## References


Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-dpm: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. *arXiv preprint arXiv:2201.06503*, 2022.

Yongxin Chen and Tryphon Georgiou. Stochastic bridges of linear systems. *IEEE Transactions on Automatic Control*, 61(2):526–531, 2015.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8188–8197, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. *arXiv preprint arXiv:2105.05233*, 2021.

Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. *arXiv preprint arXiv:2112.07068*, 2021.

Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pp. 1188–1205, 1986.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arXiv:2006.11239*, 2020.

Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numerica*, 19:209–286, 2010.

The MathWorks Inc. Matlab version: 9.13.0 (r2022b), 2022. URL `https://www.mathworks.com`.

HJ Kappen. Stochastic optimal control theory. *ICML, Helsinki, Radbound University, Nijmegen, Netherlands*, 2008.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022.

Donald E Kirk. *Optimal control theory: an introduction*. Courier Corporation, 2004.

Peter E Kloeden, Eckhard Platen, Peter E Kloeden, and Eckhard Platen. *Stochastic differential equations*. Springer, 1992.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Christian Léonard, Sylvie Rœlly, and Jean-Claude Zambrini. Reciprocal processes. a measure-theoretical point of view. 2014.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I2sb: Image-to-image schr\" odinger bridge. *arXiv preprint arXiv:2302.05872*, 2023.

Xingchao Liu, Lemeng Wu, Mao Ye, and Qiang Liu. Let us build bridges: Understanding and extending diffusion generative models. *arXiv preprint arXiv:2208.14699*, 2022.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pp. 65–84. Springer, 2003.

Stefano Peluchetti. Non-denoising forward-time diffusions. 2021.

Stefano Peluchetti. Diffusion bridge mixture transports, schr\" odinger bridge problems and generative modeling. *arXiv preprint arXiv:2304.00917*, 2023.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky Chen. Multisample flow matching: Straightening flows with minibatch couplings. *arXiv preprint arXiv:2304.14772*, 2023.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.

Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Conditional simulation using diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pp. 1792–1802. PMLR, 2022.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schr\" odinger bridge matching. *arXiv preprint arXiv:2303.16852*, 2023.

Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schr\" odinger bridges. *arXiv preprint arXiv:2302.11419*, 2023.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *arXiv e-prints*, pp. arXiv–2101, 2021.

Alexander Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free schr\" odinger bridges via score and flow matching. *arXiv preprint arXiv:2307.03672*, 2023a.

Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023b.

Jiongmin Yong and Xun Yu Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.

Qinsheng Zhang, Molei Tao, and Yongxin Chen. gddim: Generalized denoising diffusion implicit models. *arXiv preprint arXiv:2206.05564*, 2022.

Qinsheng Zhang, Jiaming Song, and Yongxin Chen. Improved order analysis and design of exponential integrator for diffusion models sampling. *arXiv preprint arXiv:2308.02157*, 2023.

# A SUPPLEMENTARY SUMMARY

We state the assumptions in Appendix.B. We provide the technique details appearing in Section.3 at Appendix.C. The details of the experiments can be found in Appendix.D. The visualization of generated figures can be found in Appendix.E.

# B ASSUMPTIONS

We will use the following assumptions to construct the proposed method. These assumptions are adopted from stochastic analysis for SGM (Song et al., 2021; Yong & Zhou, 1999; Anderson, 1982),

(i) $p_0$ and $p_1$ with finite second-order moment.

(ii) $g_t$ is continuous functions, and $|g(t)|^2 > 0$ is uniformly lower-bounded w.r.t. $t$.

(iii) $\forall t \in [0, 1]$, we have $\nabla_{\mathbf{v}} \log p_t(\mathbf{m}_t, t)$ Lipschitz and at most linear growth w.r.t. $\mathbf{x}$ and $\mathbf{v}$.

Assumptions (i) (ii) are standard conditions in stochastic analysis to ensure the existence-uniqueness of the SDEs; hence also appear in SGM analysis (Song et al., 2021).

# C TECHNIQUE DETAILS IN SECTION.3

## C.1 BROWNIAN BRIDGE AS THE SOLUTION OF STOCHASTIC OPTIMAL CONTROL

We adopt the presentation form Kappen (2008). We consider the control problem:

$$\begin{aligned} &\min_{\mathbf{v}_t} \int_{t_0}^{1} \frac{1}{2}\|\mathbf{v}_t\|_2^2 \mathrm{d}t + \frac{\mathbf{r}}{2}\|\mathbf{x}_1 - x_1\|_2^2 \\ &\text{s.t.} \quad \mathrm{d}\mathbf{x}_t = \mathbf{v}_t \mathrm{d}t, \quad \mathbf{x}_0 = x_0 \end{aligned}$$

Where $\mathbf{r}$ is the terminal cost coefficient. According to Pontryagin Maximum Principle (PMP;Kirk (2004)) recipe, one can construct the Hamiltonian:

$$H(t, \mathbf{x}, \mathbf{v}, \gamma) = -\frac{1}{2}\|\mathbf{v}_t\|_2^2 + \gamma \mathbf{v}_t$$

Hence the optimized Hamiltonian is:

$$H(t, \mathbf{x}, \mathbf{v}, \gamma)^* = \frac{1}{2}\gamma^2, \quad \text{where} \quad \mathbf{v}_t = \gamma$$

Then we solve the Hamiltonian equation of motion:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} &= \frac{\partial H^*}{\partial \gamma} = \gamma \\ \frac{\mathrm{d}\gamma}{\mathrm{d}t} &= \frac{\partial H^*}{\partial \mathbf{x}} = 0 \\ \text{where} \quad \mathbf{x}_0 = x_0 \quad &\text{and} \quad \gamma_1 = -\mathbf{r} \cdot (\mathbf{x}_1 - x_1) \end{aligned}$$

One can notice that the solution for $\gamma_t$ is the constant $\gamma_t = \gamma = -\mathbf{r} \cdot (\mathbf{x}_1 - x_1)$, hence the solution for $\mathbf{x}_t$ is $\mathbf{x}_t = \mathbf{x}_1 + \gamma t$.

$$\begin{aligned} &\gamma = -\mathbf{r}(\mathbf{x}_1 - x_1) = -\mathbf{r}(\mathbf{x}_0 + (1 - t_0)\gamma - x_1) \\ \rightarrow \quad &\mathbf{v}_t^* := \gamma = \frac{\mathbf{r}(x_1 - \mathbf{x}_0)}{1 + \mathbf{r}(1 - t_0)} \end{aligned}$$

When $\mathbf{r} \rightarrow +\infty$, we arrive the optimal control as $\mathbf{v}_t^* = \frac{x_1 - \mathbf{x}_0}{1 - t_0}$. Due to certainty equivalence, this is also the optimal control law for

$$\mathrm{d}\mathbf{x}_t = \mathbf{v}_t \mathrm{d}t + \mathrm{d}\mathbf{w}_t$$

By plugging it back into the dynamics, we obtain the well-known Brownian Bridge:

$$\mathrm{d}\mathbf{x}_t = \frac{x_1 - \mathbf{x}_t}{1-t}\mathrm{d}t + g_t \mathrm{d}\mathbf{w}_t$$

