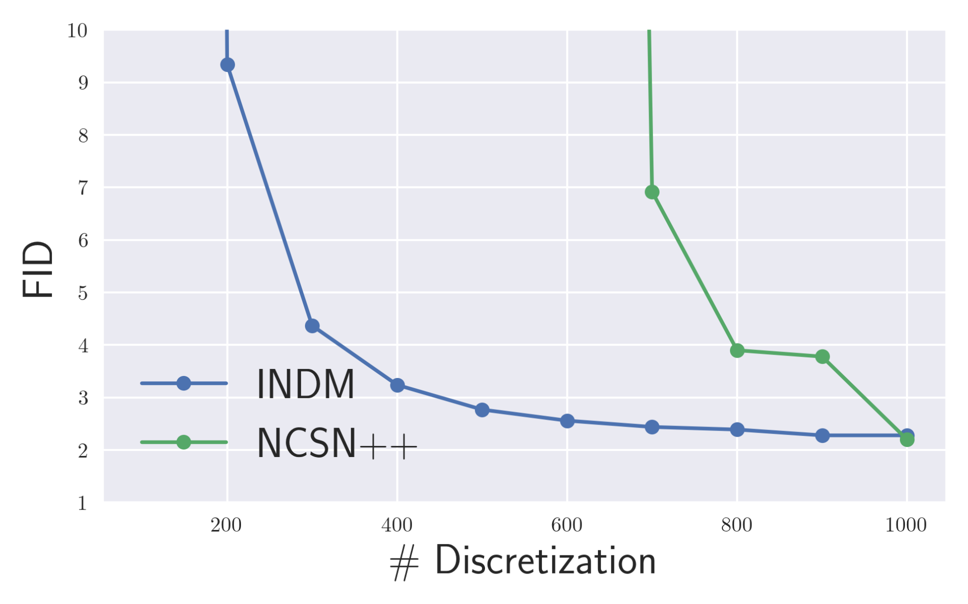


The sample quality of INDM is robust on the number of discretization steps

Robust Sampling

Characteristics of INDM

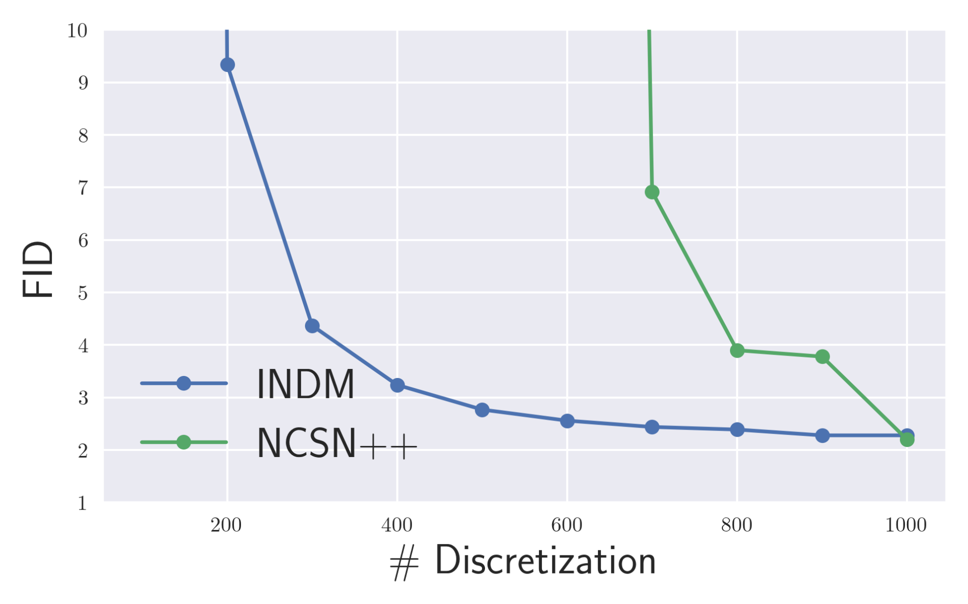
Fast Training | MLE Training | **Robust Sampling** | Image-to-Image Translation



