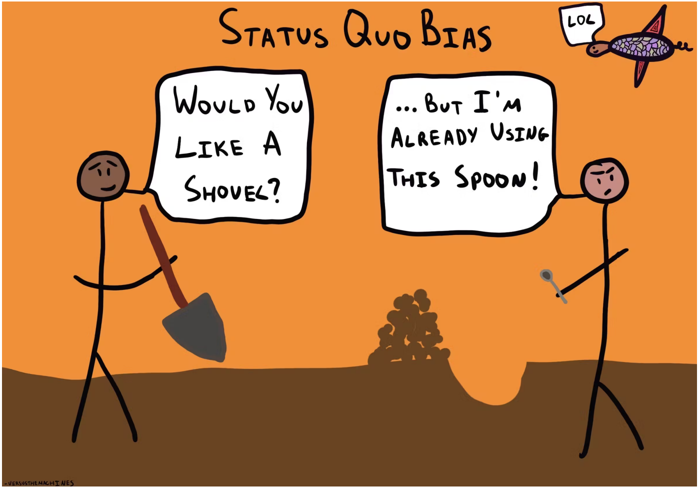
Beyond Defaults:
Is Noise Conditioning Necessary for Diffusion Models?

Junoh Kang

Computer Vision Laboratory
ECE, Seoul National University
junoh.kang@snu.ac.kr

Status quo bias

Status quo bias is people's irrational preference for maintaining current situation or state of affairs.



uEDM [Sun et al., 2025]

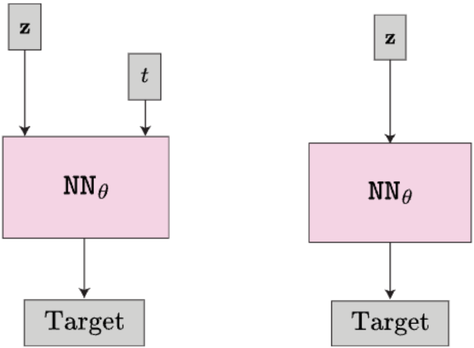
Is Noise Conditioning Necessary for Denoising Generative Models?

Qiao Sun [*1] **Zhicheng Jiang** [*1] **Hanhong Zhao** [*1] **Kaiming He** [1]

uEDM [Sun et al., 2025]

This paper examines **the necessity of noise conditioning** in denoising-based generative models.

$\mathtt{NN}_\theta(\mathbf{x}_t|t)$ v.s. $\mathtt{NN}_\theta(\mathbf{x}_t)$



uEDM [Sun et al., 2025]

The motivation of challenging the necessity of noise levels is that ”The noise level can be estimated from corrupted data”

noise level = 0.2

noise level = 0.3

uEDM [Sun et al., 2025]

Analysis on model accuracy

A loss function is defined as

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x},\epsilon,t}\left[w(t)||\mathrm{NN}_\theta(\mathbf{x}_t|t) - r(\mathbf{x},\epsilon,t)||_2^2\right] \quad (1)$$

- Data point : $\mathbf{x} \sim p_{\text{data}}$
- Noise : $\epsilon \sim p_{\text{noise}}$ (i.e. $\mathcal{N}(\mathbf{0}, \mathbf{I})$)
- Noisy image : $\mathbf{x}_t = a(t)\mathbf{x} + b(t)\epsilon \sim p_t$
- Regression target : $r(\mathbf{x},\epsilon,t) = c(t)\mathbf{x} + d(t)\epsilon$

	iDDPM, DDIM	**EDM**	**FM**
$a(t)$	$\sqrt{\bar{\alpha}(t)}$	$\frac{1}{\sqrt{t^2+\sigma_{\mathrm{d}}^2}}$	$1-t$
$b(t)$	$\sqrt{1-\bar{\alpha}(t)}$	$\frac{t}{\sqrt{t^2+\sigma_{\mathrm{d}}^2}}$	t
$c(t)$	0	$\frac{t}{\sigma_{\mathrm{d}}\sqrt{t^2+\sigma_{\mathrm{d}}^2}}$	-1
$d(t)$	1	$-\frac{\sigma_{\mathrm{d}}}{\sqrt{t^2+\sigma_{\mathrm{d}}^2}}$	1

uEDM [Sun et al., 2025]

Analysis on model accuracy

Definition 1

An **Effective target** of a model $\texttt{NN}_\theta$ is a function that model learns ideally to minimize its training loss.

