Strong Gravitational Lensing Posterior Sampling in Pixel-Space Using Diffusion Models and Recurrent Inference Machines

Guillaume Payeur [1 2 3 4] **Laurence Perreault-Levasseur** [1 2 3 5 6 4] **Gabriel Missael Barco** [1 2 3]
Yashar Hezaveh [1 2 3 5 6 4]

Abstract

Modeling galaxy-galaxy strong gravitational lenses to infer the brightness of the source galaxy and the mass distribution of the foreground galaxy is computationally challenging, particularly for high-resolution, high signal-to-noise observations. In this regime, high-dimensional representations of both the source and the foreground mass distribution are necessary to model the data down to the noise level. This inference problem has been challenging for both traditional and machine learning-based methods because of its high dimensionality and its non-linearity in the foreground mass distribution. We present a method to generate joint posterior samples of the source galaxy and foreground mass distribution as pixelated images conditioned on observations. The method combines diffusion-based generative modeling and recurrent inference machines. It can model realistic gravitational lensing simulations with background and foreground galaxies drawn from cosmological hydrodynamical simulations down to the noise level.

1. Introduction

Galaxy-galaxy strong gravitational lensing is the physical process wherein the gravitational potential of a foreground galaxy bends the trajectory of light emitted by a distant background galaxy in a way that multiple distorted images of the source are observed from Earth. Observations of these systems can answer many questions in cosmology and astrophysics. As an example, the distortions of the images of the source are tracers of the mass distribution of the foreground galaxy, allowing for the constraint of the properties of dark matter on galactic scales (see, e.g., Grillo et al. 2015; Hezaveh et al. 2016; Hsueh et al. 2020; Gilman et al. 2020; 2021; Minor et al. 2021; Anau Montel et al. 2022; Bayer et al. 2023a;b; Powell et al. 2023; Ballard et al. 2024 and see Vegetti et al. 2024 for a review article). As another example, by considering the time delay between the multiple images, it is possible to constrain the rate of expansion of the universe given by the Hubble constant (see, e.g., Shajib et al. 2020; Qi et al. 2021; Shajib et al. 2023; Colaço et al. 2025 and see Birrer et al. 2024 for a review article). As a last example, one can study galaxy formation at high redshifts by observing magnified images of very distant galaxies which would otherwise be unresolved (see, e.g., Marques-Chaves et al. 2017; 2018; Fan et al. 2019; Jacobs et al. 2019; Marques-Chaves et al. 2020; Shu et al. 2022a;b; Nightingale et al. 2025 and see Shajib et al. 2024 for a review article).

The scientific potential of gravitational lenses can be realized via lens modeling, which is the process of jointly inferring the brightness distribution of the background source and the mass distribution of the foreground galaxy using observations. Traditionally, due to the computational cost and complexity of the inference problem, this has been done assuming that the source brightness and foreground mass distribution follow simple low-dimensional parametric profiles (e.g., Bolton et al. 2008; Hezaveh et al. 2017; Savary et al. 2022; Knabel et al. 2023; Huang et al. 2026). These simplistic assumptions often bias the inference and prevent the desired information from being extracted from the data (Xu et al., 2015; Sonnenfeld, 2018; Galan et al., 2024). This is particularly true for high-resolution and high signal-to-noise ratio (SNR) observations.

In recent years, advances in machine learning have given rise to models that offer an alternative to traditional inference methods. On the one hand, diffusion-based generative models are a class of machine learning models that can generate samples from complex probability distributions $p(\boldsymbol{x})$, including in settings where $\boldsymbol{x}$ is high-dimensional and

[1]Department of physics, Université de Montréal, Montreal QC, Canada [2]Mila - Quebec AI Institute, Montreal QC, Canada [3]Ciela Institute, Institute for Astrophysics and Machine Learning, Montreal QC, Canada [4]Trottier Space Institute, McGill University, Montreal QC, Canada [5]Center for Computational Astrophysics, Flatiron Institute, New York, USA [6]Perimeter Institute for Theoretical Physics, Waterloo ON, Canada. Correspondence to: Guillaume Payeur <guillaume.payeur@umontreal.ca>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