Theorem. *If* p_g *is sample distribution*

$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

Characteristics of INDM

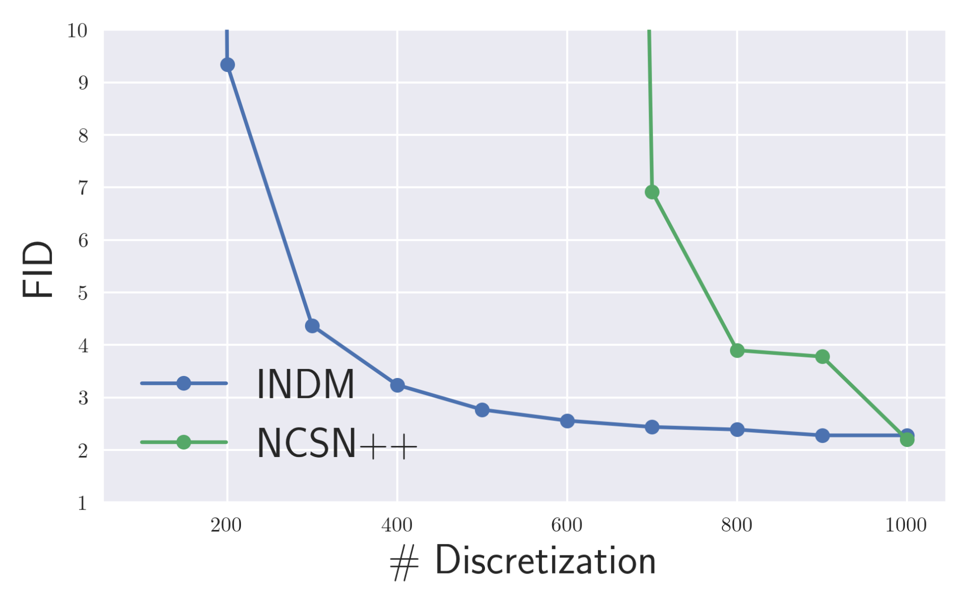


Theorem. *If* p_g *is sample distribution*

$$\|p_r - {\color{red}p_g}\|_{TV} \le E_{prior} + E_{disc} + E_{est}$$

Characteristics of INDM

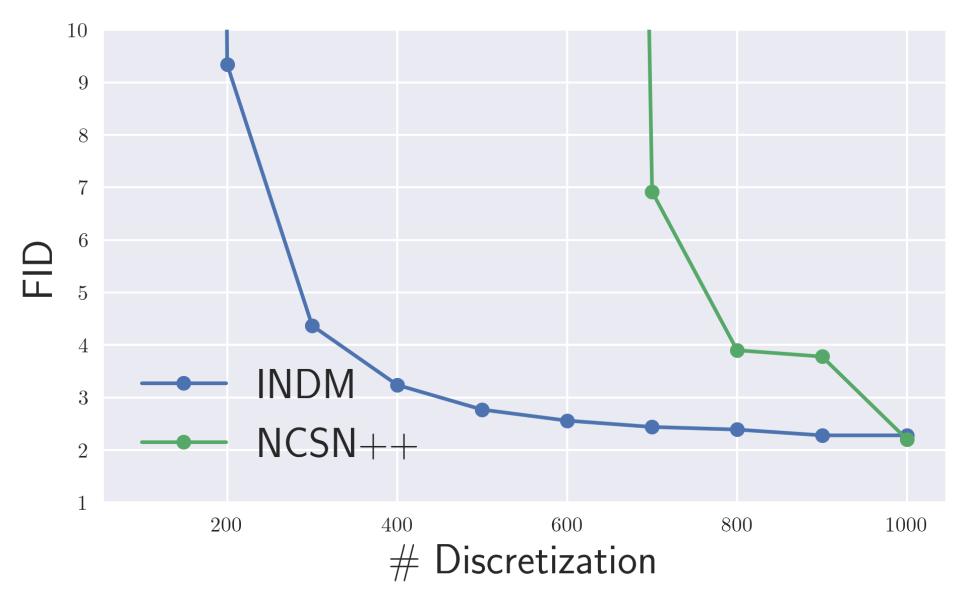
Fast Training | MLE Training | **Robust Sampling** | Image-to-Image Translation



