Diffusion Models with Learned Adaptive Noise

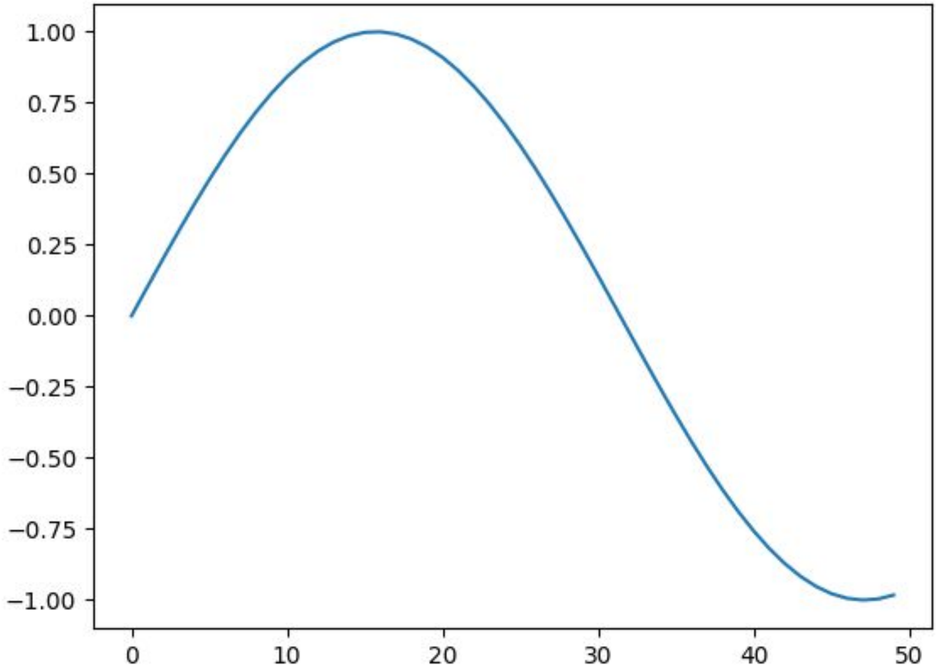

Subham Sahoo

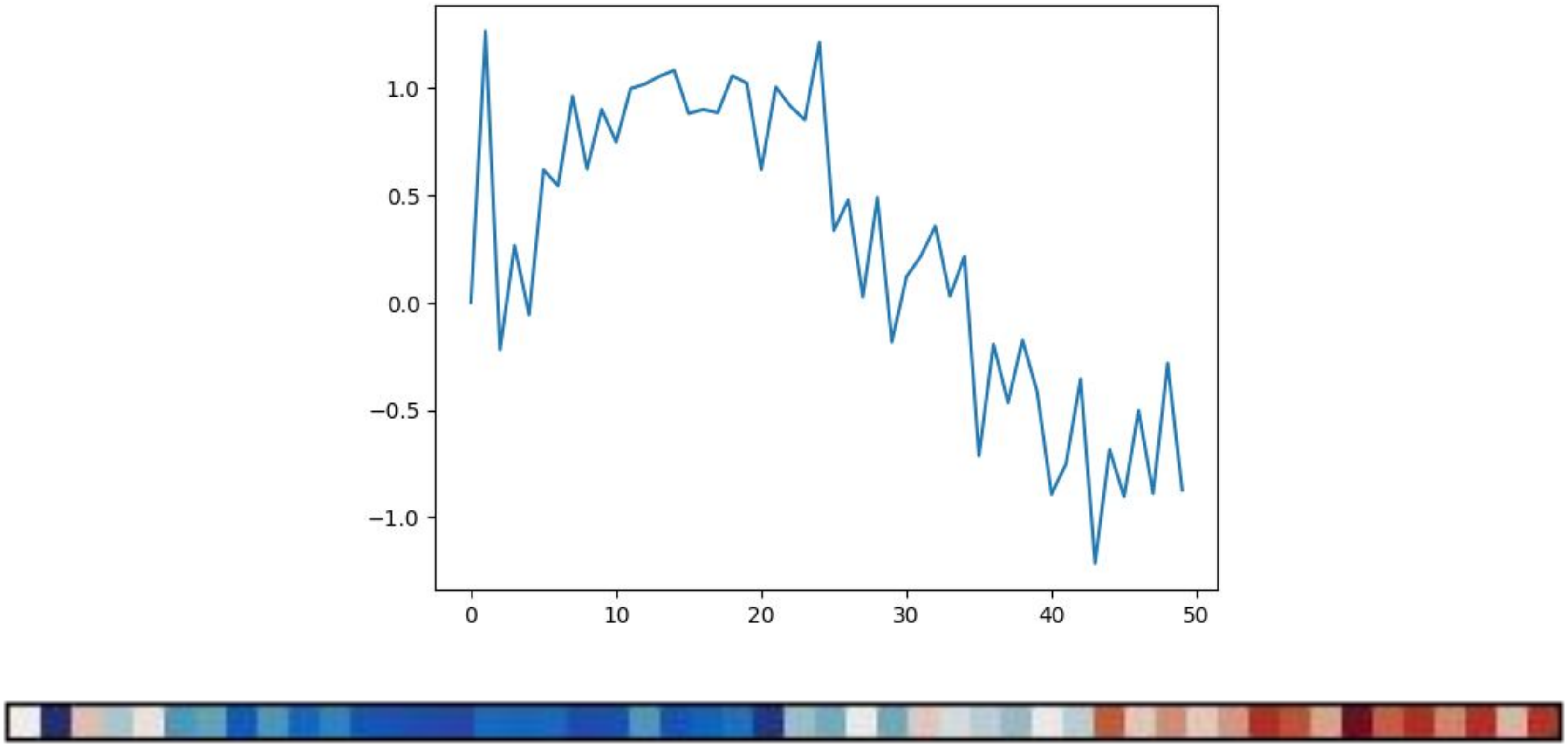

Aaron Gokaslan

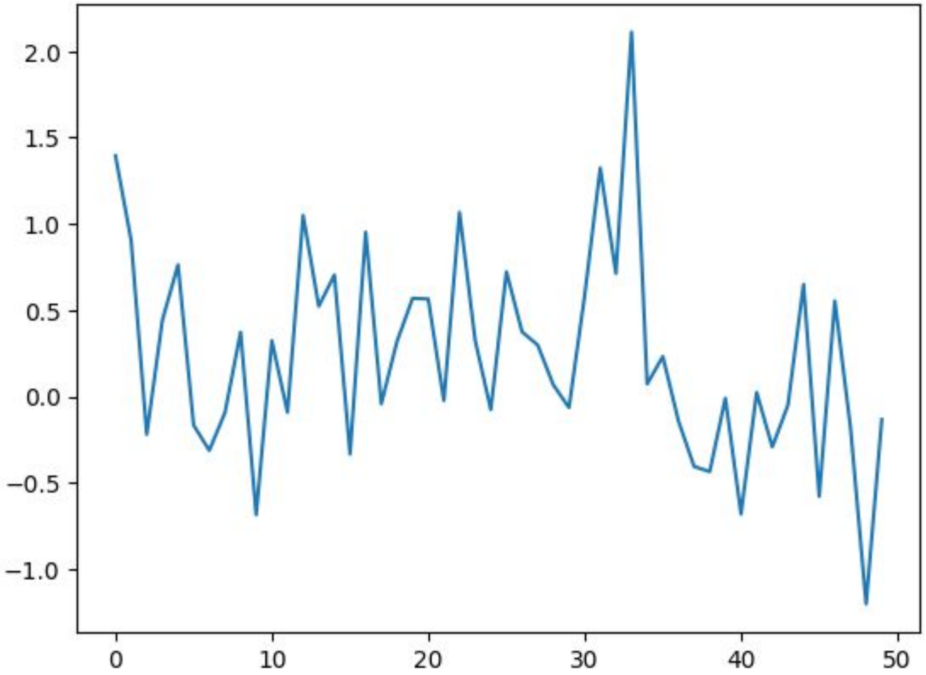

Chris De Sa


Volodymyr Kuleshov

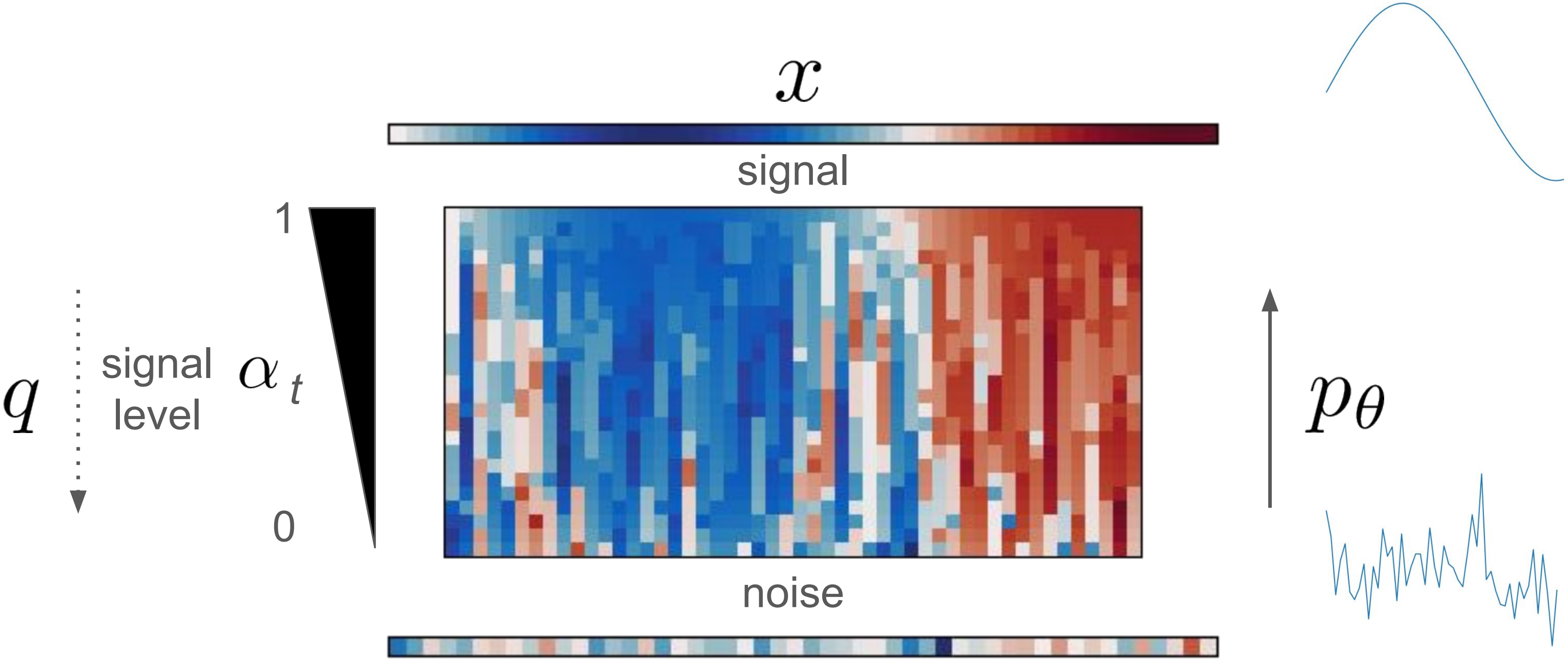
$$x \sim p_\theta(x)$$




1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
50

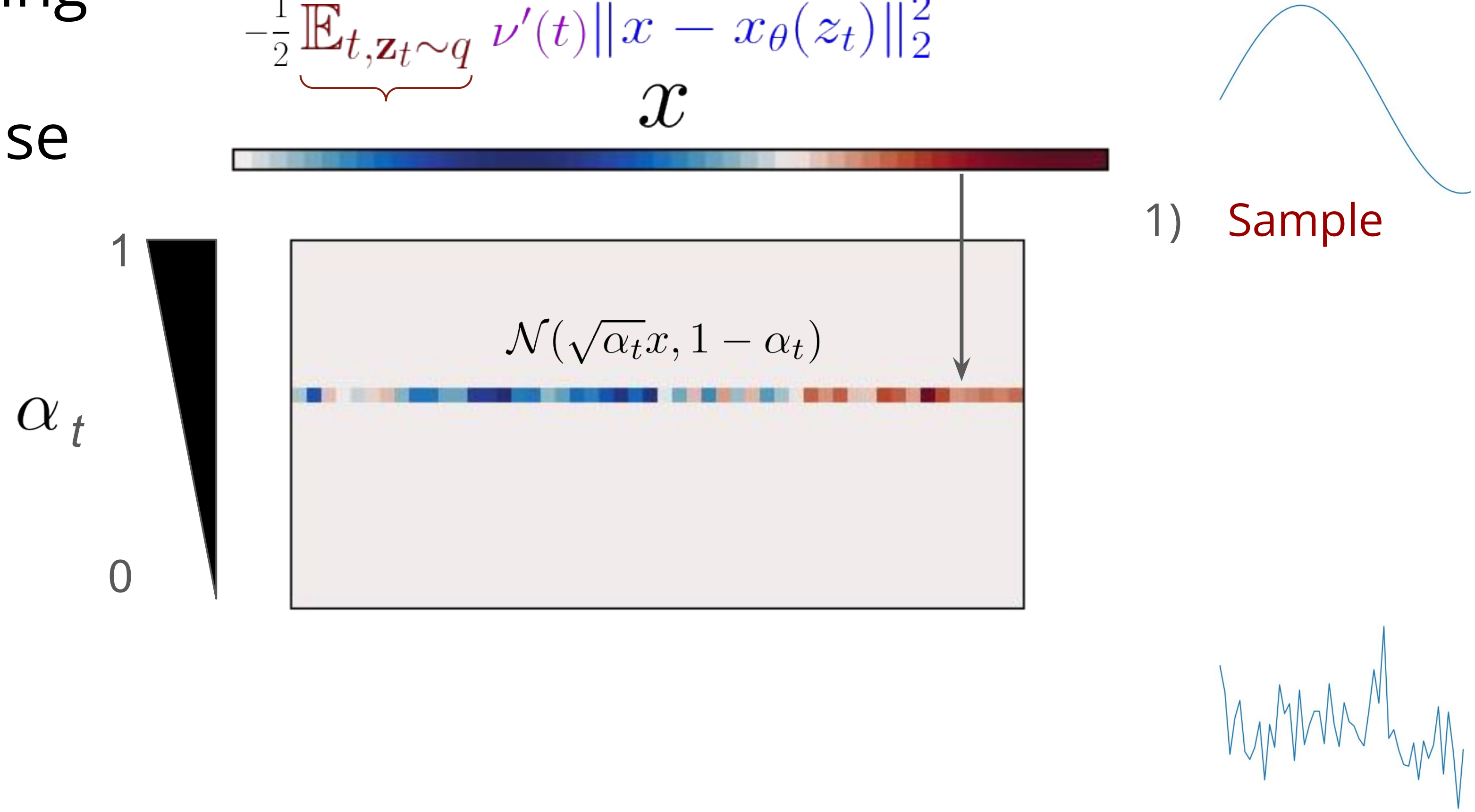

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
50

Continuous Diffusion



Learning to Denoise

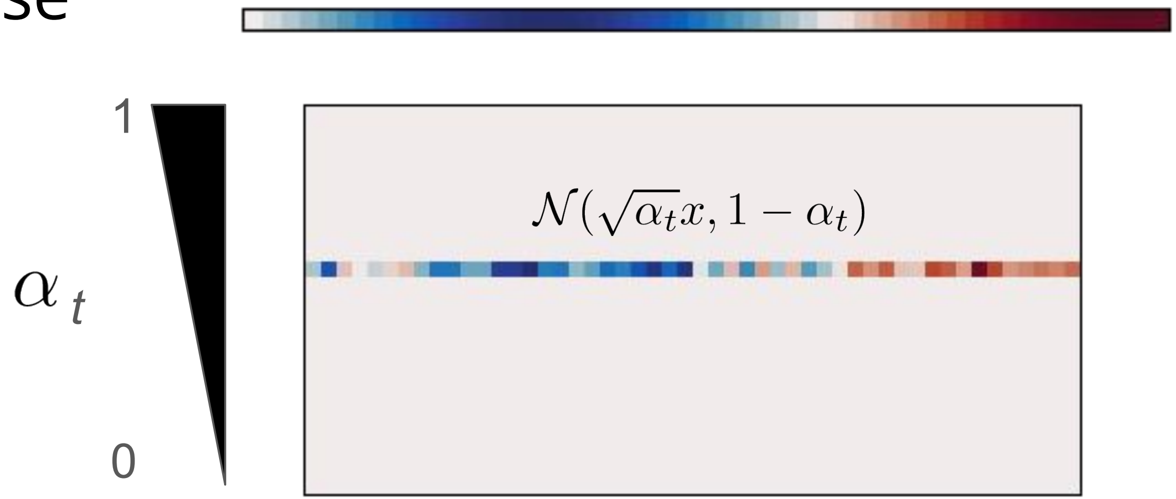
$$-\frac{1}{2}\underbrace{\mathbb{E}_{t,\mathbf{z}_t\sim q}}\ \nu'(t)\|x - x_\theta(z_t)\|_2^2$$

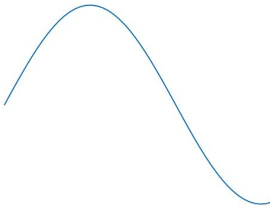


Learning to Denoise

$$-\frac{1}{2}\mathbb{E}_{t,\mathbf{z}_t\sim q}\underbrace{\nu'(t)}\|x - x_\theta(z_t)\|_2^2$$

x