### C.2 Proof of Proposition.3

**Lemma 6.** *The solution of following Lyapunov equation,*

$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^{\mathsf{T}} - \boldsymbol{g}\boldsymbol{g}^{T} \tag{12}$$

*with terminal condition*

$$\mathbf{P}_T = \begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix} \tag{13}$$

*is given by*

$$\mathbf{P}_t = \begin{bmatrix} r(t-1)^2 - \frac{1}{3}g^2(t-1)^3 & r(t-1) - \frac{1}{2}g^2(t-1)^2 \\ r(t-1) - \frac{1}{2}g^2(t-1)^2 & g^2(1-t) + r \end{bmatrix}$$

*and the inverse of $\mathbf{P}_t$ is,*

$$\mathbf{P}_t^{-1} = \frac{1}{g^2(-4r + g^2(t-1))(t-1)} \begin{bmatrix} \frac{12(r-g^2(t-1))}{(t-1)^2} & \frac{6(-2r+g^2(t-1))}{t-1} \\ \frac{6(-2r+g^2(-1+t))}{t-1} & 12r - 4g^2(t-1) \end{bmatrix}$$

*When $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\boldsymbol{g} = \begin{bmatrix} 0 \\ g \end{bmatrix}$ and $R = r\mathbf{I}$.*

*Thus,*

$$P_{10} = \frac{-12r + 6g^2(t-1)}{g^2[-4r + g^2(t-1)](t-1)^2} = \frac{-12r}{g^2[-4r + g^2(t-1)](t-1)^2} + \frac{6}{[-4r + g^2(t-1)](t-1)}$$
$$P_{11} = \frac{12r - 4g^2(t-1)}{g^2[-4r + g^2(t-1)](t-1)} = \frac{12r}{g^2[-4r + g^2(t-1)](t-1)} + \frac{-4}{[-4r + g^2(t-1)]}$$

*Proof.* One can plug in the solution of $\mathbf{P}_t$ into the Lyapunov equation $\mathbf{P}_t$ and it validates $\mathbf{P}_t$ is indeed the solution.

**Remark 7.** *Here we provide a general form when the terminal condition of the Lyapunov function is not a zero matrix. It explicitly means that it allows that the velocity does not necessarily need to converge to the exact predefined $\mathbf{v}_1$. It will have the same results as shown in the paper by setting $r = 0$.*

□

**Lemma 8.** *The state transition function $\Phi(t, s)$ of following dynamics,*

$$\mathbf{dm}_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{m}_t \mathrm{d}t$$

*is,*

$$\Phi(t, s) = \begin{bmatrix} 1 & t-s \\ 0 & 1 \end{bmatrix}$$

*Proof.* One can easily verify that such $\Phi$ satisfies $\partial\Phi/\partial t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\Phi$. □

**Lemma 9** (Chen & Georgiou (2015)). *The optimal control $\mathbf{u}_t^*$ of following problem,*

$$\mathbf{u}_t^* \in \arg\min_{\mathbf{u}_t \in \mathcal{U}} \mathbb{E}\left[\int_0^T \frac{1}{2}\|\mathbf{u}_t\|^2\right] \mathrm{d}t$$
$$s.t \quad \mathrm{d}\mathbf{m}_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{m}_t \mathrm{d}t + \mathbf{u}_t \mathrm{d}t + \mathbf{g}\mathrm{d}\mathbf{w}_t$$
$$\mathbf{m}_0 = m_0, \quad \mathbf{m}_1 = m_1$$

*is given by*

$$\mathbf{u}_t^* = -\mathbf{g}\mathbf{g}^\mathsf{T}\mathbf{P}_t^{-1}\left(\mathbf{m}_t - \Phi(t,1)\mathbf{m}_1\right)$$

*Where $\mathbf{P}_t$ follows Lyapunov equation (eq.12) with boundary condition $\mathbf{P}_1 = \mathbf{0}$. and function $\Phi(t,s)$ is the transition matrix from time-step $s$ to time-step $t$ given uncontrolled dynamics.*

*Proof.* See Chen & Georgiou (2015). □

**Proposition 10.** *The solution of the stochastic bridge problem of linear momentum system (Chen & Georgiou, 2015) is*

$$\mathbf{a}^*(\mathbf{m}_t, t) = g_t^2 P_{11}\left(\frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t} - \mathbf{v}_t\right) \quad \textit{where}: \quad P_{11} = \frac{-4}{g_t^2(t-1)}. \tag{14}$$

*Proof.* From Lemma.9, one can get the optimal control for this problem is

$$\mathbf{u}_t^* = -\mathbf{g}\mathbf{g}^\mathsf{T}\mathbf{P}_t^{-1}\left(\mathbf{m}_t - \Phi(t,1)\mathbf{m}_1\right)$$

where state transition function $\Phi$ can be obtained from Lemma.8 and $\mathbf{P}_t$ is the solution of Lyapunov equation and $\mathbf{P}_t^{-1}$ can be found in Lemma.6.

Then we have:

$$\begin{aligned}
\mathbf{u}_t^* &= -\mathbf{g}\mathbf{g}^\mathsf{T}\mathbf{P}_t^{-1}\left(\mathbf{m}_t - \Phi(t,1)\mathbf{m}_1\right) \\
&= -\mathbf{g}\mathbf{g}^\mathsf{T}\mathbf{P}_t^{-1}\mathbf{m}_t + \mathbf{g}\mathbf{g}^\mathsf{T}\mathbf{P}_t^{-1}\Phi(t,1)\mathbf{m}_1 \\
&= -\begin{bmatrix} 0 & 0 \\ 0 & g^2 \end{bmatrix}\mathbf{P}_t^{-1}\mathbf{m}_t + \mathbf{g}\mathbf{g}^\mathsf{T}\mathbf{P}_t^{-1}\begin{bmatrix} 1 & t-1 \\ 0 & 1 \end{bmatrix}\mathbf{m}_1 \\
&= -g_t^2\begin{bmatrix} 0 & 0 \\ P_{10} & P_{11} \end{bmatrix}\mathbf{m}_t + \begin{bmatrix} 0 & 0 \\ 0 & g_t^2 \end{bmatrix}\begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}\begin{bmatrix} 1 & t-1 \\ 0 & 1 \end{bmatrix}\mathbf{m}_1 \\
&= -g_t^2\begin{bmatrix} 0 & 0 \\ P_{10} & P_{11} \end{bmatrix}\mathbf{m}_t + g_t^2\begin{bmatrix} 0 & 0 \\ P_{10} & P_{11} \end{bmatrix}\begin{bmatrix} 1 & t-1 \\ 0 & 1 \end{bmatrix}\mathbf{m}_1 \\
&= -g_t^2\begin{bmatrix} 0 & 0 \\ P_{10} & P_{11} \end{bmatrix}\mathbf{m}_t + g_t^2\begin{bmatrix} 0 & 0 \\ P_{10} & P_{10}(t-1) + P_{11} \end{bmatrix}\mathbf{m}_1 \\
&= \begin{bmatrix} 0 \\ g_t^2 P_{10}(\mathbf{x}_1 - \mathbf{x}_t) + g_t^2 P_{10}(t-1)\cdot\mathbf{v}_1 + g_t^2 P_{11}(\mathbf{v}_1 - \mathbf{v}_t) \end{bmatrix} \\
&\text{Plug in } \mathbf{v}_1 := \frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t} \\
&= \begin{bmatrix} 0 \\ g_t^2 P_{11}\left(\frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t} - \mathbf{v}_t\right) \end{bmatrix}
\end{aligned}$$