Loss function of a noise-conditional model $\texttt{NN}_\theta(\mathbf{x}_t|t)$:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x},\epsilon,t}\left[w(t)||\texttt{NN}_\theta(\mathbf{x}_t|t) - r(\mathbf{x},\epsilon,t)||_2^2\right].$$

An effective target of the noise-conditional model $\texttt{NN}_\theta(\mathbf{x}_t|t)$:

$$R(\mathbf{x}_t|t) = \mathbb{E}_{(\mathbf{x},\epsilon)\sim p(\mathbf{x},\epsilon|\mathbf{x}_t,t)}\left[r(\mathbf{x},\epsilon,t)\right].$$

uEDM [Sun et al., 2025]

Analysis on model accuracy

Loss function of a noise-unconditional model $\mathtt{NN}_\theta(\mathbf{x}_t)$:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x},\epsilon,t}[w(t)||\mathtt{NN}_\theta(\mathbf{x}_t) - r(\mathbf{x}, \epsilon, t)||_2^2].$$

An effective target of the noise-unconditional model $\mathtt{NN}_\theta(\mathbf{x}_t)$:

$$R(\mathbf{x}_t) = \mathbb{E}_{t\sim p(t|\mathbf{x}_t)}[R(\mathbf{x}_t|t)].$$

uEDM [Sun et al., 2025]

Analysis on model accuracy

Ideally,

$$\mathtt{NN}_\theta(\mathbf{x}_t|t) \to R(\mathbf{x}_t|t), \tag{2}$$

$$\mathtt{NN}_\theta(\mathbf{x}_t) \to R(\mathbf{x}_t) = \mathbb{E}_{t\sim p(t|\mathbf{x}_t)}[R(\mathbf{x}_t|t)]. \tag{3}$$

If $p(t|\mathbf{x}_t)$ is close enough to Dirac delta function, the effective targets of noise-conditional and noise-unconditional models would be the same.

$$\mathrm{Var}(p(t|\mathbf{x}_t)) \to 0 \;\; \Rightarrow \;\; R(\mathbf{x}_t) \to R(\mathbf{x}_t|t)$$

uEDM [Sun et al., 2025]

Analysis on model accuracy

Theoretically, the variance of $p(t|\mathbf{x}_t)$ is

Statement 1 (Concentration of $p(t|\mathbf{z})$). *Consider a single datapoint* $\mathbf{x} \in [-1, 1]^d$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, $t \sim \mathcal{U}[0, 1]$, *and* $\mathbf{z} = (1-t)\mathbf{x} + t\boldsymbol{\epsilon}$ *(the Flow Matching case). Given a noisy image* $\mathbf{z} = (1-t_*)\mathbf{x} + t_*\boldsymbol{\epsilon}$ *produced by a given* t_*, *the variance of* t *under the conditional distribution* $p(t|\mathbf{z})$, *is:*

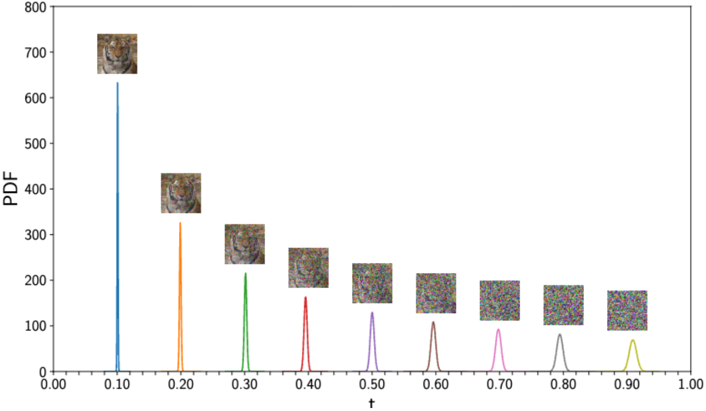
$$\mathrm{Var}_{t\sim p(t|\mathbf{z})}[t] \approx \frac{t_*^2}{2d}, \tag{9}$$