$p(\boldsymbol{x})$ is multi-modal (Song & Ermon, 2019; Ho et al., 2020; Song et al., 2020; Wong et al., 2020; Lai et al., 2025). They have been used for Bayesian analysis of high-dimensional models in the sciences, including in physics (e.g., Song et al. 2021; Adam et al. 2022; Kazerouni et al. 2022; Dia et al. 2023; Legin et al. 2023; Remy et al. 2023; Xue et al. 2023; Bourdin et al. 2024; Cuesta-Lazaro & Mishra-Sharma 2024; Wu et al. 2024; Adam et al. 2025; Andry et al. 2025; Barco et al. 2025a;b; Boruah et al. 2025; Legin et al. 2025; Riveros et al. 2025; Rozet et al. 2025; Stone et al. 2025). On the other hand, recurrent inference machines (RIMs) are a machine learning framework for solving inverse problems which proceeds by learning a data-driven iterative inference algorithm that implicitly captures the prior distribution over the model parameters. This framework has been applied to under-constrained inverse problems for which the prior distribution $p(\boldsymbol{x})$ is hard to compute or intractable (e.g., Morningstar et al. 2018; 2019; Lønning et al. 2019; Sabidussi et al. 2021; Rhea et al. 2024), including to non-linear inverse problems (e.g., Modi et al. 2021; Adam et al. 2023). In these works, the RIM produces a maximum a posteriori (MAP) estimate of the model parameters, given an observation.

Over the coming years, the Rubin Observatory Legacy Survey of Space and Time (LSST) and the Euclid Space Telescope are expected to discover over 100,000 strong gravitational lenses (Lines et al., 2025; Shajib et al., 2025). Realizing the scientific potential of this data requires lens modeling methods that leverage flexible, high-dimensional representations of both the source and the foreground mass. Because gravitational lensing depends non-linearly on the foreground mass distribution, this inference task constitutes a high-dimensional non-linear inverse problem. Such problems are typically characterized by highly complex, multimodal posterior distributions and severe parameter degeneracies that often render traditional, as well as modern, diffusion-based sampling algorithms ineffective (see, e.g., Appendix E of Sharief et al. 2026).

While methods leveraging flexible representations of the source (Warren & Dye, 2003; Suyu et al., 2006; Nightingale & Dye, 2015) and foreground mass (Koopmans, 2005; Saha et al., 2006; Vegetti et al., 2009) exist, they typically employ analytic prior distributions chosen for their simplicity. A previous work (Adam et al., 2023) discussed a method to obtain point estimates of the source and foreground mass as pixelated images which leveraged flexible data-driven priors, but sampling from the corresponding joint posterior distribution has remained an open problem.

In this work, we develop a Bayesian method to sample from the joint posterior of the source galaxy brightness and foreground mass distribution as pixelated images conditioned on observations. Our method, named Diffusion RIM (DiRIM), is based on diffusion-based generative models and RIMs. It works by using a RIM to determine scores of the posterior distribution and subsequently solving a reverse stochastic differential equation (SDE) to generate the posterior samples. This key idea is summarized in Fig.1.

In this paper, we train a DiRIM on realistic mock observations of strong gravitational lenses generated using source galaxy and foreground mass images taken from cosmological magneto-hydrodynamical simulations, and show that we can model the observations down to the noise level. This constitutes a first proof of concept regarding DiRIM's ability to solve high-dimensional non-linear inverse problems.

2. Related Works

Our work is part of a class of methods for solving inverse problems that use conditional diffusion models, which are trained to estimate the score of the posterior distribution by incorporating the noisy observations as an input to the model. This approach has been applied to inverse problems in cosmology (Legin et al., 2023; Bourdin et al., 2024), waveform generation (Chen et al., 2020), and inpainting, super-resolution and colorization (Batzolis et al., 2021; Saharia et al., 2021; 2022; Shang et al., 2023). The novelty in our approach is that it integrates RIMs into the diffusion model, enabling the use of a known and differentiable likelihood function to facilitate solving non-linear inverse problems.