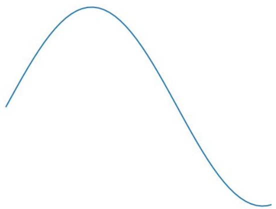


1) Sample
2) Weight by step change

$$\nu_t = \sqrt{\frac{\alpha_t}{1-\alpha_t}}$$

Learning to Denoise

$$-\frac{1}{2}\mathbb{E}_{t,\mathbf{z}_t\sim q}\ \nu'(t)\underbrace{\|x - x_\theta(z_t)\|_2^2}_{x}$$



1) Sample

2) Weight by step change

3) Reconstruct

$$\nu_t = \sqrt{\frac{\alpha_t}{1-\alpha_t}}$$

Evidence Lower BOund

$$\underbrace{\frac{1}{2}\mathbb{E}_{t,\mathbf{z}_t\sim q}\nu'(t)\|\mathbf{x}-\mathbf{x}_\theta(\mathbf{z}_t)\|_2^2}$$

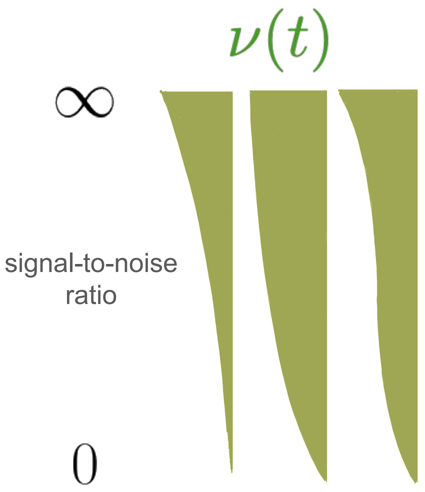
$$\log p_\theta(\mathbf{x}) \geq \max_{p\in\mathcal{P}} \mathrm{ELBO}(p,q)$$

Our Goal

$$\frac{1}{2}\mathbb{E}_{t,\mathbf{z}_t\sim q}\nu'(\mathbf{t})\|\mathbf{x}-\mathbf{x}_\theta(\mathbf{z}_t)\|_2^2$$

$$\log p_\theta(\mathbf{x}) \geq \max_{p\in\mathcal{P},q\in\mathcal{Q}} \mathrm{ELBO}(p, q)$$