when the data dimension d *satisfies* $\frac{1}{d} \ll t_*$ *and* $\frac{1}{d} \ll 1 - t_*$. *(Derivation in Appendix C.2)*

uEDM [Sun et al., 2025]

Analysis on model accuracy

Empirically, the distribution of $p(t|\mathbf{x}_t)$ is



uEDM [Sun et al., 2025]

Analysis on model accuracy

The norm between two effective targets $E(\mathbf{x}_t)$ is $1/10^3$ of $R(\mathbf{x}_t)$.

$$E(\mathbf{x}_t) = \mathbb{E}_{t \sim p(t|\mathbf{x}_t)} \left[||R(\mathbf{x}_t|t) - R(\mathbf{x}_t)||_2^2 \right]$$

Statement 2 (Error of effective regression targets). *Consider the scenario in Statement 1 and the Flow Matching case. The error defined in Eq. (10) satisfies:*

$$E(\mathbf{z}) \approx \frac{1}{2}(1 + \sigma_{\mathrm{d}}^2) \tag{11}$$

when the data dimension d satisfies $\frac{1}{d} \ll t_*$ and $\frac{1}{d} \ll 1 - t_*$. Here, σ_{d} denotes the per-pixel standard deviation of the dataset. (Derivation in Appendix C.3)

uEDM [Sun et al., 2025]

Analysis on model accuracy

Empirically, the distribution of $p(t|\mathbf{x}_t)$ is

t_*	$\mathrm{Var}_{t\sim p(t\|\mathbf{z})}[t]$		$E(\mathbf{z})$		$\|R(\mathbf{z})\|^2$
	Empirical ($\times 10^{-4}$)	Estimation ($\times 10^{-4}$)	Empirical	Estimation	Empirical
0.1	0.0143 ± 0.0002	0.0163	0.558 ± 0.005	0.628	3894 ± 87
0.3	0.1280 ± 0.0002	0.1465	0.561 ± 0.006	0.628	3953 ± 102
0.5	0.3695 ± 0.0004	0.4069	0.556 ± 0.006	0.628	3878 ± 108
0.7	0.7008 ± 0.0010	0.7975	0.564 ± 0.005	0.628	3968 ± 88
0.9	1.3085 ± 0.0007	1.3184	1.822 ± 0.245	0.628	3310 ± 71

Statement 3

Noise-conditional and unconditional samplings are expressed as

$$\mathbf{x}_{i+1} = \kappa_i \mathbf{x}_i + \eta R(\mathbf{x}_i | t_i) + \xi \epsilon_i, \tag{4}$$

$$\mathbf{x}'_{i+1} = \kappa_i \mathbf{x}'_i + \eta R(\mathbf{x}'_i) + \xi \epsilon_i. \tag{5}$$

Assuming

1. *Lipshitz* $||R(\mathbf{x}'_i | t_i) - R(\mathbf{x}_i | t_i)|| \leq L||x'_i - x_i||$,
2. $||R(\mathbf{x}'_i | t_i) - R(\mathbf{x}'_i)|| \leq \delta_i$,

$$||\mathbf{x}_N - \mathbf{x}'_N|| \leq A_0 B_0 + \ldots + A_{N-1} B_{N-1}, \tag{6}$$

where $A_i = \prod_{j=i+1}^{N-1} (\kappa_i + |\eta_i| L_i)$ *and* $B_i = |\eta_i| \delta_i$.

uEDM [Sun et al., 2025]

Training noise-unconditional model

Using EDM [Karras et al., 2022] as a backbone, they modified schedules to make model robust to noise in the absence of noise conditioning.

