CAN SCORE-BASED DIFFUSION MODELS DISCOVER PHYSICAL CONSTRAINTS?

Anonymous authors
Paper under double-blind review

ABSTRACT

Scientific data often satisfy physical constraints that are unknown a priori. In this work, we study whether a score-based diffusion model trained on scientific data can encode geometric information that may be used for constraint localization. To do so, we first show that an appropriately-scaled score function provides a projection-like displacement toward the data-support manifold; and that we can leverage this projection-like behavior to reveal geometric information about constraints that were never explicitly provided during training. Motivated by these observations, we propose a new algorithm, SCoREMaP (Score-based Constraint Residual Extraction and Manifold Projection), to localize constraints from a trained score field. Theoretically, we show that for sufficiently accurate trained score fields, the resulting constraint region converges to the manifold in Hausdorff distance. We also show that we can propagate uncertainty over the score-model parameters to obtain pointwise epistemic uncertainty for the extracted constraint information. Empirically, we show that SCoREMaP localizes physical constraints more accurately than data-only baselines across a variety of geometric, molecular, and PDE-based tasks. These results show that diffusion models implicitly learn geometric information about physical constraints.

1 INTRODUCTION

In scientific domains, data $x \in \mathbb{R}^d$ frequently adhere to equality constraints $h(x) = 0$, where h is flexible and may represent physical laws or governing dynamics. However, in many discovery science and experimental settings, such underlying governing laws may not be known a priori. Rigorously identifying and localizing these constraints from data is a known research challenge (González et al., 2021; Camps-Valls et al., 2023). Dealing with constraints is a well-known challenge in scientific machine learning more generally (Hansen et al., 2023; Cheng et al., 2025; Utkarsh et al., 2026). Concurrently, in the seemingly disparate field of generative modeling, there has been growing interest in understanding how score-based diffusion models (SBDMs) (Song et al., 2021) behave when modeling distributions p_0 that live on a lower-dimensional manifold $\mathcal{M} = \{x : h(x) = 0\}$. This setting renders invalid many of the theoretical conditions assumed for SBDMs, namely differentiability in $\mathbb{R}^d$. This leads to the phenomenon of *score explosion*: close to $\mathcal{M}$, the score $s_t(x)$ diverges to infinity due to nondifferentiability across $\mathcal{M}$ (Bortoli, 2023; Pidstrigach, 2022).

Recently, several theoretical works have developed precise rates for this score explosion (Bortoli, 2023). While pursuing the analysis of score explosion, multiple works have independently observed that, up to scaling, the score behaves like a *projection operator onto the manifold* $\mathcal{M}$ (Brosse & Dalalyan, 2026; Huang et al., 2026; Permenter & Yuan, 2024). This is true even when the SBDM has not been given explicit knowledge of the manifold's properties, e.g., through physics-informed loss terms (Bastek et al., 2024). Stanczuk et al. (2024) proposed an algorithm for extracting the intrinsic dimension m from the score field. However, to our knowledge, *no work has converted this theoretical observation into an algorithm for reliably learning physical laws from data.*

In this work, we investigate using SBDMs *not* for their traditional role as samplers, but instead as a tool for *identifying governing physics*, by harnessing known divergence near constraint surfaces. We show that following the score field naturally unveils the location of the constraint manifold $\mathcal{M}$, and thus underlying governing equations. Based on this, we introduce our novel algorithm, SCoREMaP (Score-based Constraint Residual Extraction and Manifold Projection), for pointwise

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

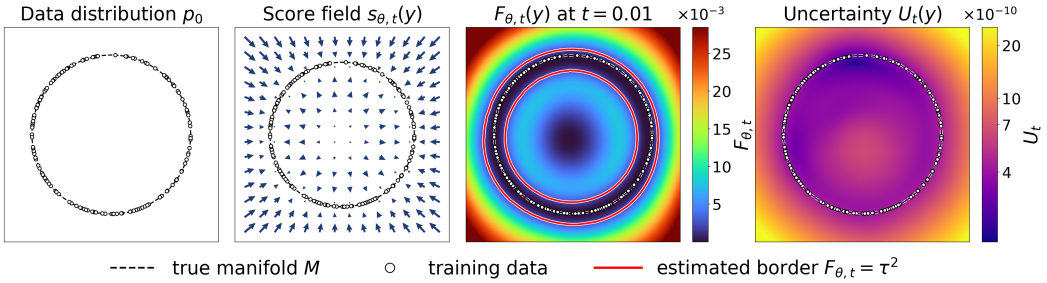


Figure 1: **Using SCoREMaP to recover a manifold with uncertainty quantification.** Demonstration of SCoREMaP on the unit circle $\mathcal{M} = S^1$ ($m = 1, d = 2$). **(Leftmost plot)** 150 samples from the data distribution p_0. **(Second plot)** We train a diffusion model on 10,000 samples from p_0, yielding a learned score field $s_{\theta,t}(y)$. We plot this score field at $t = 0.01$. Notice that the score field points orthogonally in the direction of $\mathcal{M}$. **(Third plot)** SCoREMaP postprocesses $s_{\theta,t}$ to estimate how far any point y is from $\mathcal{M}$. Dark blue indicates that y is close to $\mathcal{M}$ while dark red indicates that y is far. **(Rightmost plot)** SCoREMaP also produces uncertainty estimates on the distance from y to $\mathcal{M}$. While the third plot shows that SCoREMaP thinks that the middle of the circle might be on the manifold (dark blue), the fourth plot shows that SCoREMaP is uncertain in this claim.

detection of $\mathcal{M}$. Theoretical results prove that our procedure achieves localization of the underlying true constraints, even with no physics information injected into the training procedure. Moreover, using mature techniques in epistemic uncertainty quantification (UQ) (Daxberger et al., 2022), we propagate parameter posteriors through SCoREMaP. This allows practitioners to query points to determine constraint satisfaction and simultaneously obtain associated UQ estimates.

We illustrate our approach schematically in Figure 1. In particular, the quiver plot depicting the score field $s_{\theta,t}(y)$ intuitively depicts the projection-like behavior of diffusion models at small t: we see that the arrows of the vector field point precisely in the direction of the underlying constraint manifold. SCoREMaP harnesses this known insight and converts $s_{\theta,t}(y)$ into a scalar field $F_{\theta,t}(y)$ that is easy to analyze by practitioners. Finally, if we interpret diffusion model parameters θ as a realization of some underlying posterior distribution, as in Daxberger et al. (2022), we may derive a UQ metric $U_t(y)$, which can be similarly visualized in the rightmost plot. We highlight that even when $F_{\theta,t}(y)$ misclassifies the center of the circle as being close to the manifold (an issue we avoid by using held-out samples), $U_t(y)$ correctly assigns high uncertainty to this region.

Our precise contributions are as follows:

1. We present SCoREMaP, a **theoretically-justified algorithm for robust constraint localization** from finite queries, assuming no prior domain knowledge about the problem.
2. We apply **parametric epistemic UQ** via well-developed Laplace approximation theory to provide propagated uncertainty estimates.
3. We **demonstrate on a suite of realistic scientific tasks** that even cheaply trained score fields contain enough information to yield **meaningful improvements over data-only baselines.**

2 Related Work

The overarching goal of SCoREMaP is to bridge existing theory into an actionable procedure for physical constraint localization with UQ that may be used by scientific practitioners. To that end, we briefly summarize the most relevant prior work here.

Score explosion rates. At their core, SBDMs generally operate by following samples from a standard Gaussian distribution along trajectories induced by the score field $s_t(x)$ backwards in time. The idea is that as $t \to 0$, the score field $s_t(x)$ pushes samples towards the underlying data density. Studying the behavior of $s_{\theta,t}(x)$ at small t has gained substantial interest for understanding the sampling dynamics of diffusion models (Biroli et al., 2024; Hirono et al., 2024; Chen et al., 2024). The specialization to the manifold-constrained case has also gained substantial interest and theoretical

coverage (Pidstrigach, 2022; Bortoli, 2023; Huang et al., 2026). However, these works analyze score fields for the purpose of enhanced sampling and do not focus on downstream manifold extraction.

Non-generative applications of scores. Scores have found utility in many other areas of computational mathematics and scientific computing beyond the generative AI domain. Some notable works include dimension reduction (Brennan et al., 2022), Bayesian inference (Chandramoorthy et al., 2024), and numerical time-stepping in PDE solvers (Gal et al., 2026). In fact, the relationship between scores and data support residuals was also observed in earlier work from the denoising autoencoder literature (Alain & Bengio, 2014). This suggests the ripe potential of trained score models as a tool beyond sampling.

Learning manifold geometry from score fields. While, to our knowledge, no other work directly aims to extract physical constraint residuals from data using diffusion models, the notion of using diffusion models to elicit insights into the underlying geometry of data manifolds has been studied. The most similar work to ours uses rank-deficient second-order score field behavior to produce an actionable algorithm for estimation of intrinsic dimension m (Stanczuk et al., 2024). Manifold intrinsic dimension estimation can also be accomplished by studying the Fokker-Planck equation, which governs the dynamics of diffusion models (Kamkari et al., 2024). Other works aim to extract manifold geometry information, but focus on cases that do not necessarily induce a strict equality-constraint formulation of the support manifold and more on cases where there exists some predominantly low-dimensional structure, such as in image generation (Saito & Matsubara, 2026; Yun et al., 2024; Kadkhodaie et al., 2024).

3 Our Approach: SCoREMaP

In this section, we first establish a common vocabulary for diffusion modeling and basic manifold geometry. We then provide some theoretical motivation before formally presenting SCoREMaP. Finally, we discuss expected theoretical behavior and formalize a notion of UQ for SCoREMaP.

3.1 Preliminaries

To ensure that our paper may be understood in a self-contained manner, we briefly introduce terms needed to understand the kinds of constraints we consider.

Definition 3.1 (Nearest-Point Projection Π)**.** *The* nearest-point projector $\Pi : \mathbb{R}^d \to \mathcal{M}$ *projects ambient points* $x \in \mathbb{R}^d$ *to their nearest point on* $\mathcal{M}$*, i.e.* $\Pi(x) \in \arg\min_{y\in\mathcal{M}} \|x - y\|_2$*. Crucially,* Π *need not be unique for general nonlinear manifolds, hence the* $\in$ *notation.*

Definition 3.2 (Reach (Aamari et al., 2019))**.** *The* reach *is the largest* $r \in \mathbb{R}$ *such that for all* x *such that* $\text{dist}(x, \mathcal{M}) < r$*, the nearest-point projection* Π *is unique.*

In this work, we focus strictly on the case where $\mathcal{M}$ may be represented as a zero level set of some function $h(x), h : \mathbb{R}^d \to \mathbb{R}^{d-m}$. We interpret $h(x)$ to be the constraints defining the physics of a problem, in that every point x that adheres to the constraints satisfies $h(x) = 0$. Examples include allowable molecular geometry configurations, PDEs, or other conservation laws (Feng et al., 2024; Xu et al., 2026; Christopher et al., 2024; Bastek et al., 2024; Keegan & Ruthotto, 2026). We also specify that h must have full-rank Jacobians on $\mathcal{M}$, i.e., $\text{rank}(Dh(x)) = d - m$ for all $x \in \mathcal{M}$.

Using this vocabulary, we can define the setting we focus on:

Let $\mathcal{M}$ be an m-dimensional $\mathcal{C}^4$ manifold embedded in $\mathbb{R}^d$ which is closed, without boundary, and with reach $r > 0$ (Aamari et al., 2019), where $m < d$. Assume data $x_i \in \mathbb{R}^d$ are sampled from $x_i \sim p_0(x)$, such that

$$p_0(dz) = \rho(z) d\text{vol}_{\mathcal{M}}(z) \quad \text{and} \quad \text{supp}(p_0) = \mathcal{M},$$

where $\rho(z)$ is positive and $\mathcal{C}^3$.

The $r > 0$ requirement guarantees the existence of a tubular neighborhood $\mathcal{U}$ (the *reach tube*) around $\mathcal{M}$ such that the nearest-point projection Π is uniquely defined.

Diffusion models can be formulated as the linear Gaussian corruption $X_t = \alpha_t X_0 + \sigma_t Z$, where $X_0 \sim p_0$, $Z \sim \mathcal{N}(0, I)$, $\alpha_t, \sigma_t > 0$ for $t > 0$ and $\alpha_t^2 + \sigma_t^2 = 1$ (Ho et al., 2020). Our experiments use the standard linear schedule from Song et al. (2021) where $\beta(s) = \beta_{\text{min}} + s(\beta_{\text{max}} - \beta_{\text{min}})$, $\alpha_t = e^{-\frac{1}{2}\int_0^t \beta(s)ds}$, where $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 20$.