Our Goal



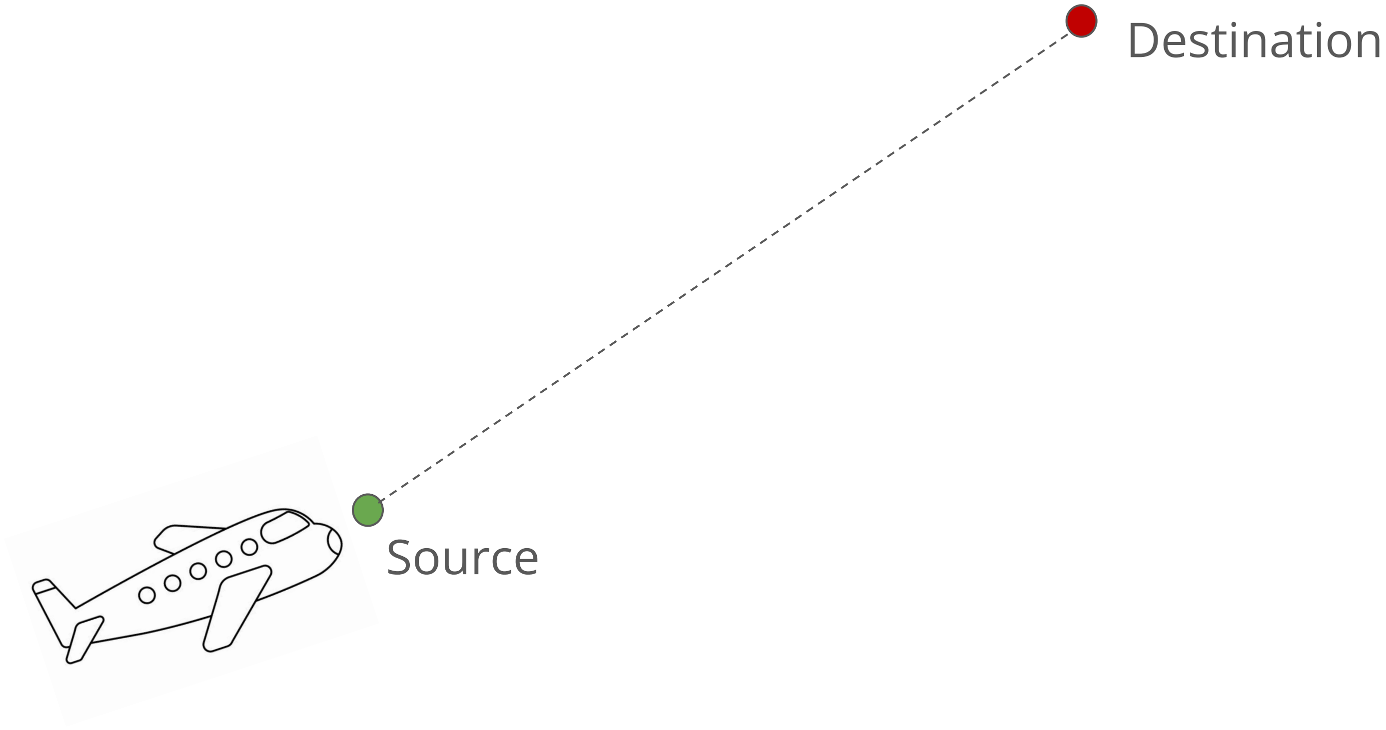
$$\frac{1}{2}\mathbb{E}_{t,\mathbf{z}_t\sim q}\nu'(\mathbf{t})\|\mathbf{x}-\mathbf{x}_\theta(\mathbf{z}_t)\|_2^2$$

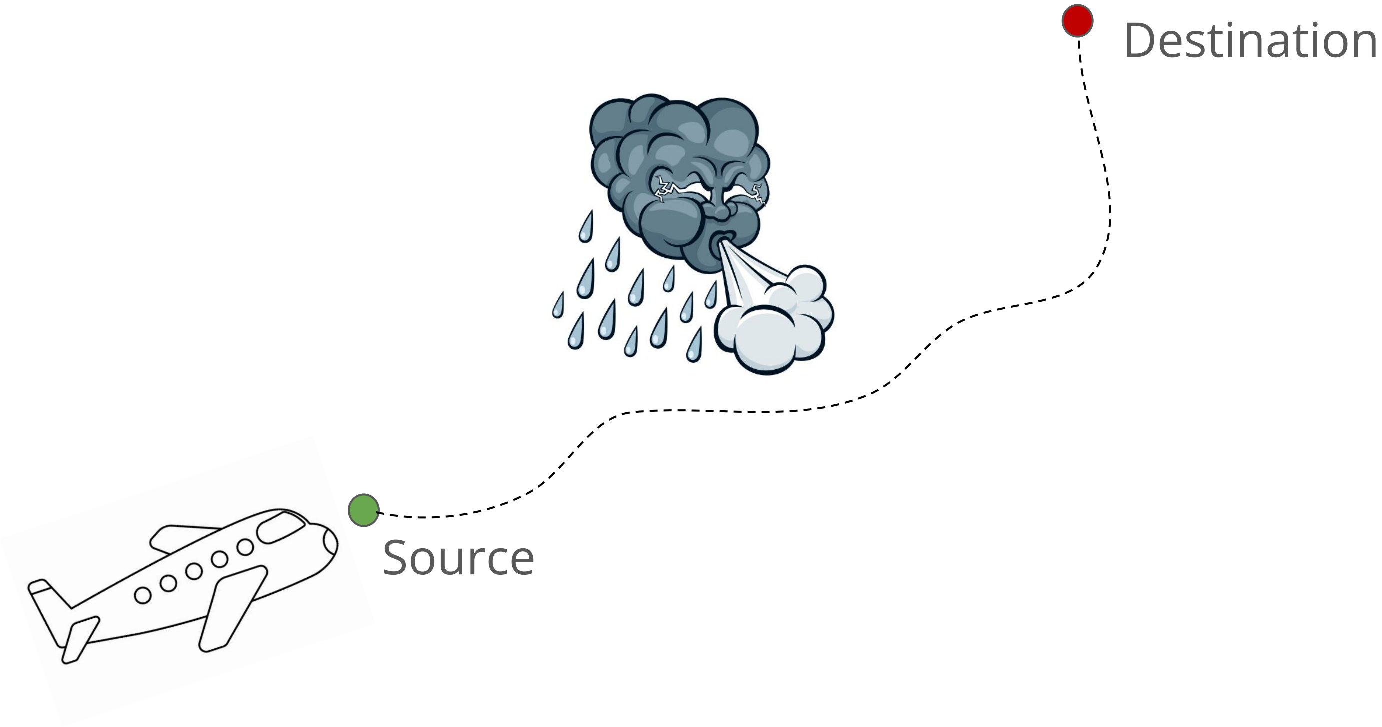
$$\max_{p\in\mathcal{P},q\in\mathcal{Q}} \text{ELBO}(p,q)$$

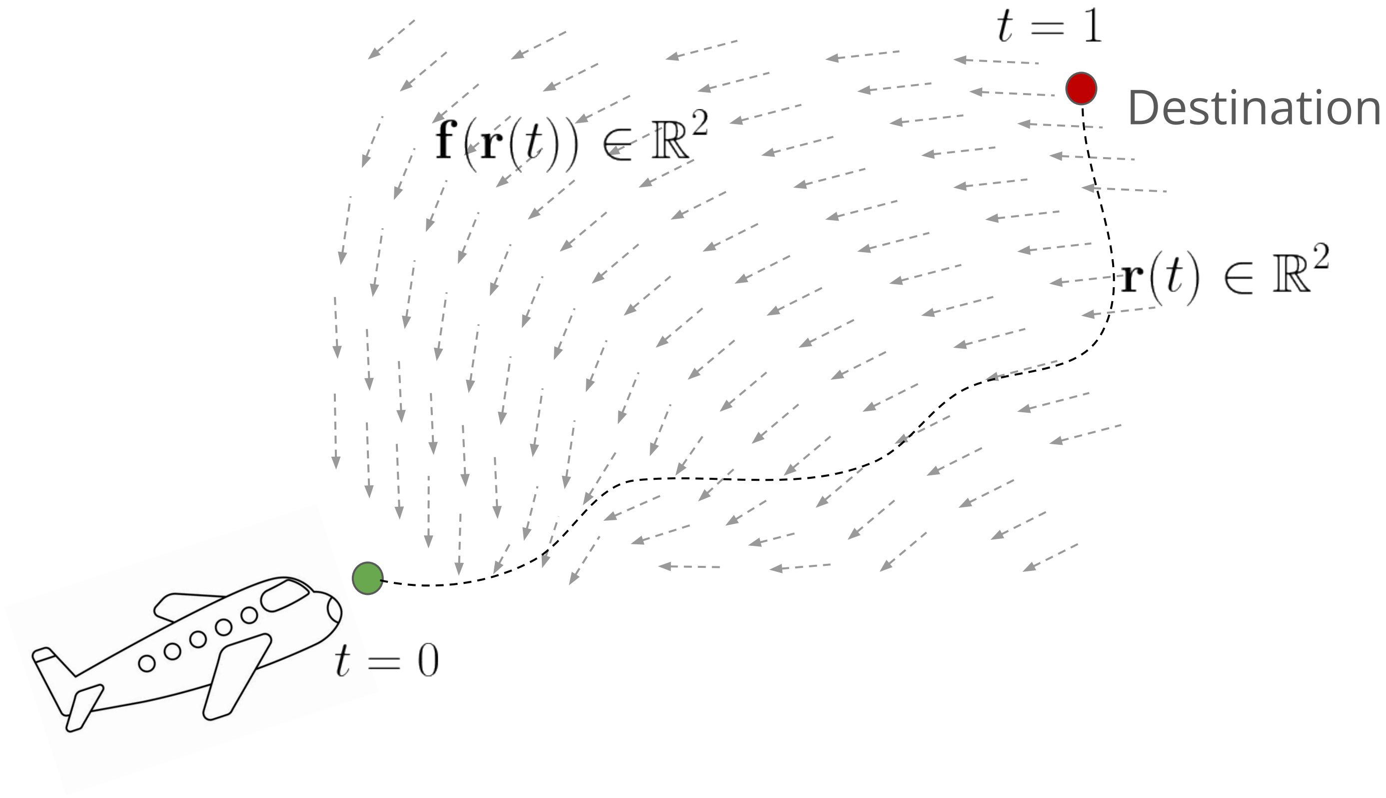
ELBO is invariant to $\nu(t)$

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. Advances in neural information processing systems, 34:21696–21707, 2021.

