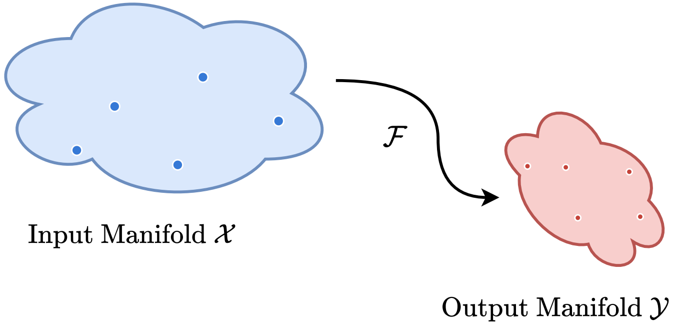
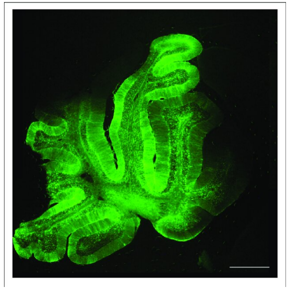
Data Manifolds as Priors for Inverse Problems: From Regularization to Representation

Jiequn Han
Center of Computational Mathematics
Flatiron Institute, Simons Foundation

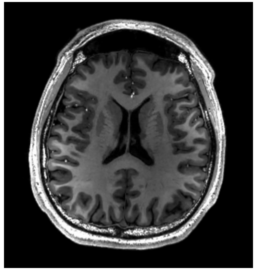
Institute for Mathematical and Statistical Innovation
June 10, 2025

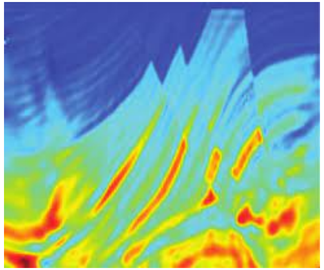


Inverse Problem

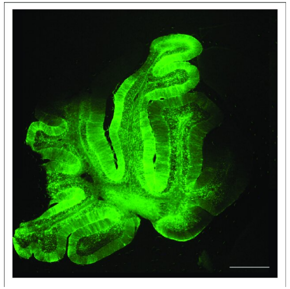


recover x from $y = \mathcal{F}(x) + \varepsilon$

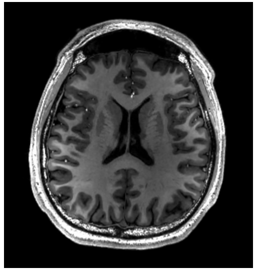


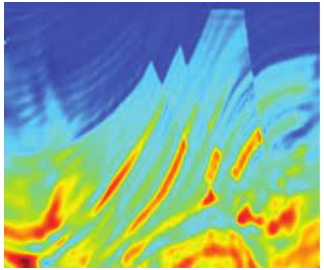


Inverse Problem



recover x from $y = \mathcal{F}(x) + \varepsilon$





Recover x from fidelity term + prior term

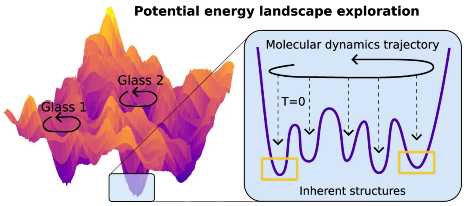
Point estimation: $\underset{x}{\operatorname{argmin}} \operatorname{dist}(\mathcal{F}(x), y) + \operatorname{Reg}(x)$

Bayesian sampling: $x \sim p(x|y) \propto p(y|\mathcal{F}(x)) p_{\text{prior}}(x)$

Priors in Inverse Problem

Recover x from fidelity term + prior term

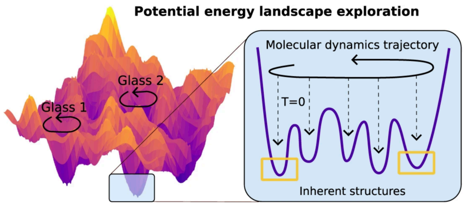
Classical priors (Tikhonov, sparsity, smoothness, etc): simple and often effective, but can fail in complex landscapes in high dimensions



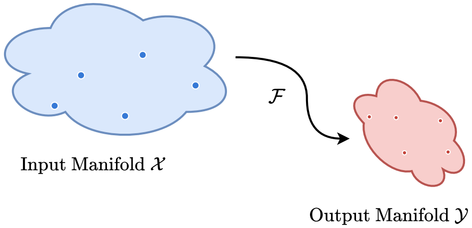
Priors in Inverse Problem

Recover x from fidelity term + prior term

Classical priors (Tikhonov, sparsity, smoothness, etc): simple and often effective, but can fail in complex landscapes in high dimensions



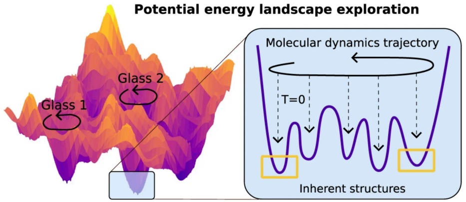
Data manifold as priors: represent data support or its distribution directly



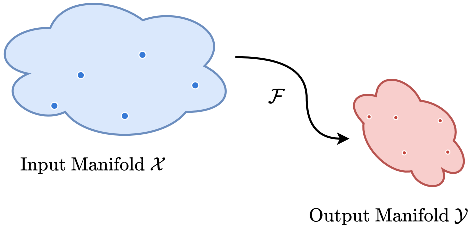
Priors in Inverse Problem

$$\text{Recover } x \text{ from } \text{fidelity term} + \text{prior term}$$

Classical priors (Tikhonov, sparsity, smoothness, etc): simple and often effective, but can fail in complex landscapes in high dimensions



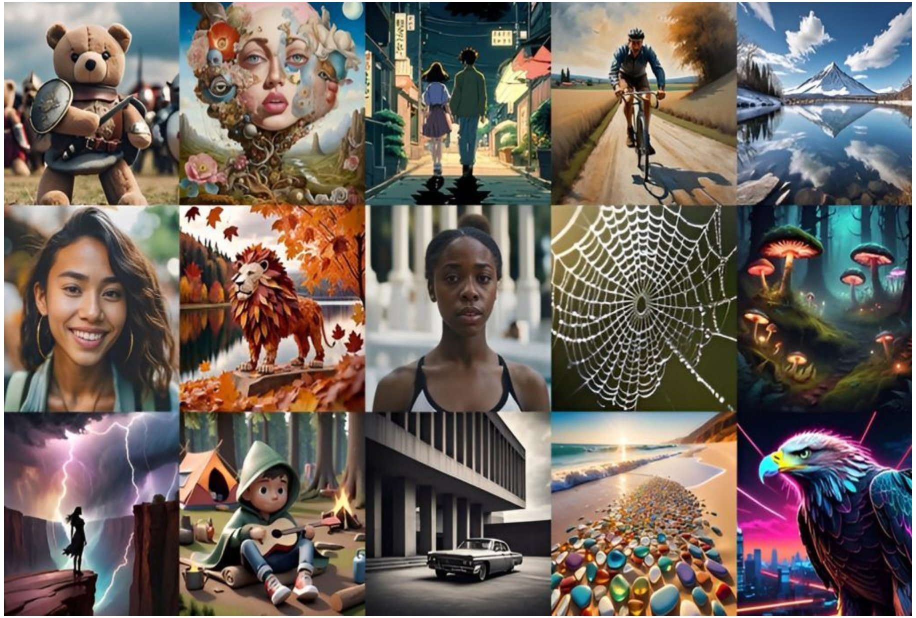
Data manifold as priors: represent data support or its distribution directly



This talk: two works showing how data manifolds help when (1) data prior or (2) fidelity term is complex

