Accurate sampling with few-step diffusion

Nikhitha Beedala, Vedant Puri, Nihali Shetty
{nbeedala, vedantpu, nsshetty} @andrew.cmu.edu
Carnegie Mellon University

1. Motivation and Overview

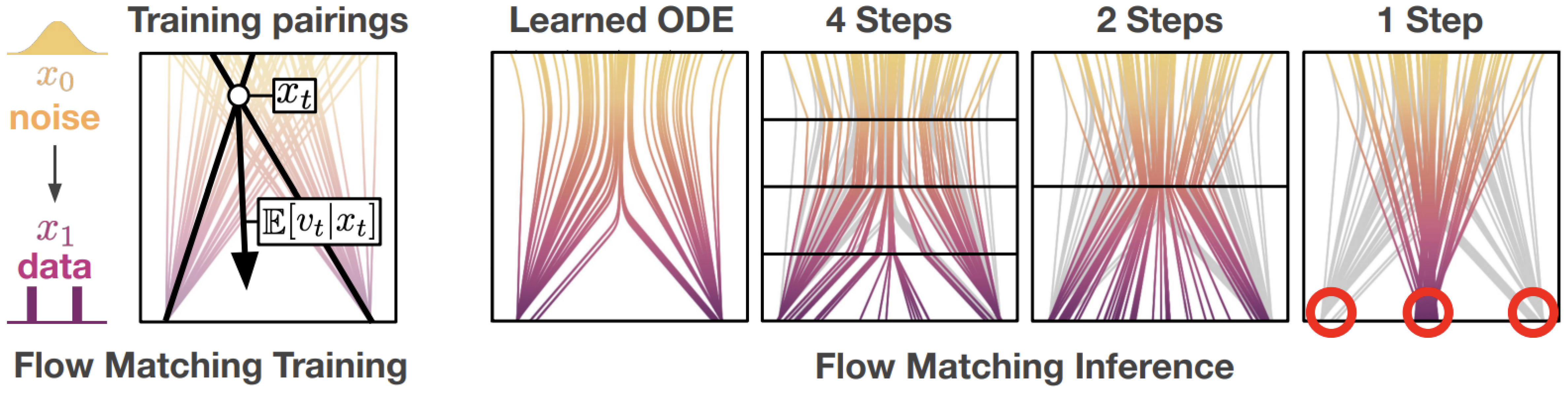
Generating samples from diffusion models involves many denoising steps leading to large latency. We build upon recent work on few-step diffusion modeling with the aim of improving sample quality. Our proposed modifications indeed improve sample quality.

2. Flow matching diffusion models

Flow matching diffusion models describe a linear interpolation (**rectified flow**) between data points $x_1 \sim \mathcal{D}$ and Gaussian noise $x_0 \sim \mathcal{N}(0, I)$ [1,2]. Flow models learn a neural network to estimate the expected velocity $\overline{v}_t = \mathbb{E}[v_t|x_t, t]$ for each sample at noise level t.

$$x_t = (1-t)x_0 + tx_1 \qquad x_{t+\Delta t} = x_t + v_t(x_t, t)\Delta t$$

$$\mathcal{L}^F(\theta) = \mathbb{E}_{x_0,x_1,t}\left[\|\overline{v}_\theta(x_t, t) - (x_1 - x_0)\|_2^2\right]$$