Illustrative Example

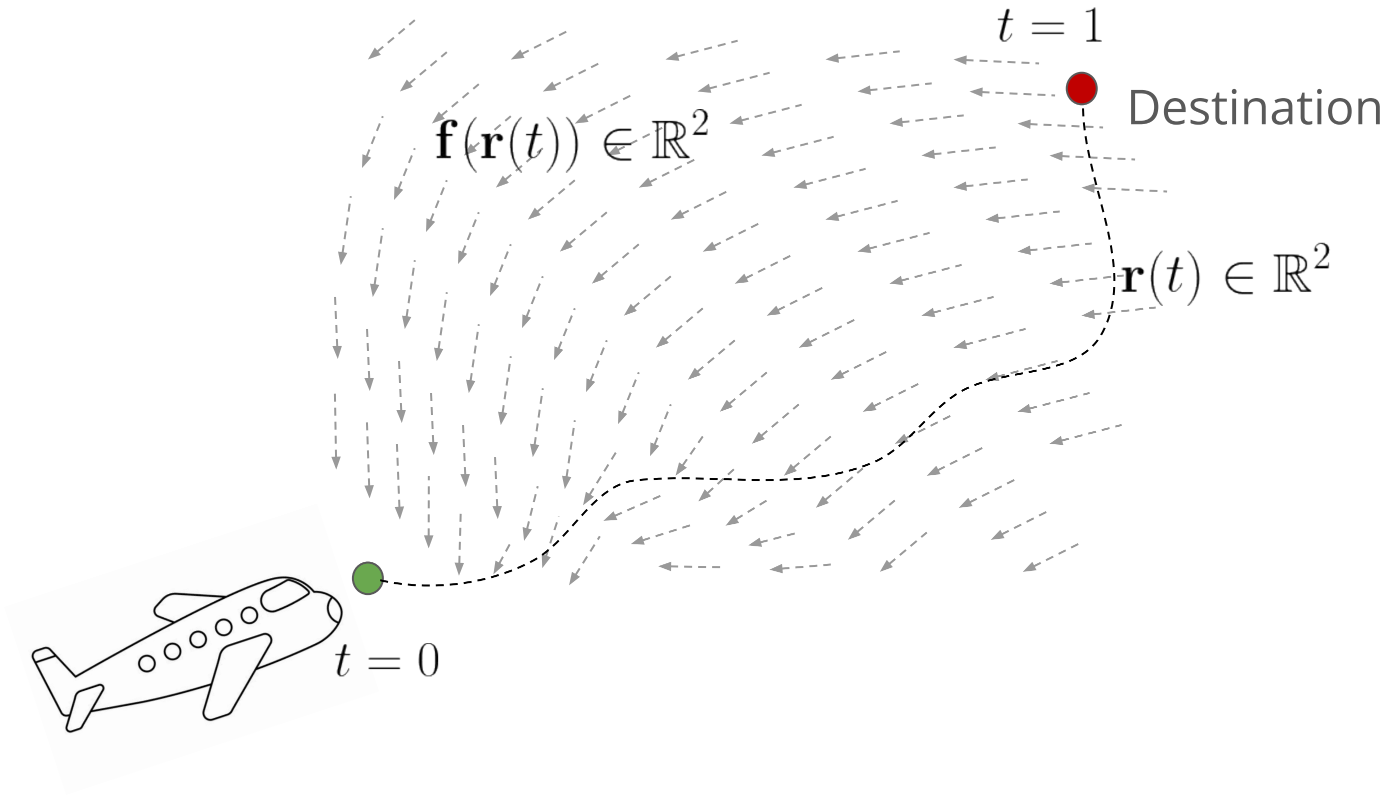

Destination
Source


Destination
Source

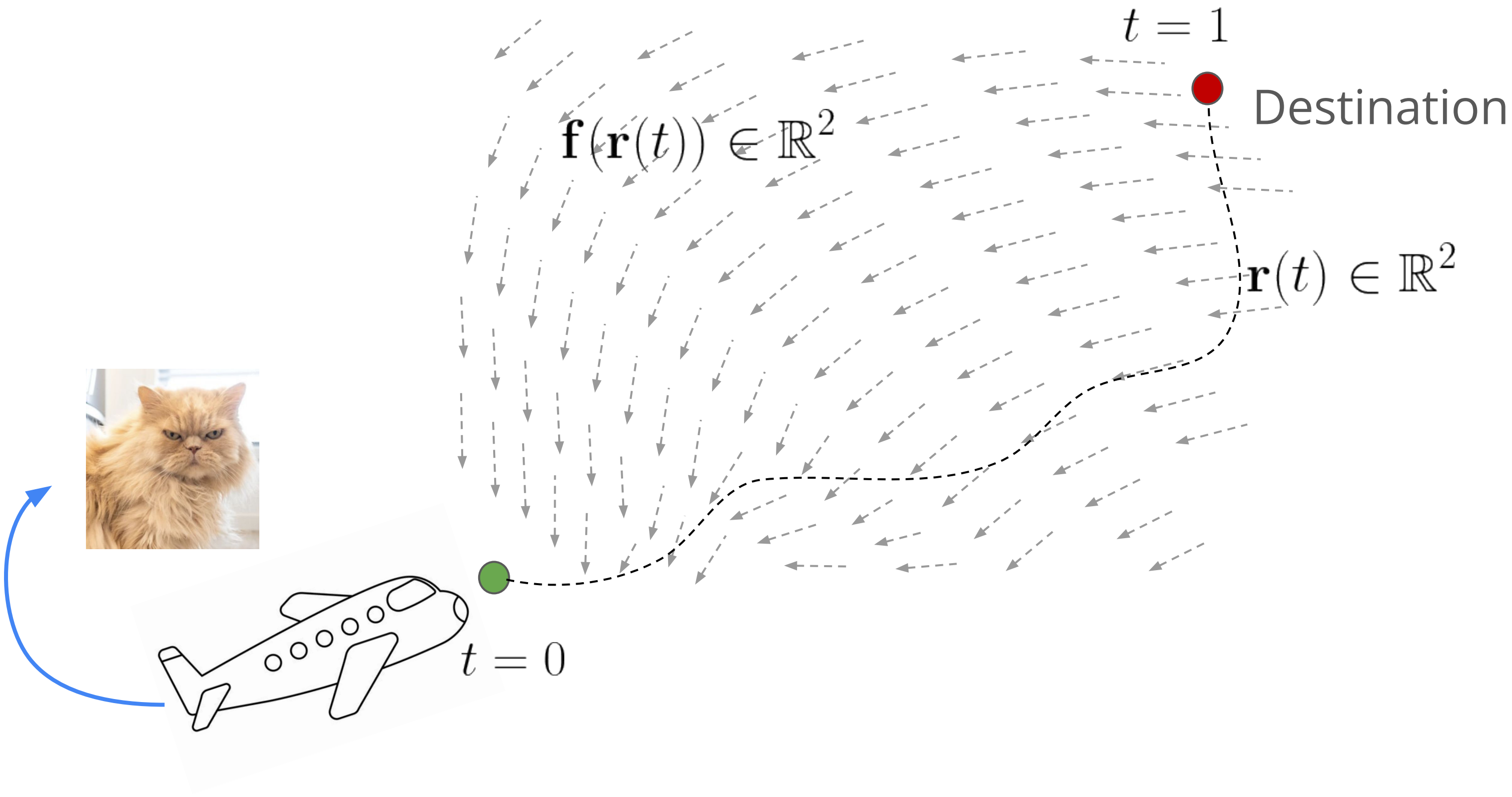


$t = 1$
Destination
$\mathbf{f}(\mathbf{r}(t)) \in \mathbb{R}^2$
$\mathbf{r}(t) \in \mathbb{R}^2$
$t = 0$

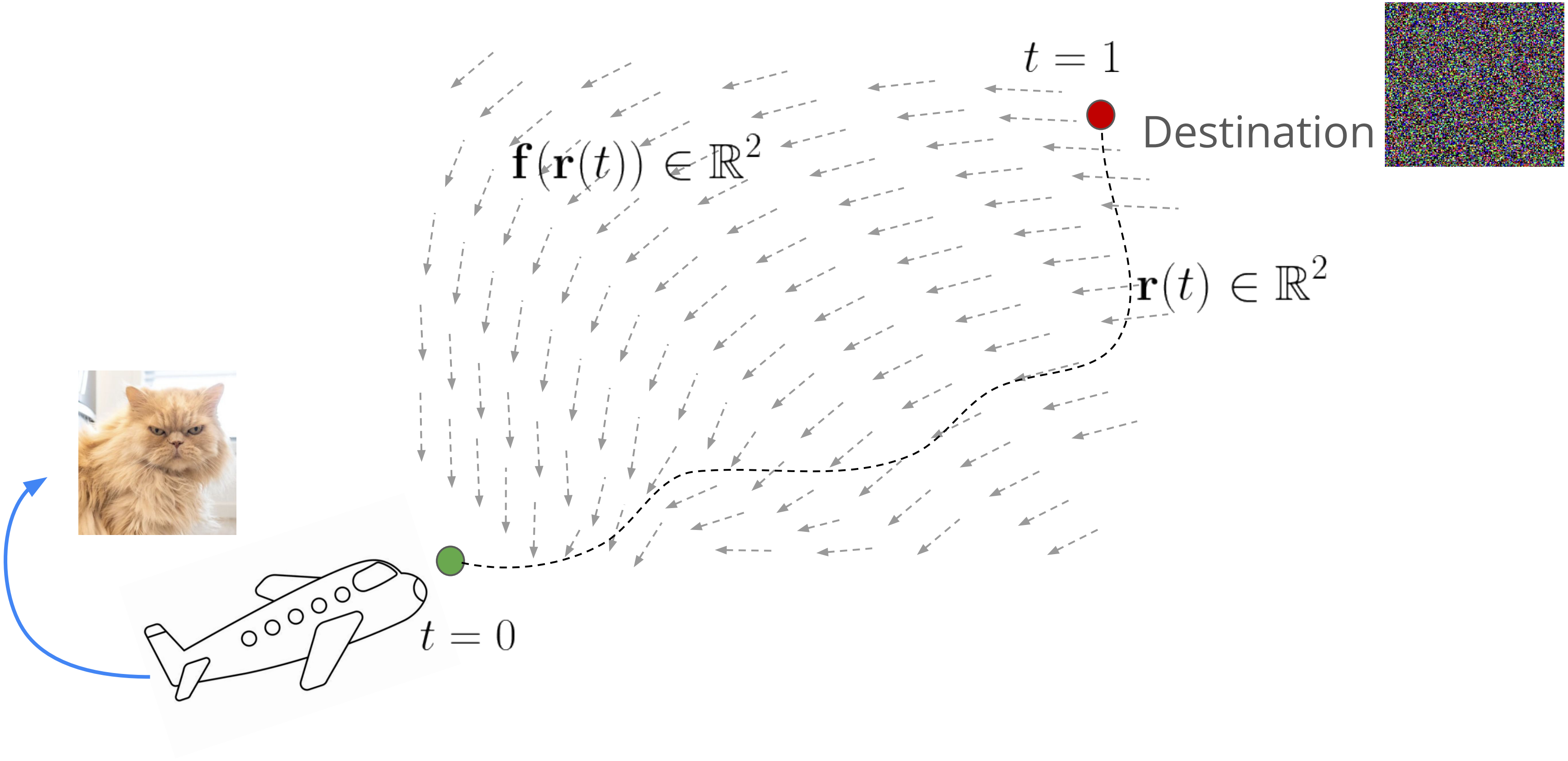
Work done = $\int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$