Let p_t denote the density associated with the law of X_t. Define the *score* as

$$s_t(x) := \nabla_x \log p_t(x)$$

and the *(population) projection estimator* as

$$\hat{\Pi}_t(x) := x + \sigma_t^2 s_t(x).$$

3.2 SCoREMaP

We now briefly discuss the key theoretical observation from the diffusion modeling literature (Huang et al., 2026; Brosse & Dalalyan, 2026) that we use to formulate SCoREMaP. Specifically, under certain conditions, $\hat{\Pi}_t$ converges to an oracle projection, *even though $\mathcal{M}$ itself is not known*.

Lemma 3.1 (Convergence to Projection-Like Behavior, (Prop. 3.1 Huang et al., 2026))**.** *Define $U_{\frac{r}{2}} := \{x : \text{dist}(x, \mathcal{M}) < \frac{r}{2}\}$. Then, for each $x \in U_{\frac{r}{2}}$, we have*

$$\hat{\Pi}_t(x) = \Pi(x) + \mathcal{O}_x(\sigma_t^2).$$

Since we have $\sigma_t^2 \to 0$ as $t \to 0$, the projection estimator recovers the true projector in the small-t limit. We verify the connection between Lemma 3.1 and Huang et al. (2026) formally in Appendix A. We use this in order to recover an approximate distance between any point and $\mathcal{M}$. Letting $\text{dist}(y, \mathcal{M}) = \|\Pi(y) - y\|_2$, we define the oracle *manifold residual field*

$$F_0(y) := \|\Pi(y) - y\|_2^2 = \text{dist}(y, \mathcal{M})^2.$$

The *(population) residual field* is

$$F_t(y) := \|\hat{\Pi}_t(y) - y\|_2^2 = \|\sigma_t^2 s_t(y)\|_2^2.$$

Finally, define the *learned residual field* as the approximation to the population field using $s_{\theta,t}(y)$:

$$F_{\theta,t}(y) := \|\sigma_t^2 s_{\theta,t}(y)\|_2^2.$$

Let $y \in U_{\frac{r}{2}}$. Then, the reverse triangle inequality combined with Lemma 3.1 yields

$$\left|\sqrt{F_t(y)} - \sqrt{F_0(y)}\right| \leq \|\hat{\Pi}_t(y) - \Pi(y)\|_2 = \mathcal{O}_y\left(\sigma_t^2\right).$$

Recalling that $\sigma_t \to 0$, this shows that $F_t(y) \to F_0(y)$ as $t \to 0$. This suggests a natural algorithm for extracting $\mathcal{M}$: simply query the score field at some probe points, scale the output in a manner suggested by existing theoretical results, take the norm, and use the resulting scalar field for manifold extraction purposes; we investigate exactly how this may be incorporated into downstream constraint discovery in Appendix F. We formalize this procedure as the SCoREMaP algorithm below. Algorithm 1 empirically evaluates and thresholds discrete query points $\mathcal{Y} \subset U_{\frac{r}{2}}$, where U is a continuous compact region within the reach tube of $\mathcal{M}$ (Aamari et al., 2019).

To theoretically analyze this procedure, we define the thresholded set $\hat{A}_{\theta,t}^U(\tau) = \{y \in U : F_{\theta,t}(y) \leq \tau^2\}$. This set covers all points that "seem close to the manifold" according to the score function. Our core theoretical result shows that, as the projection estimator $\hat{\Pi}_t(x)$ approaches the true projection $\Pi(x)$, the tresholded set correctly recovers the manifold $\mathcal{M}$.

Theorem 3.1 (SCoREMaP recovery error)**.** *Let $\mathcal{M} \subset U \subseteq \mathcal{U}$. If $\sup_{y \in U} \|\hat{\Pi}_{\theta,t}(y) - \Pi(y)\|_2 \leq \varepsilon$, then $\forall \tau \geq \varepsilon$, we have $\mathcal{M} \subseteq \hat{A}_{\theta,t}^U(\tau)$. Moreover, letting d_H denote Hausdorff distance, we have*

$$d_H(\mathcal{M}, \hat{A}_{\theta,t}^U(\tau)) \leq \tau + \varepsilon.$$

We prove Theorem 3.1 in Appendix A. To connect Theorem 3.1 and our implemented procedure, note that Algorithm 1 may be understood as returning a discrete approximation to $\hat{A}_{\theta,t}$ under finite queries, i.e. $\hat{A}_{\theta,t}^{\mathcal{Y}}(\tau) = \mathcal{Y} \cap \hat{A}_{\theta,t}^U(\tau)$.

Algorithm 1 SCoREMaP: Manifold Localization Up To Threshold τ

Require: Score model s_θ, query points $\mathcal{Y} \subset \mathbb{R}^d$, sufficiently small t, threshold τ

1: **for** $y \in \mathcal{Y}$ **do**
2: $\quad \hat{\Pi}_{\theta,t}(y) \leftarrow y + \sigma_t^2 s_{\theta,t}(y)$
3: $\quad F_{\theta,t}(y) \leftarrow \|\hat{\Pi}_{\theta,t}(y) - y\|^2$
4: **end for**
5: $\hat{A}^{\mathcal{Y}}_{\theta,t}(\tau) = \{y \in \mathcal{Y} : F_{\theta,t}(y) \leq \tau^2\}$ ▷ Optional: thresholding for $\mathcal{M}$ localization
6: **return** $\hat{A}^{\mathcal{Y}}_{\theta,t}(\tau), F_{\theta,t}$

3.3 Propagated Epistemic Uncertainty Quantification

Algorithm 1 deterministically returns $F_{\theta,t}(x)$ from $s_\theta(x)$. The theoretical results developed in the previous section show that, if the learned score field sufficiently approximates the true population score field, then the induced projection estimator is accurate. In turn, this implies that our SCoREMaP algorithm localizes the manifold. However, uncertainty in the learned model parameters also induces uncertainty in the recovered score field. Thus, to quantify this source of uncertainty, we seek epistemic uncertainty estimates for SCoREMaP.

To formalize this, we treat the learned parameters θ as a realization from a Gaussian posterior distribution π^n. That is, we let $\mathcal{D} = \{x_i\}_{i=1}^n$ represent the training data and define the *displacement* as $\Delta_{\theta,t}(y) := \hat{\Pi}_{\theta,t}(y) - y = \sigma_t^2 s_{\theta,t}(y)$ and remark that $F_{\theta,t}(y) = \|\Delta_{\theta,t}(y)\|_2^2$. For learned parameters $\hat{\theta}$, write

$$\mu_t(y) = \sigma_t^2 s_{\hat{\theta},t}(y) \quad \text{and} \quad G_t(y) = \sigma_t^2 \nabla_\theta s_{\theta,t}(y)|_{\theta=\hat{\theta}}.$$

Specifically, we assume that the posterior distribution for θ is Gaussian with mean $\mu_t(y)$ and covariance $\Sigma_{\theta,n}$, where $\Sigma_{\theta,n}$ represents the posterior covariance of the neural network parameters θ on our n samples. Linearizing with respect to θ around $\hat{\theta}$ and truncating higher-order terms yields

$$\Delta_{\theta,t}(y) \approx \mu_t(y) + G_t(y)(\theta - \hat{\theta}),$$

and propagating the Gaussian uncertainty distribution over θ into the displacement vector yields

$$\Delta_{\theta,t}(y)|\mathcal{D} \sim \mathcal{N}(\mu_t(y), C_t(y)) \quad \text{and} \quad C_t(y) := G_t(y)\Sigma_{\theta,n}G_t(y)^\top.$$

Expanding the posterior scalar field $\mathbb{E}_\theta\left[F_{\theta,t}(y)\right]$ yields

$$\mathbb{E}_\theta\left[\|\Delta_{\theta,t}(y)\|_2^2\right] = \|\mu_t(y)\|_2^2 + \text{tr}C_t(y).$$

The first term is the maximum a posteriori (MAP) displacement and is large when y is far from $\mathcal{M}$. The second is the epistemic variance, which is large when independent draws of θ from the posterior distribution lead to disagreement about if y is close to $\mathcal{M}$. We therefore use the second term, tr$C_t(y)$, as an uncertainty metric, giving $U_t(y) = \text{tr}C_t(y)$. This discussion yields Algorithm 2 for pointwise UQ analysis.

Note that this output can easily be computed without ever forming dense matrices (Daxberger et al., 2022). We use a last-layer Laplace approximation of the score field for cheap queries (Kristiadi et al., 2020). Precise UQ implementation details may be found in Appendix E.

Algorithm 2 Pointwise Uncertainty Quantification

Require: Score model $s_{\hat{\theta}}$, probe point $y \in \mathbb{R}^d$, sufficiently small t

1: Compute $G_t(y) \leftarrow \sigma_t^2 \nabla_\theta s_{\theta,t}(y)|_{\theta=\hat{\theta}}$, e.g., using automatic differentiation
2: Propagate parameter uncertainty to obtain output covariance: $C_t(y) \leftarrow G_t(y)\Sigma_{\theta,n}G_t(y)^\top$.
3: Compute $U_t(y) \leftarrow \text{tr}C_t(y)$
4: **return** $U_t(y)$

4 RESULTS

We now present empirical results of SCoREMaP on a suite of geometric and scientifically relevant tasks. The governing equations we study can be summarized categorically as *geometric constraints* (samples lie on tori or products of spheres), *matrix orthogonality constraints* (samples belong to the Stiefel manifold), *molecular geometry constraints* (samples adhere to known bond length and bond angle requirements), and *PDE constraints* (sample dynamics conform to the 1-D Burgers' equation).

We compare against multiple data-only baselines that require no training. This is done to clearly investigate the benefit of introducing a trained score model for manifold localization. For all non-PDE tasks, we compare against nearest-neighbor (NN) and k-NN distance baselines motivated by similar work (Awasthi et al., 2025; Schmitz, 2024; Alves & Yamaguchi, 2022), and we are also inspired by several similar local PCA baselines in a variety of scientific applications (Cain, 2026; Zdybał et al., 2023). We also use a baseline inspired by signed distance function (SDF) parametrizations of level sets (Liu et al., 2021), as well as standard kernel density estimation (KDE) (Chen, 2017). For the PDE task, we compare against Sparse Identification of Nonlinear Dynamical Systems (SINDy) (Brunton et al., 2016). We ablate over multiple hyperparameters and report only best-per-metric results to promote fairness to baselines. All experimental setup is described in Appendix C.

To evaluate the utility of SCoREMaP in manifold localization, we compare oracle tube inclusion for threshold-induced metrics such as IoU, AUROC, and FarFPR. To study its utility as a true surrogate residual norm for downstream extraction tasks, we present a fine-grained study of the observed correlations between $F_{\theta,t}(y)$ and the underlying ground-truth squared-distance $\text{dist}(y, \mathcal{M})^2$ in Figure 2. All $\pm$-windows indicate mean and standard deviations over ten seeds. We summarize our results according to the main empirical findings here.

Table 1: Pooled metrics over 90 task-seed evaluations (nine total non-molecular, non-PDE tasks, ten seeds). Columns titled "___ Wins" indicate the percentage of comparisons in which we strictly outperform the specified baseline for that row. We observe that we always outperform data-only baselines on all tasks with underlying sphere product and Stiefel manifold support.

Method	IoU↑	AUROC↑	FarFPR↓	IoU Wins	AUROC Wins	FarFPR Wins
NN balls	0.16	0.69	0.35	1.00	1.00	1.00
kNN RMS	0.22	0.76	0.18	1.00	1.00	1.00
LPCA (TwoNN)	0.25	0.74	0.28	1.00	1.00	1.00
LPCA (oracle dim.)	0.29	0.80	0.22	1.00	1.00	1.00
KDE	0.24	0.74	0.18	1.00	1.00	1.00
MLS	0.25	0.75	0.28	1.00	1.00	1.00
SCoREMaP (ours)	**0.87±0.08**	**1.00±0.01**	**0.01±0.03**	–	–	–

We frequently outperform data-only baselines. We first present pooled performance metrics across all non-PDE tasks and ten seeds for each task in Table 1, where results are compared against the oracle tube $\text{dist}(y, \mathcal{M}) \leq 0.04$. To make results favorable to the baselines, we present best-per-metric results over a collection of hyperparameters; this ablation is described in detail in Appendix B. All results for our approach for manifold localization are at fixed $\sigma_t = 0.05$, with similar ablation as for the baselines to identify an optimal choice of τ on a set of held-out validation probes. Our approach almost universally outperforms extensively tuned, training-data-only, general-purpose baselines. This shows that with no additional assumptions on the underlying data constraints and dynamics, our method *routinely outperforms baselines* with similarly few inductive biases. Even when O(3)-invariance is encoded into KDE and NN baselines as well for the molecular geometry tasks presented in Table 2, we still routinely outperform or are competitive against baselines.