Theorem. *If* p_g *is sample distribution*

$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

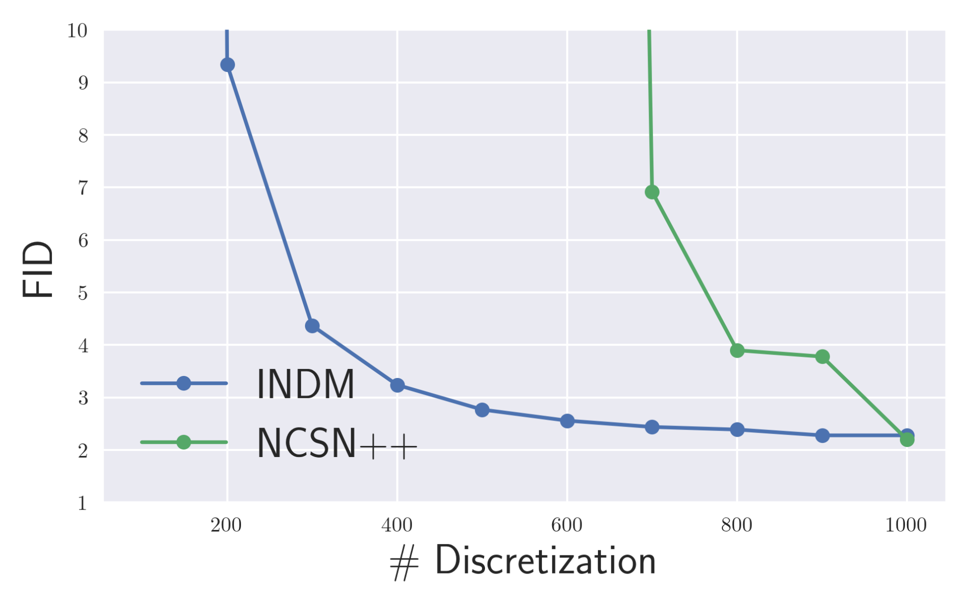
Characteristics of INDM



Theorem. *If* p_g *is sample distribution*

$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

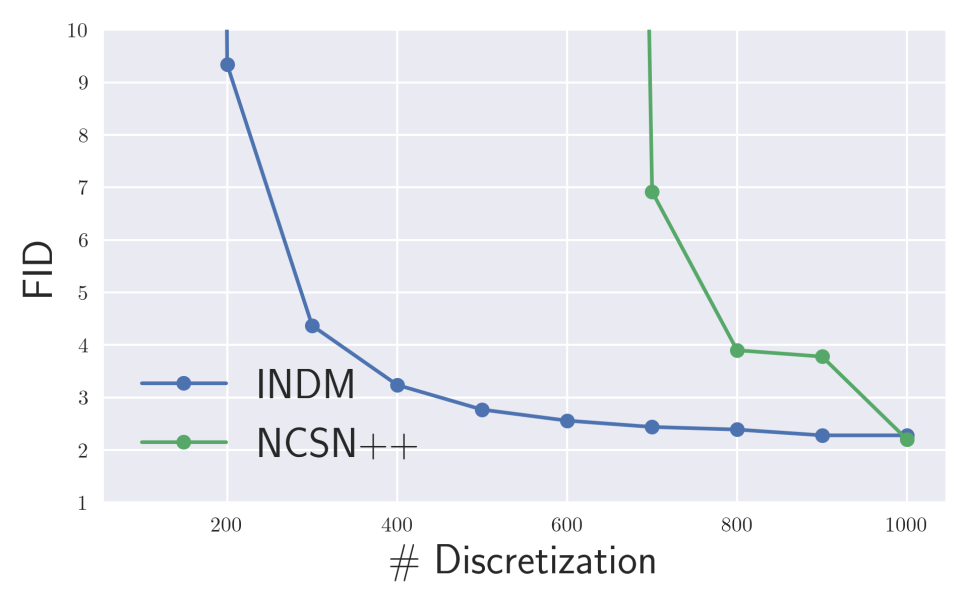
Characteristics of INDM



Theorem. *If p_g is sample distribution*

