

# Dynamics and Memorization Behaviour of Score-Based Diffusion Models

Ricardo Baptista

Statistical Sciences  
University of Toronto



ICERM Workshop: Bayesian Inverse Problems and UQ  
March 4, 2026

# Memorization in Score-Based Diffusion Models: Why Inverse Problems Expertise is Needed

Ricardo Baptista

Statistical Sciences  
University of Toronto



ICERM Workshop: Bayesian Inverse Problems and UQ  
March 4, 2026

### Memorization and Regularization in Generative Diffusion Models

Ricardo Baptista<sup>1</sup>, Agnimitras Dasgupta<sup>2</sup>, Nikola Kovachki<sup>3</sup>,  
Assad Oberai<sup>2</sup>, Andrew Stuart<sup>4</sup>

<sup>1</sup>Statistical Sciences  
University of Toronto

<sup>2</sup>Aerospace and Mechanical Engineering  
University of Southern California

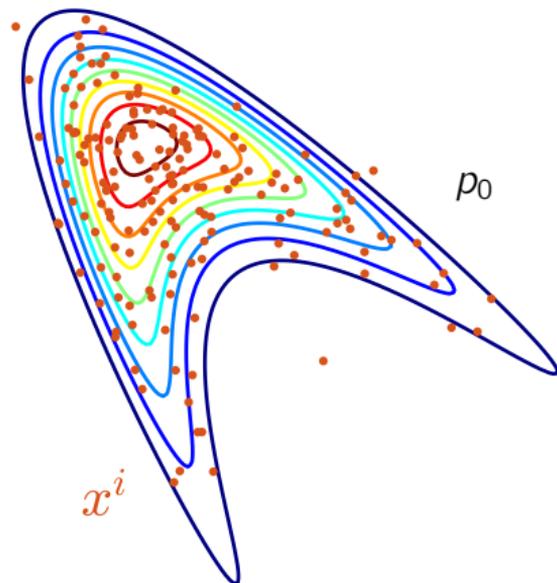
<sup>3</sup>NVIDIA Corporation

<sup>4</sup>Computing+Mathematical Sciences  
California Institute of Technology

[arXiv:2501.15785](https://arxiv.org/abs/2501.15785)

## Task of generative modeling

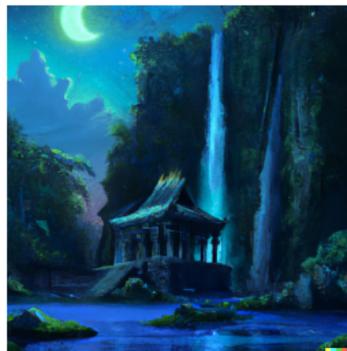
**Setting:** Collect i.i.d. samples  $\{\mathbf{x}_0^i\}_{i=1}^N$  (e.g., images, text) from probability distribution  $p_0$



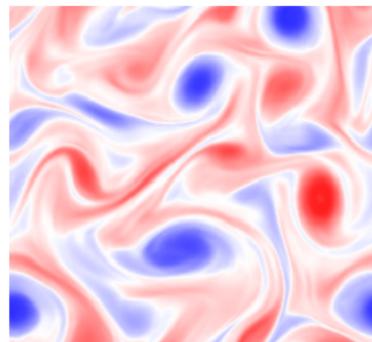
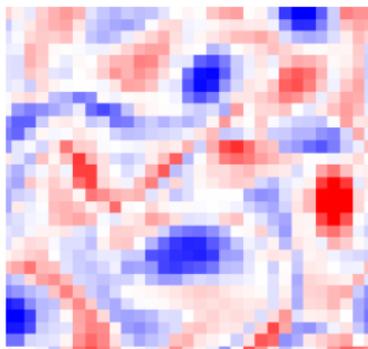
**Goal:** Generate *new* samples from  $p_0$  that are *not present in the training dataset*

# Diffusion Models Generate High-Quality Images

**Machine learning:** Prompt-to-image models (Ramesh et al., 2022)

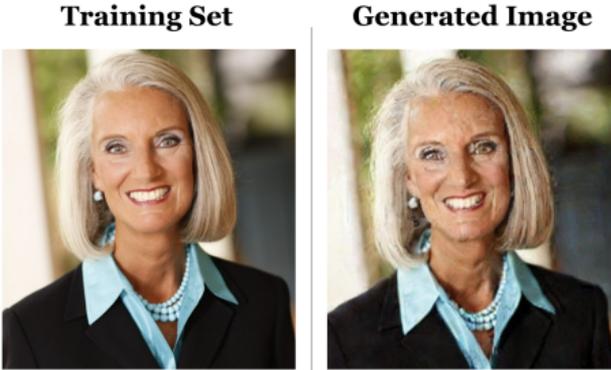


**Scientific computing:** Super-resolution inverse problems (Wan et al., 2023)



# But Diffusion Models Can Lack Diversity

Memorizing training data (Carlini et al., 2023)



Memorizing subsets of images (Somepalli et al., 2023)

Generated Image

Training Set

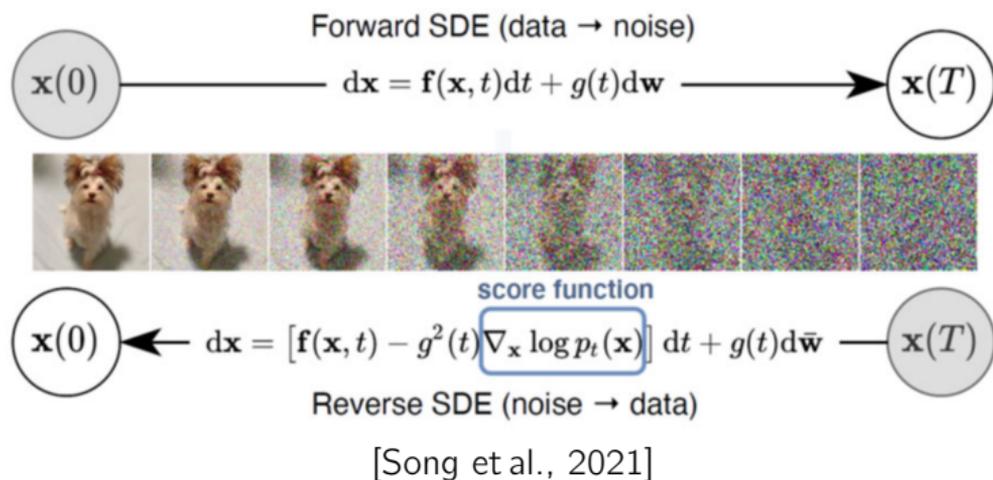


- 1 Diffusion Model Methodology
- 2 Main Theorem on Memorization
- 3 Analysis Underlying Theorem
- 4 Numerics Illustrating Theorem
- 5 Conclusions

- 1 Diffusion Model Methodology**
- 2 Main Theorem on Memorization
- 3 Analysis Underlying Theorem
- 4 Numerics Illustrating Theorem
- 5 Conclusions