Limited sensitivity to tested hyperparameters. Despite nonuniformity in scale across the different tasks, uniformly taking $\sigma_t = 0.05$ for manifold localization *performs similarly well across all tasks*, and *we use largely identical architectures with no per-task hyperparameter tuning* in all of our experiments (except for the O(3)-equivariant networks for all molecular tasks). This suggests that despite the introduction of a trained neural network, SCoREMaP has limited sensitivity to the archi-

Table 2: Performance on molecular geometry tasks. We compare against both raw baselines as well as O(3)-invariant versions using Gram matrix embeddings of samples. We routinely outperform naive baselines and have equal or competitive performance against O(3)-invariant baselines.

Task	Method	IoU↑	AUROC↑	FarFPR↓
Freely jointed chain	NN balls	0.36	0.90	0.10
	NN balls (O(3)-inv.)	0.95	**1.00**	**0.00**
	kNN RMS	0.39	0.90	0.03
	LPCA (TwoNN)	0.43	0.89	0.16
	KDE	0.38	0.91	**0.00**
	KDE (O(3)-inv.)	0.95	**1.00**	**0.00**
	MLS	0.42	0.89	0.15
	SCoREMaP (ours)	**0.97±0.00**	**1.00±0.00**	**0.00±0.00**
Fixed-angle chain	NN balls	0.14	0.85	0.11
	NN balls (O(3)-inv.)	0.90	**1.00**	**0.00**
	kNN RMS	0.17	0.88	0.05
	LPCA (TwoNN)	0.34	0.95	0.05
	KDE	0.14	0.87	0.09
	KDE (O(3)-inv.)	**0.93**	**1.00**	**0.00**
	MLS	0.35	0.95	0.06
	SCoREMaP (ours)	0.90±0.01	**1.00±0.00**	**0.00±0.00**
Three-arm star	NN balls	0.17	0.74	0.24
	NN balls (O(3)-inv.)	0.96	**1.00**	**0.00**
	kNN RMS	0.22	0.81	0.09
	LPCA (TwoNN)	0.23	0.75	0.29
	KDE	0.21	0.79	0.06
	KDE (O(3)-inv.)	0.97	**1.00**	**0.00**
	MLS	0.22	0.75	0.26
	SCoREMaP (ours)	**0.98±0.01**	**1.00±0.00**	**0.00±0.00**
Fixed-angle chain II	NN balls	0.16	0.89	0.11
	NN balls (O(3)-inv.)	**0.92**	**1.00**	**0.00**
	kNN RMS	0.21	0.90	0.07
	LPCA (TwoNN)	0.38	0.96	0.03
	KDE	0.18	0.90	0.06
	KDE (O(3)-inv.)	**0.92**	**1.00**	**0.00**
	MLS	0.37	0.97	0.03
	SCoREMaP (ours)	**0.92±0.00**	**1.00±0.00**	**0.00±0.00**

tectural choices we make in this study. This suggests the suitability of SCoREMaP out-of-the-box for constraint identification.

Outperformance of baselines even for cheaply trained models. While all results in this section are presented after a fixed number of optimization steps (20,000), we outperform data-only baselines before this stopping criterion is reached on several tasks. All models are trained on NVIDIA A100 GPUs, and the model for the longest-running task, the PDE task, trains in less than three minutes. This suggests that even inexpensively trained models can outperform baselines for physics detection.

Strong statistical correlation between $F_{\theta,t}(y)$ and dist$(y, \mathcal{M})^2$. To support our claim that $F_{\theta,t}$ reliably learns constraint residuals dist$(y, \mathcal{M})^2$ without any domain-informed prior knowledge, Figure 2 presents correlations between dist$(y, \mathcal{M})^2$ and $F_{\theta,t}$ on held-out data samples on a single seed. Notably, across all tasks, we observe nontrivial correlation across many orders of magnitude in true error, further suggesting that SCoREMaP can be a useful surrogate for dist$(y, \mathcal{M})^2$.

Behavior of UQ metric $U_t(y)$. A useful UQ metric should indicate low uncertainty for oracle-positive points within a tube around $\mathcal{M}$ and its boundary, and high uncertainty for points that are far away from the manifold. This is precisely what we observe in Figure 2. While there are some unexpected artifacts at $S^4 \times S^4$, S^8, and the Stiefel manifold tasks (where some far-away points

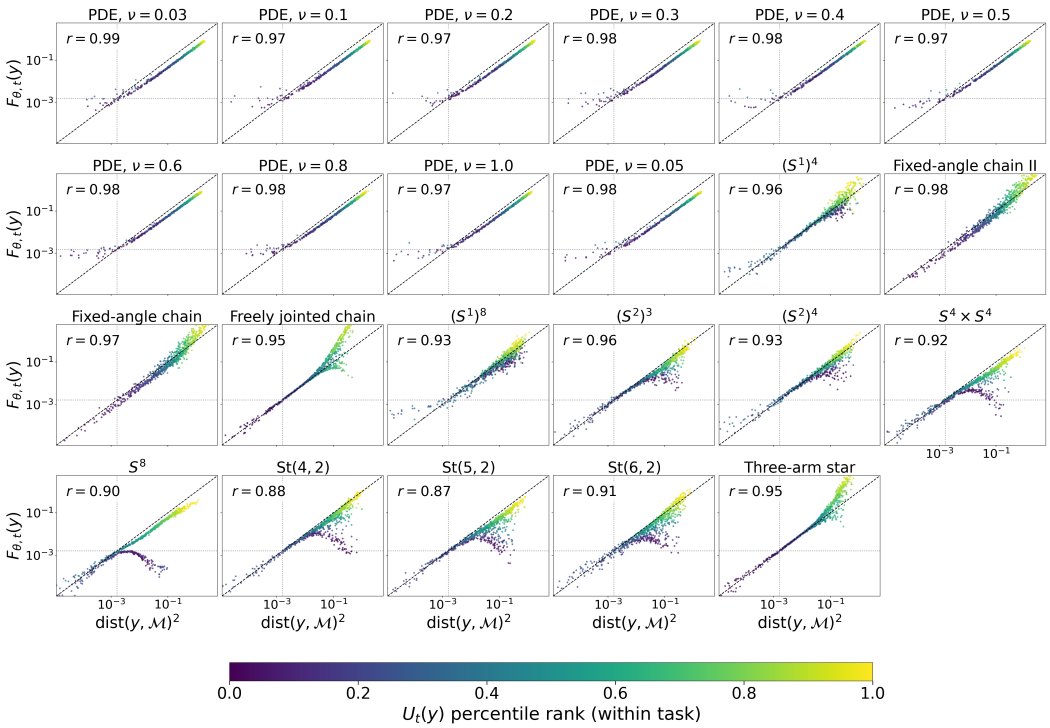


Figure 2: $F_{\theta,t}(y)$ versus $\text{dist}(y, \mathcal{M})^2$, color-coded by $U_t(y)$ percentiles across all 23 tasks. We observe a strong Pearson correlation r between $F_{\theta,t}$ and $\text{dist}(y, \mathcal{M})^2$ across all tasks. Moreover, U_t correctly classifies uncertainty across the majority of tasks: points far from $\mathcal{M}$ are classified as high-uncertainty, while points close to $\mathcal{M}$ are assigned higher confidence.

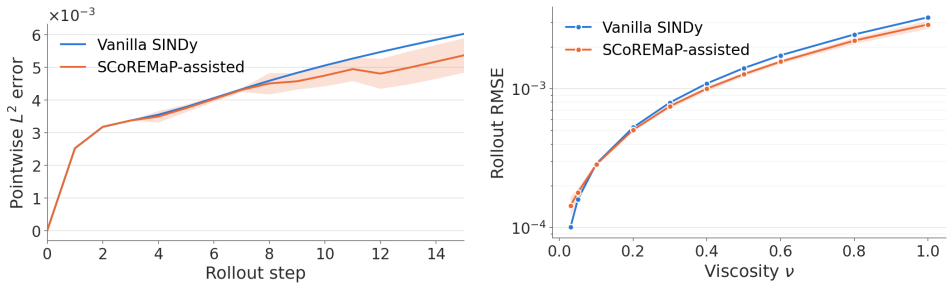
are mistakenly classified as being close to $\mathcal{M}$ with high certainty), we largely observe the expected behavior across almost all tasks. In particular, this alignment is almost perfect for the PDE tasks across all ν and the molecular geometry tasks.

Limitations. While it is expected by Theorem 3.1 that our method recovers constraints in the limit, it is not uncommon for training instability to arise as a result of score field degeneracy (Keegan & Ruthotto, 2026); therefore, score fields may not always achieve sufficient accuracy required at small t to ensure pointwise manifold localization. We anticipate regularization strategies (Baptista et al., 2025) could combat this, and also imagine that follow-up work could include an investigation into the connection between the known training dynamics for diffusion models on manifold-supported distributions (Bonnaire et al., 2025; He et al., 2026; Huang et al., 2026; Shen et al., 2026), as well as the optimal selection of τ and σ_t under minimal data assumptions. We also only compare against non-trained baselines. This choice was deliberately made to reduce hyperparameter complexity and promote fairness of comparison among baselines. Future work could study the performance of SCoREMaP against or in concert with trained techniques for constraint extraction (Liu & Tegmark, 2021; Ha & Jeong, 2021; Kaheman et al., 2020).

4.1 SCoREMaP for Improved PDE Simulation

We anticipate that SCoREMaP may be a useful data-driven tool for addressing the failure modes of traditional solvers, as well as governing-equation discovery techniques such as SINDy. To investigate how SCoREMaP can be used in concert with such techniques, we study how the learned projector $\hat{\Pi}_{\theta,t}$ can be used for correcting numerical simulations given a candidate physical law.

While raw SINDy works well in terms of overall trajectory RMSE on simple problems, numerical PDE solution techniques tend to have challenges due to compounding error over time rollout. To



(a) Per-step error.

(b) Summary over various ν.

Figure 3: SCoREMaP-assisted simulation using a SINDy-derived governing equation. In 3a, we observe an improvement in accuracy at fixed $\nu = 0.3$, while in 3b, we observe that this performance improvement is consistent with increasing viscosity.

combat this failure mode, we experiment with using $\hat{\Pi}_{\theta,t}$ for intermittent data-driven projection using an algorithm precisely defined in Appendix F.

We find that periodically projecting to the estimated manifold using SCoREMaP enhances vanilla SINDy's accuracy and appreciably reduces error later in rollout. We present results averaged across 16 initial conditions at $\nu = 0.3$ in Figure 3 and additionally ablate this result over ten possible ν in Appendix F, suggesting that our approach indeed has downstream applicability in error reduction in large-scale simulations. While vanilla SINDy is favorable at low ν, this result is unsurprising due to the known challenge of numerically solving Burgers' equation at low viscosity (Seydaoğlu, 2018). We also note that our proposed algorithm can be applied to any candidate numerical autoregressive simulation and is not tied to SINDy in particular. We anticipate that a follow-up study could use $F_{\theta,t}$ for combating known failure modes of certain PDE solvers in high-dimensional settings.

5 Conclusions

We present SCoREMaP, an algorithm for harnessing the projection-like behavior of score fields near manifolds for constraint identification. We present theory from the diffusion modeling literature which clearly motivates SCoREMaP, and we demonstrate both theoretically and empirically that SCoREMaP offers the manifold localization behavior that we expect. We also propagate epistemic UQ over neural network parameters to estimate uncertainty associated with pointwise queries. We anticipate that SCoREMaP will be highly useful in post hoc exploratory cases, where our approach for manifold localization could be used for downstream fitting of parametric governing equations.

In particular, the correlations we present in Figure 2 demonstrate that our method could be highly useful in downstream tasks where residuals are used for fitting reusable governing laws to data. Our results also suggest potential utility of our approach in scientific and engineering simulations where function evaluations are expensive to compute, offering even greater utility for our approach as a potential surrogate for the residual $\text{dist}(y, \mathcal{M})^2$ with the added benefit of uncertainty quantification.

We highlight that our baselines are based on widely used, classical machine learning techniques. We make this choice deliberately to avoid any biases that arise with tailored (possibly trained) methods in the literature. Moreover, we emphasize that domain experts *need not have knowledge of complex neural network architectures or domain-tailored techniques:* with minimal assumptions, tailoring, or compute, our method reliably recovers constraint regions across diverse tasks.

In future work, we are interested in whether $U_t(y)$ could be used for a manifold extraction variant, e.g., through a two-gate mechanism where points are thresholded according to both $F_{\theta,t}$ and U_t. Such a procedure could enable uncertainty-aware constraint extraction. We also anticipate that SCoREMaP can be combined with active learning techniques to propose an uncertainty-aware data procurement method for AI-assisted data acquisition (Settles, 2009). This would allow practitioners to tailor their limited experimental resources towards reducing uncertainty in specific regions, allowing for refined downstream physics detection and manifold localization.