$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

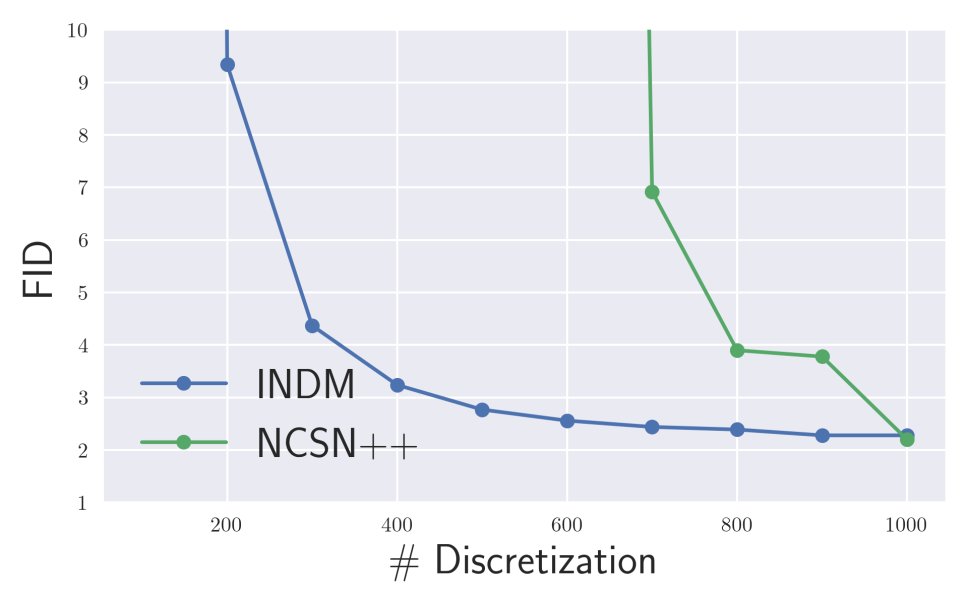
Characteristics of INDM



Theorem. *If p_g is sample distribution*

$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

Characteristics of INDM



Theorem. *If* p_g *is sample distribution*

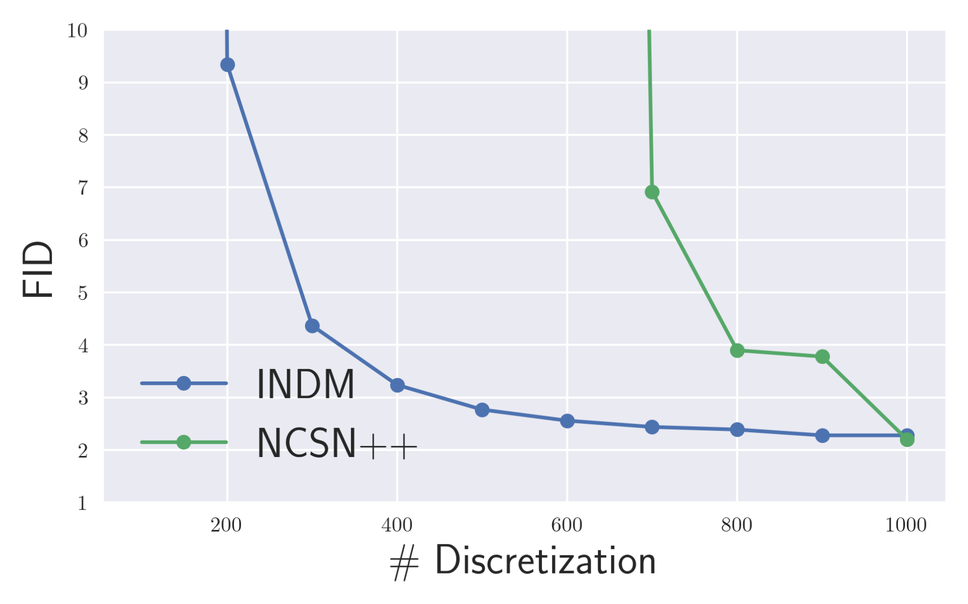
$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