□

### C.3 MEAN AND COVARIANCE OF SDE

We follow the recipe of Särkkä & Solin (2019). The mean $\boldsymbol{\mu}_t$ and variance $\boldsymbol{\Sigma}_t$ of the matrix of random variable $\mathbf{m}_t$ obey the following respective ordinary differential equations (ODEs):

$$\mathrm{d}\boldsymbol{\mu}_t = \mathbf{F}_t\boldsymbol{\mu}_t \mathrm{d}t + \mathbf{D}_t\mathrm{d}t$$
$$\mathrm{d}\boldsymbol{\Sigma}_t = \mathbf{F}_t\boldsymbol{\Sigma}_t\mathrm{d}t + [\boldsymbol{\Sigma}_t\mathbf{F}_t]^{\mathsf{T}}\,\mathrm{d}t + \mathbf{g}\mathbf{g}^{\mathsf{T}}\mathrm{d}t$$

One can solve it by numerically simulating two ODEs whose dimension is just two. Or one can use software such as Inc. (2022) to get analytic solutions.

### C.4 DERIVATION FROM SDE TO ODE FOR PHASE DYNAMICS

One can represent the dynamics in the form of,

$$\begin{bmatrix}\mathrm{d}\mathbf{x}_t\\ \mathrm{d}\mathbf{v}_t\end{bmatrix} = \begin{bmatrix}\mathbf{v}_t\\ \mathbf{F}_t\end{bmatrix}\mathrm{d}t + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & g_t\end{bmatrix}\mathrm{d}\mathbf{w}_t \quad \text{s.t} \quad \mathbf{m}_0 := \begin{bmatrix}\mathbf{x}_0\\ \mathbf{v}_0\end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{15}$$

as

$$\mathrm{d}\mathbf{m}_t = f(\mathbf{m}_t)\mathrm{d}t + \mathbf{g}_t\mathrm{d}\mathbf{w}_t$$

And its corresponding Fokker-Planck Partial Differential Equation Øksendal (2003) reads,

$$\frac{\partial p_t}{\partial t} = -\sum_d \frac{\partial}{\partial \mathbf{m}_i}[f_i(\mathbf{m}, t)p_t(\mathbf{m}_t)] + \frac{1}{2}\sum_d \frac{\partial^2}{\partial \mathbf{m}_i\mathbf{m}_j}\left[\sum_d \mathbf{g}_t\mathbf{g}_t^{\mathsf{T}} p_t(\mathbf{m}_t)\right] \tag{16}$$

According to eq.(37) in Song et al. (2020b), One can rewrite such PDE,

$$\frac{\partial p_t}{\partial t} = -\sum_d \frac{\partial}{\partial \mathbf{m}_i}\left\{f_i(\mathbf{m}_t, t)p_t(\mathbf{m}_t) - \frac{1}{2}\left[\nabla_{\mathbf{m}}\cdot(\mathbf{g}_t\mathbf{g}_t^{\mathsf{T}}) + \mathbf{g}_t\mathbf{g}_t^{\mathsf{T}}\nabla_{\mathbf{m}}\log p(\mathbf{m}_t)\right]\right\} \tag{17}$$

$$\text{due to the fact } \mathbf{g}_t \equiv \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & g_t\end{bmatrix} \tag{18}$$

$$= -\sum_d \frac{\partial}{\partial \mathbf{m}_i}\left\{f_i(\mathbf{m}_t, t)p_t(\mathbf{m}_t) - \frac{1}{2}\left[g_t^2\nabla_{\mathbf{v}}\log p(\mathbf{m}_t)\right]\right\} \tag{19}$$

Then one can get the equivalent ODE:

$$\mathrm{d}\mathbf{m}_t = \left[f(\mathbf{m}_t, t) - \frac{1}{2}g_t^2\nabla_{\mathbf{v}}\log p(\mathbf{m}, t)\right]\mathrm{d}t + \mathbf{g}_t\mathrm{d}\mathbf{w}_t \tag{20}$$

### C.5 DECOMPOSITION OF COVARIANCE MATRIX AND REPRESENTATION OF SCORE

Here we follow the procedure in Dockhorn et al. (2021). Given the covariance matrix $\boldsymbol{\Sigma}_t$, the decomposition of the positive definite symmetric matrix is,

$$\boldsymbol{\Sigma}_t = \mathbf{L}_t^{\mathsf{T}}\mathbf{L}_t \tag{21}$$

Where,

$$\mathbf{L}_t = \begin{bmatrix}L_t^{xx} & L_t^{xv}\\ L_t^{xv} & L_t^{vv}\end{bmatrix} = \begin{bmatrix}\sqrt{\Sigma_t^{xx}} & 0\\ \frac{\Sigma_t^{xv}}{\sqrt{\Sigma_t^{xx}}} & \sqrt{\frac{\Sigma_t^{xx}\Sigma_t^{vv} - \Sigma_t^{vv}}{\Sigma_t^{xx}}}\end{bmatrix} \tag{22}$$

We borrow results from Dockhorn et al. (2021), the score function reads,

$$\begin{aligned}\nabla_{\mathbf{m}}\log p(\mathbf{m}_t|\mathbf{m}_1) &= -\nabla_{\mathbf{m}_t}\frac{1}{2}(\mathbf{m}_t - \boldsymbol{\mu}_t)\boldsymbol{\Sigma}_t^{-1}(\mathbf{m}_t - \boldsymbol{\mu}_t)\\ &= -\boldsymbol{\Sigma}_t^{-1}(\mathbf{m}_t - \boldsymbol{\mu}_t)\\ &\quad \text{Cholesky decomposition of } \boldsymbol{\Sigma}_t\\ &= -\mathbf{L}^{-T}\mathbf{L}^{-1}(\mathbf{m}_t - \boldsymbol{\mu}_t)\\ &= -\mathbf{L}^{-T}\epsilon\end{aligned}$$

The form of $\mathbf{L}$ reads,

$$\mathbf{L}_t = \begin{bmatrix} \sqrt{\Sigma_t^{xx}} & 0 \\ \frac{\Sigma_t^{xv}}{\sqrt{\Sigma_t^{xx}}} & \sqrt{\frac{\Sigma_t^{xx}\Sigma_t^{vv}-(\Sigma_t^{xv})^2}{\Sigma_t^{xx}}} \end{bmatrix}$$

and the transpose inverse of $\mathbf{L}$ reads,

$$\mathbf{L}_t^{-T} = \begin{bmatrix} \frac{1}{\sqrt{(\Sigma_t^{xx}+\epsilon_{xx})}} & \frac{-\Sigma_t^{xv}}{\sqrt{(\Sigma_t^{xx})}\sqrt{(\Sigma_t^{xx})(\Sigma_t^{vv}+)-(\Sigma_t^{xv})^2}} \\ 0 & \frac{\sqrt{\Sigma_t^{xx}}}{\sqrt{(\Sigma_t^{xx})(\Sigma_t^{vv})-(\Sigma_t^{xv})^2}} \end{bmatrix}$$