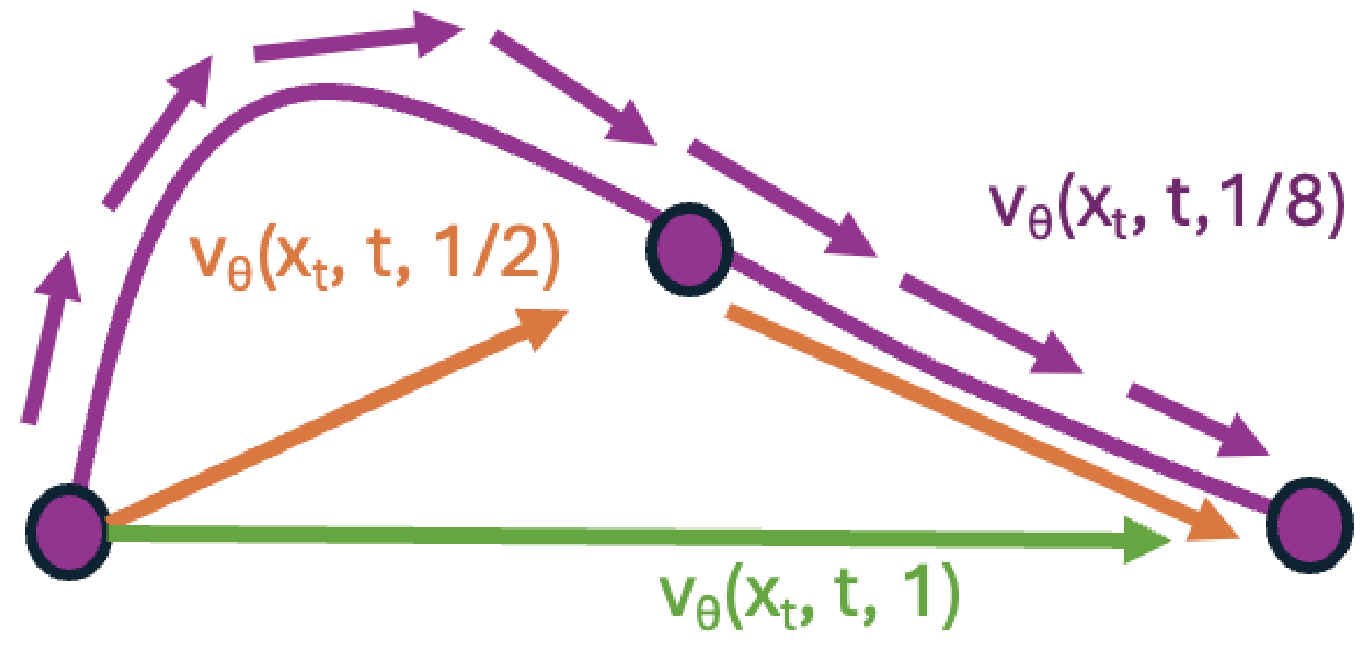


3. Conditioning velocity on denoising step size

Recently proposed shortcut diffusion models [3] **condition velocity on the intended step size** d as $\overline{v}_t = \mathbb{E}[v_t|x_t, t, d]$. This allows for **taking arbitrary sized denoising steps** as the model learns to account for curvature along the path and jump to the appropriate point. However, this model suffers from poor sample quality due to complex optimization.

$$x'_{t+\Delta t} = x_t + v_\theta(x_t, t, \Delta t)\Delta t,$$

$$\mathcal{L}^s(\theta) = \mathbb{E}_{x_0,x_1,t,\Delta t}\underbrace{\|v_\theta(x_t, t, 0) - (x_1 - x_0)\|_2^2}_{\text{flow matching objective}} + \underbrace{\|v_\theta(x_t, t, 2\Delta t) - v_{\text{target}}\|_2^2}_{\text{consistency loss}}$$



4. Proposed modification: Trigonometric Flow

Recent experimental work with trigonometric flow interpolation [2] between $x_0 \sim \mathcal{N}(0, I)$ and $x_1 \sim \mathcal{D}$ results in greater sample quality than rectified flows.