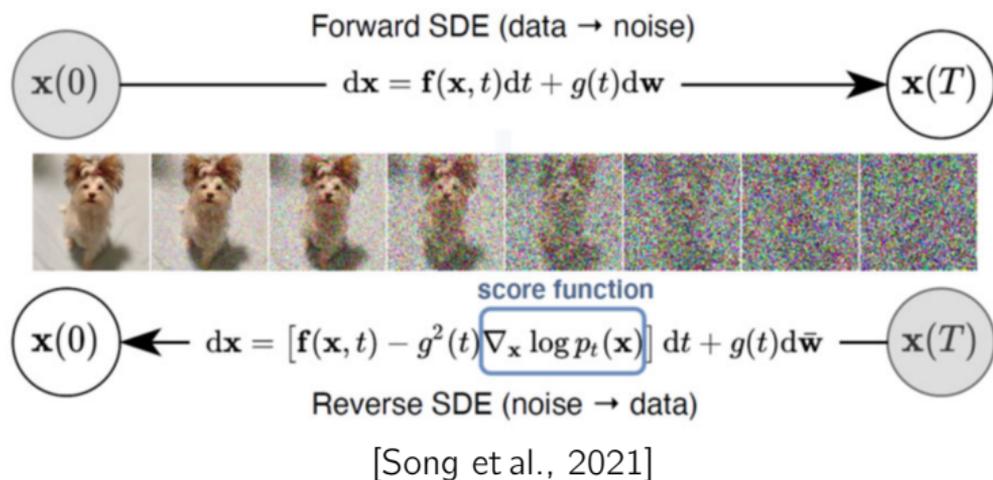
# Generative Modeling by Learning Score Functions

- ▶ **Forward process** maps data to noise at  $t = T$
- ▶ **Reverse process** maps noise to data at  $t = 0$



# Generative Modeling by Learning Score Functions

- ▶ **Forward process** maps data to noise at  $t = T$
- ▶ **Reverse process** maps noise to data at  $t = 0$



**Key Idea:** Need to learn the *score*  $\nabla_{\mathbf{x}} \log p(\mathbf{x}, t)$  of the forward process for each  $t$

**Goal:** Learn the score of  $p(\mathbf{x}, t)$

**Goal:** Learn the score of  $p(\mathbf{x}, t)$

**Example:** Variance exploding process:  $d\mathbf{x} = \sqrt{g(t)}d\mathbf{w}$

- ▶ Conditional distribution:  $p(\mathbf{x}, t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}; \mathbf{x}_0, \sigma^2(t)I_d)$  for  $\sigma^2(t) = \int_0^t g(s)ds$
- ▶ Marginal distribution:  $p(\mathbf{x}, t) = \int p(\mathbf{x}, t|\mathbf{x}_0)dp_0(\mathbf{x}_0)$

**Goal:** Learn the score of  $p(\mathbf{x}, t)$

**Example:** Variance exploding process:  $d\mathbf{x} = \sqrt{g(t)}d\mathbf{w}$

- ▶ Conditional distribution:  $p(\mathbf{x}, t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}; \mathbf{x}_0, \sigma^2(t)I_d)$  for  $\sigma^2(t) = \int_0^t g(s)ds$
- ▶ Marginal distribution:  $p(\mathbf{x}, t) = \int p(\mathbf{x}, t|\mathbf{x}_0)dp_0(\mathbf{x}_0)$

**Approach:** Denoising score-matching (Vincent, 2011)

$$\begin{aligned} & \arg \min_s \int_0^T \mathbb{E}_{\mathbf{x}} |s(\mathbf{x}, t) - \nabla_{\mathbf{x}} \log p(\mathbf{x}, t)|^2 dt \\ &= \arg \min_s \int_0^T \underbrace{\mathbb{E}_{(\mathbf{x}, \mathbf{x}_0)} |s(\mathbf{x}, t) - \nabla_{\mathbf{x}} \log p(\mathbf{x}, t|\mathbf{x}_0)|^2}_{\text{Does not explicitly depend on the data density}} dt \end{aligned}$$

- 1 Given data  $\{\mathbf{x}_0^i\}_{i=1}^N \sim p_0$ , learn score

$$s^* \in \arg \min_s \int_0^T \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\mathbf{x}} |s(\mathbf{x}, t) - \nabla_{\mathbf{x}} \log p(\mathbf{x}, t | \mathbf{x}_0^i)|^2 dt$$

- 2 Simulate the reverse process to generate new data

$$\text{SDE: } d\mathbf{x} = -g(t)s^*(\mathbf{x}, t)dt + \sqrt{g(t)}d\mathbf{w}, \quad \mathbf{x}(T) \sim \mathcal{N}(0, \sigma^2(T)I_d)$$

$$\text{ODE: } \frac{d\mathbf{x}}{dt} = -\frac{g(t)}{2}s^*(\mathbf{x}, t)$$

- 1 Diffusion Model Methodology
- 2 Main Theorem on Memorization**
- 3 Analysis Underlying Theorem
- 4 Numerics Illustrating Theorem
- 5 Conclusions

# How Does Memorization Arise?

Consider variance exploding process  $\mathbf{x}_t | \mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_0, \sigma^2(t) I_d)$

Optimal empirical score (Gu et al., 2023; Scarvelis, Borde, and Solomon, 2023)

For empirical  $p_0 = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}_0^i}$ , the score is

$$s^*(\mathbf{x}, t) = -\frac{1}{\sigma^2(t)} \sum_{i=1}^N (\mathbf{x} - \mathbf{x}_0^i) w_i(\mathbf{x}, t),$$

where  $w_i(\mathbf{x}, t) \in [0, 1]$  are normalized Gaussian weights

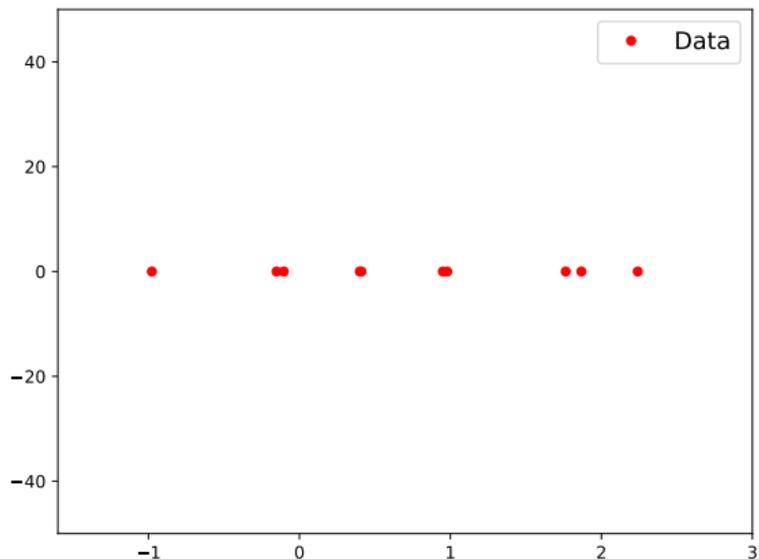
$$w_i(\mathbf{x}, t) \propto \exp\left(-\frac{|\mathbf{x} - \mathbf{x}_0^i|^2}{2\sigma^2(t)}\right)$$

## Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$

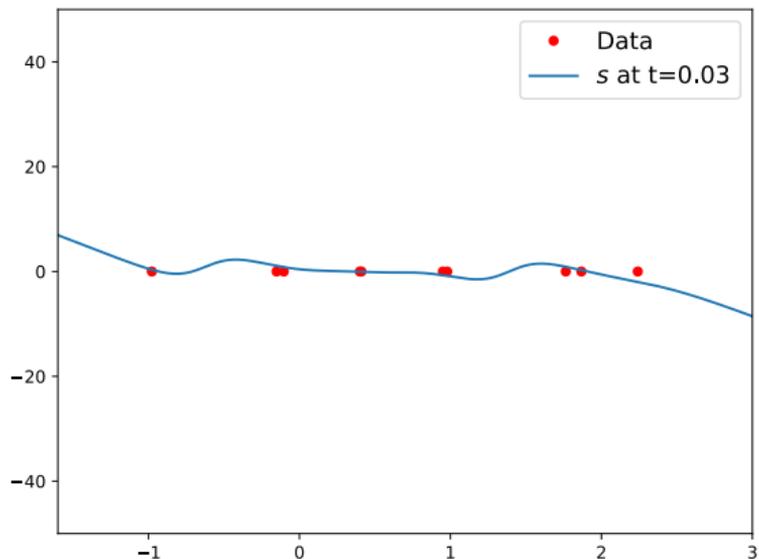
# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