3. Methods

3.1. Background

3.1.1. Strong gravitational lensing simulations

Strong gravitational lensing simulations are implemented using the thin screen approximation (Meneghetti, 2022) wherein light deflection is assumed to occur on a single plane called the lens plane. The lensed images are computed using ray tracing, with the lens plane coordinates $\boldsymbol{\theta} = \{\theta_x, \theta_y\}$ of the rays related to their source plane coordinates $\boldsymbol{\beta} = \{\beta_x, \beta_y\}$ by the lens equation

$$\boldsymbol{\beta}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}). \tag{1}$$

The reduced deflection angle $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is calculated from the dimensionless projected surface density field $\kappa(\boldsymbol{\theta})$, also referred to as convergence map, as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int d^2\boldsymbol{\theta}' \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2}. \tag{2}$$

In the simulated lensed images, the light intensity of a pixel at angular position $\boldsymbol{\theta}$ is given by the source intensity at the corresponding source plane angular position $\boldsymbol{\beta}(\boldsymbol{\theta})$. The lensed image is then convolved with a point spread function

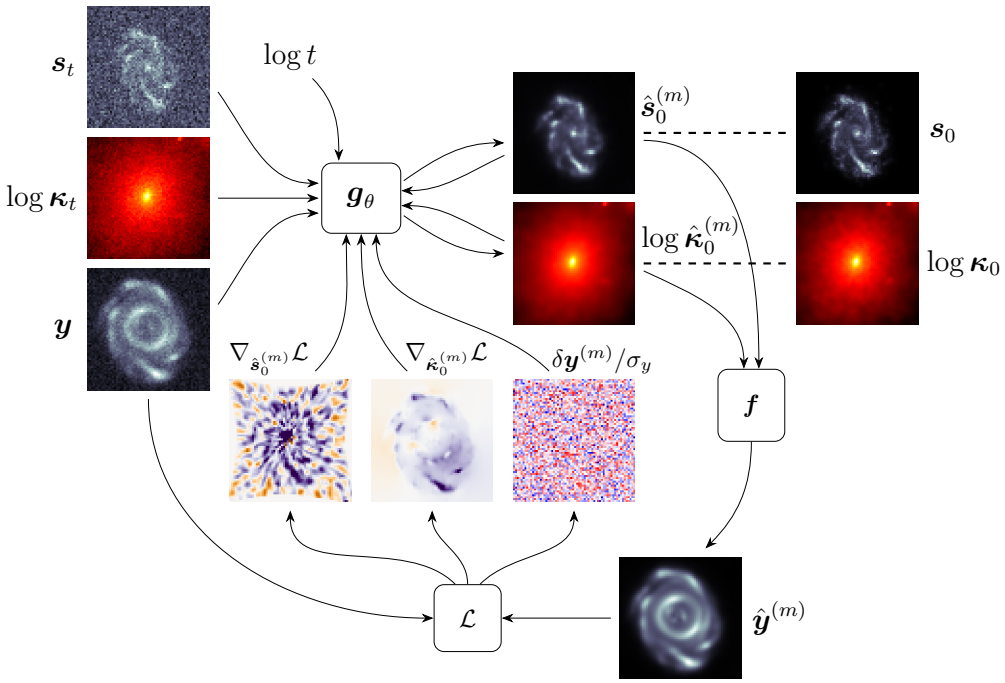

Figure 1. Graphical representation of DiRIM (Diffusion Recurrent Inference Machine) for strong gravitational lensing. A neural network $\boldsymbol{g}_\theta$ is trained to denoise noisy source galaxy images $\boldsymbol{s}_t$ and noisy convergence map images $\log \boldsymbol{\kappa}_t$ given lens observations $\boldsymbol{y}$. It iteratively updates its estimate of the denoised source and convergence map $(\hat{\boldsymbol{s}}_0, \log \hat{\boldsymbol{\kappa}}_0)$ using the Recurrent Inference Machine (RIM) framework. The dashed lines represent the loss function Eq.16 (with $\boldsymbol{x} \equiv (\boldsymbol{s}, \log \boldsymbol{\kappa})$), and the trained $\boldsymbol{g}_\theta$ is used to determine posterior scores $\nabla_{\boldsymbol{s}_t, \log \boldsymbol{\kappa}_t} \log p_t(\boldsymbol{s}_t, \log \boldsymbol{\kappa}_t | \boldsymbol{y})$ needed to obtain samples from the posterior $p(\boldsymbol{s}_t, \log \boldsymbol{\kappa}_t | \boldsymbol{y})$ via diffusion-based generative modeling.