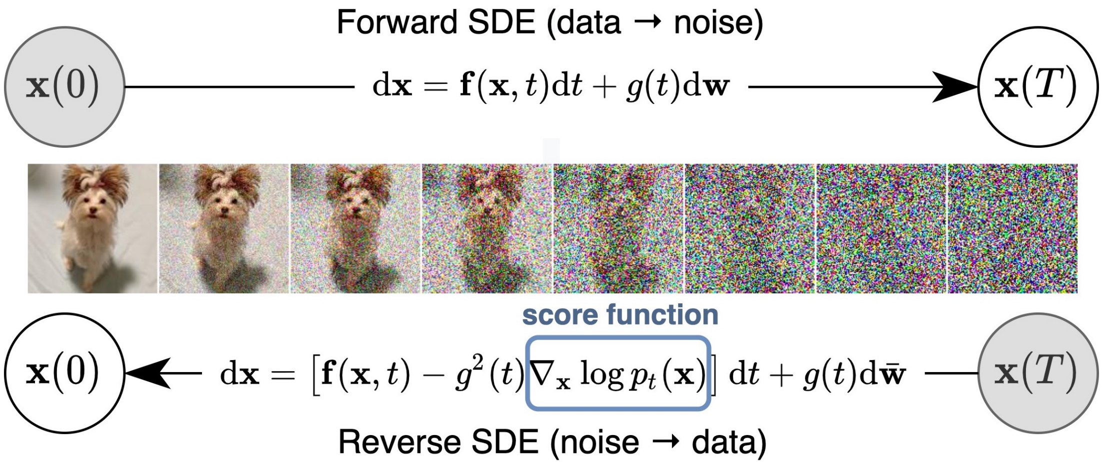
Generative Model

DALLE 3



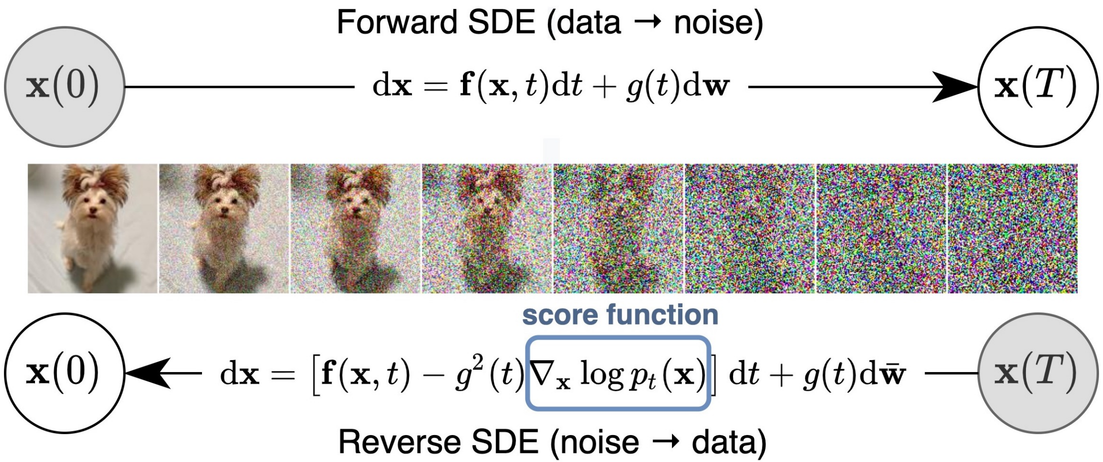
Stable Diffusion

Score-Based Diffusion and Denoising Oracles



Score-Based Generative Modeling through Stochastic Differential Equations, Song et al. (2021)

Score-Based Diffusion and Denoising Oracles



By Tweedie's formula, the time-dependent score along OU (or Heat) semigroup is equivalent to denoising oracle

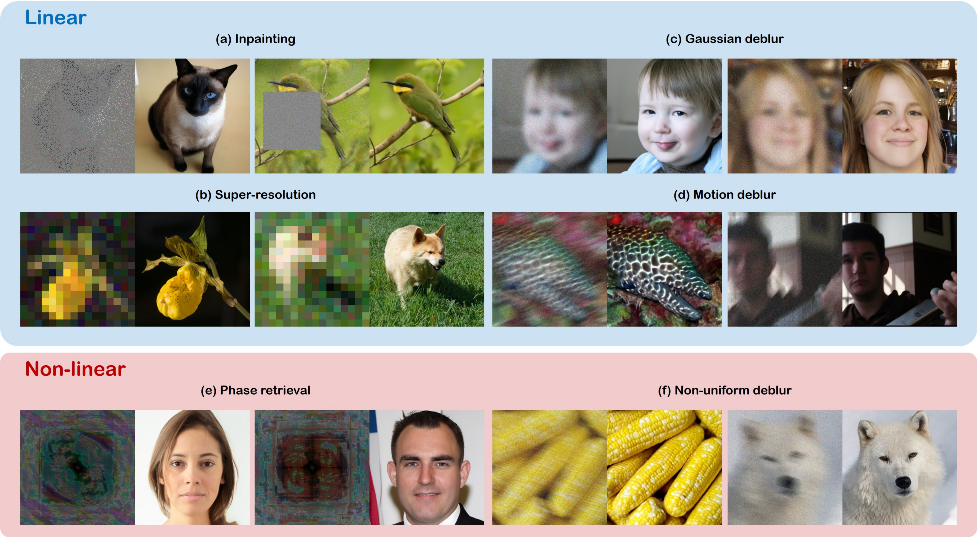
$$\mathrm{DO}_\pi(x, t) = \mathbb{E}[X | x = X + tZ], \;\; \text{where} \;\; X \sim \pi, \; Z \sim \mathcal{N}(0, I_d)$$

Score-Based Generative Modeling through Stochastic Differential Equations, Song et al. (2021)

Diffusion Posterior Sampling for Inverse Imaging Problems

$$x \sim p(x|y) \propto p(y|\mathcal{F}(x)) \boxed{p_{\text{prior}}(x)}$$

diffusion model



Diffusion Posterior Sampling for General Noisy Inverse Problems, Chung et al. (2023)

Diffusion Posterior Sampling for Inverse Imaging Problems

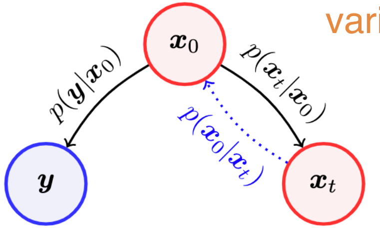
$$x \sim p(x|y) \propto p(y|\mathcal{F}(x)) \boxed{p_{\text{prior}}(x)}$$ diffusion model

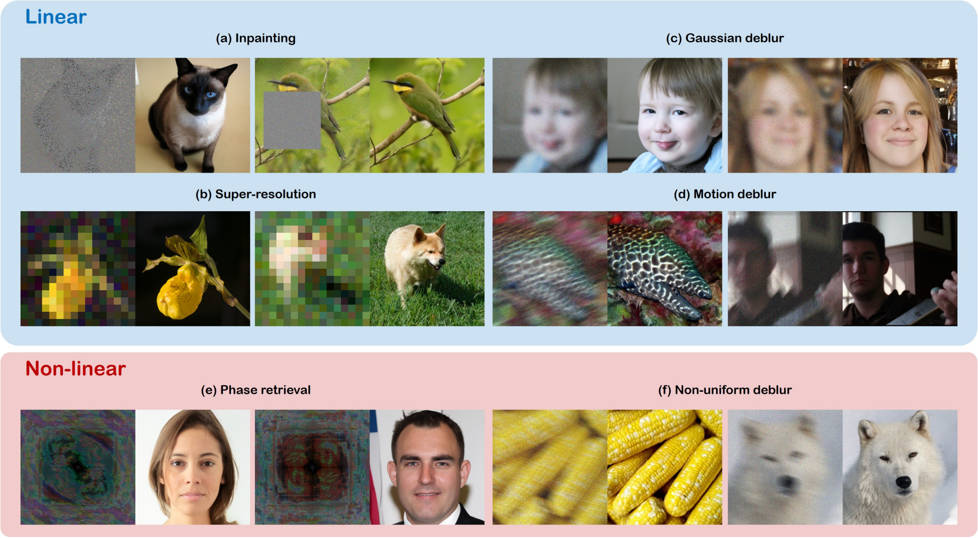
Score for prior: $\nabla_x \log p_t(x_t)$

Score for posterior:

$$\nabla_x \log p_t(x_t|y)$$
$$= \nabla_x \log p_t(x_t) + \nabla_x \log p_t(y|x_t)$$

various approximation





Diffusion Posterior Sampling for General Noisy Inverse Problems, Chung et al. (2023)

Provable Posterior Sampling

How to rigorously transfer the power of diffusion model/denoising oracle prior to sample posterior?

Provable Posterior Sampling

How to rigorously transfer the power of diffusion model/denoising oracle prior to sample posterior?

We can provably sample posterior distribution for certain linear inverse problems almost for free! *(Bruna and Han, NeurIPS 2024)*

Joan Bruna (NYU)

Problem Setup and Warmup

Given time-dependent score for OU $\mathrm{d}X_t = -X_t\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \text{ (prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \ \varepsilon \sim \gamma_d, \ \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x)\exp\{-\frac{1}{2}x^\top Qx + x^\top b\} := \mathsf{T}_{Q,b}\pi \ , \ \text{with } Q = \frac{1}{\sigma^2}A^\top A\,,\, b = -\frac{1}{\sigma^2}A^\top y$$

Problem Setup and Warmup

Given time-dependent score for OU $\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \ \text{(prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \ \varepsilon \sim \gamma_d, \ \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x) \exp\{-\frac{1}{2}x^\top Q x + x^\top b\} := \mathsf{T}_{Q,b}\pi \ , \ \text{with } Q = \frac{1}{\sigma^2}A^\top A \, , \, b = -\frac{1}{\sigma^2}A^\top y$$

Warmup: when $Q \propto \mathrm{Id}$ the task seems 'compatible' with the denoising oracle.

$$T^* = \frac{1}{2}\log(1+\sigma^2), \ \tilde{y} = e^{-T^*}y \quad \Longrightarrow \quad p(x|\tilde{y}) \overset{d}{=} p(X_0 | X_{T^*} = \tilde{y})$$

Problem Setup and Warmup

Given time-dependent score for OU $\quad \mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \text{ (prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \; \varepsilon \sim \gamma_d, \; \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x) \exp\{-\frac{1}{2}x^\top Qx + x^\top b\} := \mathsf{T}_{Q,b}\pi \;, \; \text{with } Q = \frac{1}{\sigma^2}A^\top A \,,\, b = -\frac{1}{\sigma^2}A^\top y$$