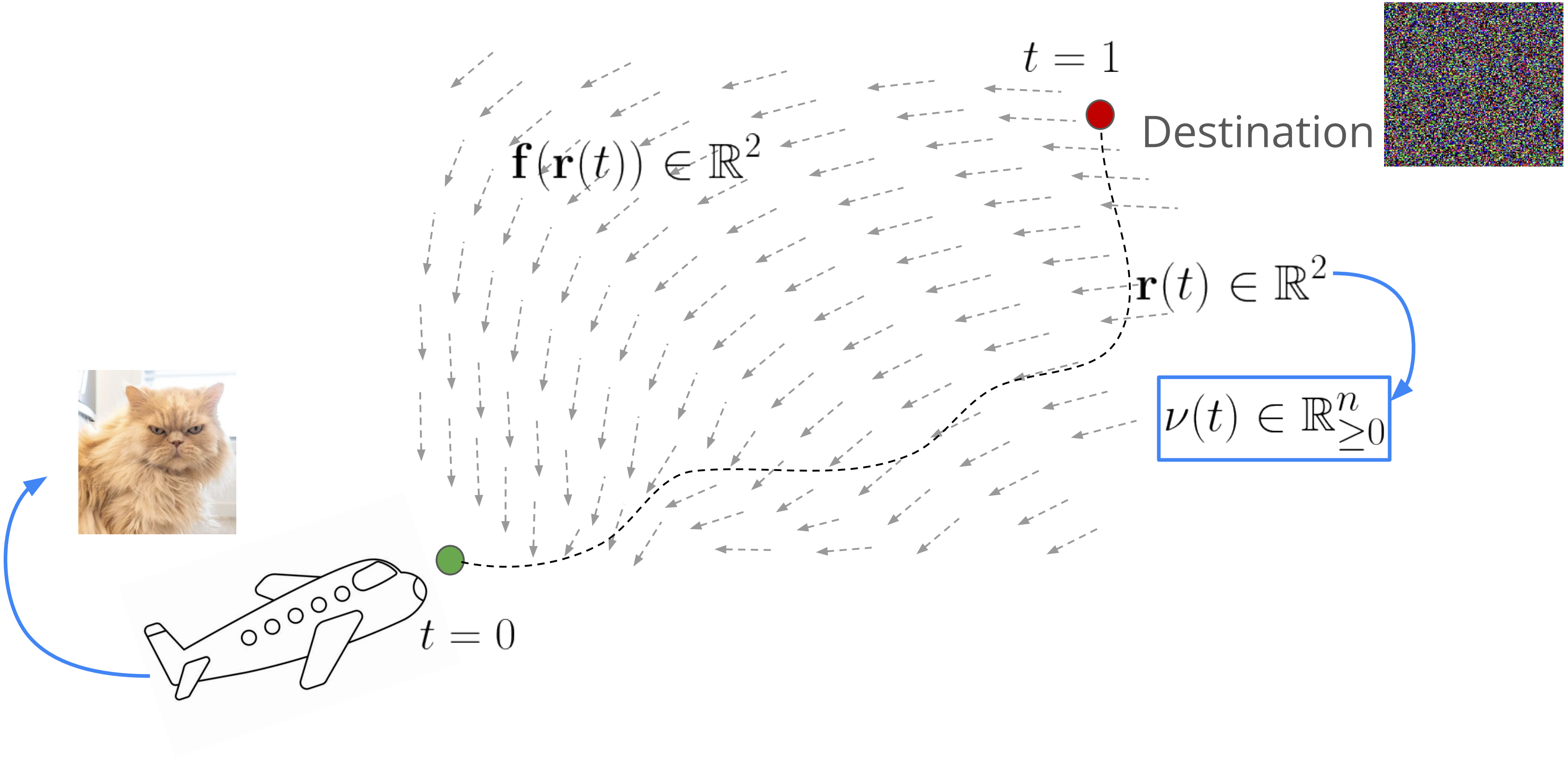
$$\text{Work done} = \int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$$



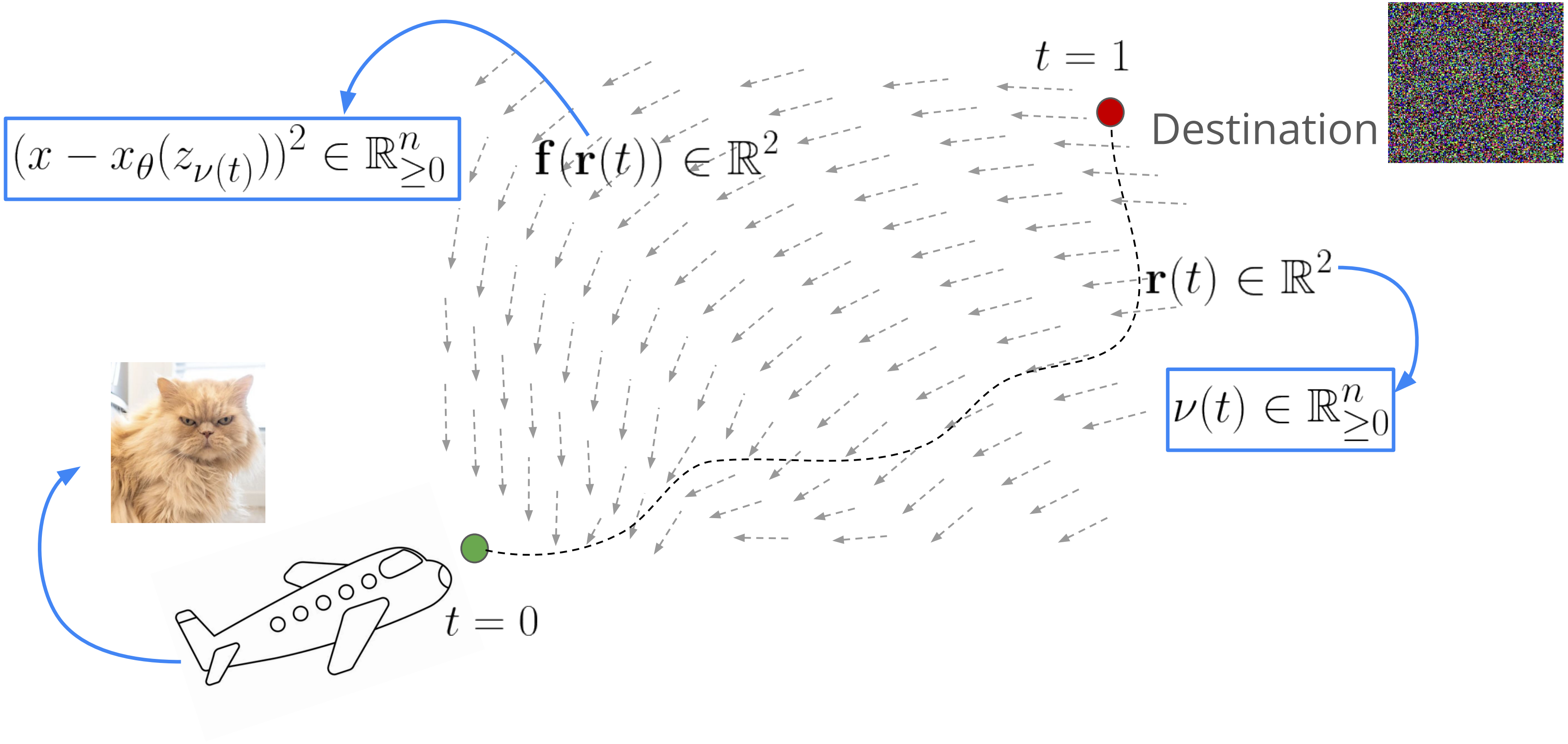
$$\text{Work done} = \int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$$