Hence, the score function reads,

$$\nabla_{\mathbf{v}} \log p(\mathbf{m}_t|\mathbf{m}_1) = -\underbrace{\frac{\sqrt{\Sigma_t^{xx}}}{\sqrt{(\Sigma_t^{xx}+\epsilon_{xx})(\Sigma_t^{vv}+\epsilon_{vv})-(\Sigma_t^{xv})^2}}}_{\ell_t}\boldsymbol{\epsilon}_1$$

### C.6 Representation of acceleration $\mathbf{a}_t$

As been shown in Proposition.3, the optimal control can be represented as,

$$\begin{aligned}
\mathbf{a}_t^* &= g_t^2 P_{11}\left(\frac{\mathbf{x}_1-\mathbf{x}_t}{1-t}-\mathbf{v}_t\right) \\
&= g_t^2 P_{11}\frac{\mathbf{x}_1}{1-t} - g_t^2 P_{11}\left(\frac{\mathbf{x}_t}{1-t}+\mathbf{v}_t\right) \\
&= g_t^2 P_{11}\frac{\mathbf{x}_1}{1-t} - g_t^2 P_{11}\left(\frac{\mu_t^x + L_t^{xx}\boldsymbol{\epsilon}_0}{1-t} + (\mu_t^v + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1)\right) \\
&= g_t^2 P_{11}\left[\left(\frac{\mathbf{x}_1-\mu_t^x}{1-t}-\mu_t^v\right) - \left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)\right] \\
&\text{solving eq.C.3 we can get : } \mathbf{x}_t = \frac{1}{3}\mathbf{x}_1 t^2(t^2-4t+6), \mathbf{v}_t = \frac{4t\mathbf{x}_1}{3}(t^2-3t+3) \\
&\text{Plug in} \mathbf{x}_t, \mathbf{v}_t \\
&= g_t^2 P_{11}\left[\left(\frac{\mathbf{x}_1 - \frac{1}{3}\mathbf{x}_1 t^2\left(6-4t+t^2\right)}{1-t} - \frac{4t\mathbf{x}_1}{3}(t^2-3t+3)\right) - \left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)\right] \\
&= g_t^2 P_{11}\left[\left(\frac{(-t^4+4t^3-6t^2+3)}{3(1-t)} - \frac{4t}{3}(t^2-3t+3)\right)\mathbf{x}_1 - \left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)\right] \\
&= g_t^2 P_{11}\left[\left(\frac{-(t-1)(t^3-3t^2+3t+3)}{3(1-t)} - \frac{4t}{3}(t^2-3t+3)\right)\mathbf{x}_1 - \left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)\right] \\
&= g_t^2 P_{11}\left[\left(\frac{(t^3-3t^2+3t+3)}{3} - \frac{1}{3}(4t^3-12t^2+12t)\right)\mathbf{x}_1 - \left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)\right] \\
&= g_t^2 P_{11}\left[(1-t)^3\mathbf{x}_1 - \left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)\right] \\
&= 4(1-t)^2\mathbf{x}_1 + g_t^2 P_{11}\left(\frac{L_t^{xx}}{1-t}\boldsymbol{\epsilon}_0 + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right)
\end{aligned}$$

### C.7 Representation of acceleration $\mathbf{a}_t$

We use

$$\lambda(t) = \frac{1}{1-t}$$

For all experiments. We admit that this might not be an optimal selection.

### C.8 NORMALIZER OF AGM-SDE AND AGM-ODE

Since the optimal control term can be represented as,

$$\mathbf{a}^*(\mathbf{m}_t, t) = 4\mathbf{x}_1(1-t)^2 - g_t^2 P_{11}\left[\left(\frac{L_t^{xx}}{1-t} + L_t^{xv}\right)\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right].$$

Then we introduce the normalizer as

$$\mathbf{z}_{SDE} = \sqrt{(4(1-t)^2 \cdot \sigma_{data})^2 + g_t^2 P_{11}\left[\left(\frac{L_t^{xx}}{1-t} + L_t^{xv}\right)^2 + (L_t^{vv})^2\right]}$$

$$\mathbf{z}_{ODE} = \sqrt{(4(1-t)^2 \cdot \sigma_{data})^2 + g_t^2 P_{11} + g_t^2 P_{11}\left(\frac{L_t^{xx}}{1-t} + L_t^{xv}\right)^2 + \left[\left(g_t^2 P_{11} L_t^{vv} - \frac{1}{2}g_t^2 \ell_t\right)^2\right]}$$

Where $\ell := \sqrt{\frac{\Sigma_t^{xx}}{\Sigma_t^{xx}\Sigma_t^{vv} - (\Sigma_t^{xv})^2}}$

### C.9 EXPONENTIAL INTEGRATOR DERIVATION

As suggested by Zhang & Chen (2022), one can write the discretized dynamics as,

$$\begin{gathered}\begin{bmatrix}\mathbf{x}_{t_{i+1}} \\ \mathbf{v}_{t_{i+1}}\end{bmatrix} = \Phi(t_{i+1}, t_i)\begin{bmatrix}\mathbf{x}_t \\ \mathbf{v}_t\end{bmatrix} + \sum_{j=0}^{r} C_{i,j}\begin{bmatrix}\mathbf{0} \\ \mathbf{s}_\theta(\mathbf{m}_{t_{i-j}}, t_{i-j})\end{bmatrix} \\ \text{Where } C_{i,j} = \int_t^{t+\delta_t} \Phi(t+\delta_t, \tau)\begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_\tau\end{bmatrix}\prod_{k\neq j}\left[\frac{\tau - t_{i-k}}{t_{i-j} - t_{i-k}}\right]\mathrm{d}\tau, \quad \Phi(t,s) = \begin{bmatrix}1 & t-s \\ 0 & 1\end{bmatrix}\end{gathered} \tag{23}$$

After plugging in the transition kernel $\Phi(t,s)$, one can easily obtain the results shown in (10).

### C.10 PROOF OF PROPOSITION.5

The estimated data point $\mathbf{x}_1$ can be represented as

$$\tilde{\mathbf{x}}_1^{SDE} = \frac{(1-t)(\mathbf{F}_t^\theta + \mathbf{v}_t)}{g_t^2 P_{11}} + \mathbf{x}_t, \text{ or } \quad \tilde{\mathbf{x}}_1^{ODE} = \frac{\mathbf{F}_t^\theta + g_t^2 P_{11}(\alpha_t \mathbf{x}_t + \beta_t \mathbf{v}_t)}{4(t-1)^2 + g_t^2 P_{11}(\alpha_t \mu_t^x + \beta_t \mu_t^v)} \tag{24}$$

for SDE and probablistic ODE dynamics respectively, and $\beta_t = L_t^{vv} + \frac{1}{2P_{11}}$,$\alpha_t = \frac{(\frac{L_t^{xx}}{1-t} + L_t^{xv}) - \beta_t L_t^{xv}}{L_t^{xx}}$.