Warmup: when $\; Q \propto \mathrm{Id} \;$ the task seems 'compatible' with the denoising oracle.

$$T^* = \frac{1}{2}\log(1+\sigma^2), \; \tilde{y} = e^{-T^*}y \quad \implies \quad p(x|\tilde{y}) \overset{d}{=} p(X_0|X_{T^*} = \tilde{y})$$

We can (1) first initialize $X_{T*} = e^{-T^*}y$ and (2) run the original reverse SDE from T^* to 0 to get the exact posterior

Problem Setup and Warmup

Given time-dependent score for OU $\quad \mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \text{ (prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \ \varepsilon \sim \gamma_d, \ \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x) \exp\{-\frac{1}{2} x^\top Q x + x^\top b\} := \mathsf{T}_{Q,b}\pi \ , \ \text{with } Q = \frac{1}{\sigma^2} A^\top A \, , \, b = -\frac{1}{\sigma^2} A^\top y$$

Warmup: when $Q \propto \mathrm{Id}$ the task seems ‘compatible’ with the denoising oracle.

$$T^* = \frac{1}{2}\log(1+\sigma^2), \ \tilde{y} = e^{-T^*} y \quad \Longrightarrow \quad p(x|\tilde{y}) \overset{d}{=} p(X_0 | X_{T^*} = \tilde{y})$$

We can (1) first initialize $X_{T^*} = e^{-T^*} y$ and (2) run the original reverse SDE from T^* to 0 to get the exact posterior

What if a general Q?

Tilted Transport for Posterior Sampling

Consider a time-varying quadratic tilt

$$\nu_t \propto \pi_t(x) \exp\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\}$$

$$\begin{cases} \dot{Q}_t = 2(I + Q_t)Q_t \ , & Q_0 = Q \\ \dot{b}_t = (I + 2Q_t)b_t \ , & b_0 = b \end{cases}$$

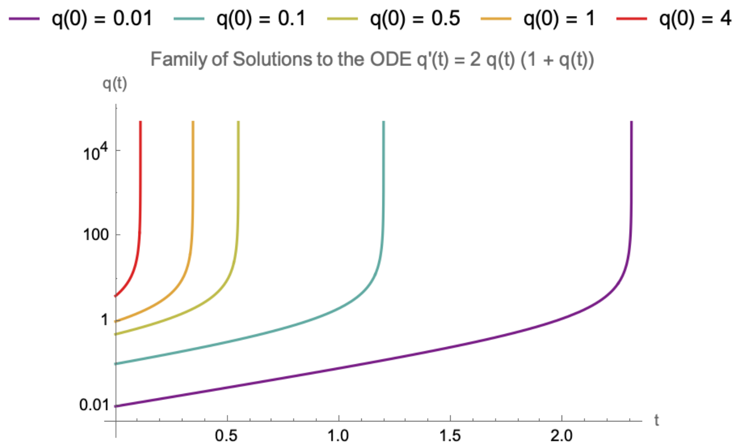
Theorem (titled transport) Assume $t < T$ such that the ODE is well-defined on $[0, t]$. By initializing $X_t \sim \nu_t$ and run the reverse SDE from t to 0, we have $X_s \sim \nu_s$ for $s \in [0, t]$, specifically, X_0 gives the desired posterior.

Tilted Transport for Posterior Sampling

Consider a time-varying quadratic tilt

$$\nu_t \propto \pi_t(x) \exp\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\}$$

$$\begin{cases} \dot{Q}_t = 2(I + Q_t)Q_t \ , & Q_0 = Q \\ \dot{b}_t = (I + 2Q_t)b_t \ , & b_0 = b \end{cases}$$



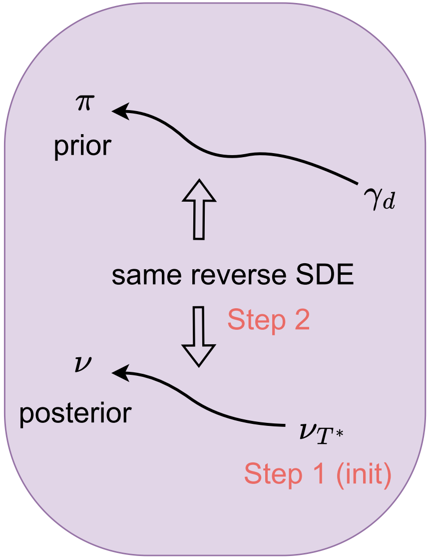
Theorem (titled transport) Assume $t < T$ such that the ODE is well-defined on $[0, t]$. By initializing $X_t \sim \nu_t$ and run the reverse SDE from t to 0, we have $X_s \sim \nu_s$ for $s \in [0, t]$, specifically, X_0 gives the desired posterior.

Tilted Transport for Posterior Sampling

Given a baseline sampling algorithm **Alg** and starting time $\tilde{T} = T^* - \epsilon$ (for stable ODE solutions), the tilted transport works in two steps:

1. Use the baseline sampling algorithm **Alg** to sample $X_{\tilde{T}}$ from $\pi_{\tilde{T}}(x)\exp\{-\frac{1}{2}x^\top Q_{\tilde{T}}x + x^\top b_{\tilde{T}}\}$

2. Run the original reverse SDE from $\tilde{T}$ to 0 to get the desired sample

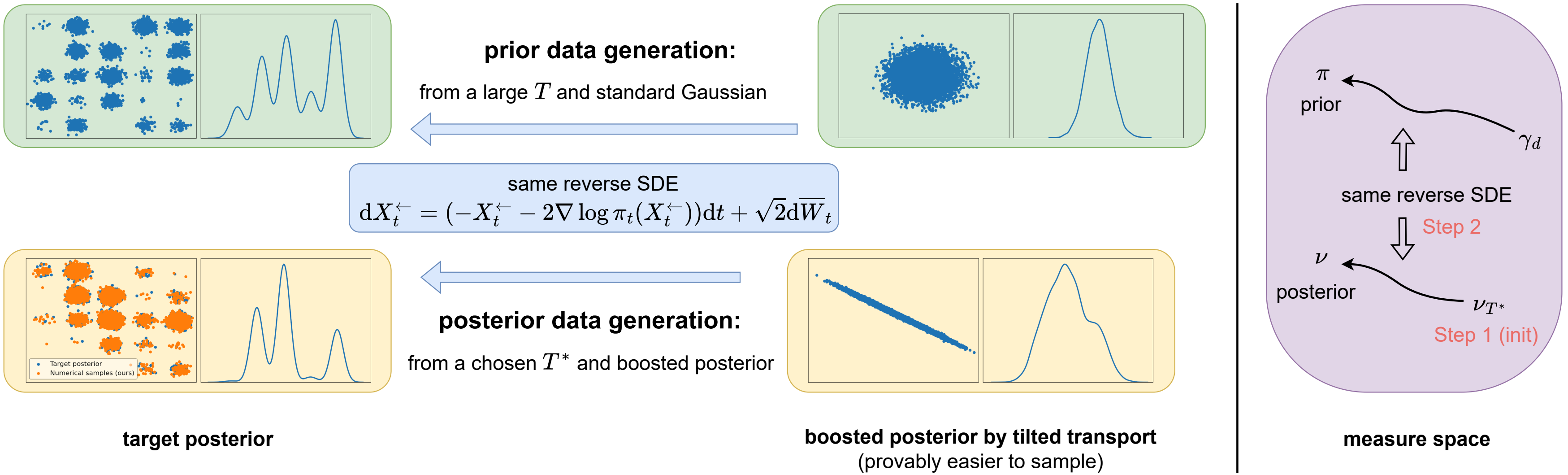


measure space

Intuition for Easier Sampling

Equivalent posterior sampling: $\nu_t \propto \boxed{\pi_t(x)}\,\boxed{\exp\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\}}$

easier prior easier likelihood



Provable Sampling

Theorem (Strong Log-Concavity of ν_T) For $t \geq 0$, let $\chi_t(\pi) := \sup_{x\in\mathbb{R}^d} \|\mathrm{Cov}[\mathsf{T}_{tI_d,tx}\pi]\|_{\mathrm{op}}$ denote the *susceptibility* of π, and let $\kappa = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ denote the condition number of Q. Then ν_{T^*} is strongly log-concave if

$$\chi_{\|Q\|}(\pi) < \|Q\|_{\mathrm{op}}^{-1} \frac{\kappa}{\kappa - 1} .$$

Sufficient condition relates

1. prior susceptibility
2. signal-to-noise ratio
3. condition of measurement

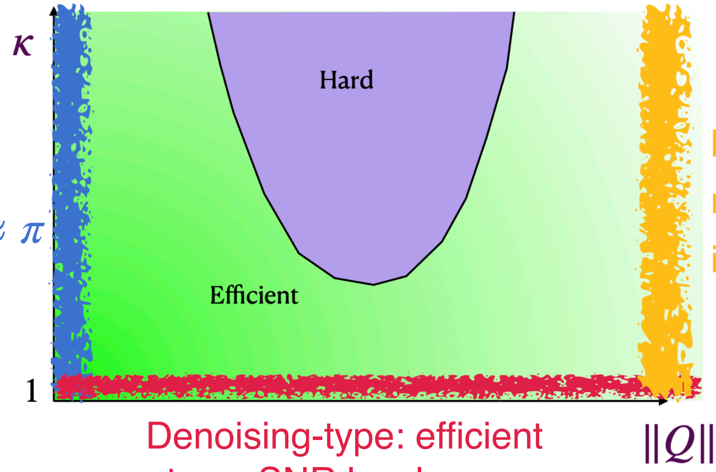
Provable Sampling

Theorem (Strong Log-Concavity of ν_T) For $t \geq 0$, let $\chi_t(\pi) := \sup_{x\in\mathbb{R}^d} \|\mathrm{Cov}[\mathsf{T}_{tI_d,tx}\pi]\|_{\mathrm{op}}$ denote the *susceptibility* of π, and let $\kappa = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ denote the condition number of Q. Then ν_{T^*} is strongly log-concave if