AI USE STATEMENT

AI and LLM tools (Codex, Claude) were used in a collaborative manner for code development and discussion of research ideas. The authors take full responsibility for all ideas put forth in this paper.

ETHICS STATEMENT

Our work uses only synthetic data with underlying known manifolds for evaluation. Due to this, we do not anticipate any ethical conflicts of interest due to dataset usage. We have written this work to contribute to the broader community goal of harnessing AI for discovery science, a scientific aim with great anticipated societal benefit.

REPRODUCIBILITY STATEMENT

Our code implements all algorithms presented in this paper in a manner that is flexible for user datasets of interest and includes ablations using multiple random seeds. Our code also precisely presents experiment-to-figure mappings to make it clear how each figure in the paper was generated. We will present the code upon acceptance.

REFERENCES

Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the Reach of a Manifold, April 2019. URL http://arxiv.org/abs/1705.04565. arXiv:1705.04565 [math].

Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *The Journal of Machine Learning Research*, 15(1):3563–3593, January 2014. ISSN 1532-4435. URL https://dl.acm.org/doi/10.5555/2627435.2750359.

Alexandre Alves and C. H. Yamaguchi. Reconstruction of missing resonances combining nearest neighbors regressors and neural network classifiers. *The European Physical Journal C*, 82(8): 746, August 2022. ISSN 1434-6052. doi: 10.1140/epjc/s10052-022-10714-1. URL https://doi.org/10.1140/epjc/s10052-022-10714-1.

Anshul Awasthi, Sujaya Grace, Ramakant Yadav, Lalita Bhanu Murthy Neti, and Surya Shankar Dan. Physics-Informed Data-Driven KNN Methodology of Modeling NCFETs for Superior Circuit Performance. *IEEE Transactions on Electron Devices*, 72(11):5878–5884, November 2025. ISSN 1557-9646. doi: 10.1109/TED.2025.3611913. URL https://ieeexplore.ieee.org/document/11192768.

Ricardo Baptista, Agnimitra Dasgupta, Nikola B. Kovachki, Assad Oberai, and Andrew M. Stuart. Memorization and Regularization in Generative Diffusion Models, March 2025. URL http://arxiv.org/abs/2501.15785. arXiv:2501.15785 [cs.LG].

Jan-Hendrik Bastek, WaiChing Sun, and Dennis M. Kochmann. Physics-Informed Diffusion Models, May 2024. URL http://arxiv.org/abs/2403.14404. arXiv:2403.14404.

Giulio Biroli, Tony Bonnaire, Valentin de Bortoli, and Marc Mézard. Dynamical regimes of diffusion models. *Nature Communications*, 15(1):9957, November 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-54281-3. URL https://www.nature.com/articles/s41467-024-54281-3.

Tony Bonnaire, Raphaël Urfin, Giulio Biroli, and Marc Mezard. Why Diffusion Models Don't Memorize: The Role of Implicit Dynamical Regularization in Training. October 2025. URL https://openreview.net/forum?id=BSZqpqgqM0.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis, May 2023. URL http://arxiv.org/abs/2208.05314. arXiv:2208.05314 [stat].

Michael Brennan, Ricardo Baptista, and Youssef Marzouk. Dimension reduction via score ratio matching. November 2022. URL https://openreview.net/forum?id=YAN97j2NmGT.

Nicolas Brosse and Arnak Dalalyan. *Boundary-layer asymptotics for Gaussian-smoothed singular measures*. July 2026.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, April 2016. doi: 10.1073/pnas.1517384113. URL https://www.pnas.org/doi/10.1073/pnas.1517384113.

Jericho Cain. Manifold learning for source separation in confusion-limited gravitational-wave data. *Classical and Quantum Gravity*, 43(13):135031, July 2026. ISSN 0264-9381. doi: 10.1088/1361-6382/ae83ff. URL https://doi.org/10.1088/1361-6382/ae83ff.

Gustau Camps-Valls, Andreas Gerhardus, Urmi Ninad, Gherardo Varando, Georg Martius, Emili Balaguer-Ballester, Ricardo Vinuesa, Emiliano Diaz, Laure Zanna, and Jakob Runge. Discovering causal relations and equations from data. *Physics Reports*, 1044:1–68, December 2023. ISSN 0370-1573. doi: 10.1016/j.physrep.2023.10.005. URL https://www.sciencedirect.com/science/article/pii/S0370157323003411.

Nisha Chandramoorthy, Florian T. Schaefer, and Youssef M. Marzouk. Score Operator Newton transport. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, pp. 3349–3357. PMLR, April 2024. URL https://proceedings.mlr.press/v238/chandramoorthy24a.html.

Defang Chen, Zhenyu Zhou, Can Wang, Chunhua Shen, and Siwei Lyu. On the trajectory regularity of ODE-based diffusion sampling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *ICML'24*, pp. 7905–7934, Vienna, Austria, July 2024. JMLR.org.

Yen-Chi Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187, January 2017. ISSN 2470-9360. doi: 10.1080/24709360.2017.1396742. URL https://doi.org/10.1080/24709360.2017.1396742.

Austin Cheng, Alston Lo, Kin Lee, Santiago Miret, and Alán Aspuru-Guzik. *Stiefel Flow Matching for Moment-Constrained Structure Elucidation*. December 2024.

Chaoran Cheng, Boran Han, Danielle C. Maddix, Abdul Fatir Ansari, Andrew Stuart, Michael W. Mahoney, and Yuyang Wang. Gradient-free generation for hard-constrained systems. In *The Thirteenth International Conference on Learning Representations*, pp. 100510–100539, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/f9d7d6c695bc983fcfb5b70a5fbdfd2f-Abstract-Conference.html.

Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained Synthesis with Projected Diffusion Models, November 2024. URL http://arxiv.org/abs/2402.03559. arXiv:2402.03559 [cs].

Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace Redux – Effortless Bayesian Deep Learning, March 2022. URL http://arxiv.org/abs/2106.14806. arXiv:2106.14806 [cs.LG].

Elena Facco, Maria d'Errico, Alex Rodriguez, and Alessandro Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific Reports*, 7(1):12140, September 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-11873-y. URL https://www.nature.com/articles/s41598-017-11873-y.

Berthy T. Feng, Ricardo Baptista, and Katherine L. Bouman. Neural Approximate Mirror Maps for Constrained Diffusion Models, June 2024. URL http://arxiv.org/abs/2406.12816. arXiv:2406.12816.

Eshed Gal, Eldad Haber, and Uri Ascher. Score-Based Stabilization for Time-Dependent Problems, July 2026. URL http://arxiv.org/abs/2607.25119. arXiv:2607.25119 [cs.LG].

David González, Francisco Chinesta, and Elías Cueto. Learning non-Markovian physics from data. *Journal of Computational Physics*, 428:109982, March 2021. ISSN 0021-9991. doi: 10.1016/j.jcp.2020.109982. URL https://www.sciencedirect.com/science/article/pii/S0021999120307567.

Aditi Gupta, Raphael A. Meyer, Yotam Yaniv, Elynn Chen, and N. Benjamin Erichson. Quantifying Epistemic Uncertainty in Diffusion Models. February 2026. URL https://openreview.net/forum?id=2mAIPSwn1C.

Seungwoong Ha and Hawoong Jeong. Discovering invariants via machine learning. *Physical Review Research*, 3(4):L042035, December 2021. doi: 10.1103/PhysRevResearch.3.L042035. URL https://link.aps.org/doi/10.1103/PhysRevResearch.3.L042035.

J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, April 1982. ISSN 0033-8419. doi: 10.1148/radiology.143.1.7063747.

Derek Hansen, Danielle C. Maddix, Shima Alizadeh, Gaurav Gupta, and Michael W. Mahoney. Learning physical models that can respect conservation laws. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12469–12510. PMLR, 2023. URL https://proceedings.mlr.press/v202/hansen23b.html.

Ye He, Yitong Qiu, and Molei Tao. Diffusion Model's Generalization Can Be Characterized by Inductive Biases toward a Data-Dependent Ridge Manifold. May 2026. URL https://openreview.net/forum?id=58NKd4lWfI.

Yuji Hirono, Akinori Tanaka, and Kenji Fukushima. Understanding diffusion models by Feynman's path integral. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *ICML'24*, pp. 18324–18351, Vienna, Austria, July 2024. JMLR.org.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, December 2020. URL http://arxiv.org/abs/2006.11239. arXiv:2006.11239.

Wei Huang, Andi Han, Mingyuan Bai, Huanjian Zhou, Qixin Zhang, Taiji Suzuki, and Kenji Fukumizu. Provably Learning Diffusion Models under the Manifold Hypothesis: Collapse and Refine, May 2026. URL http://arxiv.org/abs/2605.20235. arXiv:2605.20235 [cs.LG].

Zahra Kadkhodaie, Florentin Guth, Eero P. Simoncelli, and Stéphane Mallat. Generalization in diffusion models arises from geometry-adaptive harmonic representations, April 2024. URL http://arxiv.org/abs/2310.02557. arXiv:2310.02557 [cs.CV].

Kadierdan Kaheman, J. Nathan Kutz, and Steven L. Brunton. Sindy-pi: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2242):20200279, 10 2020. ISSN 1364-5021. doi: 10.1098/rspa.2020.0279. URL https://doi.org/10.1098/rspa.2020.0279.

Hamidreza Kamkari, Brendan Leigh Ross, Rasa Hosseinzadeh, Jesse Cresswell, and Gabriel Loaiza-Ganem. A geometric view of data complexity: Efficient local intrinsic dimension estimation with diffusion models. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 38307–38354. Curran Associates, Inc., 2024. doi: 10.52202/079017-1211. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/43ab1646052dab79731f5d70bf40f6dc-Paper-Conference.pdf.

Katherine Keegan and Lars Ruthotto. Manifold-aware perturbations for constrained generative modeling. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=UPiyC9W4ms.

Agustinus Kristiadi, Matthias Hein, and Philipp Hennig. Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks, July 2020. URL http://arxiv.org/abs/2002.10118. arXiv:2002.10118 [stat.ML].

Shi-Lin Liu, Hao-Xiang Guo, Hao Pan, Peng-Shuai Wang, Xin Tong, and Yang Liu. Deep Implicit Moving Least-Squares Functions for 3D Reconstruction. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1788–1797, June 2021. doi:

10.1109/CVPR46437.2021.00183. URL https://ieeexplore.ieee.org/document/9578427/.

Ziming Liu and Max Tegmark. Machine Learning Conservation Laws from Trajectories. *Physical Review Letters*, 126(18):180604, May 2021. doi: 10.1103/PhysRevLett.126.180604. URL https://link.aps.org/doi/10.1103/PhysRevLett.126.180604.

Frank Permenter and Chenyang Yuan. Interpreting and improving diffusion models from an optimization perspective. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *ICML'24*, pp. 40461–40483, Vienna, Austria, July 2024. JMLR.org.

Jakiw Pidstrigach. Score-Based Generative Models Detect Manifolds, October 2022. URL http://arxiv.org/abs/2206.01018. arXiv:2206.01018 [stat].

Shinnosuke Saito and Takashi Matsubara. Extracting the Data Manifold from Diffusion Models via a Score-Based Non-Conformal Riemannian Metric. May 2026. URL https://openreview.net/forum?id=cwtcd7nGyu.

Tony Schmitz. Physics-informed KNN milling stability model with process damping effects. *Journal of Manufacturing Processes*, 120:1124–1129, June 2024. ISSN 1526-6125. doi: 10.1016/j.jmapro.2024.04.090. URL https://www.sciencedirect.com/science/article/pii/S1526612524004511.

Burr Settles. Active learning literature survey. 2009. URL https://api.semanticscholar.org/CorpusID:324600.

Muaz Seydaoğlu. An accurate approximation algorithm for Burgers' equation in the presence of small viscosity. *Journal of Computational and Applied Mathematics*, 344:473–481, December 2018. ISSN 0377-0427. doi: 10.1016/j.cam.2018.05.063. URL https://www.sciencedirect.com/science/article/pii/S0377042718303467.

Zebang Shen, Ya-Ping Hsieh, and Niao He. Manifold Generalization Provably Proceeds Memorization in Diffusion Models. March 2026. URL https://openreview.net/forum?id=O9StcESgB5#discussion.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations, February 2021. URL http://arxiv.org/abs/2011.13456. arXiv:2011.13456 [cs].

Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *ICML'24*, pp. 46412–46440, Vienna, Austria, July 2024. JMLR.org.

Utkarsh, Danielle C. Maddix, Ruijun Ma, Michael W. Mahoney, and Yuyang Wang. End-to-end probabilistic framework for learning with hard constraints. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=teE4pl9ftK.

Jian Xu, Yanning Wu, Delu Zeng, John Paisley, and Qibin Zhao. The Right Measure for Physics-Constrained Generation: A Co-Area Correction for Posterior-Consistent PDE Inverse Problems, June 2026. URL http://arxiv.org/abs/2606.04804. arXiv:2606.04804 [cs.LG].