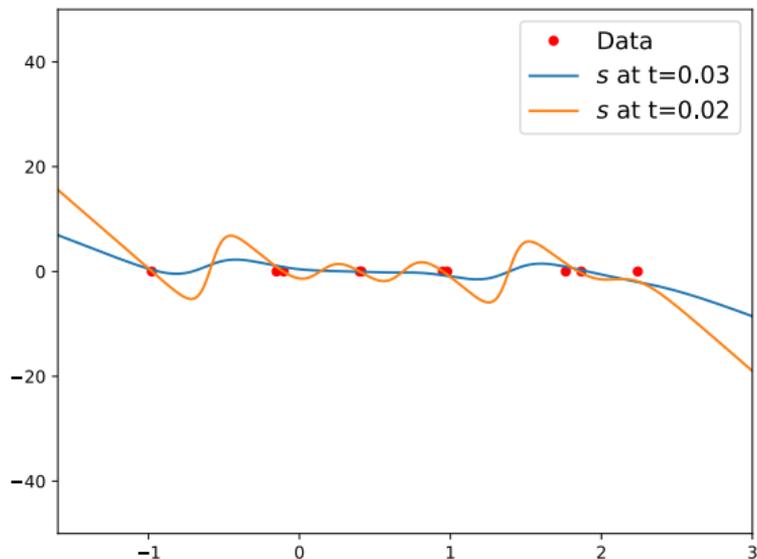
# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



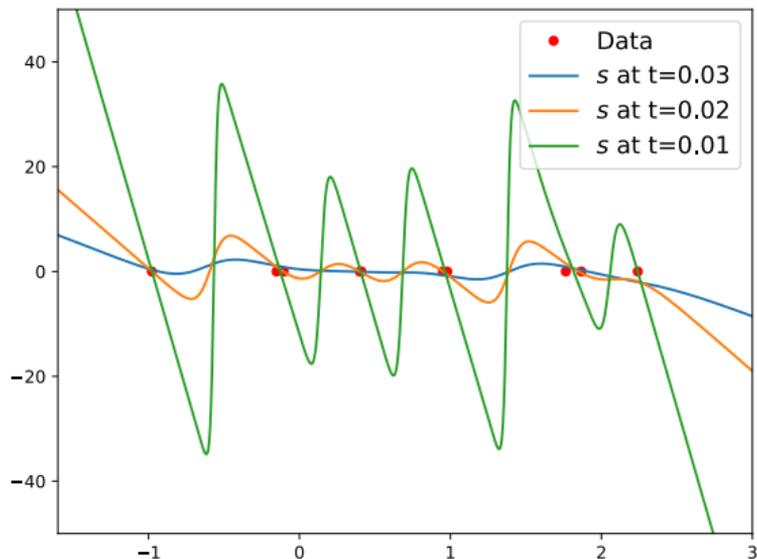
# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



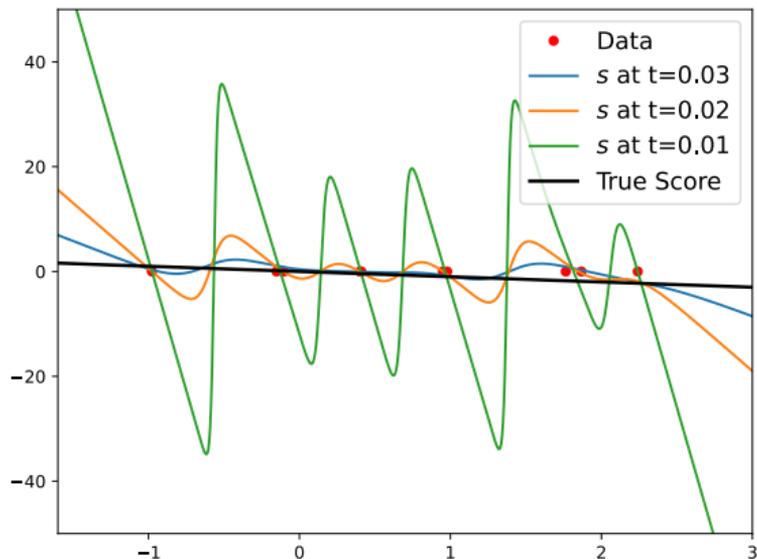
# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



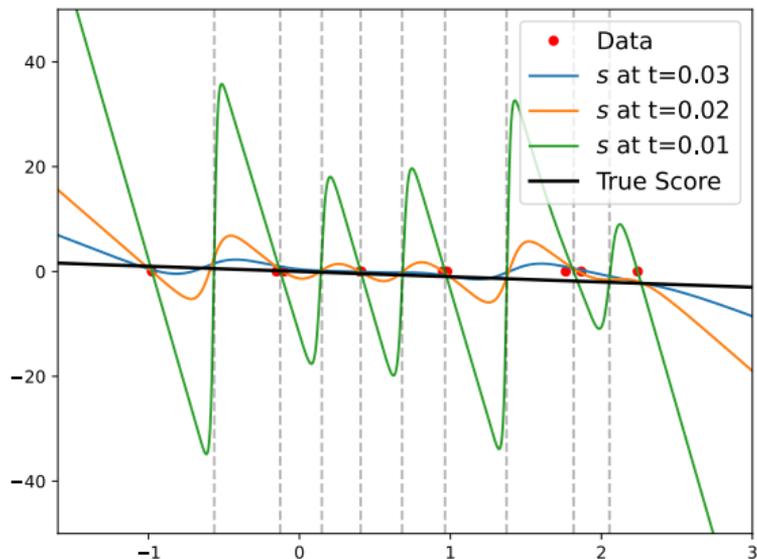
# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



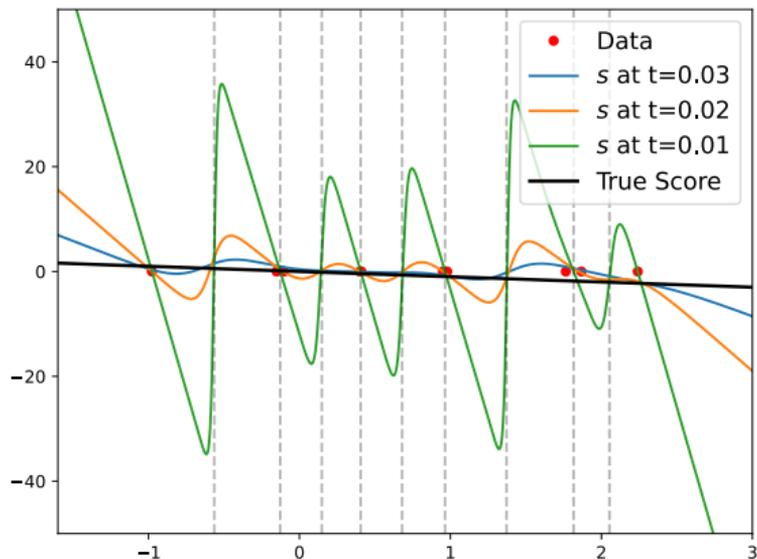
# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



# Limiting Behaviour of the Empirical Score

For  $\mathbf{x}$  near  $\mathbf{x}_0^i$  and  $t \rightarrow 0$ , weights collapse and the score is  $s^*(\mathbf{x}, t) \rightarrow -\frac{\mathbf{x} - \mathbf{x}_0^i}{\sigma^2(t)}$



Score's behaviour depends on the Voronoi partitioning of the data points

$$V(\mathbf{x}_0^i) \equiv \{\mathbf{x} \in \mathbb{R}^d \text{ s.t. } |\mathbf{x} - \mathbf{x}_0^i| < |\mathbf{x} - \mathbf{x}_0^\ell|, \ell \neq i\}$$