$$x_t = \cos\left(\frac{\pi}{2}x_0\right) + \sin\left(\frac{\pi}{2}\right)x_1$$

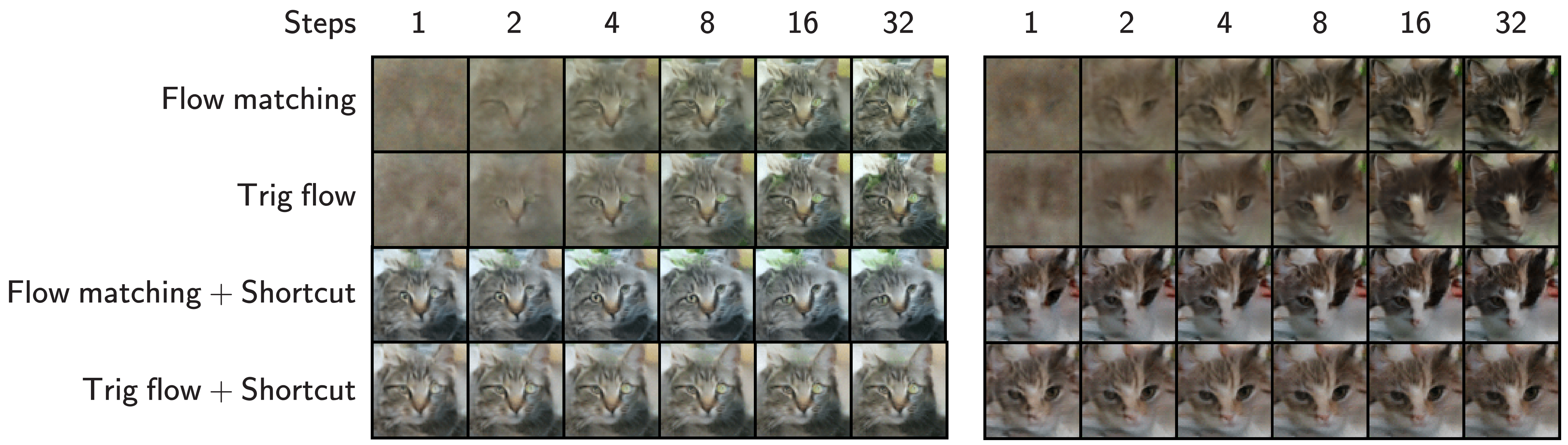
$$x_{t+\Delta t} = x_t \cos\left(\frac{\pi}{2}\Delta t\right) + v_t \sin\left(\frac{\pi}{2}\Delta t\right)/\frac{\pi}{2}$$

$$\mathcal{L}^T(\theta) = \mathbb{E}_{x_0,x_1,t}\left[\left\|v_\theta(x_t, t) - \frac{\pi}{2}(\cos\left(\frac{\pi}{2}\right)x_1 - \sin\left(\frac{\pi}{2}\right)x_0)\right\|_2^2\right]$$

In this project, we formulate shortcut diffusion models in context of trigonometric flow interpolation. Deriving the training objective and sampling scheme for trigonometric flow with and without shortcut model are deliverables for this project.

5. Numerical experiments

We train a flow matching model, trigonometric flow model with and without the shortcut formulation on AFHQ cats dataset. The backbone model is a U-Net with 9 m parameters.



Comparison of FID scores for different number of denoising steps

Diffusion method \ Steps	1	2	4	8	16	32
Flow matching [1]	331	124	77	83	88	90
Trigonometric flow [2]	322	125	70	66	73	82
Flow matching + Shortcut [3]	77	70	69	68	69	72
Trigonometric flow + Shortcut **ours**	**69**	**64**	**65**	**64**	**64**	**66**

6. Conclusions

In this project, we have formulated shortcut diffusion models in context of trigonometric flow interpolation. We have derived the training objective and sampling scheme for trigonometric flow with and without the shortcut model formulation. For the task of few-step generation, our trig flow formulation for shortcut model surpasses all baselines in terms of sample accuracy. This work is publicly available at `https://github.com/vpuri3/FastDiffusion.py`.

7. References

[1] Liu X et al. ICLR. 2023.
[2] Lu C et al. arXiv:2410.1108. 2024.
[3] Frans K et al. arXiv:2410.12557. 2024.