Zeyu Yun, Galen Chuang, Derek Dong, and Yubei Chen. Denoising for Manifold Extrapolation. November 2024. URL https://openreview.net/forum?id=tG75AqtP6U#discussion.

Kamila Zdybał, Giuseppe D'Alessio, Antonio Attili, Axel Coussement, James C. Sutherland, and Alessandro Parente. Local manifold learning and its link to domain-based physics knowledge. *Applications in Energy and Combustion Science*, 14:100131, June 2023. ISSN 2666-352X. doi: 10.1016/j.jaecs.2023.100131. URL https://www.sciencedirect.com/science/article/pii/S2666352X23000201.

A Proofs of Statements in Section 3

Proof of Lemma 3.1. Proposition 3.1 of Huang et al. (2026) gives

$$s_t(x) = \frac{\Pi(x) - x}{\sigma_t^2} + \mathcal{O}_x(1).$$

Multiplying both sides by σ_t^2 and adding x gives the desired result. □

For the following proof, recall that the Hausdorff distance d_H for nonempty sets $A, B \subset \mathbb{R}^d$ is

$$d_H(A, B) := \max\{\sup_{a\in A} \inf_{b\in B} \|a - b\|_2, \sup_{b\in B} \inf_{a\in A} \|b - a\|_2\}. \tag{1}$$

Proof of Theorem 3.1. Let $x \in \mathcal{M}$. We have

$$F_{\theta,t}(x) = \|\hat{\Pi}_{\theta,t}(x) - x\|_2^2 = \|\hat{\Pi}_{\theta,t}(x) - \Pi(x)\|_2^2 \le \varepsilon^2 \le \tau^2 \implies \mathcal{M} \subseteq \hat{A}^U_{\theta,t}(\tau).$$

A consequence of this is that $\sup_{a\in\mathcal{M}} \inf_{b\in\hat{A}_{\theta,t}} \|a - b\|_2 = 0$, so it only remains to compute the second directed term in equation 1:

$$d_H(\mathcal{M}, \hat{A}_{\theta,t}) = \sup_{b\in\hat{A}_{\theta,t}} \inf_{a\in\mathcal{M}} \|b - a\|_2 = \sup_{b\in\hat{A}_{\theta,t}} \text{dist}(b, \mathcal{M}).$$

Let $y \in \hat{A}^U_{\theta,t}(\tau)$. Then, the desired statement is accomplished by applying the triangle inequality:

$$\text{dist}(y, \mathcal{M}) = \|y - \Pi(y)\|_2 \le \|y - \hat{\Pi}_{\theta,t}(y)\|_2 + \|\hat{\Pi}_{\theta,t}(y) - \Pi(y)\|_2 \le \tau + \varepsilon.$$

□

B Baselines

We conduct extensive ablations so that results are as favorable towards the baselines as possible. Moreover, all of the following baselines are adjusted to be "comparable" to our method of identifying a pointwise scalar and forming an estimated target tube around it. For uniformity in the following definitions, let $x_{(j)}(y)$ denote the jth-closest training point to y, and write $r_j(y) = \|y - x_{(j)}(y)\|_2$. Moreover, let the collection of k-nearest neighbors be $N_k(y) = \{x_{(1)}(y), ..., x_{(k)}(y)\}$. All algorithms expect the training data $X = \{x_j\}_{j=1}^n$ a set of validation probes formed as in Appendix C.4, Y^{val}, corresponding labels z^{val}, and a set of test probes Y^{test} as input. We index the baseline type as b and (baseline-specific) hyperparameters as η, and denote their scalar statistic as $S_{b,\eta}(y)$. Additional baseline-specific inputs are included in the Input sections for each respective algorithm.

To calibrate the thresholds, we evaluate $S_{b,\eta}$ at each η option on the validation probes Y^{val}. We form a list of candidate thresholds

$$\mathcal{G}_{b,\eta} = \{\text{Quantile}_\beta(\{S_{b,\eta}(y) : y \in Y^{\text{val}}\}) : \beta \in \{0.01, 0.03, ..., 0.99\}\}.$$

We choose γ^*_η for each η based on its performance on the IoU metric when evaluated on validation probes. The selection of η^* is made based on performance according to a specific metric on test probes. This procedure induces $\hat{z}_b(y) = \mathbb{1}\{S_{b,\eta^*}(y) \le \gamma^*_{\eta^*}\}$ as the pointwise classifier.

Nearest-neighbor balls: The union of balls of radius γ^* induces the localization, with statistic

$$S_{\text{NN}}(y) = r_1(y).$$

k-NN distance tubes: We calibrate over $k \in \{4, 8, 16, 32, 64, 128\}$. The statistic is

$$S_{k\text{-NN},k}(y) = \left(\frac{1}{k} \sum_{x\in N_k(y)} \|y - x\|_2^2\right)^{\frac{1}{2}}.$$

Tangent-patch local PCA: With intrinsic dimensions estimated using TwoNN (Facco et al., 2017), tangent patches formed from nearest-neighbor batches of q points are used to localize the manifold. The calibration is done over $q \in \{16, 32, 64, 128, 256\}$. In particular, given an estimated intrinsic dimension $\hat{m}$ obtained with TwoNN (or an oracle intrinsic dimension if available), take each query point y, select its nearest q points in the training data, compute their mean $\overline{x}_q(y)$, and construct a matrix $V_{q,\hat{m}}(y) \in \mathbb{R}^{d\times\hat{m}}$ of their leading principal directions at y. Then,

$$S_{\text{LPCA},q}(y) = \|(I - V_{q,\hat{m}}(y)V_{q,\hat{m}}(y)^\top)(y - \overline{x}_q(y))\|_2$$

Kernel Density Estimation (KDE): We also evaluate against a standard KDE-based approach for density estimation, where calibration is done over various possible bandwidths $h \in \{0.02, 0.05, 0.1, 0.2, 0.5, 1.0\}$. We fit

$$\hat{p}_h(y) \leftarrow \frac{1}{n}\sum_{i=1}^{n}(2\pi h^2)^{-\frac{d}{2}} e^{-\frac{\|y-x_i\|_2^2}{2h^2}}$$

and then write the statistic

$$S_{\text{KDE},h}(y) = -\log \hat{p}_h(y).$$

MLS estimation. We also compare against a method inspired by fitting signed distance functions. Originally, the method of Liu et al. (2021) learns oriented SDFs, which are not necessarily readily available in the settings we consider. Therefore, we adapt their algorithm to our approach by locally fitting a Gaussian-weighted tangent plane and computing an unsigned normal residual using moving least squares (MLS). Using candidate neighborhood sizes $k \in \{16, 32, 64, 128, 256\}$, write

$$w_i(y) \leftarrow \frac{\exp\frac{-\|x_i-y\|^2}{2r_k(y)^2}}{\sum_{x_j\in N_k(y)}\exp\frac{-\|x_j-y\|^2}{2r_k(y)^2}}, x_i \in N_k(y).$$

From this, we form $\mu_k(y) \leftarrow \sum_{x_i\in N_k(y)} w_i(y)x_i$ and compute the weighted covariance matrix

$$B_k(y) = \sum_{x_i\in N_k(y)} w_i(y)(x_i - \mu_k(y))(x_i - \mu_k(y))^\top.$$

Then, write $V_{k,\hat{m}}(y)$ be the matrix of the leading $\hat{m}$ eigenvectors of $B_k(y)$, where $\hat{m}$ is obtained using Facco et al. (2017). Then, the statistic is

$$S_{\text{MLS},k}(y) = \|(I - V_{k,\hat{m}}(y)V_{k,\hat{m}}(y)^\top)(y - \mu_k(y))\|_2.$$

SINDy: For the PDE task, we use the SINDy baseline which is tailored to learning time-varying dynamics. We ablate over multiple SINDy sparsity thresholds $\lambda \in \{0.0003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3\}$ and write $\hat{\Phi}_\lambda$ be the one-step map induced by SINDy at this sparsity threshold. Then, denoting a probe trajectory as $Y = (y^0, ..., y^{T-1})$, the statistic is

$$S_{\text{SINDy},\lambda}(Y) = \left(\sum_{n=0}^{T-2}\|y^{n+1} - \hat{\Phi}_\lambda(y^n)\|_2^2\right)^{\frac{1}{2}}.$$

C Experimental Setup

C.1 Dataset Details

We present tasks in order of increasing manifold complexity as well as scientific realism.

Geometric tasks. We first study constraints induced by products of spheres (e.g., tori). Write $x = (x^{(1)}, ..., x^{(J)})$ and $x^{(j)} \in \mathbb{R}^{m_j+1}$. Denote S^{m_j} as a sphere of dimension m_j. Generally, let

$$\mathcal{M} = \prod_{j=1}^{J} S^{m_j} \subset \mathbb{R}^d, \qquad h_j(x) = \|x^{(j)}\|_2^2 - 1.$$

In all cases, the distribution follows $U_j \sim \text{Uniform}(S^{m_j}), X_0 = (U_1, ..., U_J)$.

Matrix generation tasks. We also study various Stiefel manifolds:

$$\mathcal{M} = \mathrm{St}(p, 2) = \{X \in \mathbb{R}^{p\times 2} : X^\top X = I_2\}, \qquad h(X) = \mathrm{vech}(X^\top X - I_2) \in \mathbb{R}^3.$$

Generation of matrices lying on the Stiefel manifold has emerged in certain chemistry tasks (Cheng et al., 2024). We experiment on $p \in \{4, 5, 6\}$, with samples having uniform distribution on $\mathcal{M}$.

Molecular geometry tasks. We also conduct experiments on a suite of synthetic molecular geometry tasks over 3-D coordinates of N atoms $r_i \in \mathbb{R}^3$. To remove translation information, we vectorize the distances between atoms (the bonds) and store these bond vectors as samples, where the total number of bonds is B and samples are thus $x = (b_1, ..., b_B) \in (\mathbb{R}^3)^B$. Note that the bonds are assumed to have a fixed ordering. All of the following tasks have length constraint $h_i^{\text{len}} = \|b_i\|_2^2 - \ell^2$. For tasks including an angle constraint, the forward-bond convention yields $h_i^{\text{ang}}(x) = b_i^\top b_{i+1} + \ell^2 \cos\alpha = 0$. We use fixed bond length $\ell = 0.7$ and angle $\alpha = 109.5°$. For these tasks, we use an O(3)-equivariant network to encode rotational and reflection equivariance.

1. **Freely Jointed Chain** $(B = 2)$: Atoms form a sequential chain constrained only by bond lengths, yielding manifold $\mathcal{M} = (\ell S^2)^B$. Samples are distributed as $b_i = \ell U_i, U_i \sim \mathrm{Unif}(S^2)$.
2. **Fixed-Angle Chain** $(B = 3)$**:** Same as above but with an angle constraint α over atom triplets. Distributions are formed by taking torsion angles to be uniform on $[-\pi, \pi)$, and x is additionally randomly rotated as $b_i \mapsto Rb_i$, where $R \in \mathbb{R}^{3\times 3}$ is a uniformly random rotation matrix.
3. **Three-Arm Star** $(B = 3)$**:** A star-shaped molecule where each bond-vector points from the central atom to a single leaf atom. The constraint is governed by $(\ell S^2)^3$ with bond angles as free variables, and the directions are again uniformly sampled on S^2.
4. **Fixed-Angle Chain II** $(B = 3)$**:** Here, $\mathcal{M}$ is the same as above, but torsion angles are sampled with density $q(\phi) = \frac{1-\cos(3\phi)}{2\pi}$, independently followed by uniform (Haar) random rotation.

PDE task. We also present a simple PDE simulation task with $(m, d) = (32, 512)$ inspired by both the 1-D Burgers' equation task in (Feng et al., 2024) and recent interest in PDE-constrained diffusion modeling (Bastek et al., 2024; Xu et al., 2026). Specifically, we consider solutions of

$$\partial_t u + u\partial_s u = \nu \partial_{ss} u, \quad \text{where } (s, t) \in [0, 2\pi) \times [0, 0.30], \tag{2}$$

under periodic spatial boundary conditions. Our discretization is $m = 32$ and $T = 16$. We use $\Delta s = \frac{2\pi}{m}$, $\Delta t = 0.02$, and $\ell = 1.0$. Let $D_1, D_2 \in \mathbb{R}^{m\times m}$ denote the standard periodic centered first- and second-order difference matrices. Define the semi-implicit time step $\Phi(v) = (I - \nu\Delta t D_2)^{-1}[v - \Delta t v \odot (D_1 v)]$. Samples x are temporally flattened trajectories conforming to equation 2, i.e., $x = U := (u^0, ..., u^{T-1}) \in \mathbb{R}^{mT} = \mathbb{R}^{512}$. The constraint $h(x) : \mathbb{R}^{mT} \to \mathbb{R}^{m(T-1)}$ is then

$$[h(U)]_n = u^{n+1} - \Phi(u^n) \quad \text{for } n = 0, ..., T-2.$$

We induce a trajectory distribution by sampling initial conditions. Let $C \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. Then, let

$$u^0 | C \sim \mathcal{N}\big((-1)^C a[\sin(s_0),\ \cdots,\ \sin(s_{m-1})]^\top, K\big), \text{ where}$$
$$K_{jk} := b^2 \exp\left[-\frac{2\sin^2(\frac{s_j - s_k}{2})}{\ell^2}\right] + \epsilon \mathbb{1}_{\{j=k\}}$$
$$a := 0.22,\ b := 0.45,\ \epsilon := 10^{-5},\ u^{n+1} := \Phi(u^n).$$

Note that K is positive definite by construction, so the law of u^0 admits a positive density over its m degrees of freedom. The Gaussian mixture simply makes this density nonuniform.