**Today:** Study the reverse process with the empirical score and  $g(t) \propto t$

**Today:** Study the reverse process with the empirical score and  $g(t) \propto t$

**Reverse-time ODE:**

- ▶ Integrate  $\mathbf{x}$  backward-in-time starting from  $\mathbf{x}(T) \sim \mathcal{N}(0, \sigma^2(T)I_d)$

$$\frac{d\mathbf{x}}{dt} = -\frac{g(t)}{2}s^*(\mathbf{x}, t) = \frac{1}{t} \sum_{i=1}^N (\mathbf{x} - \mathbf{x}_0^i) w_i(\mathbf{x}, t)$$

**Limit Points:** For any initial condition  $\mathbf{x}(T)$  :

$$\lim_{t \rightarrow 0^+} \mathbf{x}(t) = \mathbf{x}^*$$

**Limit Points:** For any initial condition  $\mathbf{x}(T)$  :

$$\lim_{t \rightarrow 0^+} \mathbf{x}(t) = \mathbf{x}^*$$

Main Theorem [BDKOS 2025]

Limit points  $\mathbf{x}^*$  are one the data points  $\mathbf{x}_0^i$  or on the boundaries of the Voronoi cells

**Limit Points:** For any initial condition  $\mathbf{x}(T)$  :

$$\lim_{t \rightarrow 0^+} \mathbf{x}(t) = \mathbf{x}^*$$

Main Theorem [BDKOS 2025]

Limit points  $\mathbf{x}^*$  are one the data points  $\mathbf{x}_0^i$  or on the boundaries of the Voronoi cells

Key Lemma: Dynamics Get Trapped

Consider subsets of the Voronoi cells,  $\delta > 0$  separated from the boundary:

$$V^\delta(\mathbf{x}_0^i) \subset V(\mathbf{x}_0^i)$$

If  $\mathbf{x}(t^*) \in V^\delta(\mathbf{x}_0^i)$  for some time  $t^*(N, \delta)$ , then

$$\mathbf{x}(t) \in V^\delta(\mathbf{x}_0^i), \quad \forall 0 < t \leq t^*$$

- 1 Diffusion Model Methodology
- 2 Main Theorem on Memorization
- 3 Analysis Underlying Theorem**
- 4 Numerics Illustrating Theorem
- 5 Conclusions

## Reverse-time ODE:

- ▶ Integrate  $\mathbf{x}$  backward-in-time starting from  $\mathbf{x}(1) \sim \mathcal{N}(0, I)$

$$\frac{d\mathbf{x}}{dt} = -\frac{g(t)}{2} s^*(\mathbf{x}, t) = \frac{1}{t}(\mathbf{x} - x_N(\mathbf{x}, t)), \quad x_N(\mathbf{x}, t) := \sum_{i=1}^N \mathbf{x}_0^i w_i(\mathbf{x}, t)$$

## Reverse-time ODE:

- ▶ Integrate  $\mathbf{x}$  backward-in-time starting from  $\mathbf{x}(1) \sim \mathcal{N}(0, I)$

$$\frac{d\mathbf{x}}{dt} = -\frac{g(t)}{2} s^*(\mathbf{x}, t) = \frac{1}{t}(\mathbf{x} - x_N(\mathbf{x}, t)), \quad x_N(\mathbf{x}, t) := \sum_{i=1}^N \mathbf{x}_0^i w_i(\mathbf{x}, t)$$

## Change of variables: $s = -\log(t)$

- ▶ integrate  $\mathbf{y}(s) = \mathbf{x}(e^{-s})$  forward-in-time starting from  $\mathbf{y}(0) \sim \mathcal{N}(0, I)$

$$\frac{d\mathbf{y}}{ds} = -(\mathbf{y} - y_N(\mathbf{y}, s)), \quad y_N(\mathbf{y}, s) := \sum_{i=1}^N \mathbf{x}_0^i w_i(\mathbf{y}, e^{-s})$$

## Collapse Onto the Data Points

**Main idea:** approximate dynamics within Voronoi cell

$$\begin{aligned}\frac{d\mathbf{y}}{ds} &= -(\mathbf{y} - y_N(\mathbf{y}, s)) \\ &= -(\mathbf{y} - \mathbf{x}_0^i) - (\mathbf{x}_0^i - y_N(\mathbf{y}, s))\end{aligned}$$

## Collapse Onto the Data Points

**Main idea:** approximate dynamics within Voronoi cell

$$\begin{aligned}\frac{d\mathbf{y}}{ds} &= -(\mathbf{y} - y_N(\mathbf{y}, s)) \\ &= -(\mathbf{y} - \mathbf{x}_0^i) - (\mathbf{x}_0^i - y_N(\mathbf{y}, s))\end{aligned}$$

**Reasons:** From cell invariance, the weights  $w_i(\mathbf{y}, s) \propto \exp(-e^{2s}|\mathbf{y} - \mathbf{x}_0^i|^2)$  for  $\mathbf{y} \in V^\delta(\mathbf{x}_0^i)$  are

$$w_i(\mathbf{y}, s) \approx 1, \quad w_\ell(\mathbf{y}, s) \approx 0$$

## Collapse Onto the Data Points

**Main idea:** approximate dynamics within Voronoi cell

$$\begin{aligned}\frac{d\mathbf{y}}{ds} &= -(\mathbf{y} - y_N(\mathbf{y}, s)) \\ &= -(\mathbf{y} - \mathbf{x}_0^i) - (\mathbf{x}_0^i - y_N(\mathbf{y}, s))\end{aligned}$$

**Reasons:** From cell invariance, the weights  $w_i(\mathbf{y}, s) \propto \exp(-e^{2s}|\mathbf{y} - \mathbf{x}_0^i|^2)$  for  $\mathbf{y} \in V^\delta(\mathbf{x}_0^i)$  are

$$w_i(\mathbf{y}, s) \approx 1, \quad w_\ell(\mathbf{y}, s) \approx 0$$

The nonlinear part of dynamics is small:

$$\mathbf{x}_0^i - y_N(\mathbf{y}, s) = (1 - w_i(\mathbf{y}, s))\mathbf{x}_0^i + \sum_{\ell \neq i} w_\ell(\mathbf{y}, s)\mathbf{x}_0^\ell \approx 0$$

## Collapse Onto the Data Points

**Main idea:** approximate dynamics within Voronoi cell

$$\begin{aligned}\frac{d\mathbf{y}}{ds} &= -(\mathbf{y} - y_N(\mathbf{y}, s)) \\ &= -(\mathbf{y} - \mathbf{x}_0^i) - (\mathbf{x}_0^i - y_N(\mathbf{y}, s))\end{aligned}$$

**Reasons:** From cell invariance, the weights  $w_i(\mathbf{y}, s) \propto \exp(-e^{2s}|\mathbf{y} - \mathbf{x}_0^i|^2)$  for  $\mathbf{y} \in V^\delta(\mathbf{x}_0^i)$  are

$$w_i(\mathbf{y}, s) \approx 1, \quad w_\ell(\mathbf{y}, s) \approx 0$$

The nonlinear part of dynamics is small:

$$\mathbf{x}_0^i - y_N(\mathbf{y}, s) = (1 - w_i(\mathbf{y}, s))\mathbf{x}_0^i + \sum_{\ell \neq i} w_\ell(\mathbf{y}, s)\mathbf{x}_0^\ell \approx 0$$