⇩

$$E_{disc}(INDM) < E_{disc}(DDPM)$$

Characteristics of INDM

Fast Training MLE Training **Robust Sampling** Image-to-Image Translation



Theorem. *If p_g is sample distribution*

$$\|p_r - p_g\|_{TV} \leq E_{prior} + E_{disc} + E_{est}$$

$$E_{disc}(INDM) < E_{disc}(DDPM)$$

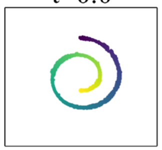
Robust Sampling

Characteristics of INDM

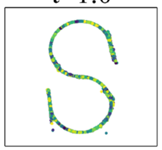
Fast Training | MLE Training | Robust Sampling | **Image-to-Image Translation**

Data Space

t=0.0



t=1.0



Conditional Diffusion Model
Or ≥ 2 Diffusion Models

Dog ↔ Cat

Characteristics of INDM

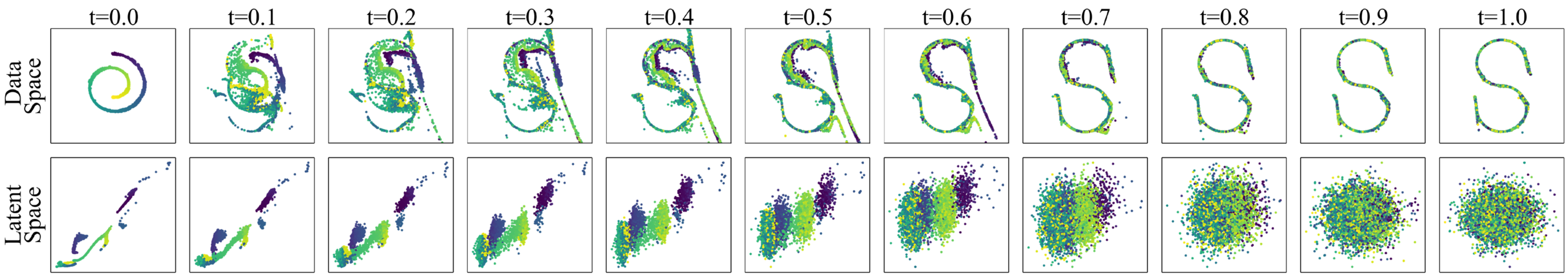
Fast Training | MLE Training | Robust Sampling | **Image-to-Image Translation**