(PSF), and Gaussian additive instrument noise is applied to produce a simulated observation. This defines a forward model

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{\kappa}) + \boldsymbol{\eta}, \tag{3}$$

where $\boldsymbol{s}$ is the source brightness distribution, $\boldsymbol{\kappa}$ is the convergence map, $\boldsymbol{f}$ is the forward operator consisting of ray tracing and PSF convolution, $\boldsymbol{\eta}$ is the additive observational noise and $\boldsymbol{y}$ is the simulated observation.

When $\boldsymbol{s}$ and $\boldsymbol{\kappa}$ are represented by pixelated images, the deflection angles and source brightnesses at the required angular positions are obtained using bilinear interpolation of the pixelated deflection angle images and pixelated source brightness images, respectively.

The forward operator $\boldsymbol{f}$ is implemented in CAUSTICS (Stone et al., 2024), a GPU-accelerated, automatically differentiable gravitational lensing simulator.

3.1.2. DIFFUSION MODELS

In their formulation as score-based models, diffusion models can be described using SDEs (Song et al., 2020). In this description, the forward SDE

$$d\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)dt + g(t)d\boldsymbol{w}(t), \tag{4}$$

is defined, where $\boldsymbol{w}(t)$ is the standard Brownian motion, $\boldsymbol{f}(\boldsymbol{x}, t)$ is the drift coefficient, and $g(t)$ is the diffusion coefficient. The forward SDE is chosen so that when evolved over the time interval $t \in [0, T]$, it transports samples from $p(\boldsymbol{x})$ into samples from a Gaussian distribution[1].

In order to generate samples from $p(\boldsymbol{x})$, one generates samples from the Gaussian distribution and transports them back in time from $t = T$ to $t = 0$ by solving the corresponding reverse-time SDE (Anderson, 1982)

$$d\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2 \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})]dt + g(t)d\bar{\boldsymbol{w}}(t), \tag{5}$$

where $\bar{\boldsymbol{w}}$ is the reverse time standard Brownian motion and $p_t(\boldsymbol{x})$ is the probability density of $\boldsymbol{x}$ at time t. Doing so requires knowing the quantity $\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$, known as the score of $p_t(\boldsymbol{x})$, for all $t \in [0, T]$.

The score of $p_t(\boldsymbol{x})$ is approximated using deep learning. Given a dataset $\{\boldsymbol{x}^{(i)}\}_{i=1}^N$ of samples $\boldsymbol{x}^{(i)} \sim p(\boldsymbol{x})$, one uses a neural network $\boldsymbol{s}_\theta(\boldsymbol{x}_t, t)$ with parameters θ trained by minimizing the denoising score-matching loss function

[1]For the choice of SDE in practical implementations, the SDE transports samples from $p(\boldsymbol{x})$ into samples from a distribution that is only approximately Gaussian.

(Hyvärinen, 2005; Vincent, 2011; Song et al., 2020)

$$L_{\text{DSM}} = \mathbb{E}_{t\sim\mathcal{U}[0,T],\boldsymbol{x}_0\sim p(\boldsymbol{x}),\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_0)}\{ \lambda(t)\|\boldsymbol{s}_\theta(\boldsymbol{x}_t,t) - \nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{x}_t|\boldsymbol{x}_0)\|_2^2\}. \quad (6)$$

It can be shown that in the limit of infinite data and model capacity, the set of parameters θ^* minimizing the loss function L_{DSM} satisfy $s_{\theta^*}(\boldsymbol{x},t) = \nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x})$ for almost all $\boldsymbol{x}$ and t (Song et al., 2020). Therefore, a trained $\boldsymbol{s}_\theta(\boldsymbol{x},t)$ can be used to solve the reverse SDE Eq.5.