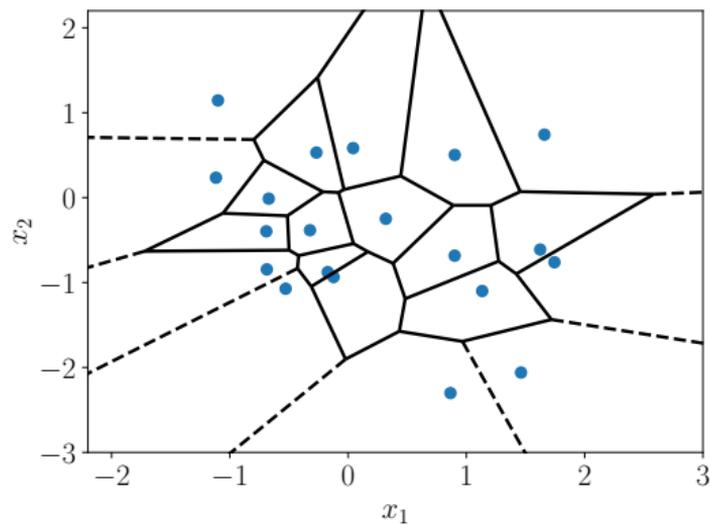
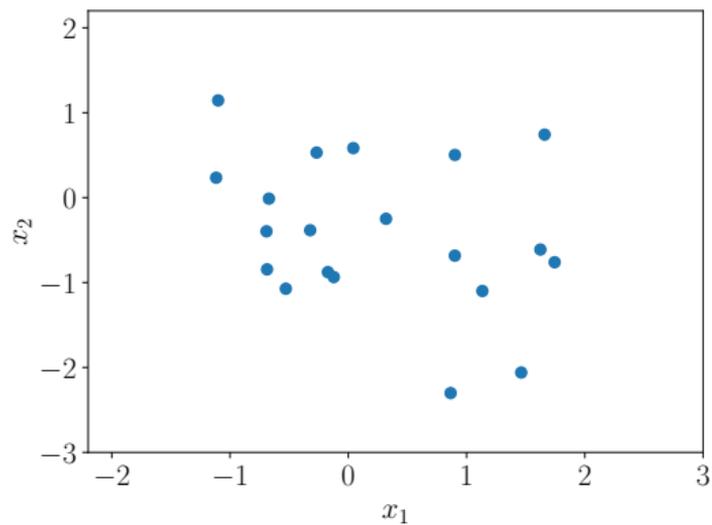
**Takeaway:** Within cells, the dynamics are linear

$$\frac{d(\mathbf{y} - \mathbf{x}_0^i)}{ds} = \frac{d\mathbf{y}}{ds} \approx -(\mathbf{y} - \mathbf{x}_0^i)$$

- 1 Diffusion Model Methodology
- 2 Main Theorem on Memorization
- 3 Analysis Underlying Theorem
- 4 Numerics Illustrating Theorem**
- 5 Conclusions

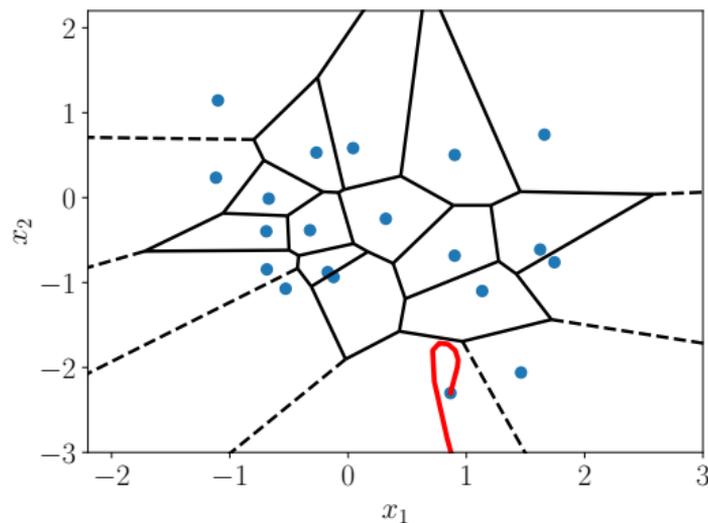
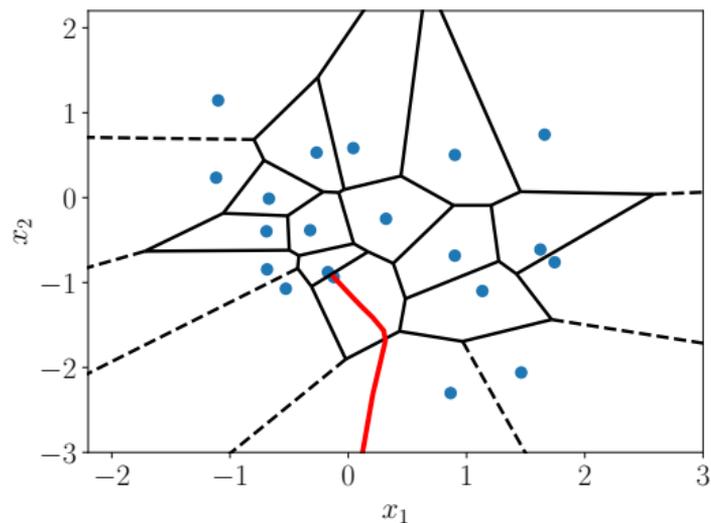
# Data with Voronoi Tesslations

Data:  $N = 20$  i.i.d. samples from  $\mathcal{N}(0, I_2)$



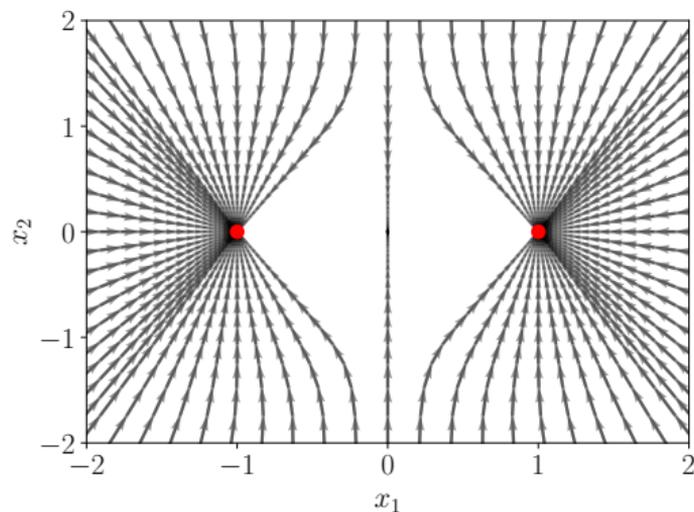
# ODE Dynamics with Voronoi Tessellations

Integrate ODE with empirical score using  $N = 20$  i.i.d. samples from  $\mathcal{N}(0, I_2)$



# Trajectories Can Remain On Hyper-Planes

- ▶ Most trajectories collapse onto the  $N = 2$  data points (red)
- ▶ Some trajectories remain on Voronoi boundaries



**Tikhonov-regularized score matching problem:**

$$s_{\text{reg}}^* \in \arg \min_s \int_0^T \mathbb{E} |s(\mathbf{x}, t) - \nabla \log p(\mathbf{x}, t)|^2 + \gamma^2(t) \mathbb{E} |s(\mathbf{x}, t)|^2 dt.$$

**Tikhonov-regularized score matching problem:**

$$s_{\text{reg}}^* \in \arg \min_s \int_0^T \mathbb{E} |s(\mathbf{x}, t) - \nabla \log p(\mathbf{x}, t)|^2 + \gamma^2(t) \mathbb{E} |s(\mathbf{x}, t)|^2 dt.$$