Work done = $\int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$

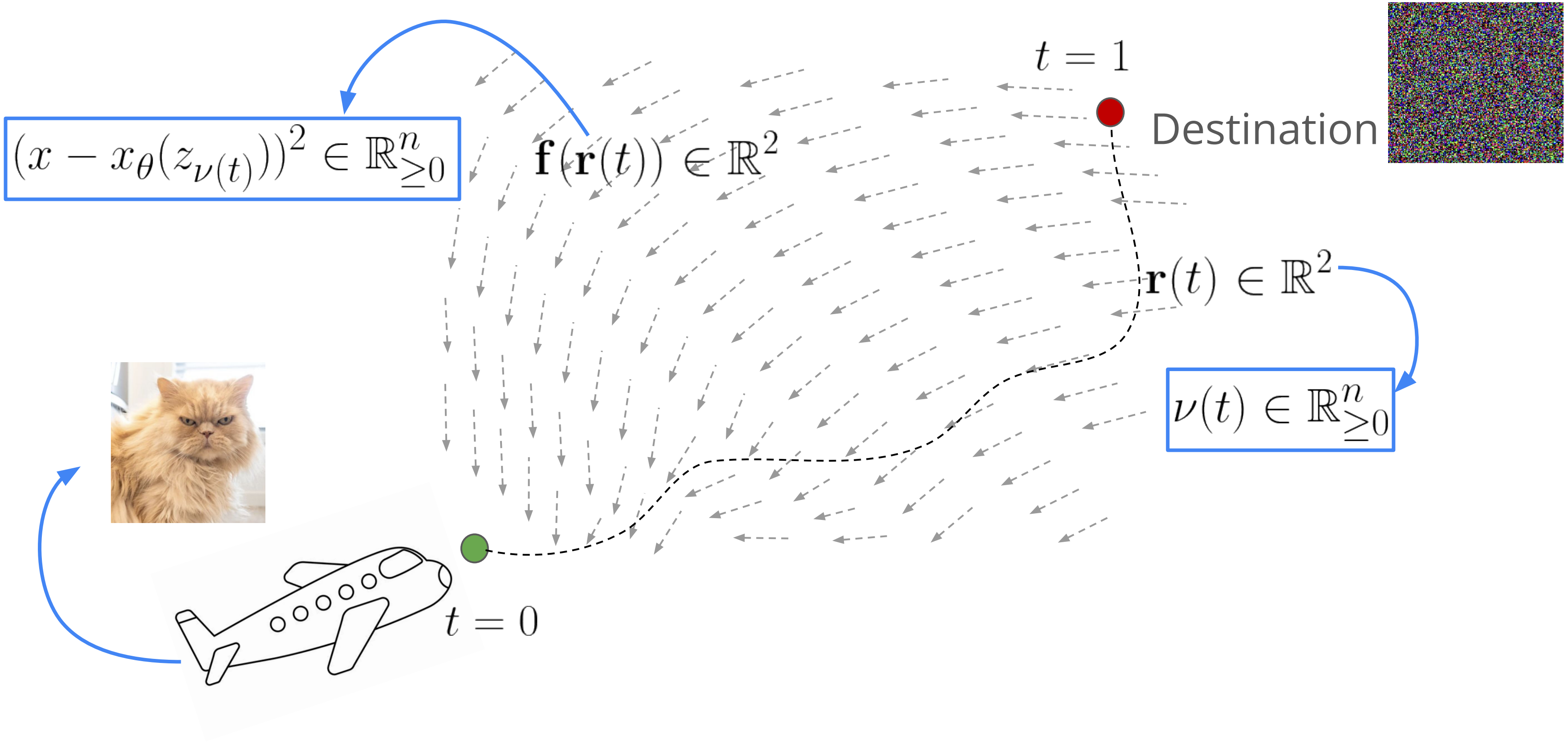


Work done = $\int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$

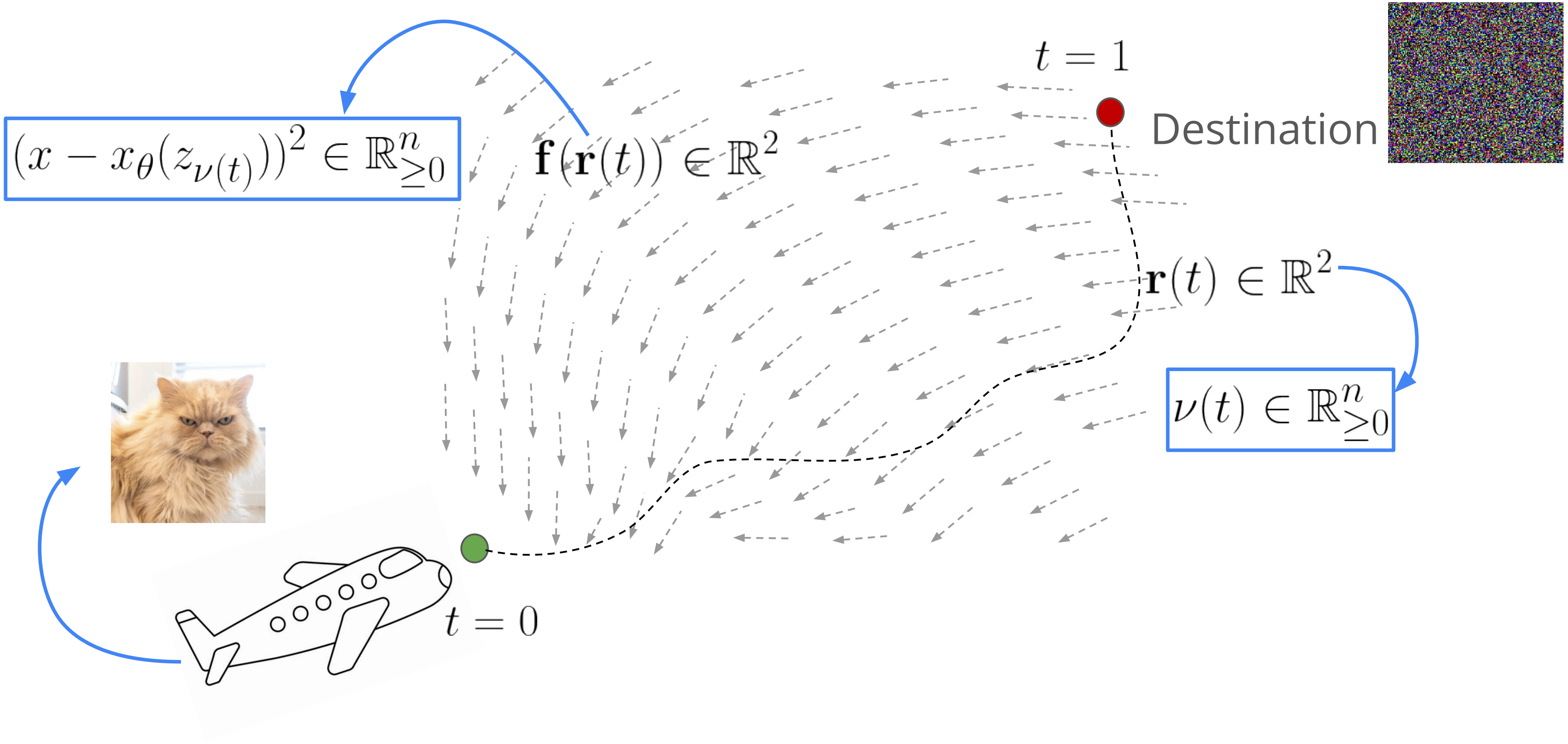


Work done = $\int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$

$-\frac{1}{2}\int_{t=0}^{t=1} (x - x_\theta(z_{\nu(t)}))^2 \odot \frac{d}{dt}\nu(t)dt$ NELBO



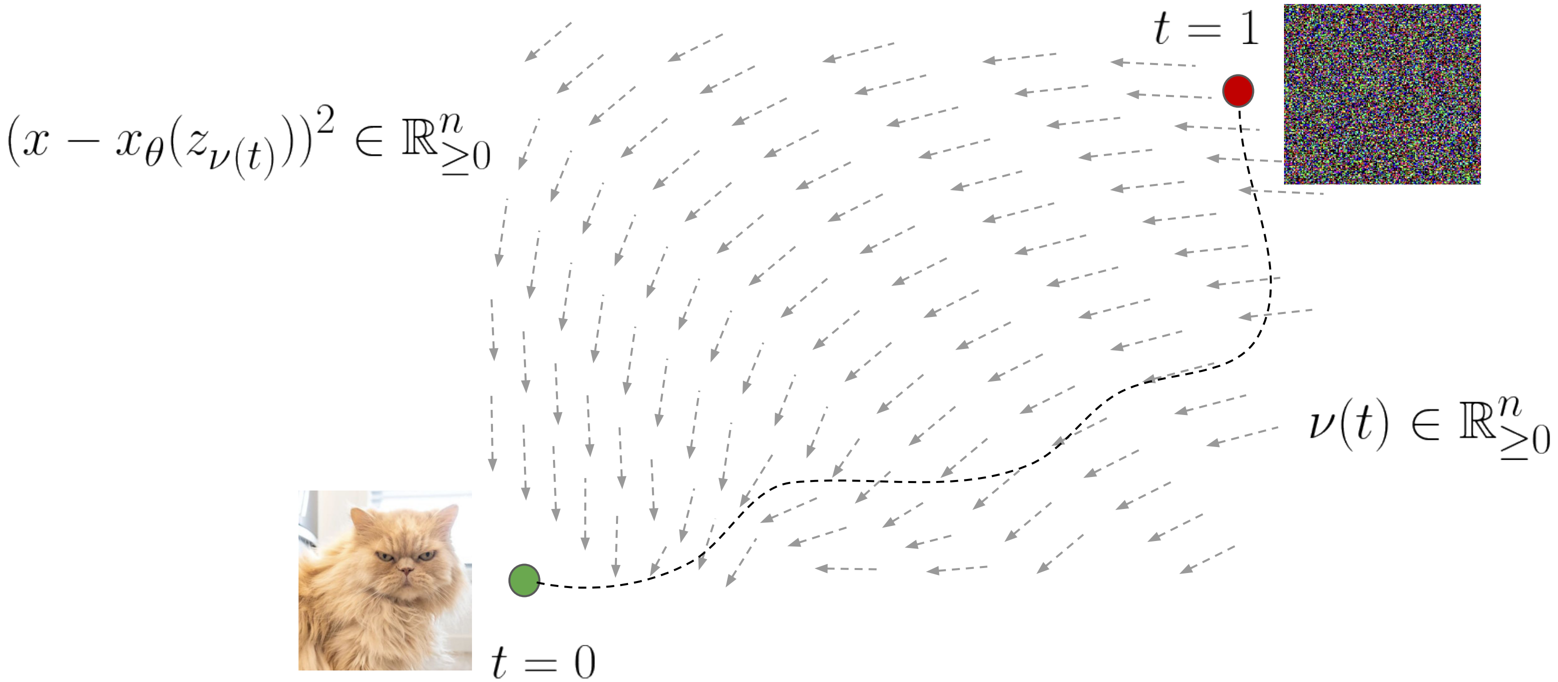
Work done = $\int_{t=0}^{t=1} \mathbf{f}(\mathbf{r}(t)) \odot \frac{d}{dt}\mathbf{r}(t)dt$ $\quad -\frac{1}{2}\int_{t=0}^{t=1} (x - x_\theta(z_{\nu(t)}))^2 \odot \frac{d}{dt}\nu(t)dt$ NELBO



Multivariate Learned Adaptive Noise (MuLAN)