C.2 Architecture and Training Details

To improve numerical stability at small t, we use an $\varepsilon_\theta(x, t)$-style objective (Ho et al., 2020). Specifically, the output of the score model is divided by σ_t in the final forward pass. The loss is then

$$\mathcal{L}(\theta) = \mathbb{E}\left[\|\sigma_t s_{\theta,t}(X_t) + Z\|_2^2\right],$$

where $X_t = \alpha_t X_0 + \sigma_t Z$. We also use log-uniform sampling on $\sigma_t \in [0.006, 0.2]$ since SCoREMaP only requires queries at small t. In practice, however, we clamp t if it is less than or equal to 10^{-3}, so the effective range of σ_t is then roughly $[0.0105, 0.2]$. All tasks use an AdamW optimizer, 32 Fourier features for time conditioning, weight decay of 10^{-4}, and learning rate 3×10^{-4}.

Setting	Non-PDE Tasks	PDE Task
Network architecture	Four 256-width MLP layers w/ SiLU activations	6 dilated residual convolutional layers w/ 64 channels and SiLU activations
Gradient clipping	Norm 1.0	Same
Batch size	8,192	512
Training samples	10,000	Same

Table 3: Training and architecture setup.

C.3 Evaluation Metrics

We use the binary labels $z_i = \mathbb{1}\{y_i \in A_\tau\}$, where A_τ denotes the oracle tube of radius τ around $\mathcal{M}$. The Intersection over Union (IoU) is given by

$$IoU = \frac{TP}{TP + FP + FN},$$

where TP indicates "True Positive," FP indicates "False Positive," and FN indicates "False Negative." Area Under The Receiver Operating Characteristic curve (AUROC) is the usual classification metric Hanley & McNeil (1982). The Far False-Positive Rate (FarFPR) is a metric of outward leakage from the extraction tube: specifically, for $\delta > 0$, define the set of far points as $\mathcal{F} = \{i : \text{dist}(y_i, \mathcal{M}) \geq \delta\}$. Then, the FarFPR metric is the rate at which far-from-$\mathcal{M}$ points are mistakenly classified as being near $\mathcal{M}$:

$$\text{FarFPR} = \frac{\left|\{i \in \mathcal{F} : y_i \in \hat{A}^{\mathcal{Y}}_{\theta,t}(\tau)\right|}{|\mathcal{F}|}.$$

We use $\delta = 0.0625$ throughout all tasks in this paper.

C.4 Probe Point Generation

Given unlimited computational resources, the naive approach for probe point selection would simply be to construct a spatial grid on a fine mesh of $\mathbb{R}^d$. However, this is inappropriate for two reasons. First, evaluating $s_{\theta,t}(x)$ on an evenly-spaced fine grid for large d is prohibitively computationally expensive and may arbitrarily inflate performance, as the vast majority of points in $\mathbb{R}^d$ will lie nowhere near $\mathcal{M}$. Moreover, as suggested by Figure 1, evaluating globally will result in spurious extracted regions at the center of certain $\mathcal{M}$. Thus, we construct probe points in an unbiased manner from a held-out test dataset. Then, a suitable search region is constructed by randomly perturbing the test samples at various scales of Gaussian noise. Letting I denote the holdout test data index, we sample $Z \sim \mathcal{N}(0, I_d)$ and $R \sim \text{Unif}[0, 1]$ and let a single probe point be $Y = X_I^{\text{holdout}} + \frac{R}{\sqrt{d}} Z$.

C.5 O(3)-Equivariant Neural Network

Let $X \in \mathbb{R}^{B \times 3}$ such that the bond vectors form the rows, and take its Gram matrix $G = XX^\top$. Let $A_\theta(G(X), t)$ be an MLP that takes in the upper triangular entries of $G(X)$ (since $G(X)$ is symmetric) and returns a $B \times B$ matrix. Let the O(3)-equivariant score be $v_\theta(X, t) = A_\theta(G(X), t)X$.

Lemma C.1 (O(3)-Equivariance of Constructed Score Network)**.** *For all orthogonal matrices (i.e., rotations and reflections) $R \in O(3)$, we have*

$$v_\theta(XR^\top, t) = v_\theta(X, t)R^\top.$$

Proof. We have $G(XR^\top) = XR^\top RX^\top = XX^\top = G(X)$, so the network will always produce the same output. Thus, $v_\theta(XR^\top, t) = A_\theta(G(X), t)XR^\top = v_\theta(X, t)R^\top$. □

D Additional Experimental Results

In this section, we provide detailed ablations and experimental results. For space, we only provide ablations over the smallest five ν choices for the PDE task.

D.1 Improvement over Optimization Steps

This section presents performance evaluations at checkpoints each taken 100 optimizer steps apart. All baselines use hyperparameters that perform best in terms of the specified metric on a held-out validation set. We uniformly use $\sigma_t = 0.05$ for all checkpoint evaluations.

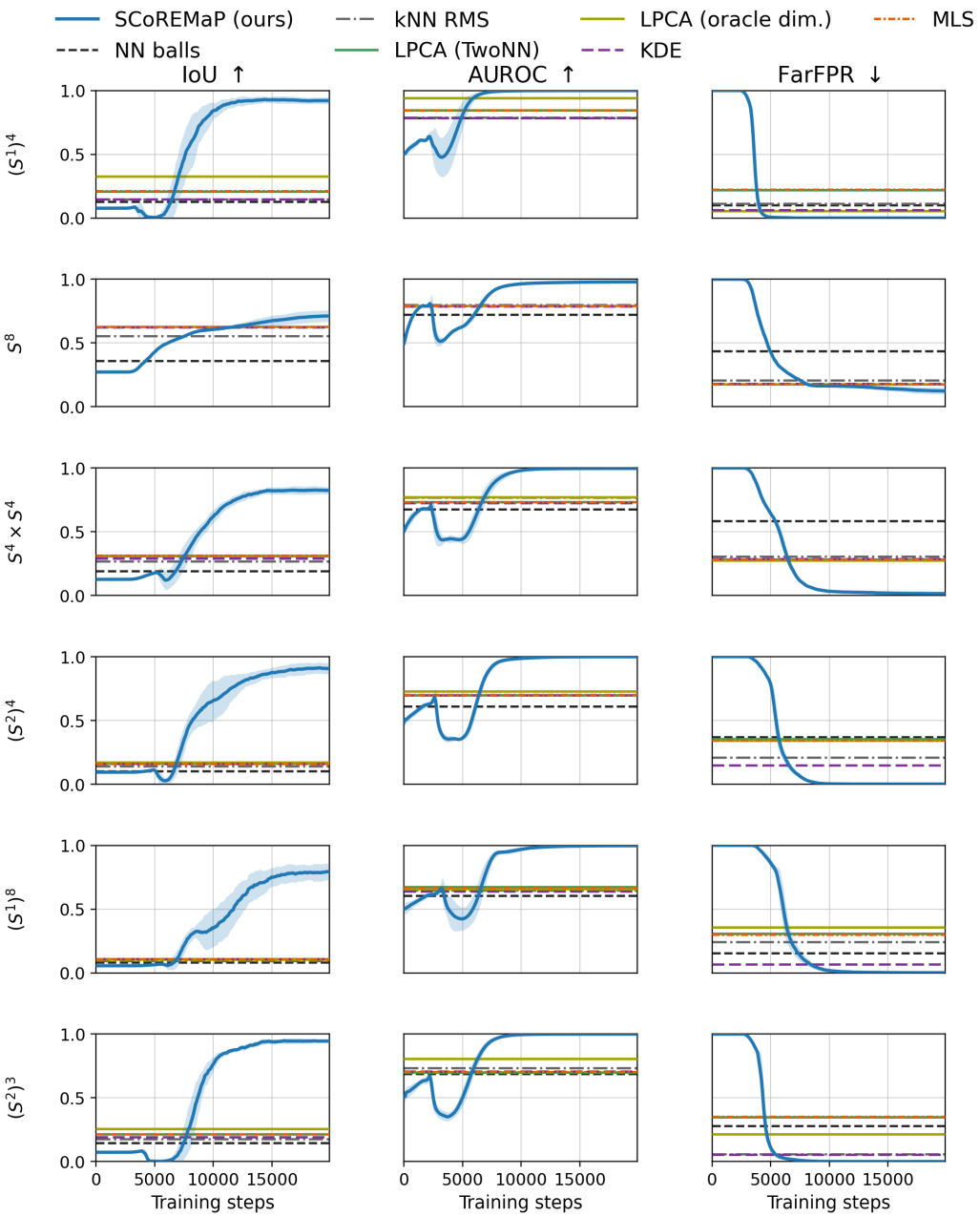


Figure 4: Spheres.

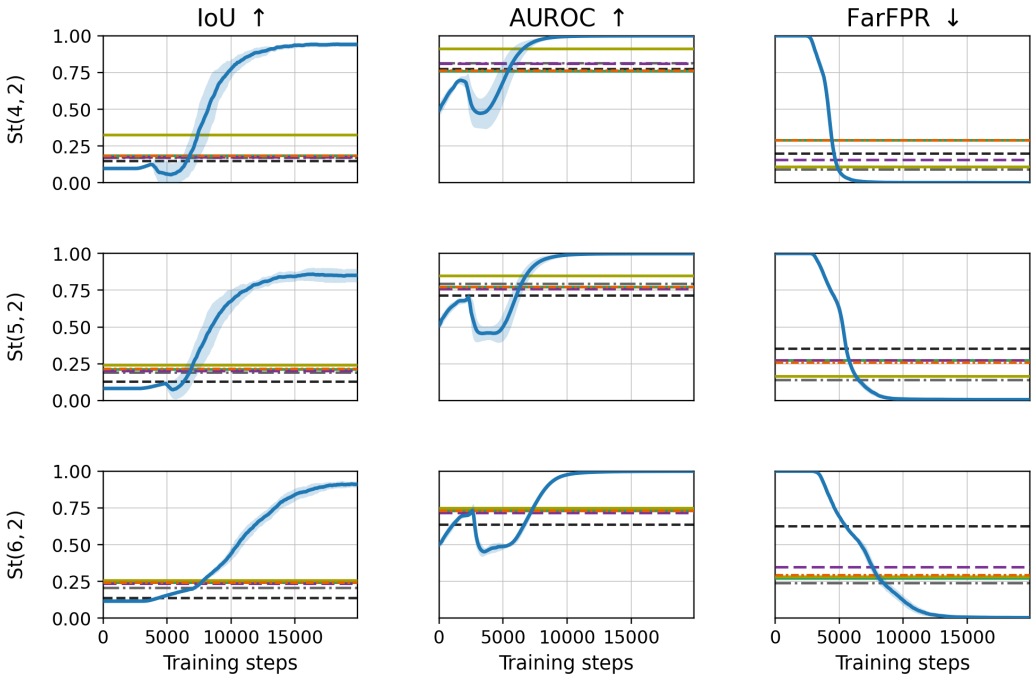


Figure 5: Matrix tasks.