$$\chi_{\|Q\|}(\pi) < \|Q\|_{\mathrm{op}}^{-1} \frac{\kappa}{\kappa - 1} .$$

Sufficient condition relates

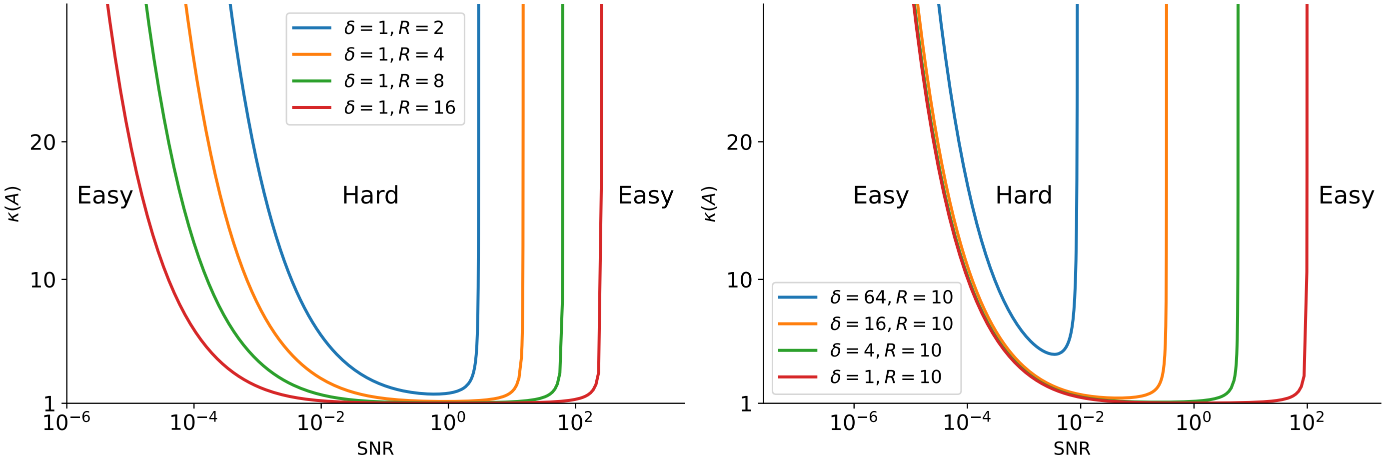
1. prior susceptibility
2. signal-to-noise ratio
3. condition of measurement



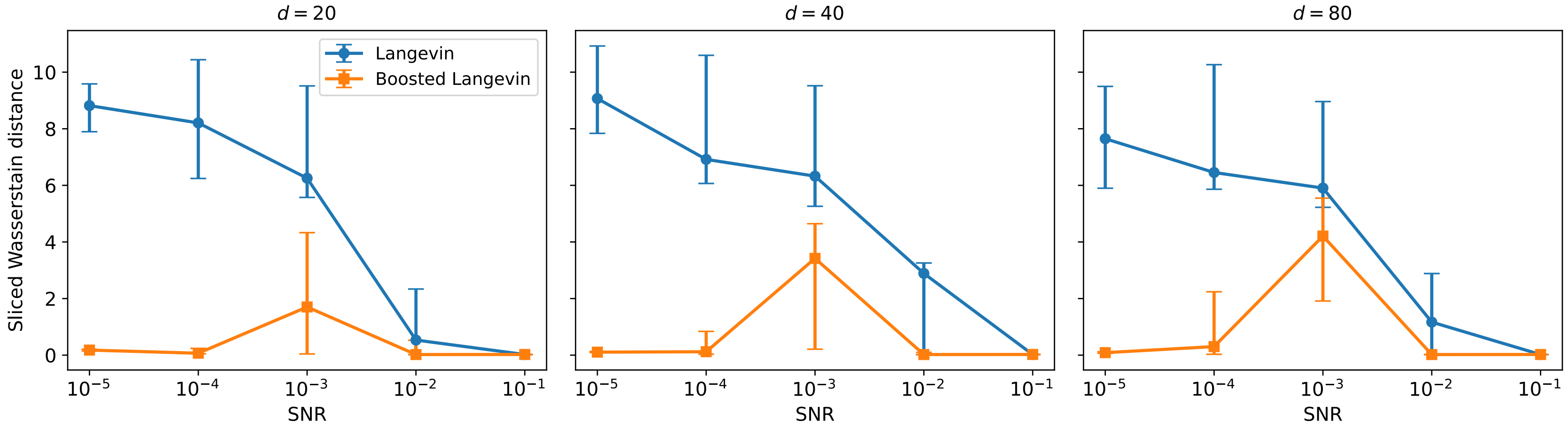
Provable Sampling

Corollary (tilted transport for Gaussian mixtures) Let $\pi = \mu \star \gamma_\delta$ and $\mathrm{diam}(\mathrm{supp}(\mu)) \leq R$, then ν_{T^*} is strongly log-concave if $(\mathrm{SNR} := \lambda_{\min}(Q) = \lambda_{\min}(A)^2/\sigma^2)$

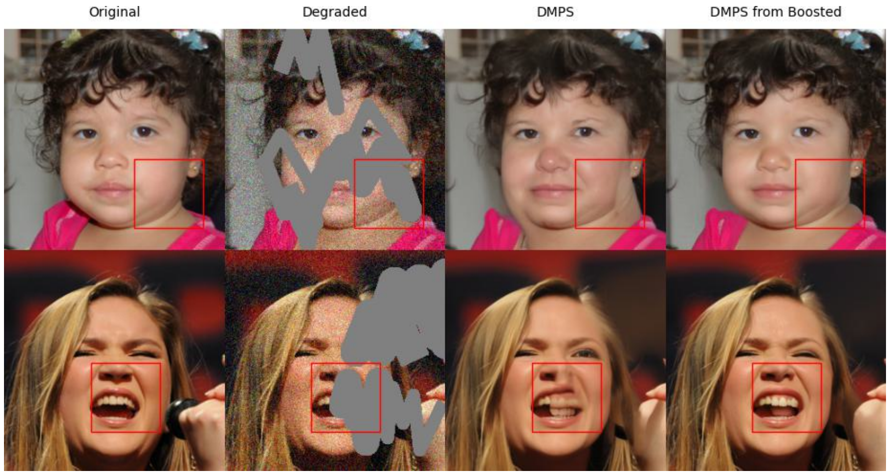
$$\frac{(1 + \delta \mathrm{SNR}^2)(\delta \kappa(A)^2 + \mathrm{SNR}^{-2})}{\kappa(A)^2 - 1} > R^2 \; .$$



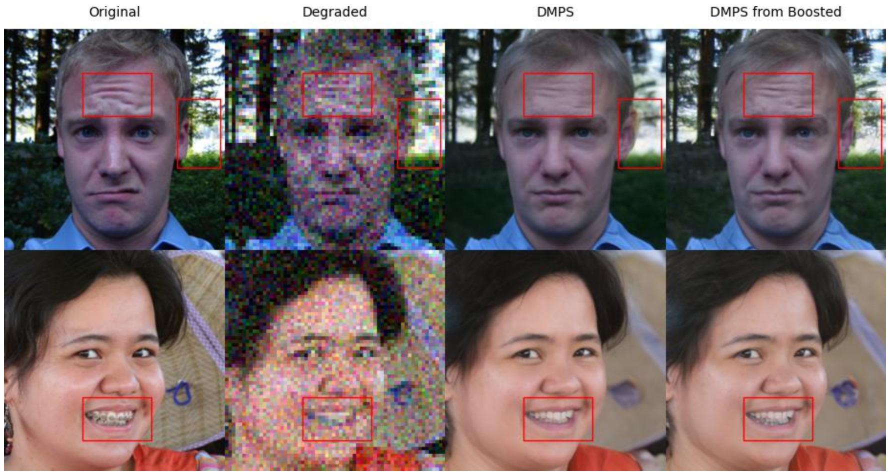
Provable Sampling (cont.)