When a forward model mapping model parameters $\boldsymbol{x}$ to observations $\boldsymbol{y}$ (e.g., Eq.3 for which $\boldsymbol{x} = \{\boldsymbol{s},\boldsymbol{\kappa}\}$) is defined, this diffusion framework can be extended to generate samples from the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ for any $\boldsymbol{y}$. This is done by solving the reverse-time SDE

$$d\boldsymbol{x} = [f(\boldsymbol{x},t) - g(t)^2\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x}|\boldsymbol{y})]dt + g(t)d\bar{\boldsymbol{w}}(t), \quad (7)$$

which requires knowing the score $\nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x}|\boldsymbol{y})$. This score can be approximated using a neural network $\boldsymbol{s}_\theta(\boldsymbol{x},\boldsymbol{y},t)$ by minimizing the conditional denoising score-matching loss function

$$L_{\text{CDSM}} = \mathbb{E}_{t\sim\mathcal{U}[0,T],\boldsymbol{x}_0\sim p(\boldsymbol{x}),\boldsymbol{y}\sim p(\boldsymbol{y}|\boldsymbol{x}_0),\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_0)}\{ \lambda(t)\|\boldsymbol{s}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) - \nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)\|_2^2\}. \quad (8)$$

One can show that the set of parameters θ^* minimizing L_{CDSM} satisfy, in the limit of infinite data and model capacity, $\boldsymbol{s}_{\theta^*}(\boldsymbol{x},\boldsymbol{y},t) = \nabla_{\boldsymbol{x}}\log p_t(\boldsymbol{x}|\boldsymbol{y})$ for almost all $\boldsymbol{x}$, $\boldsymbol{y}$, and t (Batzolis et al., 2021).

In practice, a popular choice of forward SDE is the variance exploding (VE) SDE (Song et al., 2020)

$$d\boldsymbol{x} = \sqrt{\frac{d[\sigma^2(t)]}{dt}}d\boldsymbol{w}, \quad (9)$$

where $\sigma(t)$ is a monotonically increasing function satisfying $\sigma(0) \approx 0$. In this case, the loss function L_{CDSM} can be written as

$$L_{\text{CDSM}} = \mathbb{E}_{t\sim\mathcal{U}[0,T],\boldsymbol{x}_0\sim p(\boldsymbol{x}),\boldsymbol{y}\sim p(\boldsymbol{y}|\boldsymbol{x}_0),\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_0)}\{ \frac{\lambda(t)}{\sigma^4(t)}\|\boldsymbol{g}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) - \boldsymbol{x}_0\|_2^2\}, \quad (10)$$

where $\boldsymbol{g}_\theta$ is related to $\boldsymbol{s}_\theta$ by

$$\boldsymbol{s}_\theta = \frac{\boldsymbol{g}_\theta - \boldsymbol{x}_t}{\sigma^2(t)}, \quad (11)$$

(see Appendix A.1 for details). Eq.10 implies that $\boldsymbol{g}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t)$ can be trained as a denoising model recovering $\boldsymbol{x}_0$ given $\boldsymbol{x}_t$. This is represented graphically in Fig.2.

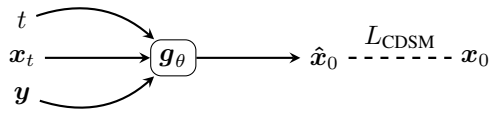

Figure 2. Graphical representation of conditional denoising score matching. A neural network $\boldsymbol{g}_\theta$ is trained to denoise noisy model parameters $\boldsymbol{x}_t$ generated by passing prior samples $\boldsymbol{x}_0$ through the forward SDE Eq.4. The neural network does so given t and observations $\boldsymbol{y}$ generated from $\boldsymbol{x}_0$ by the forward model. It outputs an estimate $\hat{\boldsymbol{x}}_0$ of $\boldsymbol{x}_0$, and the dashed line represents the loss function Eq.10.

3.1.3. Recurrent inference machines

Recurrent Inference Machines (RIMs) are a machine learning framework for solving inverse problems. The task in inverse problems is to reconstruct a signal $\boldsymbol{x}$ from a noisy observation $\boldsymbol{y}$ generated via a forward process

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\eta}, \quad (12)$$

where $\boldsymbol{f}$ is the forward operator and $\boldsymbol{\eta}$ is additive observational noise.