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x},\epsilon,t}\left[w(t) || \underbrace{c_{\text{skip}}(t)\mathbf{x}_t + c_{\text{out}}(t)\mathrm{NN}_\theta(c_{\text{in}}(t)\mathbf{x}_t | t)}_{\text{Denoiser}} - \mathbf{x}||_2^2 \right],$$

where $c_{\text{skip}}(t) = \sigma_d^2/(t^2 + \sigma_d^2)$,

$$c_{\text{out}}(t) = t \cdot \sigma_d / \sqrt{t^2 + \sigma_d^2} \Rightarrow 1,$$

$$c_{\text{in}}(t) = 1/\sqrt{t^2 + \sigma_d^2} \Rightarrow 1/\sqrt{t^2 + 1}.$$

uEDM [Sun et al., 2025]

Results

CIFAR10-unconditional

model	sampler	NFE	FID w/ t	$\rightarrow$	FID w/o t
iDDPM	SDE	500	3.13	$\rightarrow$	5.51
iDDPM ($\mathbf{x}$-pred)	SDE	500	5.64	$\rightarrow$	6.33
DDIM	ODE	100	3.99	$\rightarrow$	40.90
	SDE	100	8.07	$\rightarrow$	10.85
	SDE	1000	3.18	$\rightarrow$	5.41
ADM	SDE	250	2.70	$\rightarrow$	5.27
EDM	Heun	35	**1.99**	$\rightarrow$	3.36
	Euler	50	2.98	$\rightarrow$	4.55
FM (1-RF)	Euler	100	3.01	$\rightarrow$	2.61
	Heun	99	2.87	$\rightarrow$	2.63
	RK45	$\sim$127	2.53	$\rightarrow$	2.63
iCT	-	2	2.59	$\rightarrow$	3.57
ECM	-	2	2.57	$\rightarrow$	3.27
uEDM (Sec. 5)	Heun	35	2.04	$\rightarrow$	**2.23**

uEDM [Sun et al., 2025]

Results

- Even without noise conditioning, most experiments show comparative results
- Especially for Flow Matching, they outperform without noise conditioning.
- DDIM with ODE sampling suffers from severe degradation.

uEDM [Sun et al., 2025]

Results

Recall

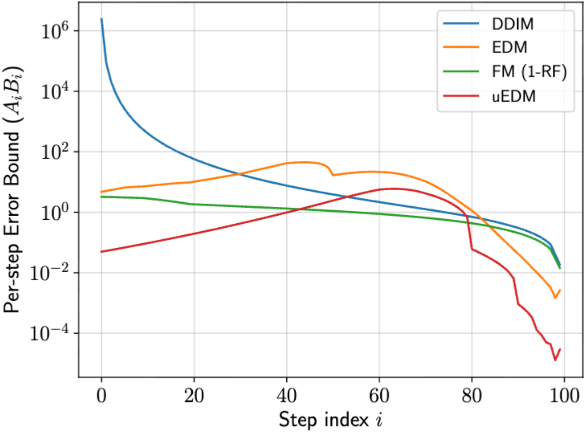
$$||\mathbf{x}_N - \mathbf{x}'_N|| \le A_0 B_0 + \ldots + A_{N-1} B_{N-1}, \quad (7)$$

where $A_i = \prod_{j=i+1}^{N-1}(\kappa_i + |\eta_i| L_i)$ and $B_i = |\eta_i| \delta_i$.