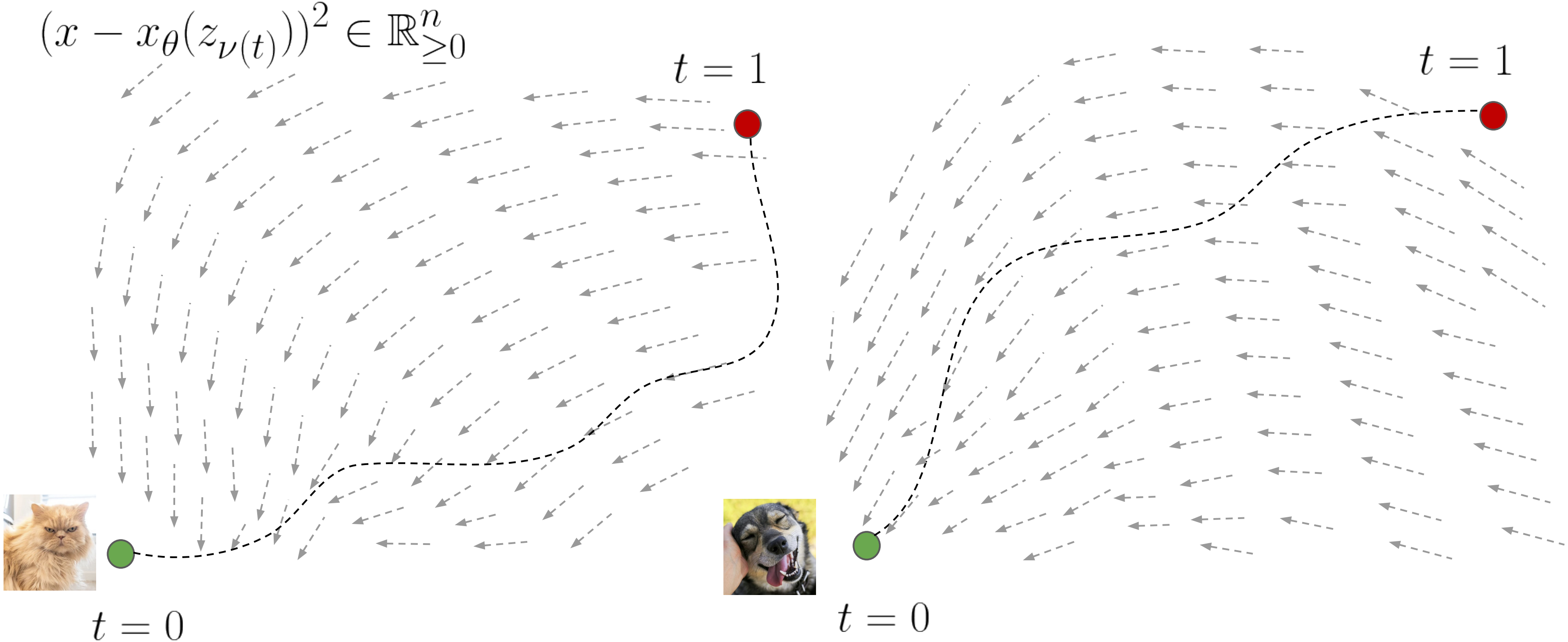
ELBO improving Noise Schedules

Property 1: **Multivariate**



ELBO improving Noise Schedules

Property 2: **Adaptive**



MuLAN

$$\nu_\phi(\mathbf{c}, t) \in \mathbb{R}^n_{\geq 0} \quad \mathbf{c} \sim q_\phi(.|x) \quad \mathbf{c} \in \mathbb{R}^m$$

ϕ Learnable parameters of the noise schedule

Properties of MuLAN

1. Multivariate
2. Learned
3. Adaptive

MuLAN

Check out the paper to see why $\mathbf{c}$ **can't** be a deterministic function of $\boldsymbol{x}$

$$\nu_\phi(\mathbf{c}, t) \in \mathbb{R}^n_{\geq 0} \quad \mathbf{c} \sim q_\phi(.|x) \quad \mathbf{c} \in \mathbb{R}^m$$

ϕ Learnable parameters of the noise schedule

Properties of MuLAN
1. Multivariate
2. Learned
3. Adaptive

MuLAN

$$\nu_\phi(\mathbf{c}, t) \in \mathbb{R}^n_{\geq 0} \qquad \mathbf{c} \sim \boxed{q_\phi(.|x)} \qquad \mathbf{c} \in \mathbb{R}^m$$

Properties of MuLAN

1. Multivariate
2. Learned
3. Adaptive
4. Discrete Latent

Backpropagation through Combinatorial Algorithms: Identity with Projection Works.
Subham S. Sahoo, Anselm Paulus*, Marin Vlastelica, Vit Musil, Volodymyr Kuleshov, Georg Martius.*
International Conference on Learning Representations (ICLR - 2023), 2023.

MuLAN

NELBO: $-\frac{1}{2}\int_{t=0}^{t=1}(x - x_\theta(z_{\nu_\phi(\mathbf{c},t)}))^2 \odot \frac{d}{dt}\nu_\phi(\mathbf{c},t)dt \;\; + \mathrm{D}_{\mathrm{KL}}(q_\phi(\mathbf{c}|x)\|p_\theta(\mathbf{c}))$

$\mathbf{c} \sim q_\phi(.|x) \quad \mathbf{c} \in \mathbb{R}^m \qquad \nu_\phi(\mathbf{c}, t) \in \mathbb{R}^n_{\geq 0}$

Properties of MuLAN

1. Multivariate
2. Learned
3. Adaptive
4. Discrete Latent

Backpropagation through Combinatorial Algorithms: Identity with Projection Works.
Subham S. Sahoo, Anselm Paulus*, Marin Vlastelica, Vit Musil, Volodymyr Kuleshov, Georg Martius.*
International Conference on Learning Representations (ICLR - 2023), 2023.

Generation

noise

$t = 1$

$\mathbf{z}_{t=1} \sim \mathcal{N}(0, I)$

Generation

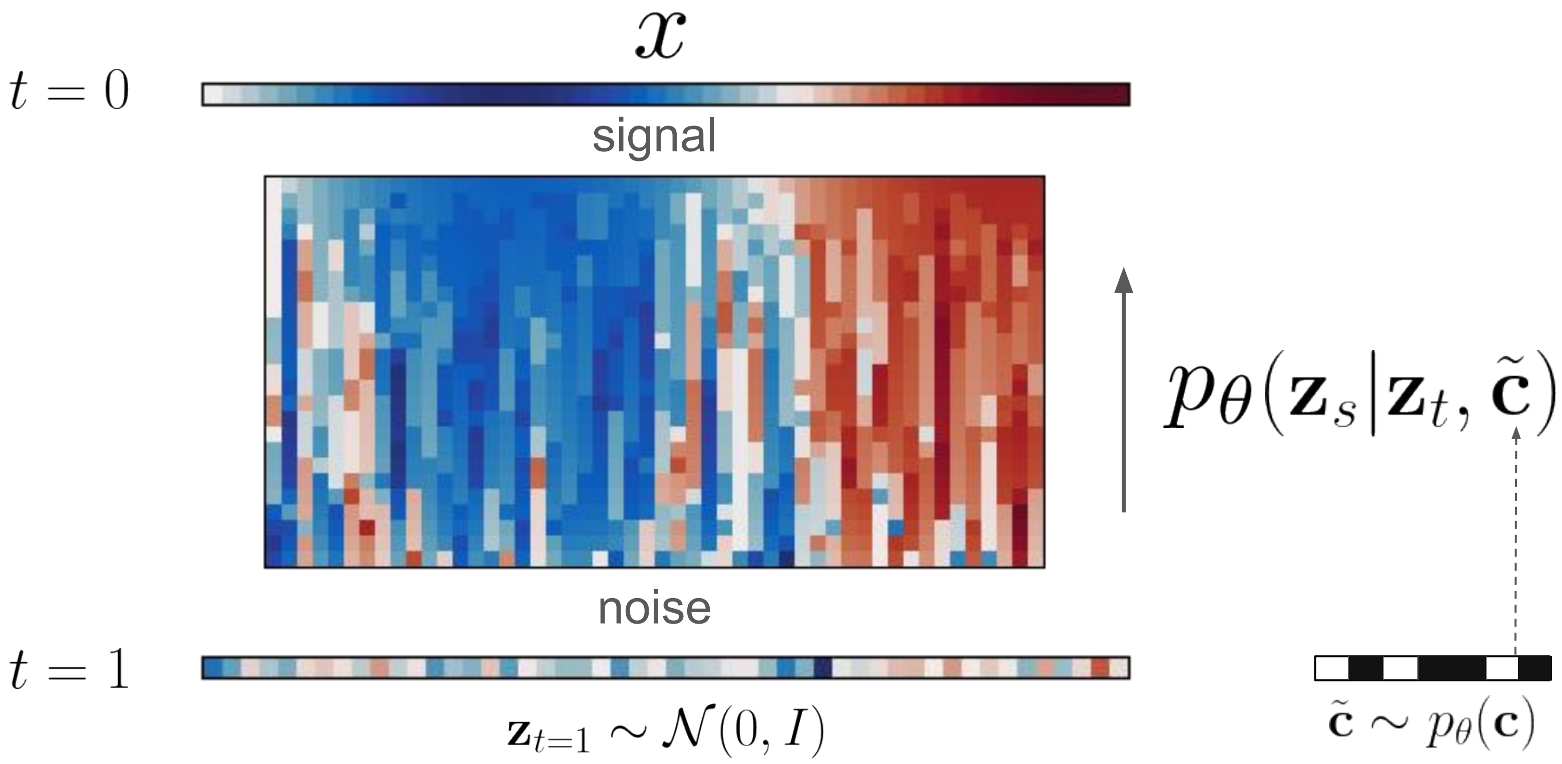
noise

$t = 1$

$\mathbf{z}_{t=1} \sim \mathcal{N}(0, I)$

$\tilde{\mathbf{c}} \sim p_\theta(\mathbf{c})$

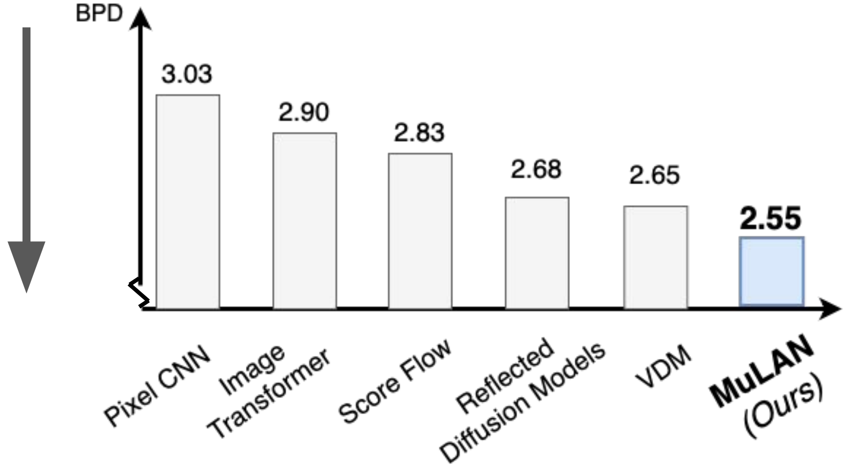
Generation



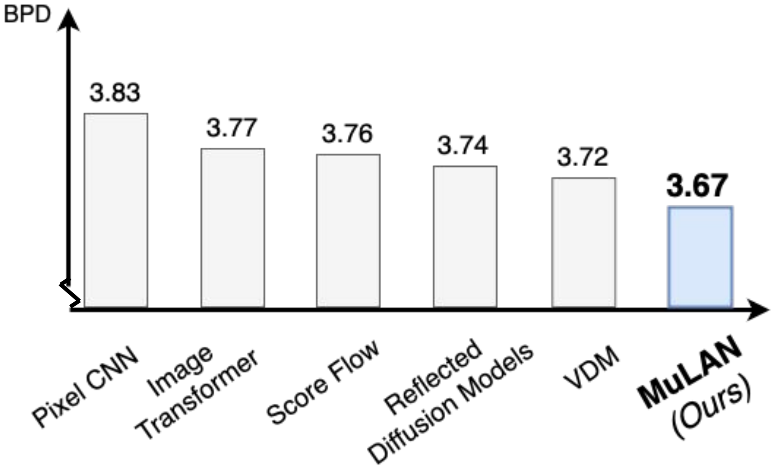
x
$t = 0$
signal
noise
$t = 1$
$p_\theta(\mathbf{z}_s | \mathbf{z}_t, \tilde{\mathbf{c}})$
$\mathbf{z}_{t=1} \sim \mathcal{N}(0, I)$
$\tilde{\mathbf{c}} \sim p_\theta(\mathbf{c})$