*Proof.* It is easy to derive the representation of $\mathbf{x}_1$ of the SDE due to the fact that the network is essentially estimating:

$$\begin{gathered}\mathbf{F}_t^\theta \approx g_t^2 P_{11}\left(\frac{\mathbf{x}_1 - \mathbf{x}_t}{1-t} - \mathbf{v}_t\right) \\ \Leftrightarrow \mathbf{x}_1 \approx \frac{(1-t)(\mathbf{F}_t^\theta + \mathbf{v}_t)}{g_t^2 P_{11}} + \mathbf{x}_t\end{gathered}$$

It will become slightly more complicated for probabilistic ODE cases. We notice that

$$\begin{gathered}\mathbf{m}_t = \boldsymbol{\mu}_t + \mathbf{L}\epsilon \\ \Leftrightarrow \quad \mathbf{x}_t = \mu_t^x + L_t^{xx}\boldsymbol{\epsilon}_1, \quad \mathbf{v}_t = \mu_t^v + L_t^{xv}\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\end{gathered}$$

In probabilistic ODE case, the force term can be represented as,

$$\mathbf{F}(\mathbf{m}_t, t) = 4\mathbf{x}_1(1-t)^2 - g_t^2 P_{11}\left[\left(\frac{L_t^{xx}}{1-t} + L_t^{xv}\right)\boldsymbol{\epsilon}_0 + L_t^{vv}\boldsymbol{\epsilon}_1\right] - \frac{1}{2}g_t^2 \ell \boldsymbol{\epsilon}_1$$

In order to use linear combination of $\mathbf{x}_t$ and $\mathbf{v}_t$ to represent $\mathbf{F}$ one needs to match the stochastic term in $\mathbf{F}_t$ by using

$$\alpha_t L_t^{xx} + \beta_t L_t^{xv} = \underbrace{\frac{L_t^{xx}}{1-t} + L_t^{xv}}_{A_t},$$

$$\beta_t L_t^{vv} = \underbrace{L_t^{vv} + \frac{1}{2P_{11}}}_{B_t}.$$

The solution can be obtained by:

$$\beta_t = \frac{B_t}{L_t^{vv}}$$
$$\alpha_t = \frac{A_t - \beta_t L_t^{xv}}{L_t^{xx}}$$

□

By subsitute it back to $\mathbf{F}_t$, one can get:

$$\begin{aligned}\mathbf{F}(\mathbf{m}_t, t) &= 4\mathbf{x}_1(1-t)^2 - g_t^2 P_{11}\left[\alpha_t(\mathbf{x}_t - \mu_t^x) + \beta_t(\mathbf{v}_t - \mu_t^v)\right] \\ &= \left[4(1-t)^2 + g_t^2 P_{11}(\alpha_t \mu_t^x + \beta_t \mu_t^v)\right]\mathbf{x}_1 - g_t^2 P_{11}\left[\alpha_t \mathbf{x}_t + \beta_t \mathbf{v}_t\right] \\ \Leftrightarrow \mathbf{x}_1 &= \frac{\mathbf{F}_t^\theta + g_t^2 P_{11}(\alpha_t \mathbf{x}_t + \beta_t \mathbf{v}_t)}{4(t-1)^2 + g_t^2 P_{11}(\alpha_t \mu_t^x + \beta_t \mu_t^v)}\end{aligned}$$

# D Experimental Details

**Training:** We stick with hyperparameters introduced in the section.4. We use AdamW(Loshchilov & Hutter, 2017) as our optimizer and Exponential Moving Averaging with the exponential decay rate of 0.9999. We use 8 $\times$ Nvidia A100 GPU for all experiments. For further, training setup, please refer to Table.6.

Table 6: Additional experimental details

| dataset | Training Iter | Learning rate | Batch Size | network architecture |
|---|---|---|---|---|
| toy | 0.05M | 1e-3 | 1024 | ResNet(Dockhorn et al., 2021) |
| CIFAR-10 | 0.5M | 1e-3 | 512 | NCSN++(Karras et al., 2022) |
| AFHQv2 | 0.5M | 1e-3 | 512 | NCSN++(Karras et al., 2022) |
| ImageNet-64 | 0.8M | 2e-4 | 512 | ADM(Dhariwal & Nichol, 2021) |

**Sampling:** For Exponential Integrator, we choose the multistep order $w = 2$ consistently for all experiments. Different from previous work (Dockhorn et al., 2021; Karras et al., 2022; Zhang et al., 2023), we use quadratic timesteps scheme with $\kappa = 2$:

$$t_i = \left(\frac{N-i}{N} t_0^{\frac{1}{\kappa}} + \frac{i}{N} t_N^{\frac{1}{\kappa}}\right)^\kappa$$

Which is opposite to the classical DM. Namely, the time discretization will get larger when the dynamics is propagated close to data. For numerical stability, we use $t_0 = 1E{-}5$ for all experiments. For $NFE = 5$, we use $t_N = 0.5$ and $NFE = 10$, $T_N = 0.7$. For the rest of the sampling, we use $t_N = 0.999$.

Due to the fact that EDM(Karras et al., 2022) is using second-order ODE solver, in practice, we allow it to have an extra one NFE as reported for all the tables.

# E Additional Figures

We demonstrate the samples for different datasets with varying NFE.