Sampling				
κ_i	$\boxed{\sqrt{\frac{\bar{\alpha}_{i+1}}{\bar{\alpha}_i}} > 1}$	$\sqrt{\frac{\bar{\alpha}_{i+1}}{\bar{\alpha}_i}}$	$\boxed{\sqrt{\frac{\sigma_d^2+t_i^2}{\sigma_d^2+t_{i+1}^2}}\left(1-\frac{t_i(t_i-t_{i+1})}{t_i^2+\sigma_d^2}\right) \sim 1}$	0
η_i	$\frac{1}{\sqrt{1-\bar{\alpha}_i}}\left(\sqrt{\frac{\bar{\alpha}_i}{\bar{\alpha}_{i+1}}}-\sqrt{\frac{\bar{\alpha}_{i+1}}{\bar{\alpha}_i}}\right)$	$\sqrt{1-\bar{\alpha}_{i+1}}-\sqrt{\frac{\bar{\alpha}_{i+1}}{\bar{\alpha}_i}(1-\bar{\alpha}_i)}$	$\frac{\sigma_d(t_i-t_{i+1})}{\sqrt{(t_i^2+\sigma_d^2)(t_{i+1}^2+\sigma_d^2)}}$	$t_{i+1}-t_i$
ζ_i	$\sqrt{\left(1-\frac{\bar{\alpha}_i}{\bar{\alpha}_{i+1}}\right)\frac{1-\bar{\alpha}_{i+1}}{1-\bar{\alpha}_i}}$	0	0	0
Schedule $t_{0\sim N}$	$t_i = \frac{N-i}{N}\cdot T$	$t_i = \frac{N-i}{N}\cdot T$	$t_i = \left(t_{\max}^{\frac{1}{\rho}}+\frac{i}{N}\left(t_{\min}^{\frac{1}{\rho}}-t_{\max}^{\frac{1}{\rho}}\right)\right)^{\rho}$	$t_i = 1-\frac{i}{N}$

uEDM [Sun et al., 2025]

Results

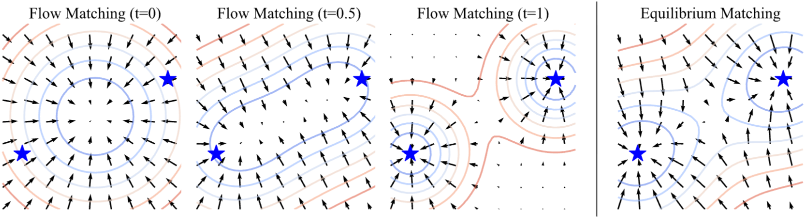


Model	accum. bound	FID w/ $t \rightarrow$ w/o t
DDIM	3e6	3.99 $\rightarrow$ 40.90
EDM	1e3	2.34 $\rightarrow$ 3.80
FM (1-RF)	1e2	3.01 $\rightarrow$ 2.61
uEDM (Sec. 5)	1e2	2.62 $\rightarrow$ 2.66

uEDM [Sun et al., 2025]

Conclusion

We can interpret diffusion models as learning p_t, or some varying energy function $\{E(x, t)\}_t$. Without noise conditioning, the model learns a single energy function $E(x)$, which aligns to classical EBM.



EqM [Wang and Du, 2025]

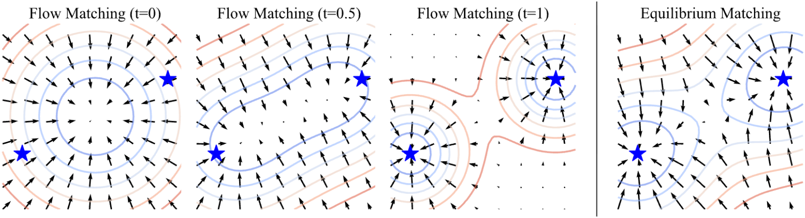
EQUILIBRIUM MATCHING: GENERATIVE MODELING WITH IMPLICIT ENERGY-BASED MODELS

Runqian Wang
MIT
raywang4@mit.edu

Yilun Du
Harvard University
ydu@seas.harvard.edu

EqM [Wang and Du, 2025]

EqM aims to train diffusion models **without noise condition**.



In other perspective, EqM models **fixed energy landscape** (or corresponding gradient field), and is similar to EBMs.