Imaging Problems

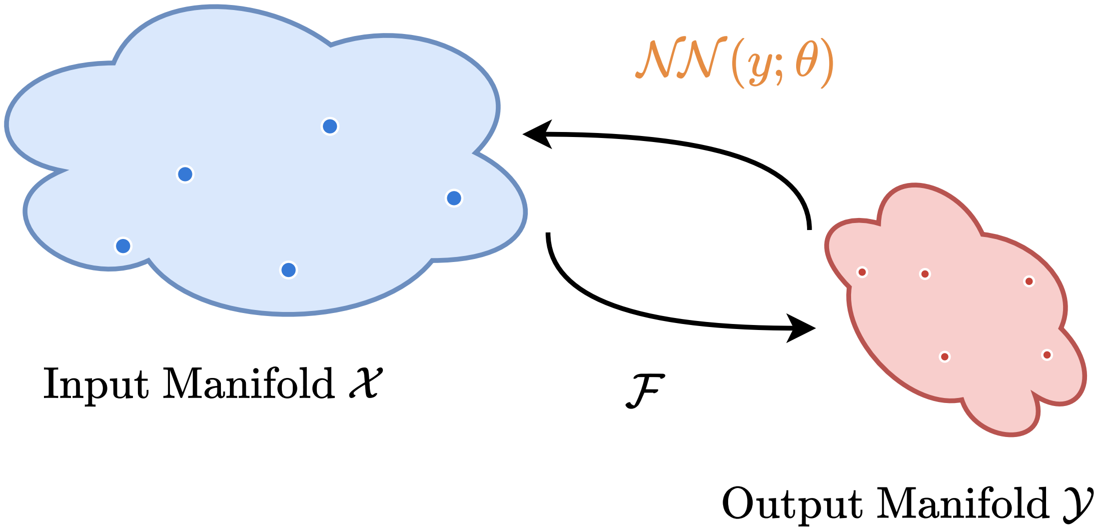


inpainting

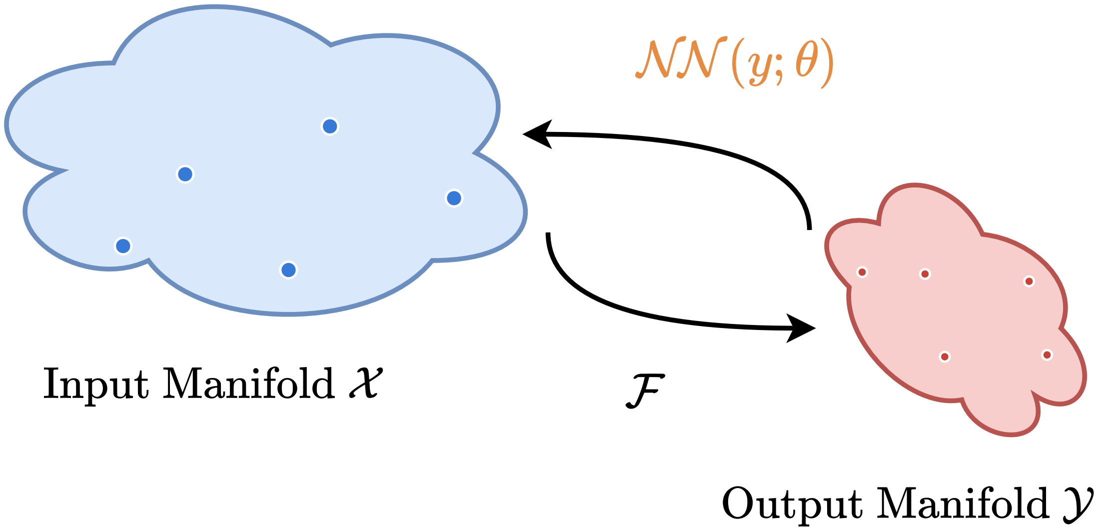


deblur

Operator Learning for Inverse Map



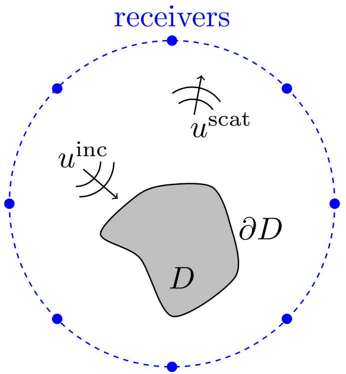
Operator Learning for Inverse Map



Highlight: the success of pretraining highly depends on the data prior complexity!

Non-convexity in Inverse Scattering

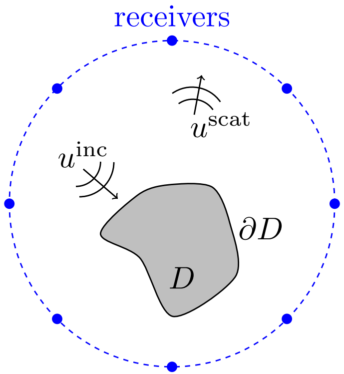
$$\Delta u^{\text{scat}} + k^2 u^{\text{scat}} = 0, \quad \text{in } \mathbb{R}^2 \setminus \overline{D}$$



High-frequency waves are needed to recover small-scale features.

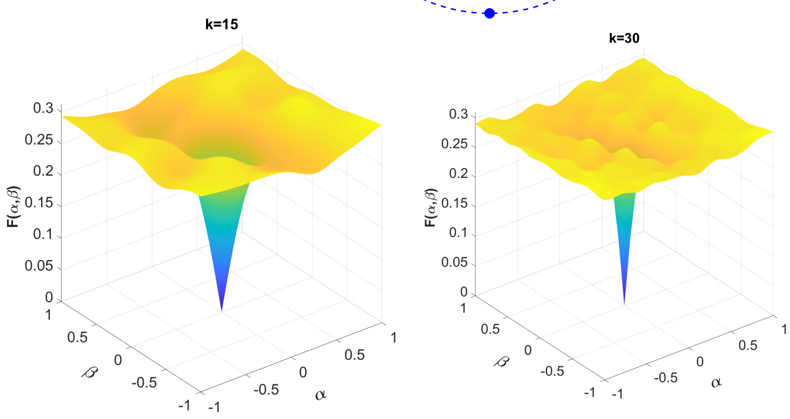
Non-convexity in Inverse Scattering

$$\Delta u^{\text{scat}} + k^2 u^{\text{scat}} = 0, \quad \text{in } \mathbb{R}^2 \setminus \overline{D}$$



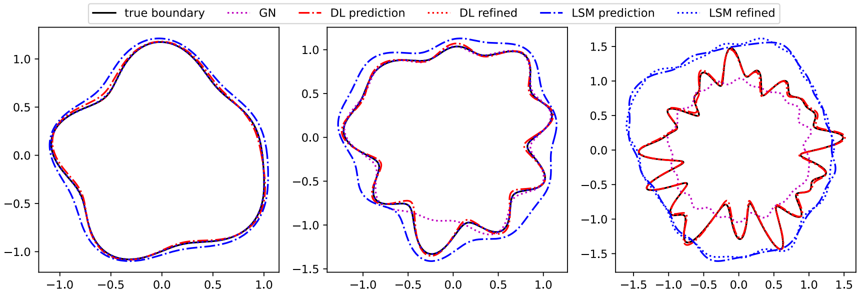
High-frequency waves are needed to recover small-scale features.

However, as frequency increases, the loss landscape becomes more non-convex, with more bad local minima.

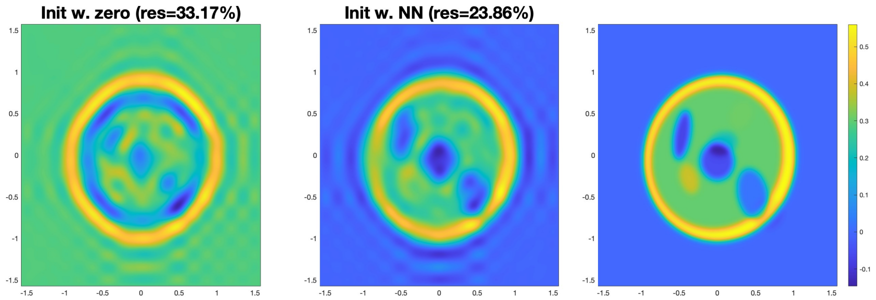


Neural Network Warm-Start

Inverse obstacle:



Inverse medium:



A neural network warm-start approach for the inverse acoustic obstacle scattering problem, JCP (2023)