## E.1 CIFAR-10

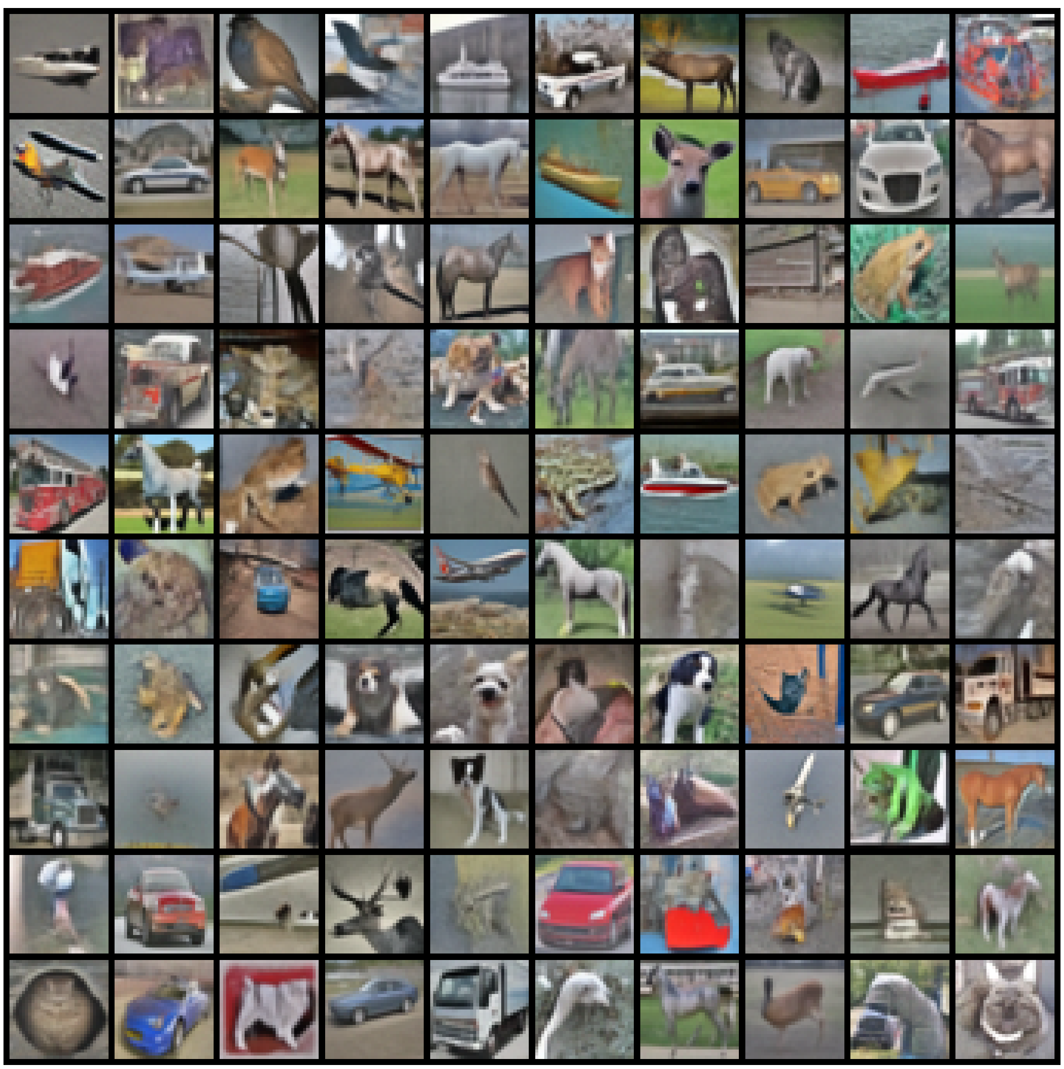

Figure 7: AGM-ODE Uncurated CIFAR-10 samples with NFE=5

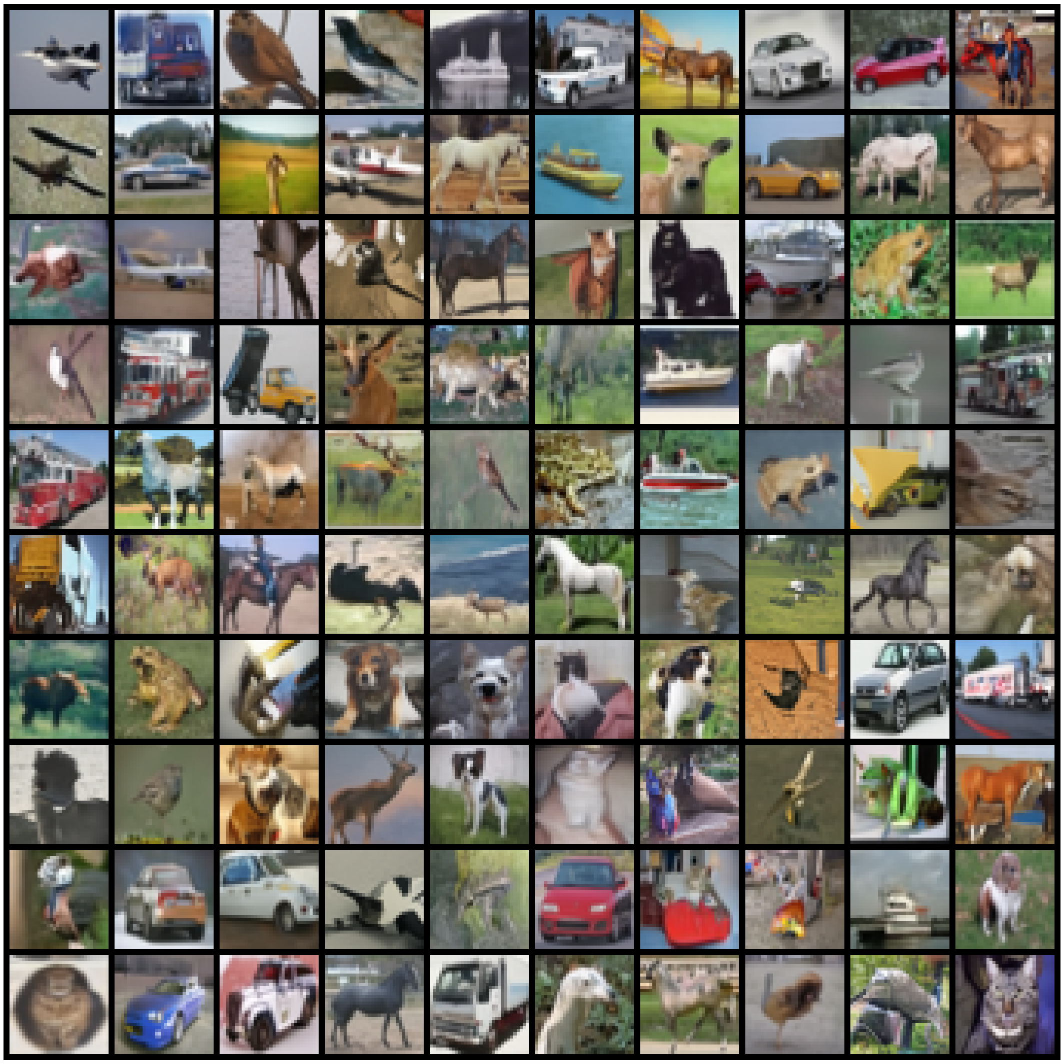

Figure 8: AGM-ODE Uncurated CIFAR-10 samples with NFE=10

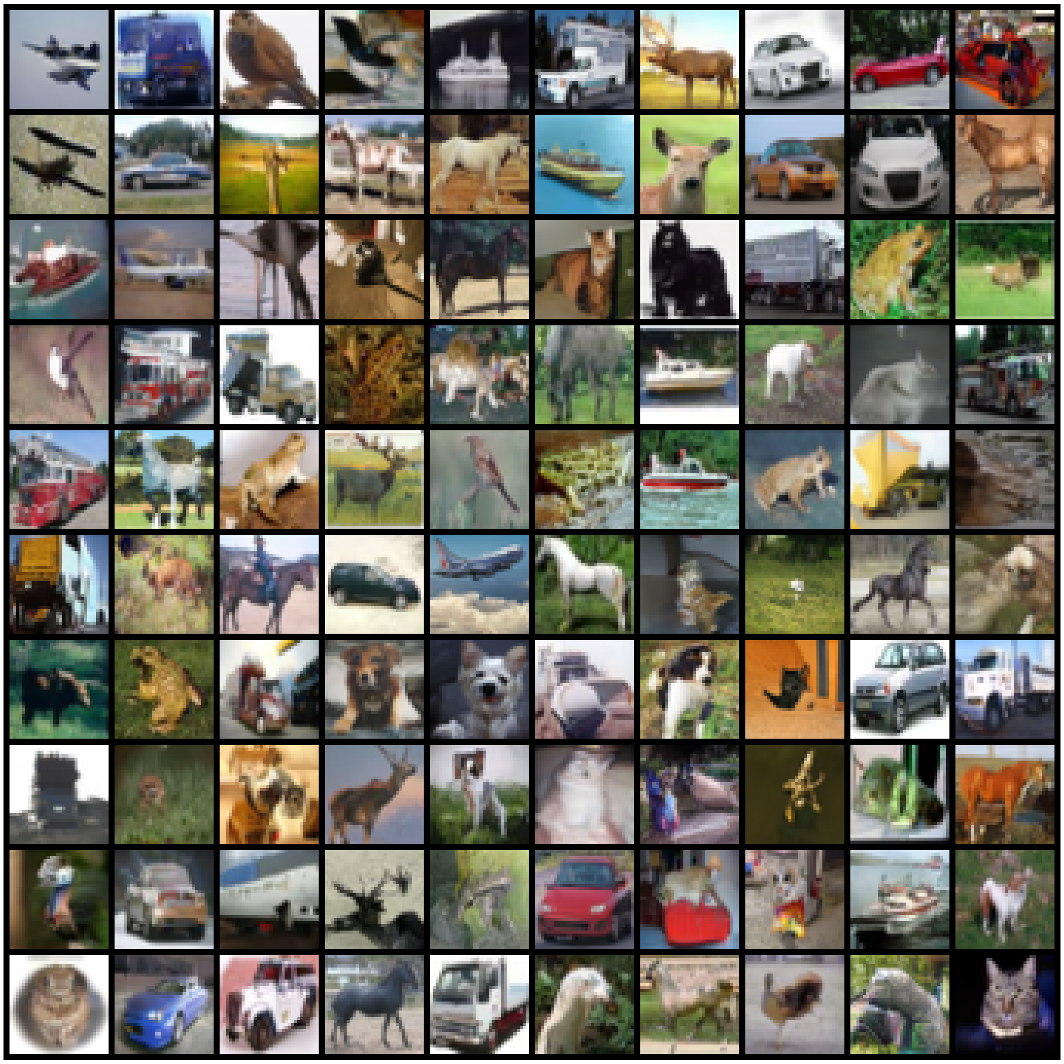

Figure 9: AGM-ODE Uncurated CIFAR-10 samples with NFE=20