EBM to flow matching

EBMs define probability density via an energy function $E(\mathbf{x})$, and the probability is defined by

$$p(\mathbf{x}) = \frac{\exp(-E(\mathbf{x}))}{Z}, \text{ where } Z = \int \exp(-E(\mathbf{x}))d\mathbf{x}.$$

EBMs are trained with

$$\mathcal{L}_{\text{EBM}} = -\mathbb{E}_{p_{\text{data}}}[E(\mathbf{x})] - \log Z,$$

and the calculation of $\log Z$ is usually intractable.

EqM [Wang and Du, 2025]

Training

Objective of Flow matching

$$\mathcal{L}_{\text{FM}} = ||f_\theta(\mathbf{x}_t, t) - (\mathbf{x} - \epsilon)||_2^2 \tag{8}$$

Objective of EqM

$$\mathcal{L}_{\text{EqM}} = ||f_\theta(\mathbf{x}_\gamma) - (\epsilon - \mathbf{x})c(\gamma)||_2^2. \tag{9}$$

- EqM learns gradient unlike flow mathcing.
- **EqM designs $c(1) = 0$ so that data point is at local minima.**

EqM [Wang and Du, 2025]

Training

Objective of EqM-E (learning explicit energy)

$$\mathcal{L}_{\text{EqM-E}} = ||\nabla g_\theta(\mathbf{x}_\gamma) - (\epsilon - \mathbf{x})c(\gamma)||_2^2, \tag{10}$$

where $g_\theta(\mathbf{x}_\gamma) = \mathbf{x}_\gamma \cdot f(\mathbf{x}_\gamma)$ or $g_\theta(\mathbf{x}_\gamma) = -\frac{1}{2}||f(\mathbf{x}_\gamma)||_2^2$.

EqM [Wang and Du, 2025]

Sampling

EqM generates samples via optimization on the learned landscape. The name **Equilibrium** comes from the fact that the model learns a static system where the **data points are the low-energy** stable states, and sampling is the process of seeking this equilibrium.

Gradient Descent

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \eta \nabla E(\mathbf{x}_k)$$

- This can be also interpreted as ODE solving.

Nesterov Accelerated Gradient

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \eta \nabla E(\mathbf{x}_k + \mu(\mathbf{x}_k - \mathbf{x}_{k-1}))$$

- Samplings can be adaptively done: e.g. $||\nabla E(\mathbf{x}_k)||_2 < thres$.

EqM [Wang and Du, 2025]

Analysis

Statement 1 (Learned Gradient at Ground-Truth Samples)**.** *Let f be an Equilibrium Matching model with $c(1) = 0$, and let $x^{(i)}$ be a ground-truth sample in $\mathbb{R}^d$. Assume perfect training, i.e., f exactly minimizes the training objective. Then, in high-dimensional settings, we have:*

$$\|f(x^{(i)})\|_2 \approx 0.$$

where $x^{(i)}$ is an arbitrary sample from the training dataset. In other words, Equilibrium Matching assigns ground-truth images with approximately 0 gradient. (Derivation in Appendix C.1)

- $E(\mathbf{x}) \approx 0$ for $\mathbf{x} \in \mathcal{D}$.

EqM [Wang and Du, 2025]

Analysis

Statement 2 (Property of Local Minima). *Let f be an Equilibrium Matching model with* $c(1) = 0$, *and let $\hat{x}$ be an arbitrary local minimum where $f(\hat{x}) = 0$. Assume perfect training, i.e., f exactly minimizes the training objective. Then, in high-dimensional settings, we have:*

$$P(\hat{x} \in \mathcal{X}) \approx 1.$$

where $\mathcal{X}$ is the ground-truth dataset. In other words, all local minima are approximately samples from the ground-truth dataset. (Derivation in Appendix C.2)

- $E(\mathbf{x}) = 0 \Rightarrow \mathbf{x} \in \mathcal{D}$ with probability 1.

Combining statement 1 and 2, all local minima are approximately samples from the dataset.