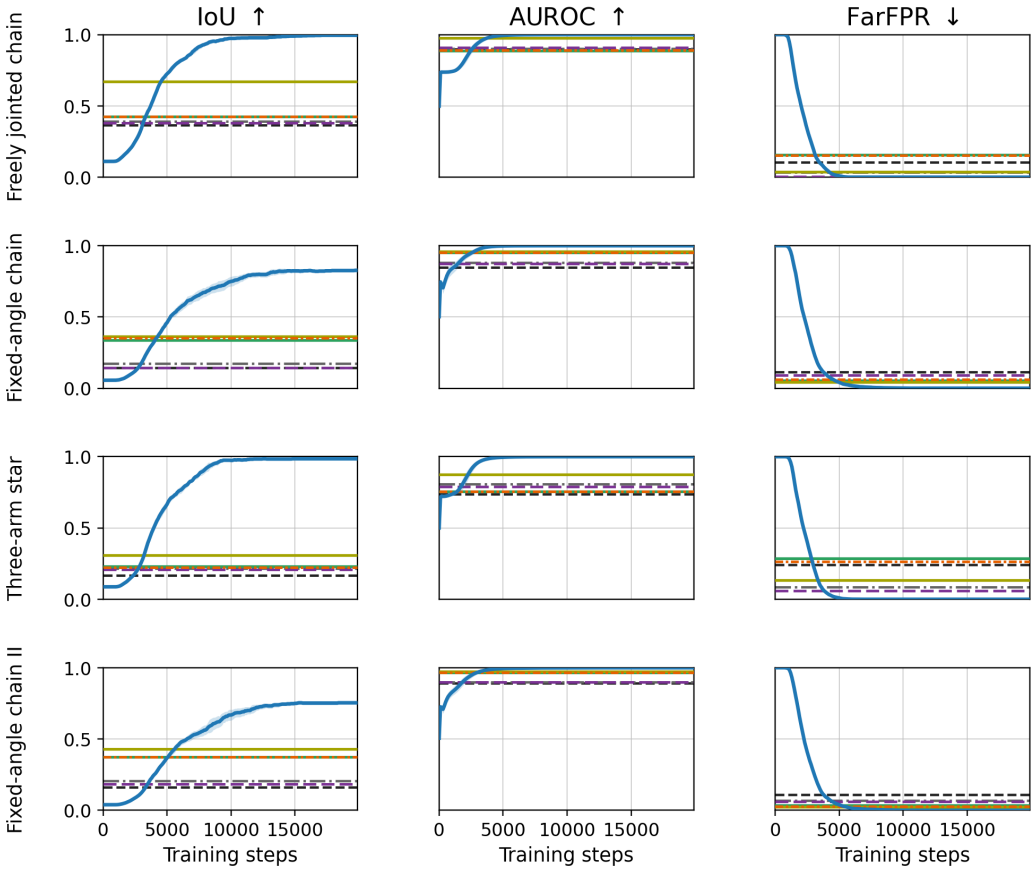


Figure 6: Molecular tasks.

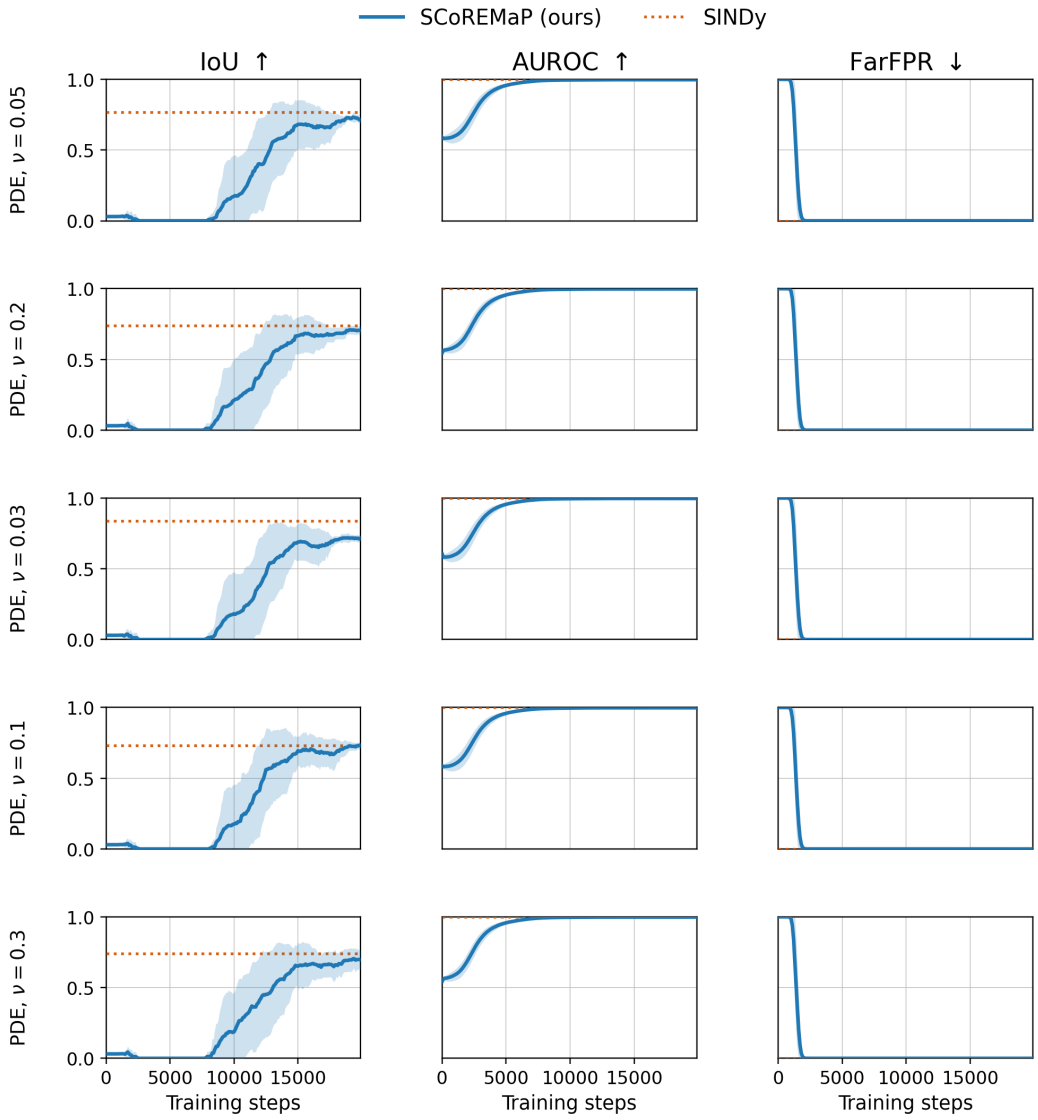


Figure 7: PDE task.

D.2 Robustness to Choice of σ_t

In this section, we evaluate sensitivity of $F_{\theta,t}$ to the diffusion time t. As expected, for all tasks, higher τ generally ensures that the thresholded region encompasses $\mathcal{M}$. However, at large σ_t (i.e., larger t), the localization region is inaccurate since the score field is modeling a blurred p_0.

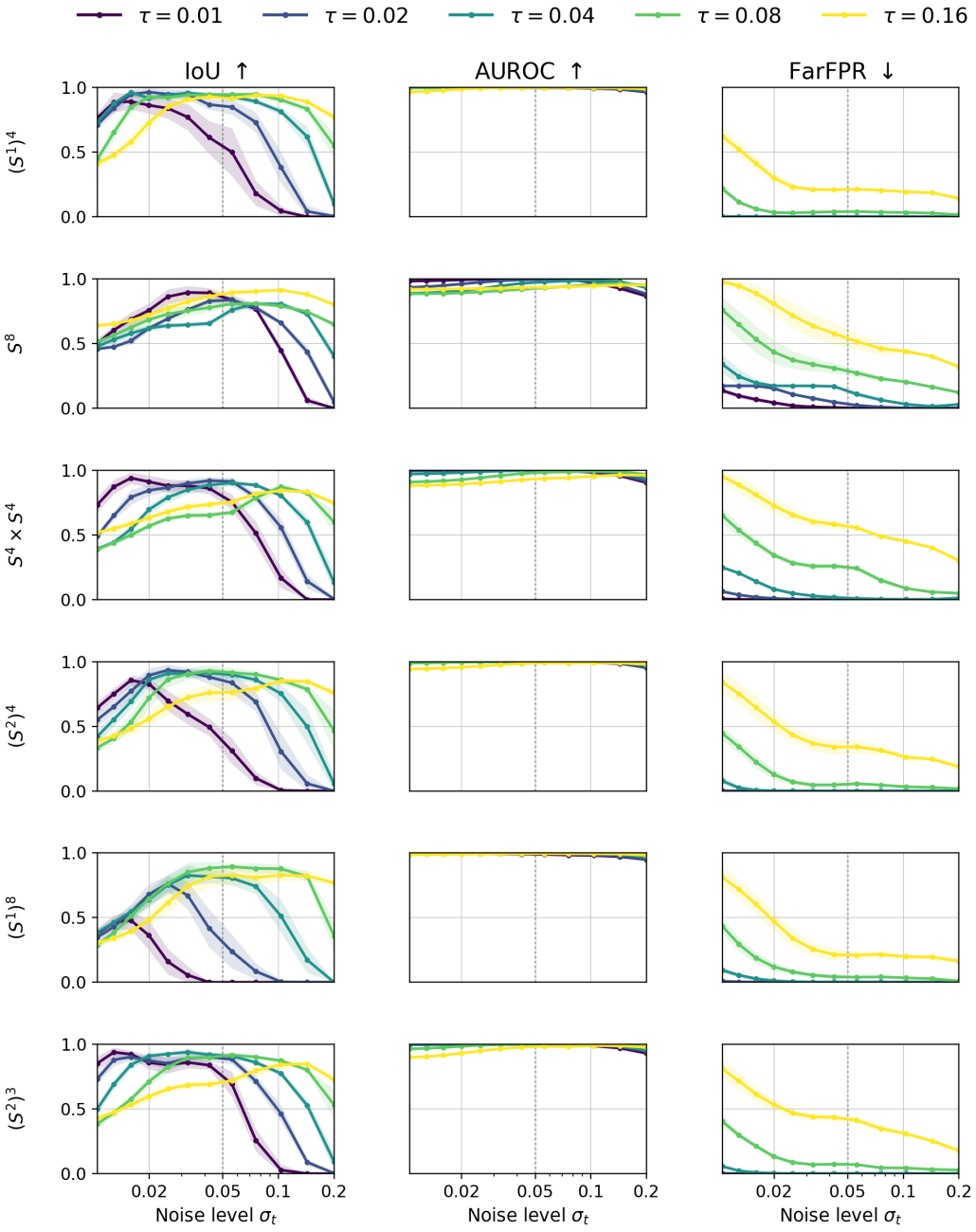


Figure 8: Sphere tasks.

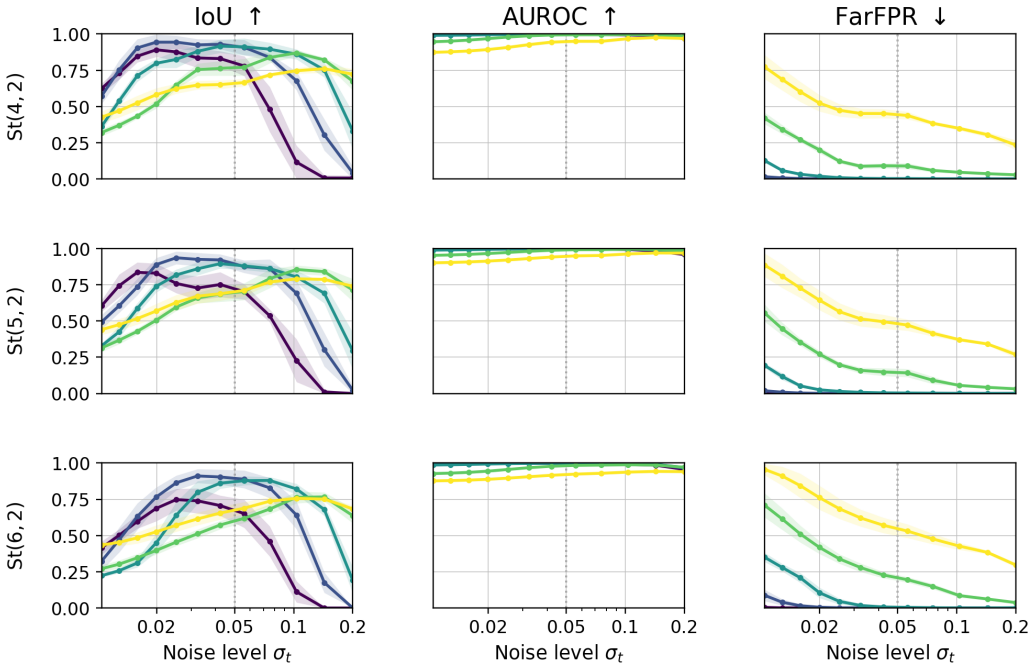


Figure 9: Matrix tasks.