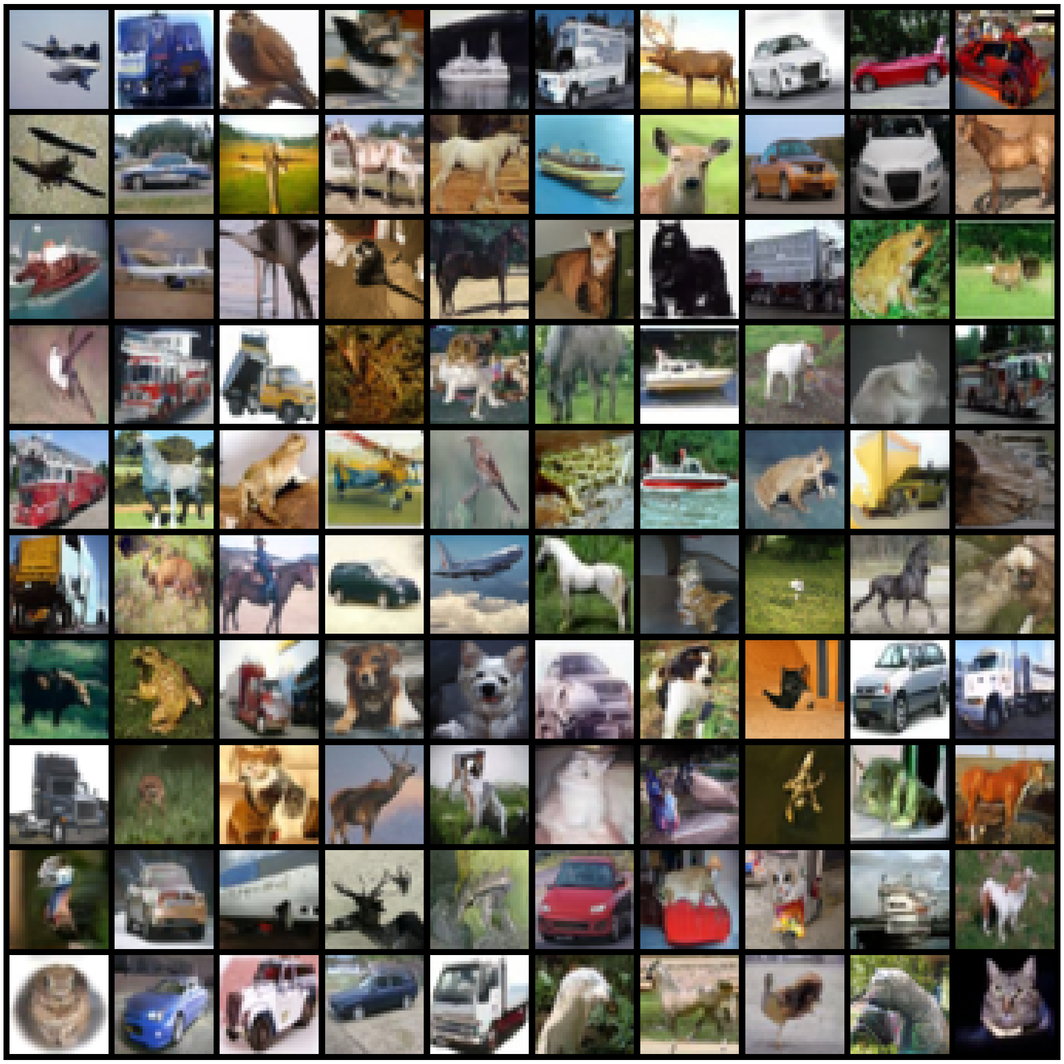

Figure 10: AGM-ODE Uncurated CIFAR-10 samples with NFE=50

### E.2 AFHQv2

Figure 11: AGM-ODE Uncurated CIFAR-10 samples with NFE=5

Figure 12: AGM-ODE Uncurated CIFAR-10 samples with NFE=10

Figure 13: AGM-ODE Uncurated CIFAR-10 samples with NFE=20