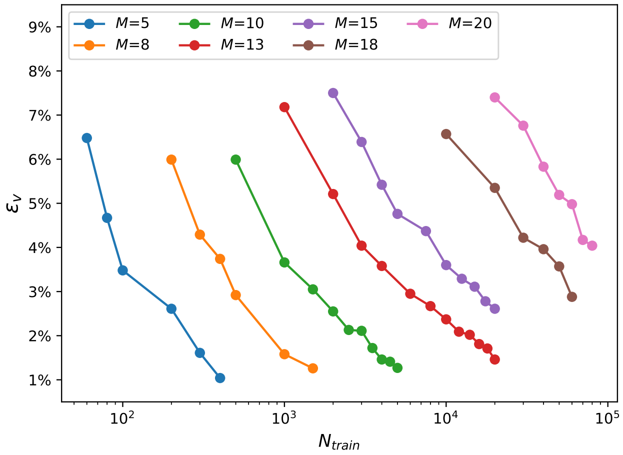
Mo Zhou

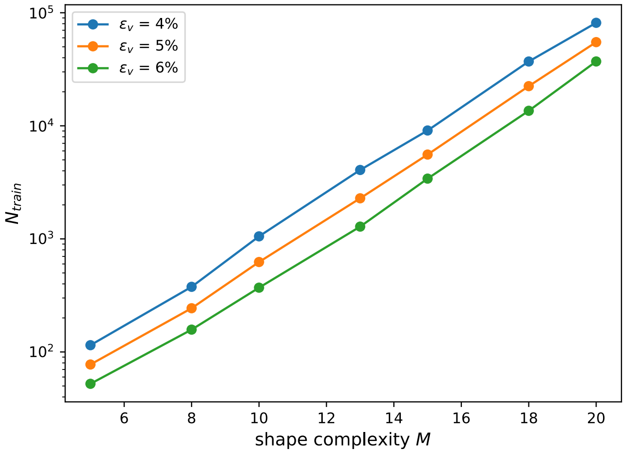
Manas Rachh

Carlos Borges

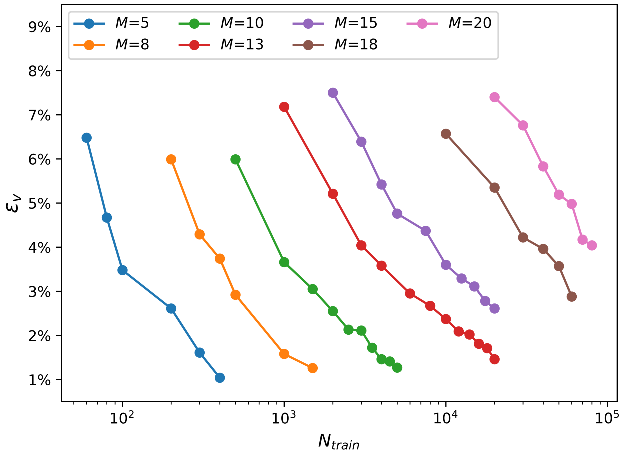
Leslie Greengard

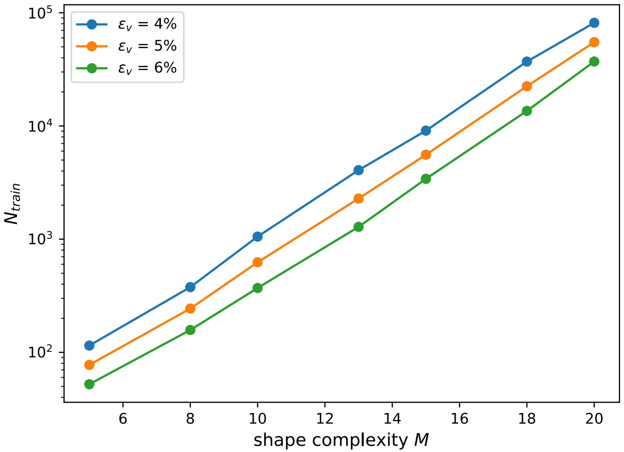
How Much Can We Scale?





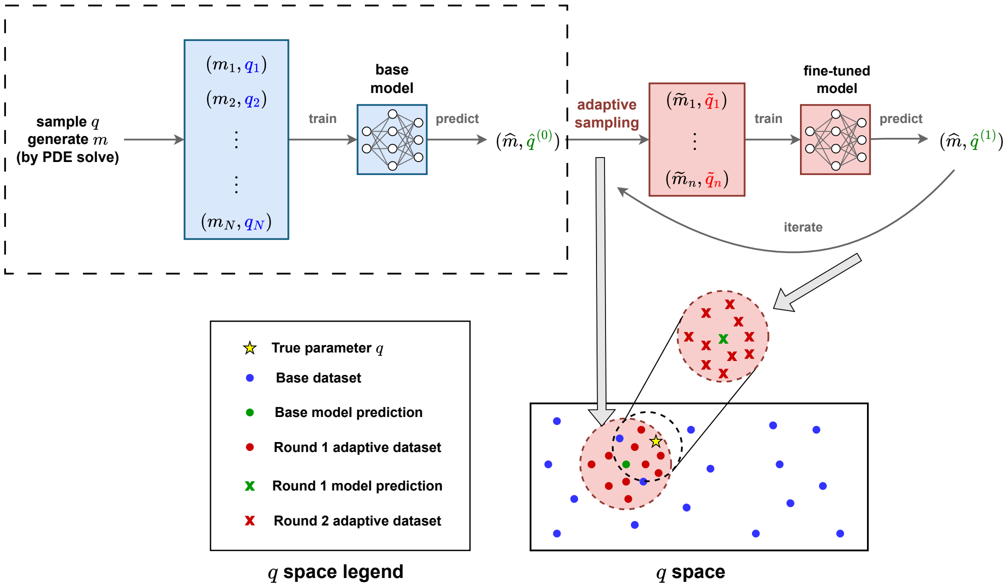
How Much Can We Scale?

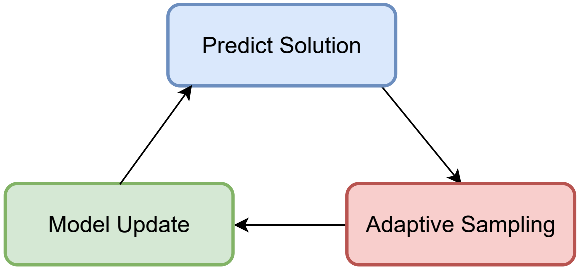




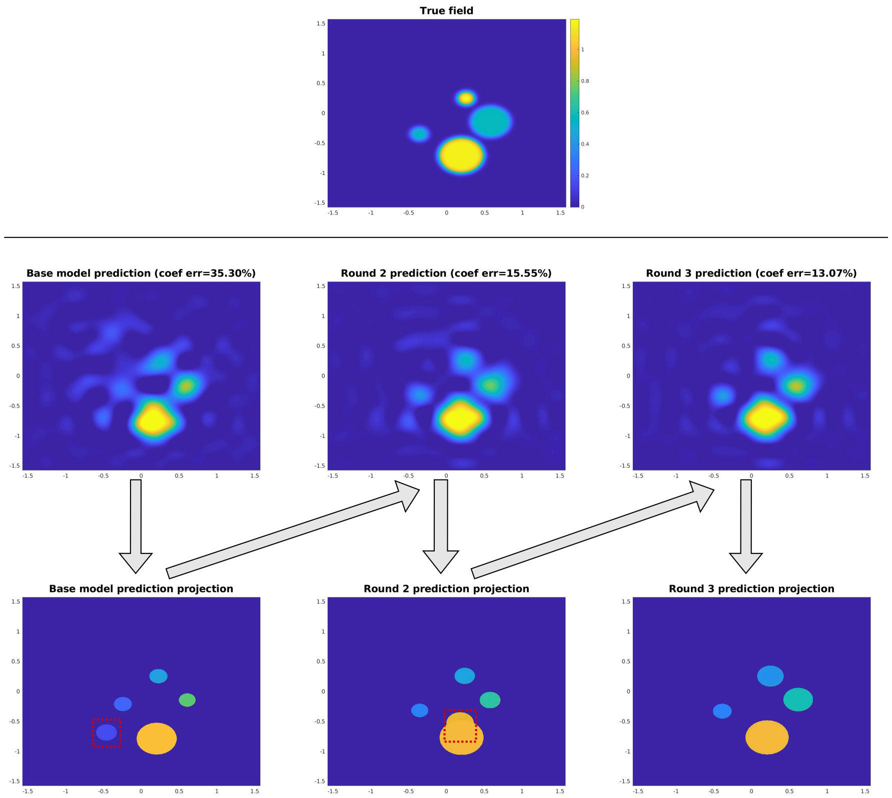
We need exponentially many samples of training data in terms of shape complexity/ frequency - A purely data-driven method is doomed to limited success