Experiments

Likelihood Evaluation



CIFAR-10



ImageNet-32

Convergence Speed

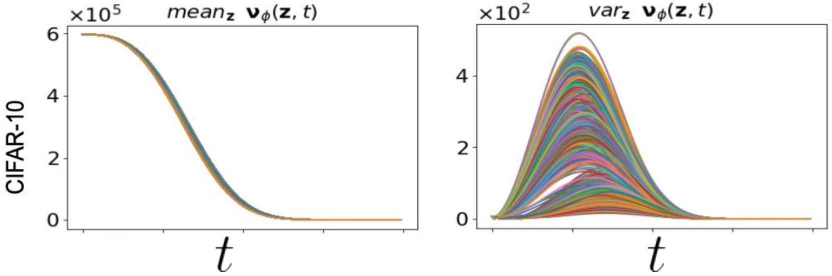
CIFAR-10

Model	Steps	VLB (↓)
VDM [18]	10M	2.65
+ MULAN	**2M**	2.65

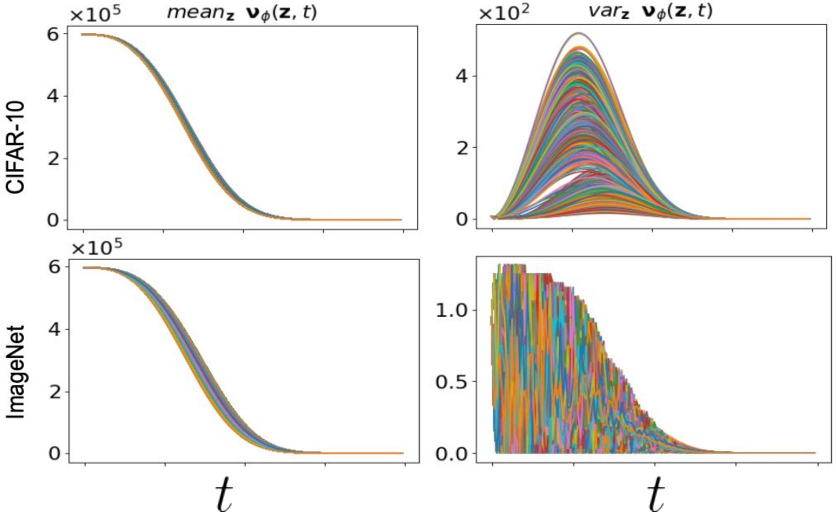
5x faster convergence

[18] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. Advances in neural information processing systems, 34:21696–21707, 2021.

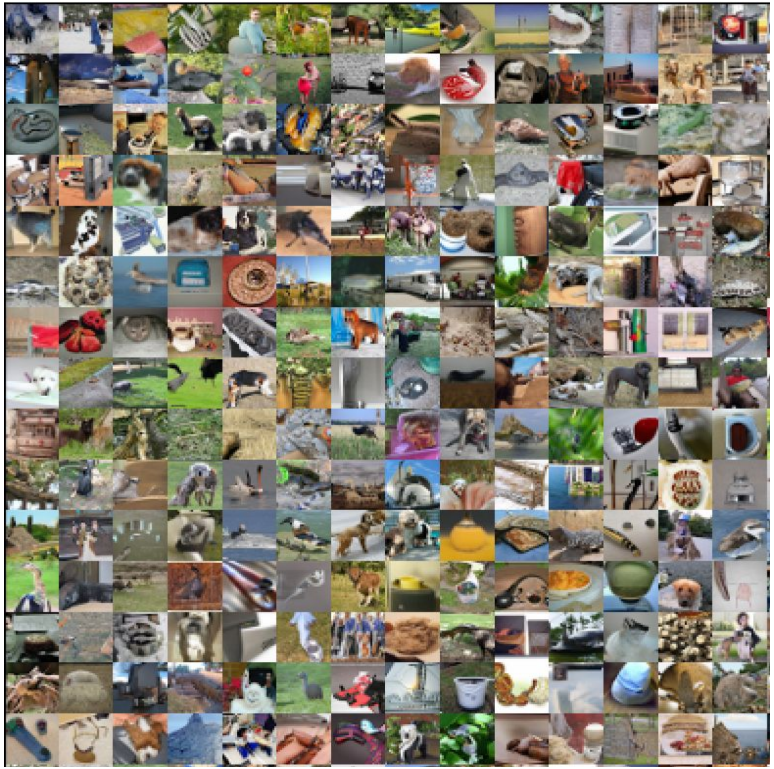
Noise Schedule Visualizations



MuLAN generated samples

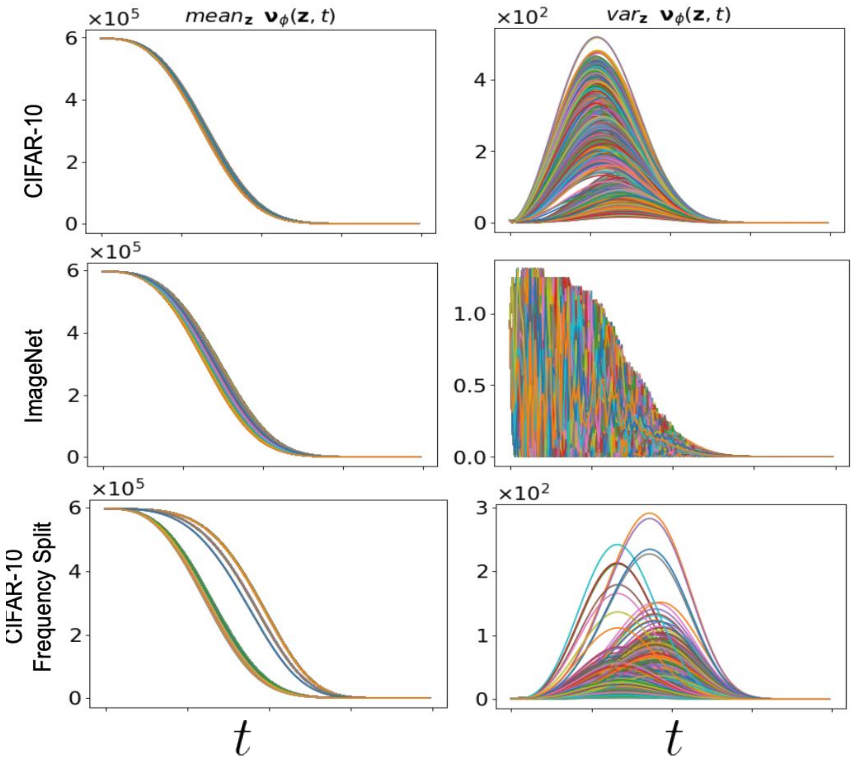


Noise Schedule Visualizations

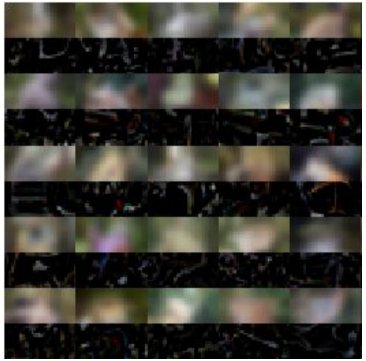


MuLAN generated samples

Noise Schedule Visualizations

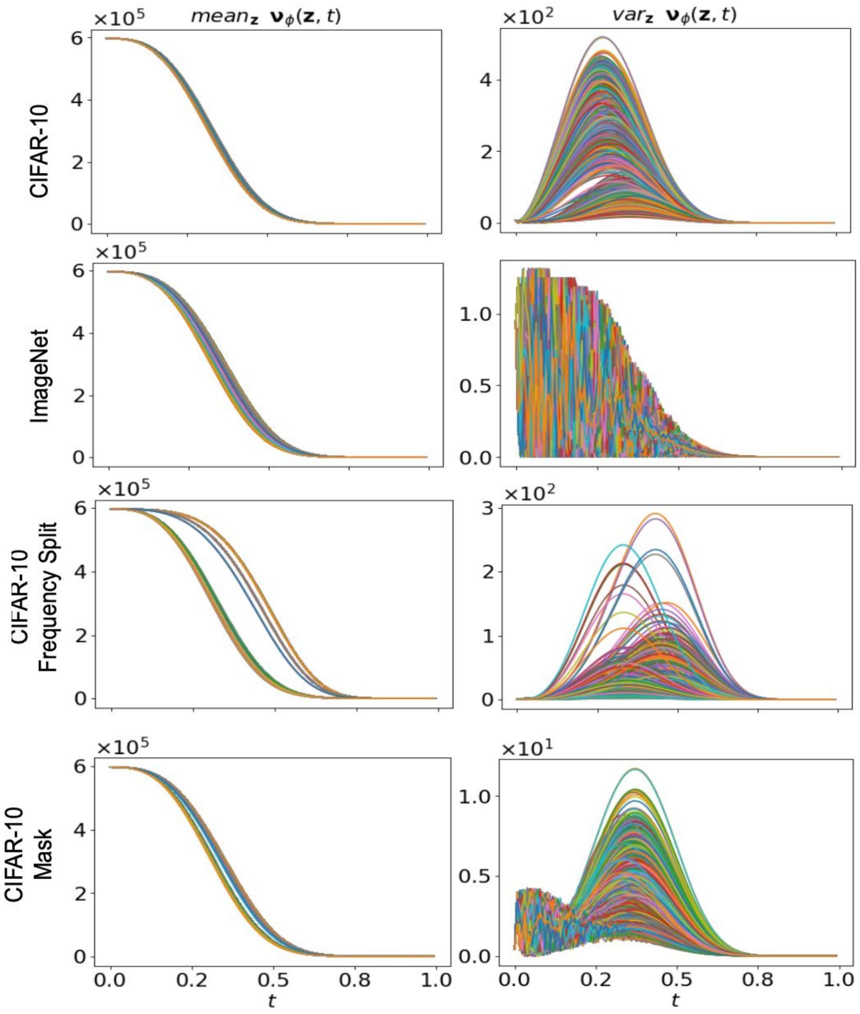


Train Samples



MuLAN generated samples

Noise Schedule Visualizations



Train Samples

MuLAN generated samples