~~Conditional Diffusion Model~~
~~Or ≥ 2 Diffusion Models~~

Dog ↔ Cat

Characteristics of INDM

Fast Training | MLE Training | Robust Sampling | **Image-to-Image Translation**

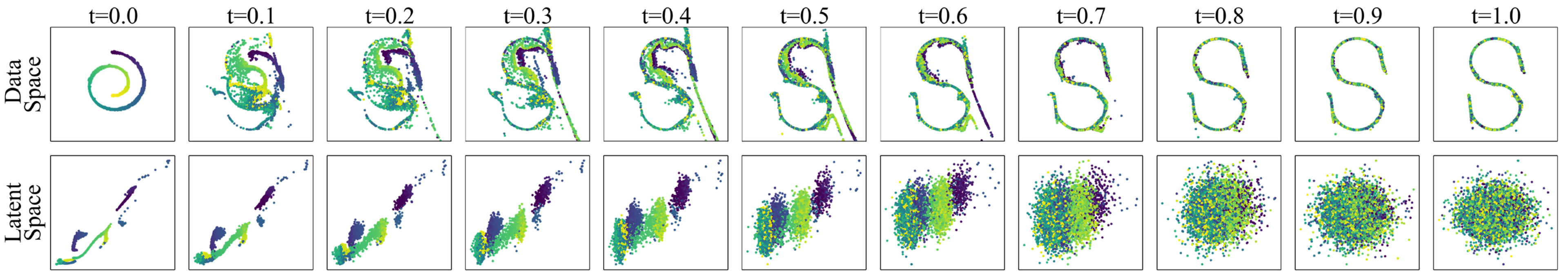


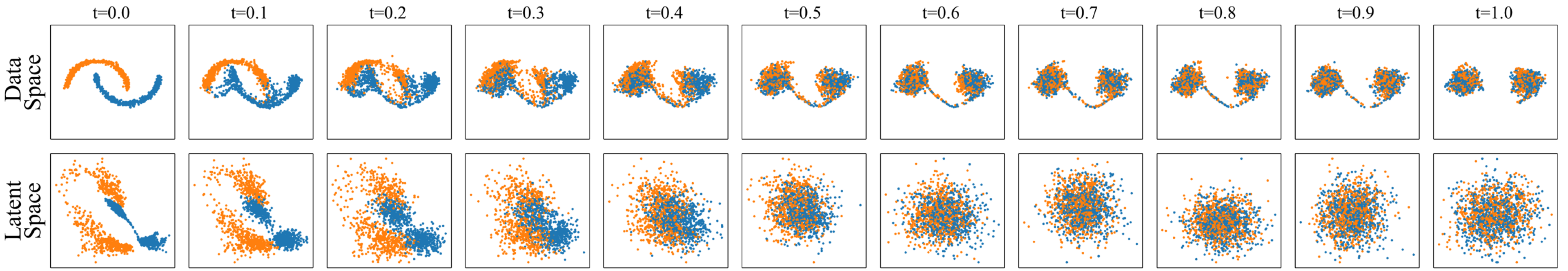
~~Conditional Diffusion Model~~
~~Or ≥ 2 Diffusion Models~~