Instance-Wise Adaptive Sampling

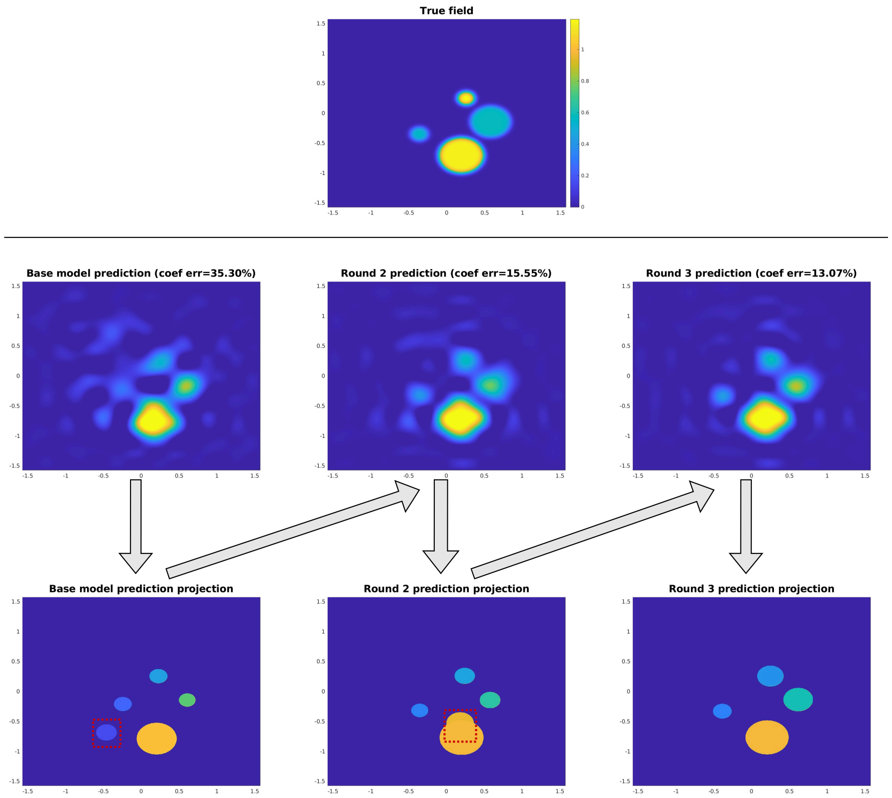




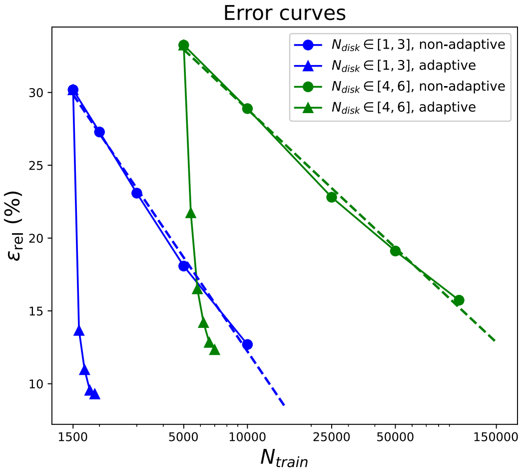
Disk Prior


True field
Base model prediction (coef err=35.30%)
Round 2 prediction (coef err=15.55%)
Round 3 prediction (coef err=13.07%)
Base model prediction projection
Round 2 prediction projection
Round 3 prediction projection

Disk Prior

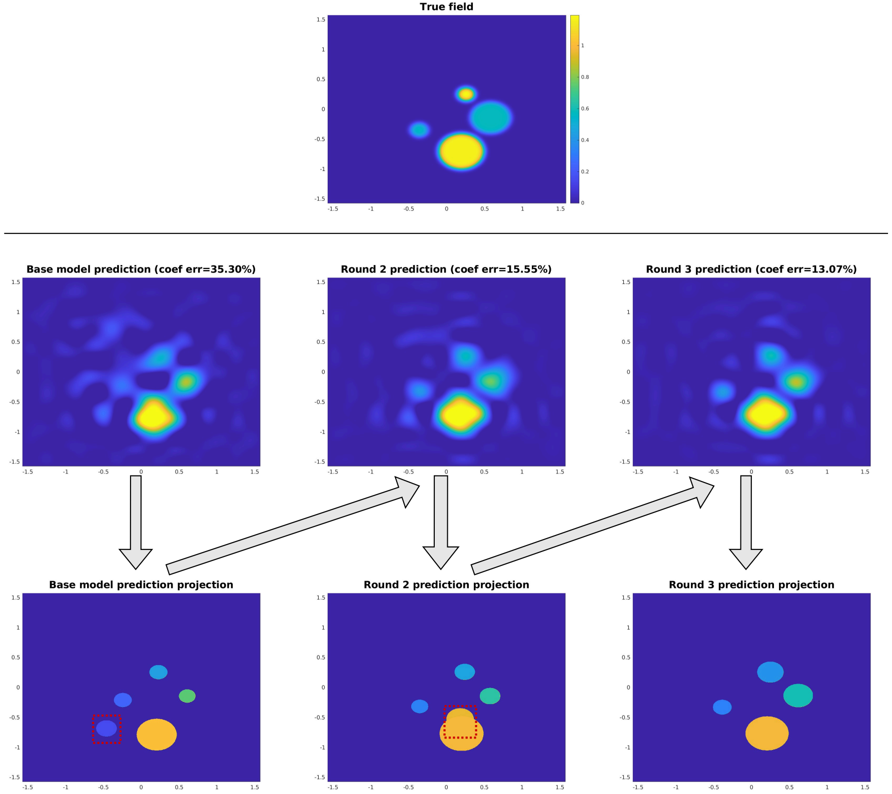


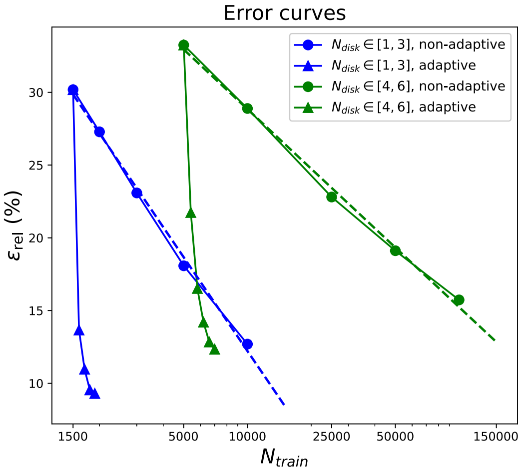
True field
Base model prediction (coef err=35.30%)
Round 2 prediction (coef err=15.55%)
Round 3 prediction (coef err=13.07%)
Base model prediction projection
Round 2 prediction projection
Round 3 prediction projection

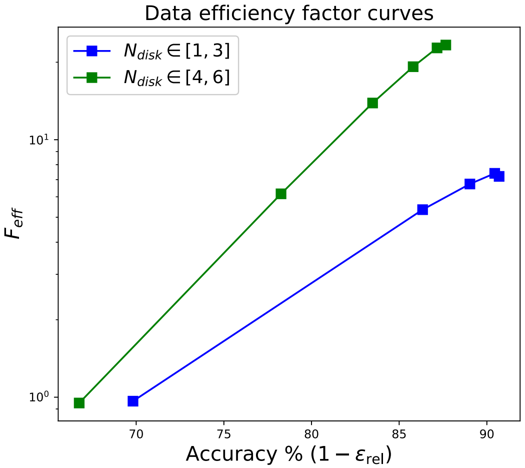


Error curves
$N_{disk} \in [1,3]$, non-adaptive
$N_{disk} \in [1,3]$, adaptive
$N_{disk} \in [4,6]$, non-adaptive
$N_{disk} \in [4,6]$, adaptive
ε_{rel} (%)
N_{train}

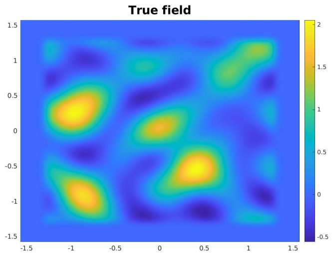
Disk Prior

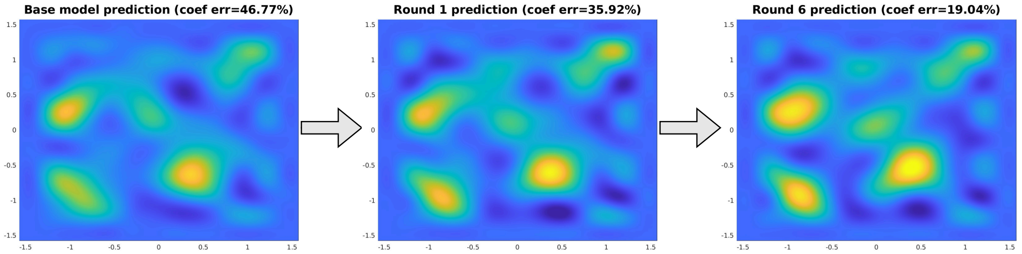

True field
Base model prediction (coef err=35.30%)
Round 2 prediction (coef err=15.55%)
Round 3 prediction (coef err=13.07%)
Base model prediction projection
Round 2 prediction projection
Round 3 prediction projection


Error curves
$N_{disk} \in [1, 3]$, non-adaptive
$N_{disk} \in [1, 3]$, adaptive
$N_{disk} \in [4, 6]$, non-adaptive
$N_{disk} \in [4, 6]$, adaptive
ε_{rel} (%)
N_{train}

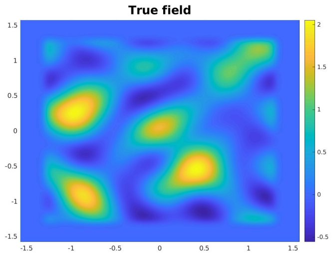

Data efficiency factor curves
$N_{disk} \in [1, 3]$
$N_{disk} \in [4, 6]$
F_{eff}
Accuracy % $(1 - \varepsilon_{rel})$