EqM [Wang and Du, 2025]

Results

Class conditional ImageNet 256x256, 250 steps.

model	method	FID
StyleGAN-XL	GAN	2.30
VDM++	Diffusion	2.12
DiT-XL/2	Diffusion	2.27
SiT-XL/2	FM	2.06
EqM-XL/2	**EqM**	**1.90**

EqM [Wang and Du, 2025]

Results

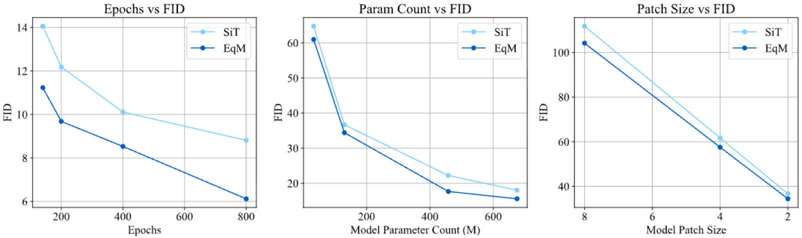
Class conditional ImageNet 256x256, 250 steps.

model	sampler	η	μ	FID
SiT-XL/2	Euler (ODE)	0.0040	-	2.10
SiT-XL/2	Heun (SDE)	0.0040	-	2.06
EqM-XL/2	Euler (ODE)	0.0017	-	1.93
EqM-XL/2	GD	0.0017	-	1.93
EqM-XL/2	NAG-GD	0.0017	0.3	**1.90**

EqM [Wang and Du, 2025]

Results

Scalability



Ablations the design of $c(\gamma)$ (experiment with $\lambda = 1$)

$c(\gamma)$	constant	linear	truncated			piecewise	
a	-	-	0.5	0.8	0.9	0.8	0.8
b	-	-	-	-	-	0.8	1.4
FID	40.81	50.47	38.98	**38.34**	41.22	38.84	38.75

EqM [Wang and Du, 2025]

Results

Ablations on noise conditioning (experiment with $\lambda = 4$)

noise conditioning $c(\gamma)$	yes const	yes trunc	no const	no trunc
FID	36.68	41.89	40.81	**32.85**

- $c(\gamma) = c$ is flow matching.

EqM [Wang and Du, 2025]

Results

Explicit energy model

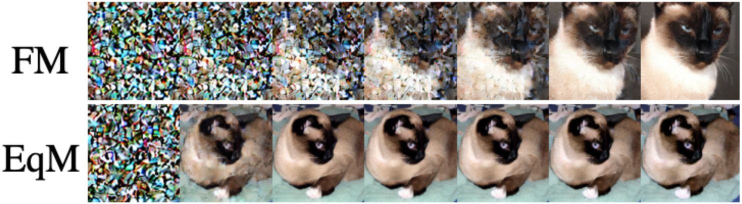
energy model	FID
none	57.54
dot product	73.40
L_2 norm	75.53

- Can be used for OOD detection.

EqM [Wang and Du, 2025]

The reason why EqM can have better performance than flow matching might be

- We might spend too much time on the noisy latent for flow matching.



- Selecting time steps in diffusion models affects sample quality. The model automatically select time steps for EqM.

EqM [Wang and Du, 2025]

The other advantages of EqMs are (my opinion)

- Code is simple as it does not require $\bar{\alpha}_i$ during sampling
- Less restrictions on added noise.

I am looking for the reason why this was not attempted before diffusion models. This is much similar to EBMs which are classics.

Random thoughts

- We might have to revisit conventions and examine their necessity. This can make modeling and implementation simpler.
- Adding functionality is one direction of the research. Making something simpler is also possible direction of the research.
- Studying classics might be helpful. Recently, ”The Principles of Diffusion Models” has been released from Ermon laboratory.

Reference I

Karras, T., Aittala, M., Aila, T., and Laine, S. (2022).
Elucidating the design space of diffusion-based generative models.
NeurIPS.

Sun, Q., Jiang, Z., Zhao, H., and He, K. (2025).
Is noise conditioning necessary for denoising generative models?
arXiv preprint arXiv:2502.13129.

Wang, R. and Du, Y. (2025).
Equilibrium matching: Generative modeling with implicit energy-based models.
arXiv preprint arXiv:2510.02300.