Dog ↔ Cat

Characteristics of INDM

Fast Training   MLE Training   Robust Sampling   Image-to-Image Translation





Summary of Contribution

INDM is a Nonlinear Diffusion Model

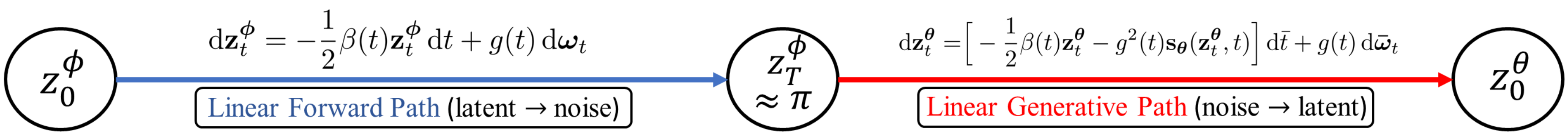
Future Works of INDM

APPLIED ARTIFICIAL INTELLIGENCE LAB

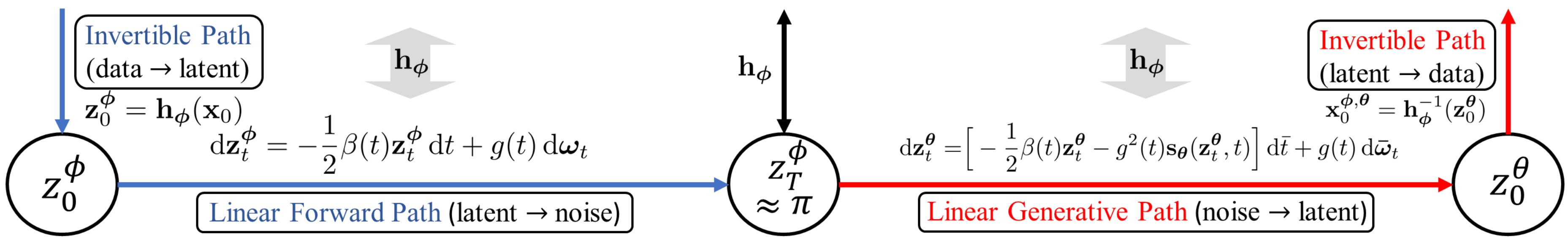
- The motivation of nonlinear diffusion in high-dimensional dataset is not sufficient.
- The invertible transformation is modeled by a flow network, which is the speed/performance bottleneck after all.
- The destined variable of INDM is not a standard Gaussian in general, and this difference could arise a qualitatively different behavior.
- The nonlinearity is purely subject to the optimization, and the behavior of the trained forward diffusion is not investigated or controllable, so far.
- The drift and volatility coefficients are highly entangled with a flow model of which flexibility is potentially limited.
- The scope of nonlinearity needs to be examined more clearly.
- The nonlinear diffusion has not been tested for the higher-dimensional dataset, such as ImageNet-256.
- The flow seems not take any role other than colorization, and further research on the role of flow network remains.
- The model works better with the pre-training of linear diffusions.
- The further analysis on why INDM fails to converge, if we use Glow-based flows instead of ResNet-based flows, is left.
- Whether or not the essential input information is retained longer than the linear diffusion with INDM to make it use in the meaningful latent extraction.

Thank you

Implicit Nonlinear Diffusion Model

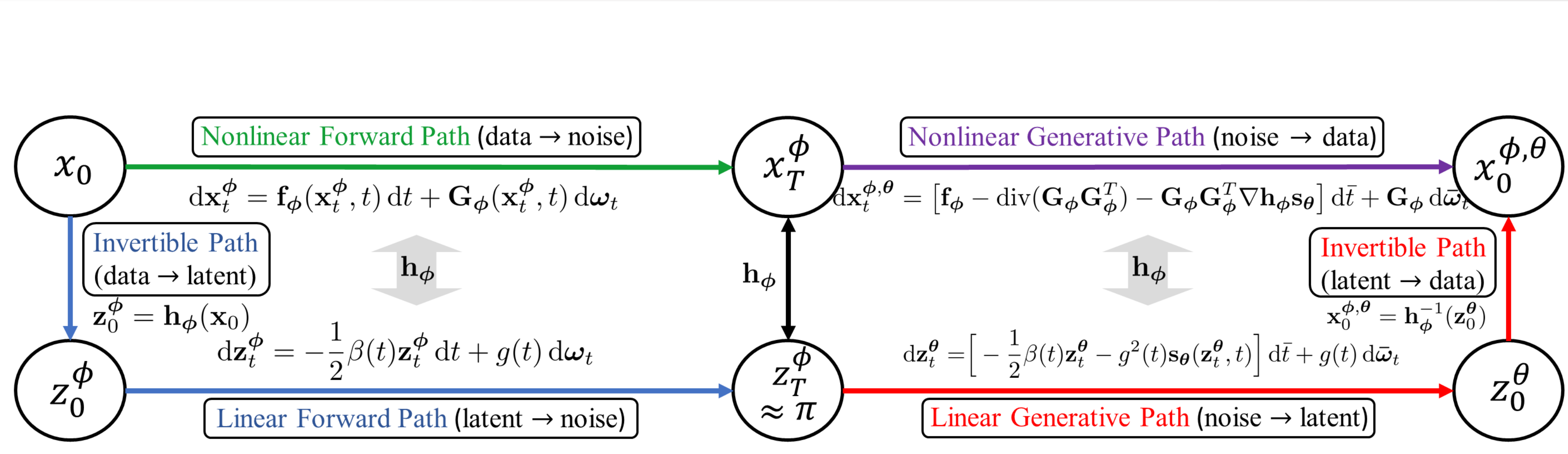


Linear Diffusion on **Latent** Space



Linear Diffusion on **Latent** Space + *Invertible* **Transformation**

Implicit Nonlinear Diffusion Model



Nonlinear Diffusion on **Data** Space =
Linear Diffusion on **Latent** Space + *Invertible* **Transformation**