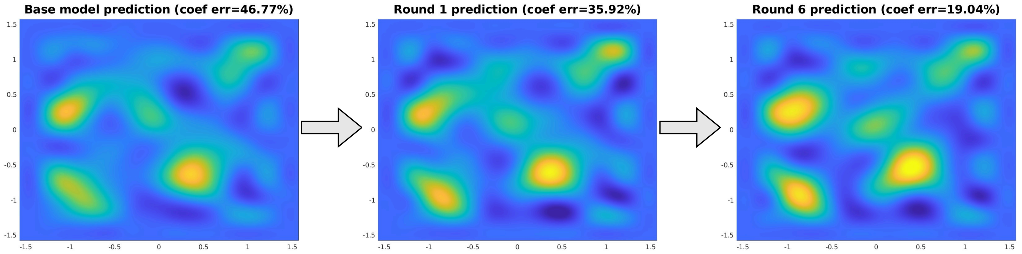
Fourier Prior

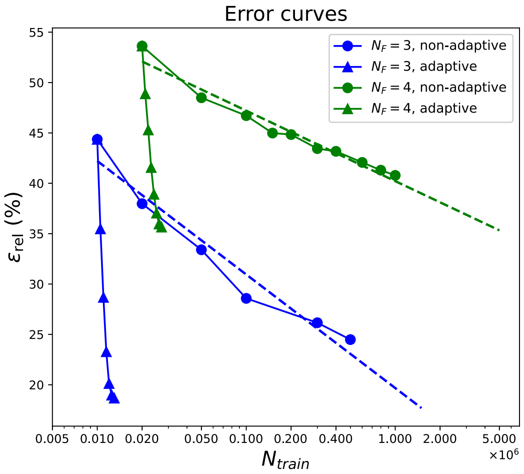




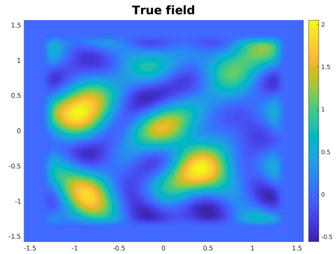
Fourier Prior

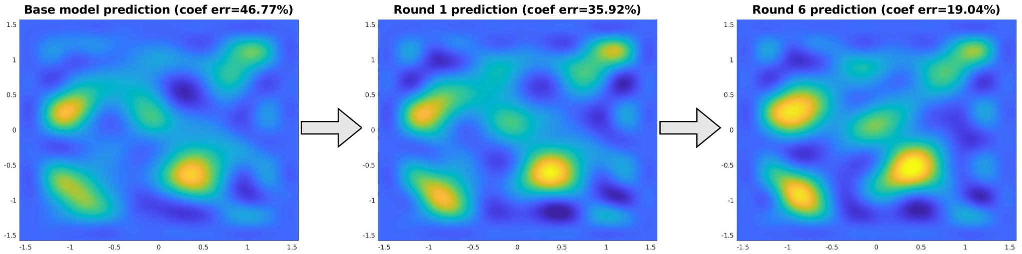


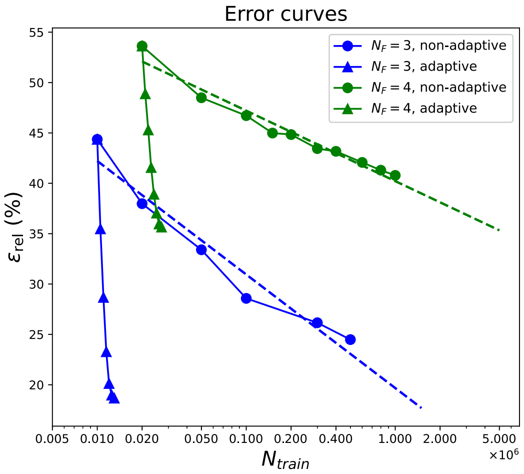


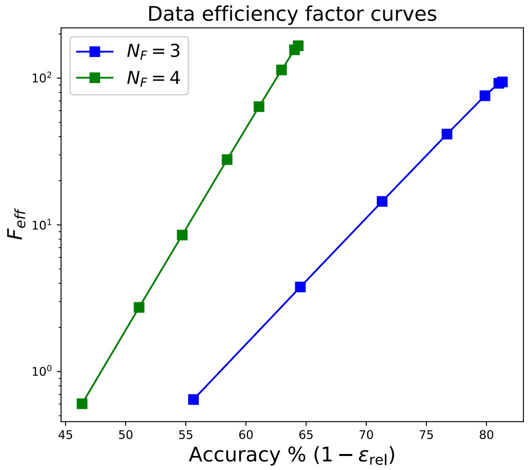


Fourier Prior

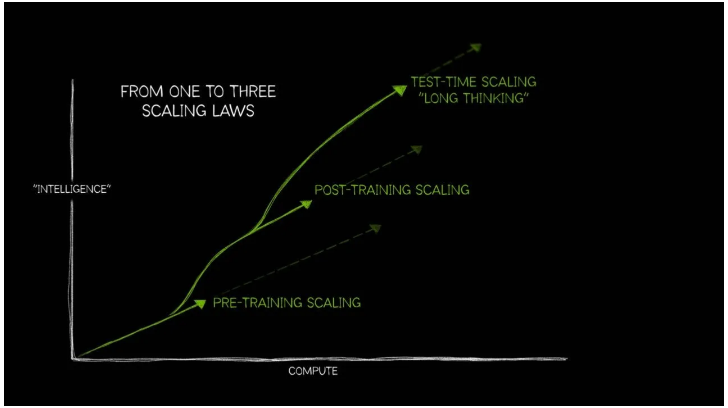


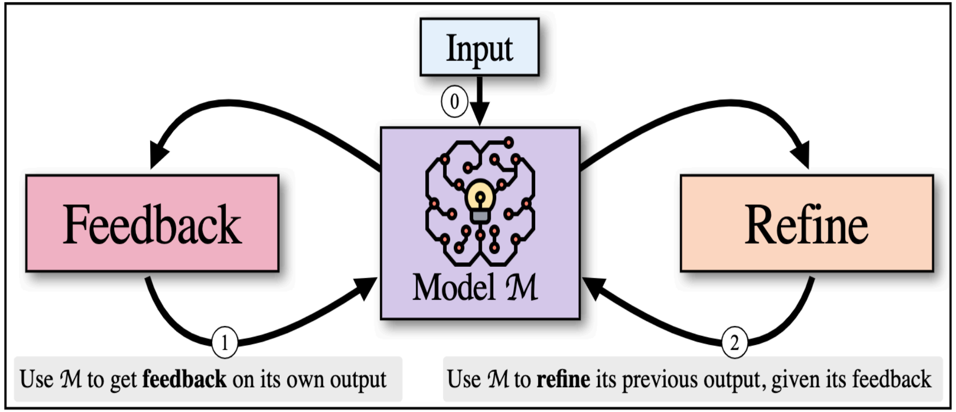






Between Pre-training and Inference-Time Scaling

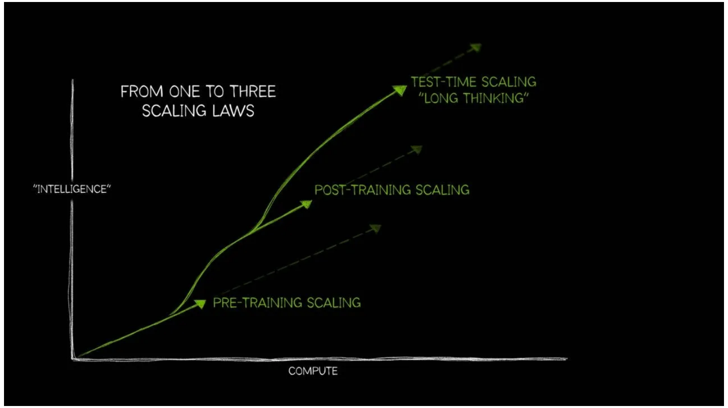




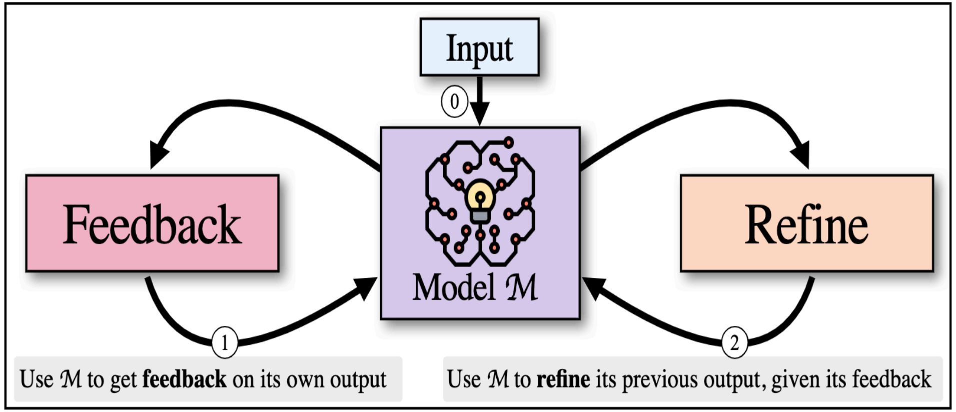
Navidia GTC AI Conference for 2025, Jensen Huang

Self-Refine: Iterative Refinement with Self-Feedback, Madaan et al. (2023)

Between Pre-training and Inference-Time Scaling



Navidia GTC AI Conference for 2025, Jensen Huang



Self-Refine: Iterative Refinement with Self-Feedback, Madaan et al. (2023)

How to distribute computation across the scientific machine learning pipeline?

Summary

- Data manifolds offer richer prior structure for inverse problems, bridging geometry and representation
- Given data manifold, adaptive sampling improves learning efficiency for supervised-learning approach
- Generative modeling provides huge opportunities for real-world complex inverse problems

Thanks for your attention