## Optimal score function

For empirical  $p_0$  with  $\gamma^2(t)\sigma^2(t) = c$ , the score is

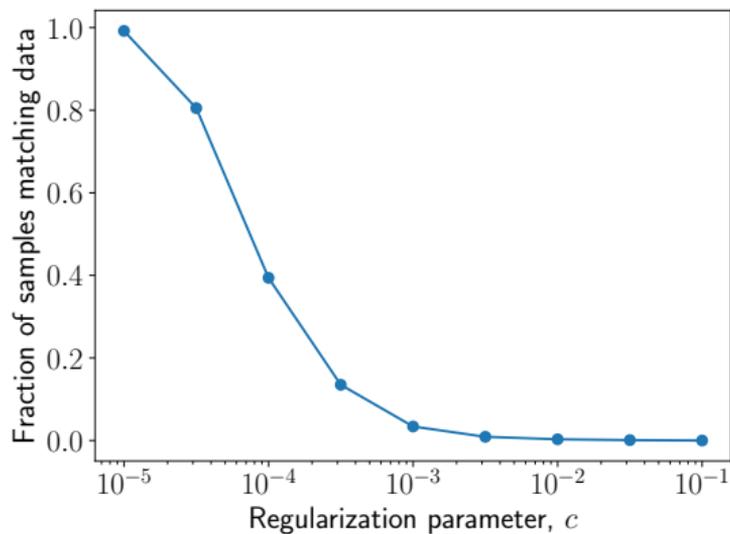
$$s_{\text{reg}}^*(\mathbf{x}, t) = -\frac{1}{\sigma^2(t) + c} \sum_{i=1}^N (\mathbf{x} - \mathbf{x}_0^i) w_i(\mathbf{x}, t)$$

The score is bounded as  $t \rightarrow 0$ :

$$s_{\text{reg}}^*(\mathbf{x}, t) \rightarrow \frac{-(\mathbf{x} - \mathbf{x}_0^i)}{c}$$

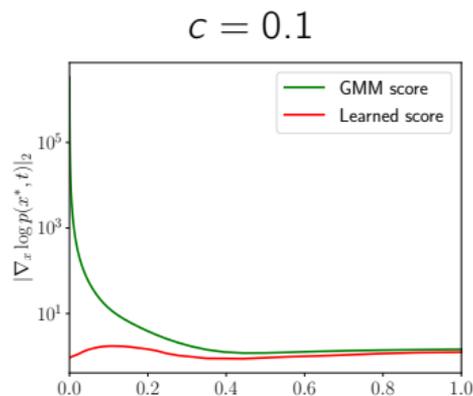
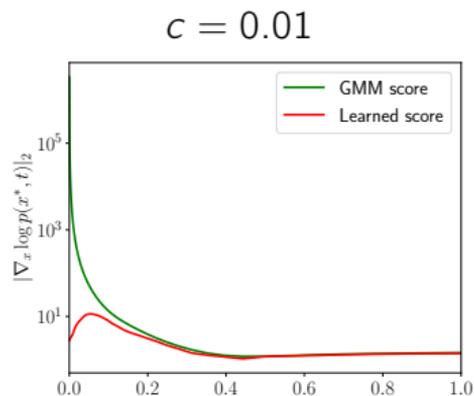
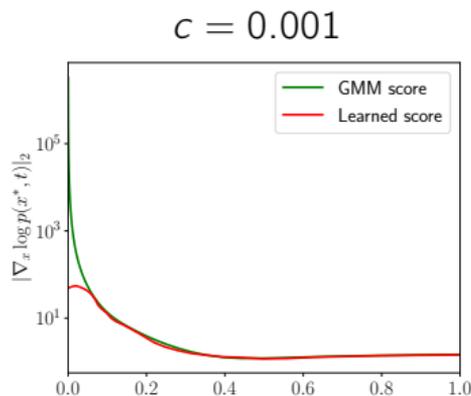
# Memorization Versus Regularization using Tikhonov

- ▶ Evaluated the fraction of 2000 generated samples  $\mathbf{x}(0)$  that match the data samples



**Takeaway:** Increasing Tikhonov regularization on Gaussian mixture prevents memorization

# Regularization and Learned Score using Tikhonov

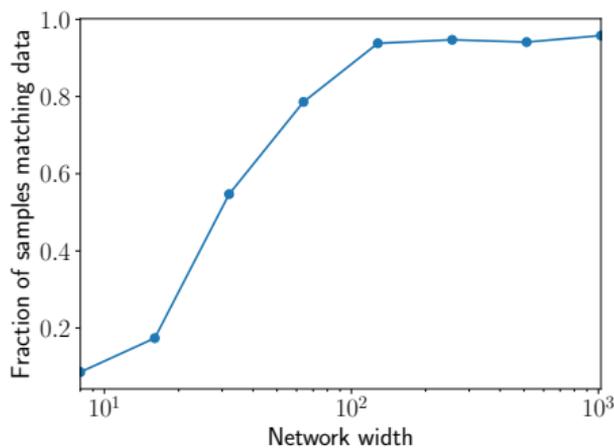
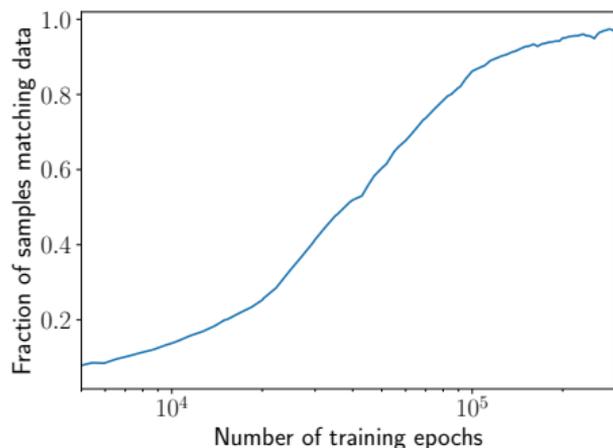


Time-dependence of the learned score  $s(x^*, t)$  at fixed  $x^*$

**Takeaway:** Increasing Tikhonov regularization reduces singular behaviour in Gaussian mixture

# Memorization Versus Regularization using Neural Networks

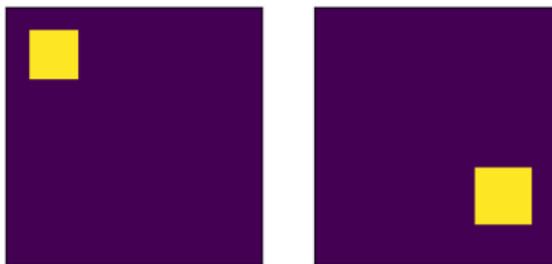
- ▶ Parameterized the score using a three-layer feedforward NN
- ▶ Study increased training iterations and model parameters (NN width)



**Takeaway:** Early stopping in training and under-parameterization avoids memorization

## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process

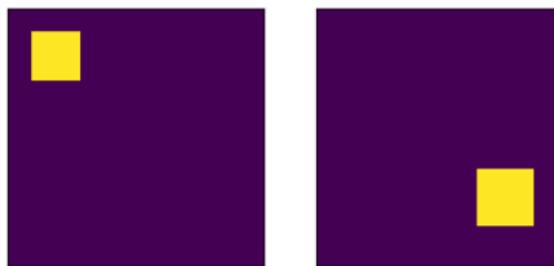


Training data

# Collapse Is Observed With Images

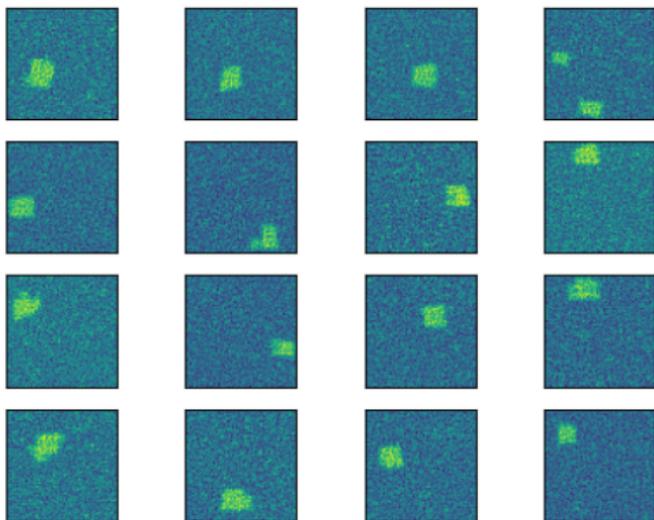
## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process