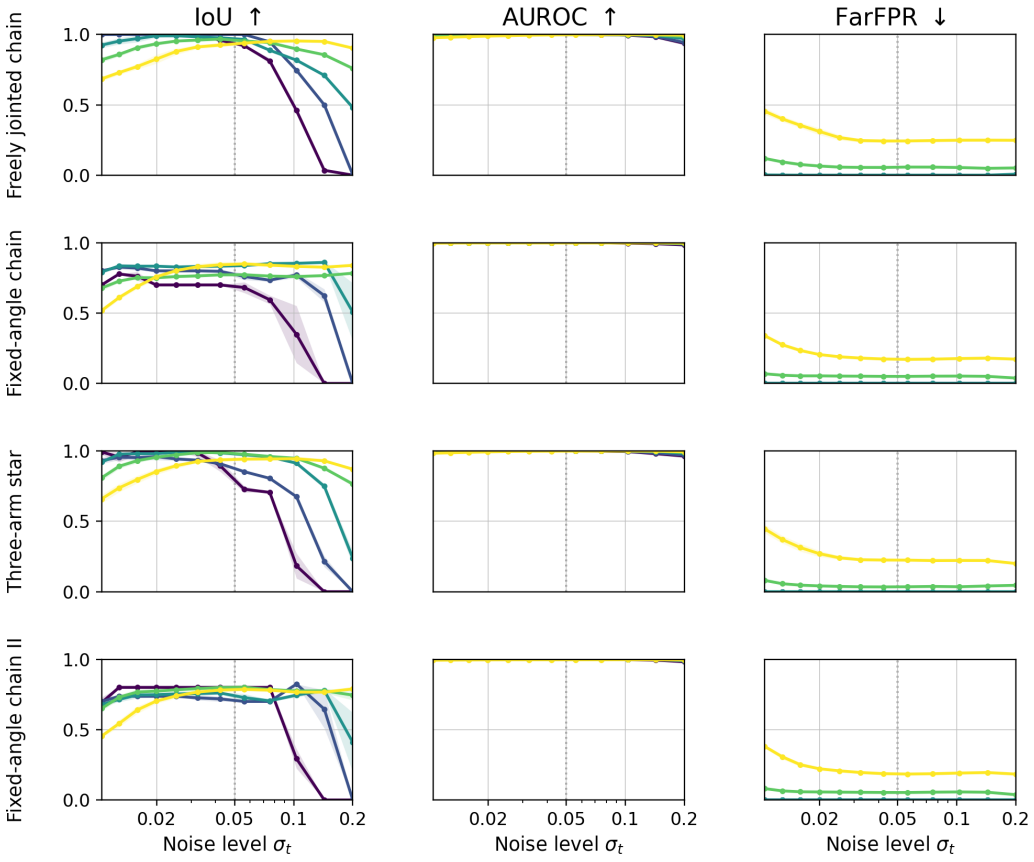


Figure 10: Molecular tasks.

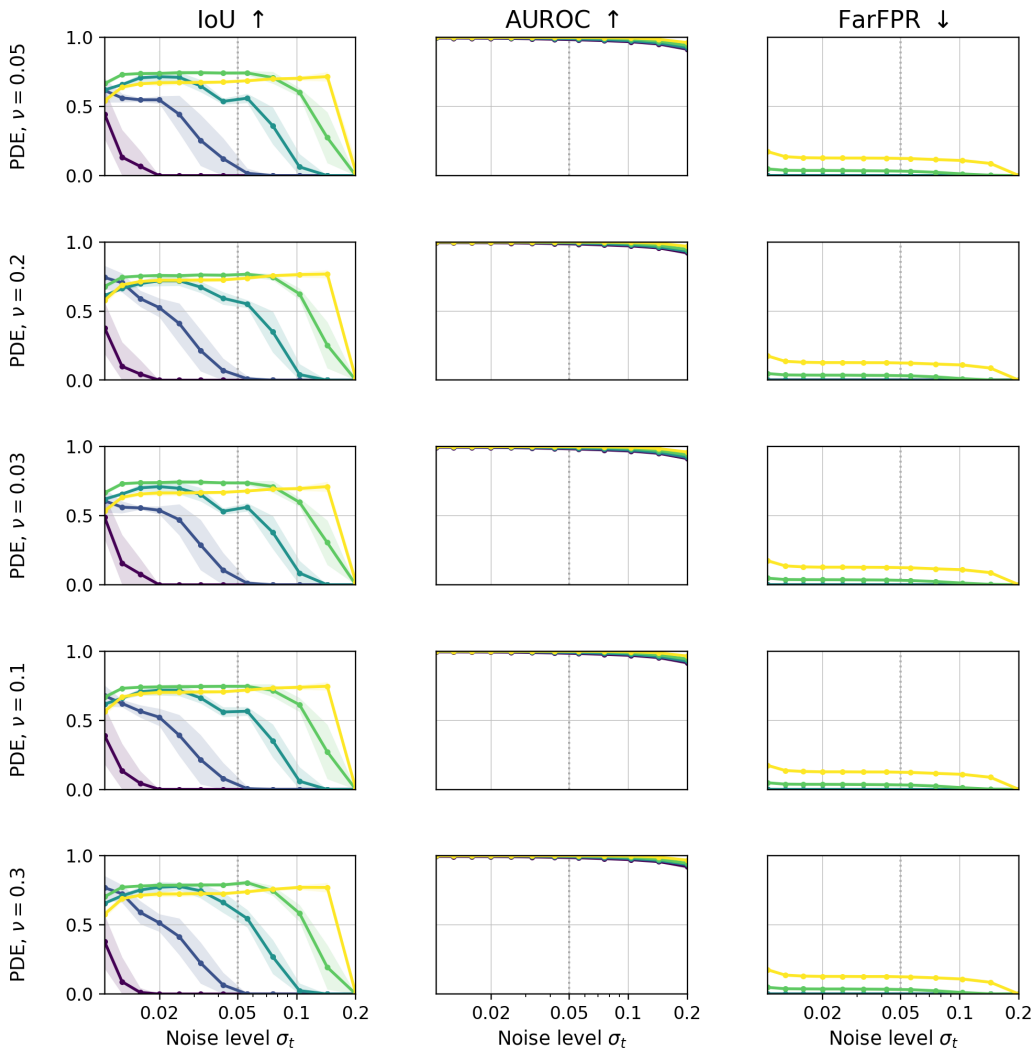


Figure 11: PDE task.

E EPISTEMIC UQ IMPLEMENTATION

For all UQ experiments presented in this paper, we use a last-layer Laplace approximation (LLLA), which is motivated by prior results (Daxberger et al., 2022; Gupta et al., 2026; Kristiadi et al., 2020). Algorithm 3 documents the implementation of LLLA that we use. We denote the collection of last-layer parameters as θ_{last}. We let $\beta_n = n = 10{,}000$ in all experiments, as we use the same number of training data samples for every task and seed. We also let $\lambda = 1$. To fit the covariance, we use 256 points sampled from the data density p_0. We also maintain 768 test probes.

Algorithm 3 Last-layer Laplace approximation for SCoREMaP UQ.

Require: Holdout calibration points $\{x_j\}_{j=1}^{n_c}$, likelihood scale β_n, precision λ

for $j = 1, ..., n_c$ **do**
  Compute last-layer Jacobian $J_j = \frac{\partial s_{\theta,t}(x_j)}{\partial \theta_{\text{last}}}$ using automatic differentiation
end for ▷ Get per-parameter variance
for parameter index ξ **do**
  $P_\xi \leftarrow \lambda + \beta_n \sum_j \sum_a J_j[a, \xi]^2$
  $\text{Var}(\theta_\xi) \leftarrow P_\xi^{-1}$
end for
for probe y_i **do**
  $J_i = \frac{\partial s_{\theta,t}(y_i)}{\partial \theta_{\text{last}}}$
  $U_t(y_i) \leftarrow \sigma_t^4 \sum_\xi \text{Var}(\theta_\xi) \|J_i[:, \xi]\|_2^2$ ▷ Compute propagated variance
end for
return U_t

F SCOREMAP FOR ENHANCED NUMERICAL PDE TIME ROLLOUT

We present our SCoREMaP-assisted PDE implementation for improved dynamics recovery in Algorithm 4. Results at various ν are presented in Figure 12.

Algorithm 4 SCoREMaP-Assisted PDE Time Rollout

Require: Candidate one-time-step map $\hat{\Phi}$ (e.g. from raw SINDy), initial condition u^0, projection interval $k > 0$, diffusion model s_θ, time t (or corresponding variance σ_t)

$\hat{u}^0 \leftarrow u^0$
for $n = 1, ..., T-1$ **do**
  $\hat{u}^n \leftarrow \hat{\Phi}(\hat{u}^{n-1})$ ▷ e.g. with vanilla SINDy rollout
  if $n \bmod k = 0$ and $n < T-1$ **then** ▷ snap back only every k steps
    $v^0, ..., v^n \leftarrow \hat{u}^0, ..., \hat{u}^n$
    for $j = n+1, ..., T-1$ **do**
      $v^j \leftarrow \hat{\Phi}(v^{j-1})$ ▷ rollout to full candidate trajectory
    end for
    $V \leftarrow \text{flatten}(v^0, ..., v^{T-1})$
    $\hat{\Pi}_t \leftarrow V + \sigma_t^2 s_\theta(V, t)$ ▷ Projection from Tweedie's formula with trained score model
    $(\hat{u}_V^0, ..., \hat{u}_V^{T-1}) \leftarrow \text{unflatten}(\hat{\Pi}_t(V))$
    $\hat{u}^n \leftarrow \hat{u}_V^n$ ▷ replace the state at time index n with estimated projected state
  end if
end for
return $\hat{U} = (\hat{u}^0, ..., \hat{u}^{T-1})$

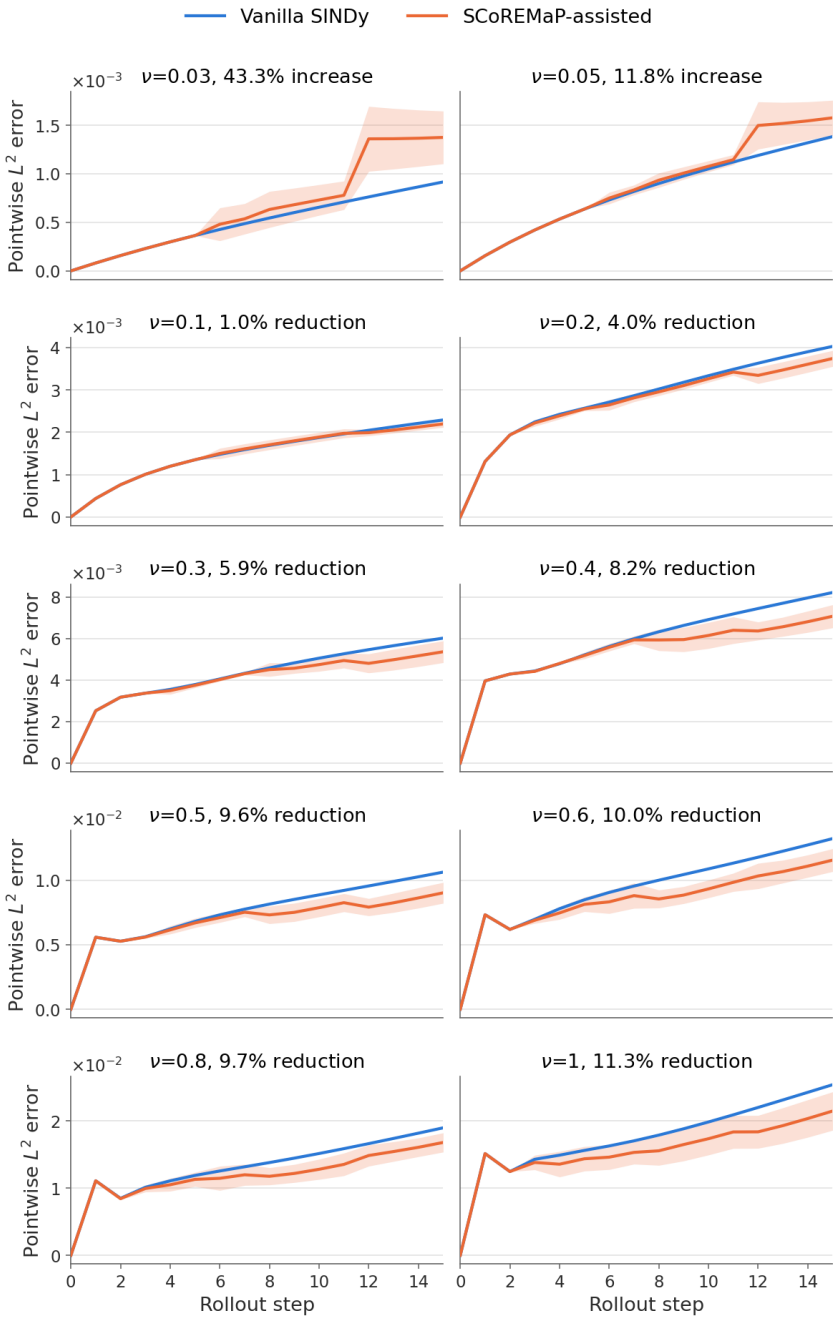


Figure 12: Rollout analysis at various ν. We observe the expected improvement of SCoREMaP-assisted SINDy rollout as we increase the viscosity.