Figure 14: AGM-ODE Uncurated CIFAR-10 samples with NFE=50

### E.3 IMAGENET-64

Figure 15: AGM-ODE Uncurated CIFAR-10 samples with NFE=10

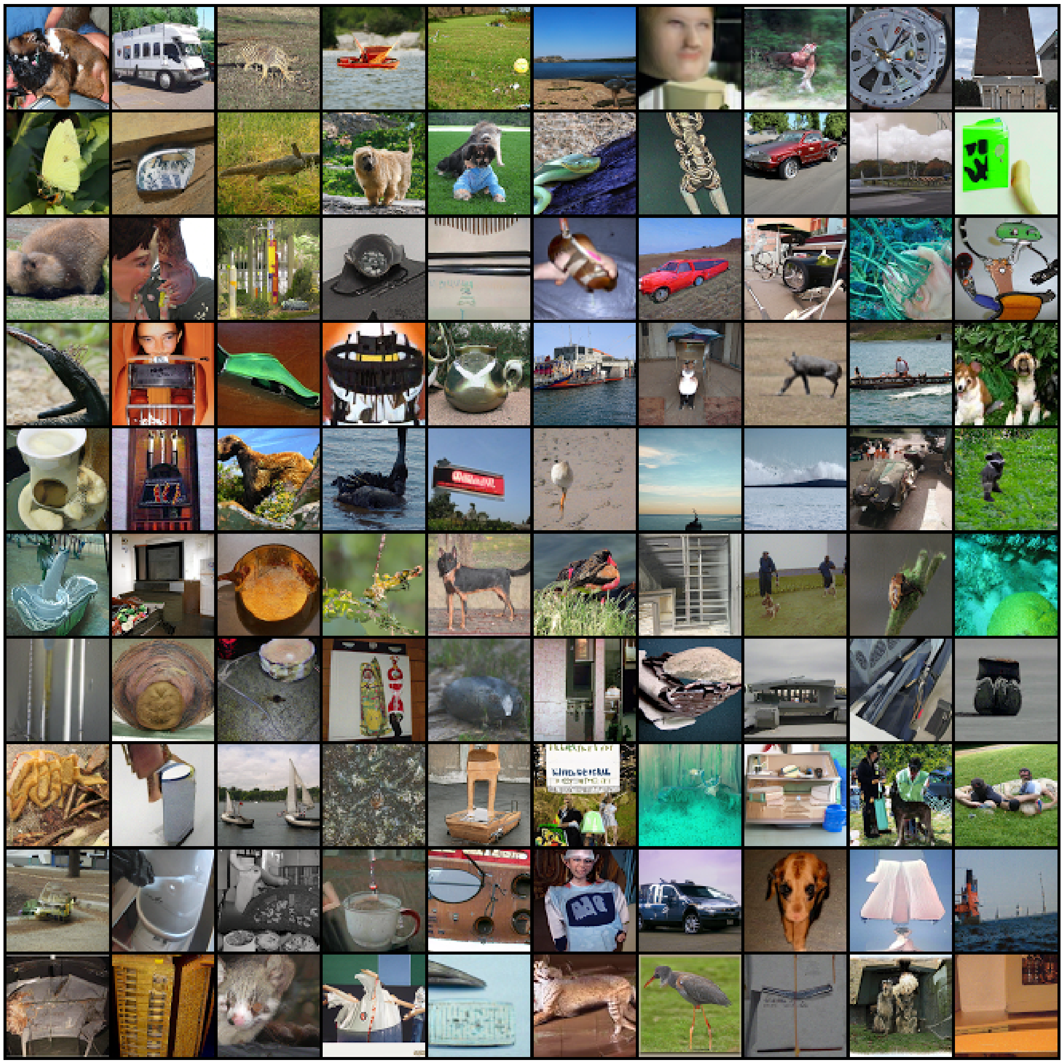

Figure 16: AGM-ODE Uncurated CIFAR-10 samples with NFE=20

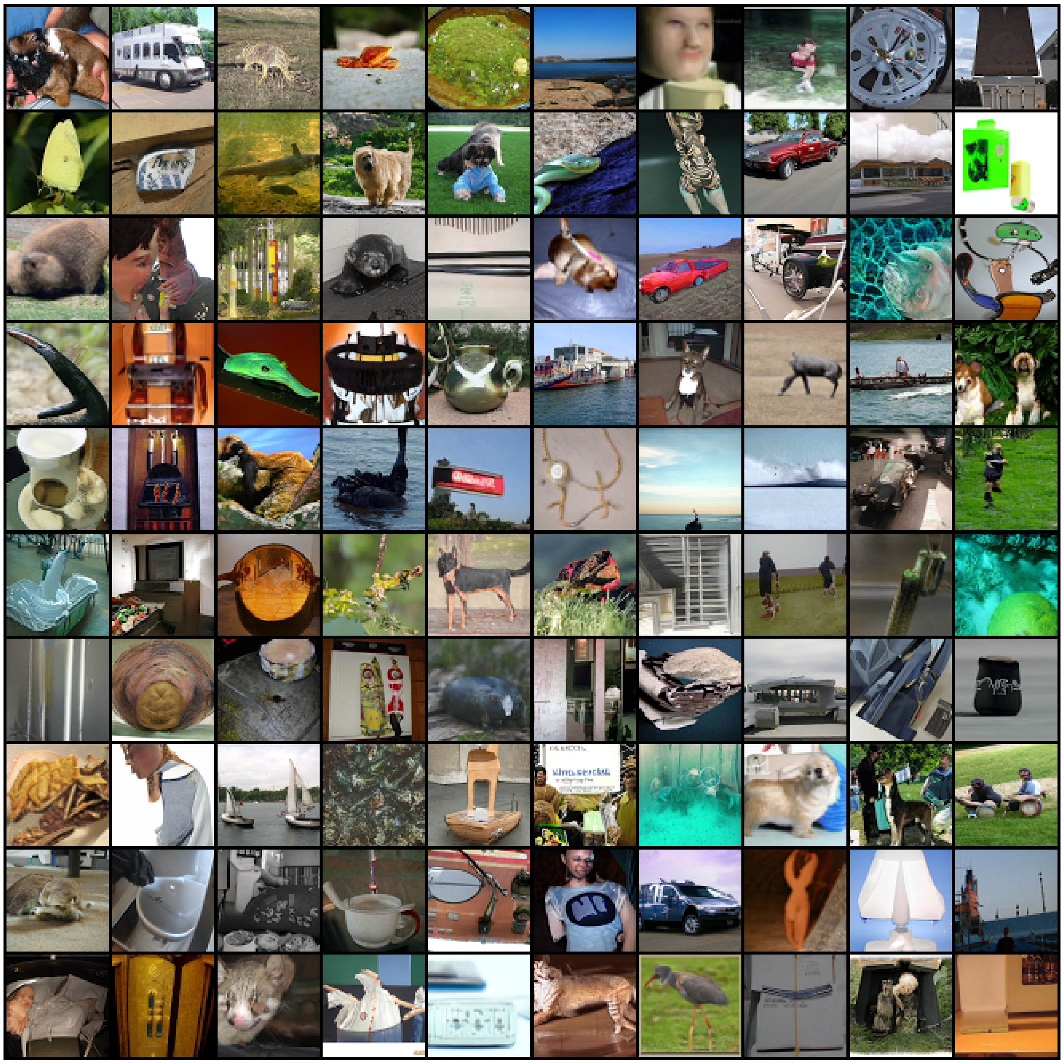

Figure 17: AGM-ODE Uncurated CIFAR-10 samples with NFE=50