Training data

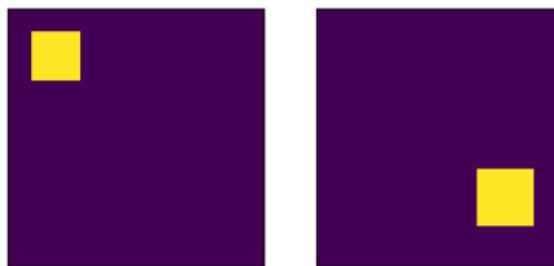
2k epochs



# Collapse Is Observed With Images

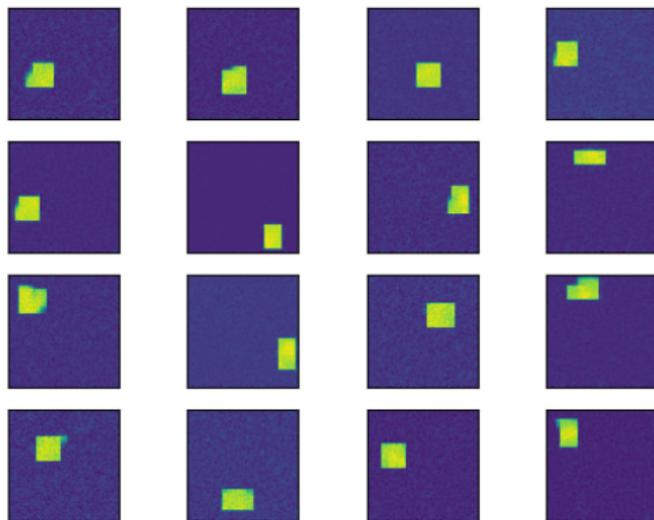
## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process



Training data

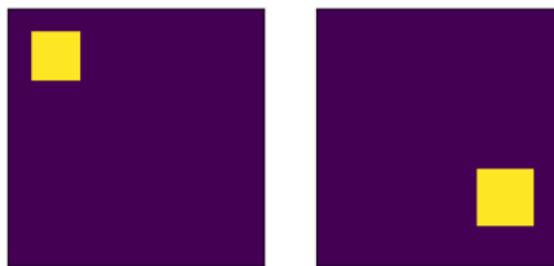
4k epochs



# Collapse Is Observed With Images

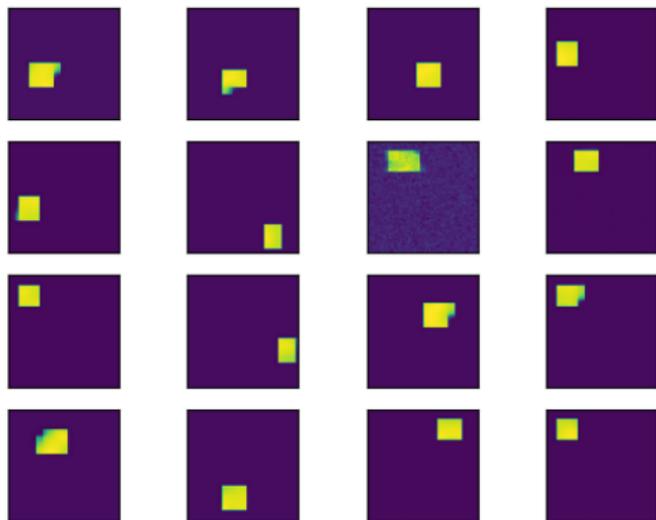
## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process



Training data

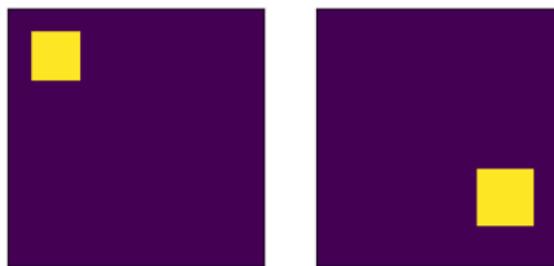
6k epochs



# Collapse Is Observed With Images

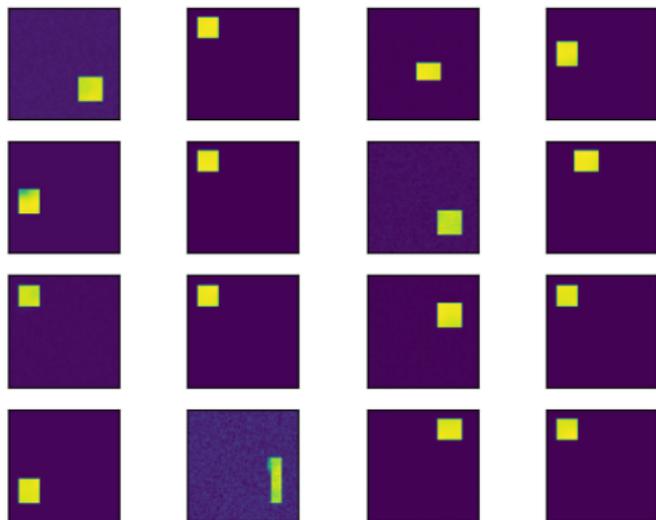
## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process



Training data

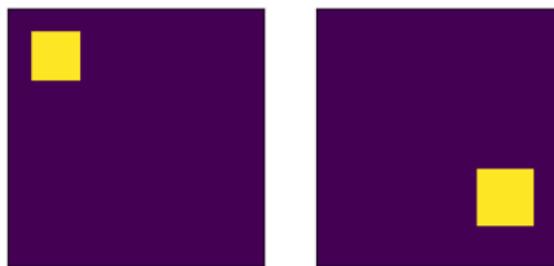
10k epochs



# Collapse Is Observed With Images

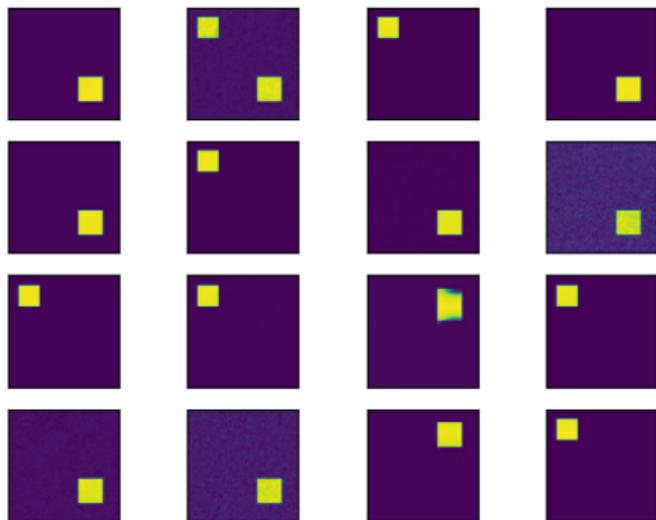
## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process



Training data

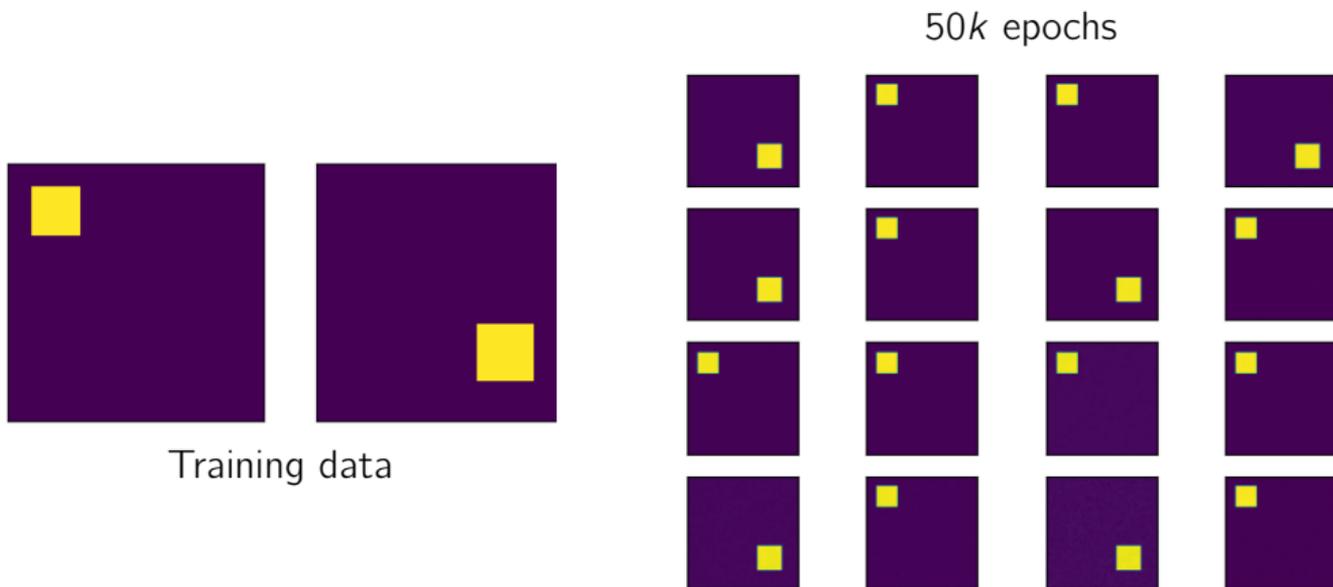
20k epochs



# Collapse Is Observed With Images

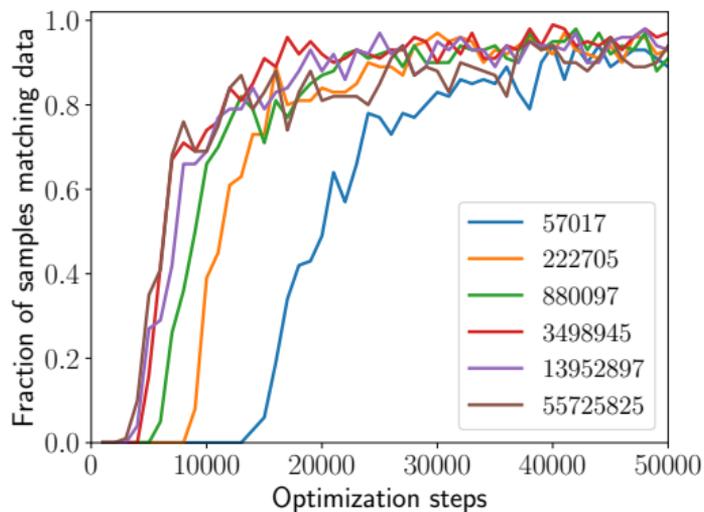
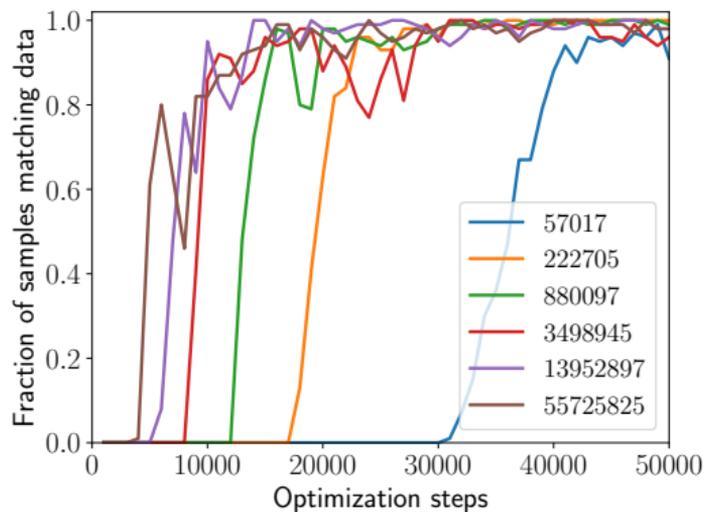
## Imaging example:

- ▶ Learned score function using EDM model (Karras et al., 2022) with U-Net architecture
- ▶ Training set of  $N = 2$  images of small squares embedded in empty background
- ▶ Generated samples after each epoch with fixed noise process



**Takeaway:** Early stopping of training is one way to prevent data collapse

# Collapse Is Observed With Images



Fraction of memorized samples. Legend indicates number of parameters in a U-Net model for the score. The left plot uses  $N = 2$  training samples while the right plot uses  $N = 8$ .

**Takeaway:** Using fewer model parameters also prevents memorization

- 1 Diffusion Model Methodology
- 2 Main Theorem on Memorization
- 3 Analysis Underlying Theorem
- 4 Numerics Illustrating Theorem
- 5 Conclusions**

## Main ideas

- ▶ Empirical score function has closed form expression
- ▶ Limit points of dynamics with empirical score contain data and Voronoi boundaries
- ▶ Dynamics are constrained by Voronoi tessellation

## Future work

- ▶ Dynamics under explicit regularization in unconditional and conditional settings
- ▶ Building data-adaptive regularizers

## Main ideas

- ▶ Empirical score function has closed form expression
- ▶ Limit points of dynamics with empirical score contain data and Voronoi boundaries
- ▶ Dynamics are constrained by Voronoi tessellation

## Future work

- ▶ Dynamics under explicit regularization in unconditional and conditional settings
- ▶ Building data-adaptive regularizers

**Thank You**

# References I

-  Baptista, Ricardo et al. (2025). “Memorization and Regularization in Generative Diffusion Models”. In: *arXiv preprint arXiv:2501.15785*.
-  Carlini, Nicolas et al. (2023). “Extracting training data from diffusion models”. In: *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 5253–5270.
-  Gu, Xiangming et al. (2023). “On Memorization in Diffusion Models”. In: *arXiv:2310.02664*.
-  Karras, Tero et al. (2022). “Elucidating the design space of diffusion-based generative models”. In: *Advances in Neural Information Processing Systems 35*, pp. 26565–26577.
-  Li, Sixu et al. (2024). “A good score does not lead to a good generative model”. In: *arXiv:2401.04856*.
-  Ramesh, Aditya et al. (2022). “Hierarchical text-conditional image generation with CLIP latents”. In: *arXiv:2204.06125* 1.2, p. 3.
-  Scarvelis, Christopher et al. (2023). “Closed-form diffusion models”. In: *arXiv:2310.12395*.
-  Somepalli, Gowthami et al. (2023). “Understanding and mitigating copying in diffusion models”. In: *Advances in Neural Information Processing Systems 36*, pp. 47783–47803.



Vincent, Pascal (2011). “A connection between score matching and denoising autoencoders”. In: *Neural computation* 23.7, pp. 1661–1674.



Wan, Zhong Yi et al. (2023). “Debias coarsely, sample conditionally: Statistical downscaling through optimal transport and probabilistic diffusion models”. In: *Advances in Neural Information Processing Systems* 36, pp. 47749–47763.