In their original implementation (Putzky & Welling, 2017), RIMs aim to determine the maximum a posteriori (MAP) solution of the inverse problem defined by

$$\boldsymbol{x}_{\text{MAP}} = \max_{\boldsymbol{x}} \log p(\boldsymbol{x}|\boldsymbol{y}). \quad (13)$$

Given the gradient of the negative log likelihood function, $\nabla_{\boldsymbol{x}}\mathcal{L}$, where $\mathcal{L} \equiv -\log p(\boldsymbol{y}|\boldsymbol{x})$, RIMs determine $\boldsymbol{x}_{\text{MAP}}$ within a recurrent neural network framework. In detail, $\boldsymbol{x}_{\text{MAP}}$ is approximated recursively via the update equations

$$\begin{aligned}\hat{\boldsymbol{x}}^{(m+1)} &= \hat{\boldsymbol{x}}^{(m)} + \boldsymbol{g}_\theta(\hat{\boldsymbol{x}}^{(m)}, \boldsymbol{h}^{(m)}, \nabla_{\hat{\boldsymbol{x}}^{(m)}}\mathcal{L}) \\ \boldsymbol{h}^{(m+1)} &= \boldsymbol{g}^*_\theta(\hat{\boldsymbol{x}}^{(m)}, \boldsymbol{h}^{(m)}, \nabla_{\hat{\boldsymbol{x}}^{(m)}}\mathcal{L}),\end{aligned} \quad (14)$$

where $\boldsymbol{g}_\theta$ is a neural network with parameters θ and $\boldsymbol{g}^*_\theta$ is the update model for the hidden state variables $\boldsymbol{h}$, which constitute the RIM's memory and allow it to track its progression through the refinement iterations $m \in \{1,\cdots,M\}$. This is represented graphically in Fig.3.

The RIM parameters θ are determined by minimizing the mean squared error loss function

$$L_{\text{RIM}} = \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}),\boldsymbol{y}\sim p(\boldsymbol{y}|\boldsymbol{x})}\left\{\sum_{m=1}^{M} w_m\|\hat{\boldsymbol{x}}^{(m)} - \boldsymbol{x}\|_2^2\right\}, \quad (15)$$

where $\hat{\boldsymbol{x}}^{(m)}$ is the prediction of the RIM at refinement iteration m, $\boldsymbol{x}$ is the true parameters and w_m is a set of RIM iteration loss weights. During training, the RIM learns an iterative inference algorithm tailored to the considered inverse problem without the need to explicitly specify the prior distribution $p(\boldsymbol{x})$.

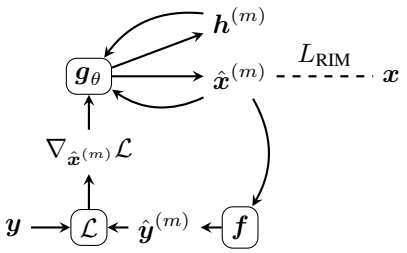

Figure 3. Graphical representation of RIMs. A neural network $\boldsymbol{g}_\theta$ is trained to solve inverse problems by producing point estimates $\hat{\boldsymbol{x}}$ of the model parameters $\boldsymbol{x}$ that produced an observation $\boldsymbol{y}$ through the forward model $\boldsymbol{f}$. The neural network iteratively updates its predictions $\hat{\boldsymbol{x}}$ using gradients of the negative log likelihood function $\mathcal{L} \equiv -\log p(\boldsymbol{y}|\hat{\boldsymbol{x}})$. The dashed line represents the loss function Eq.15.

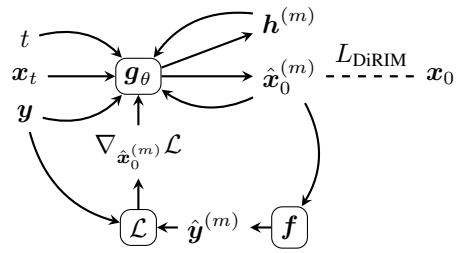

Figure 4. Graphical representation of DiRIM. A neural network $\boldsymbol{g}_\theta$ is trained to denoise noisy model parameters $\boldsymbol{x}_t$ given t and observations $\boldsymbol{y}$. It does so by iteratively updating its prediction $\hat{\boldsymbol{x}}_0$ using the RIM framework. The dashed line represents the loss function Eq.16.

3.2. DiRIM

This section introduces our DiRIM framework and how we apply it to the problem of strong gravitational lensing.

3.2.1. DiRIM framework

The purpose of DiRIM is to combine the ability of diffusion models to sample from complex high-dimensional distributions and the ability of RIMs to solve non-linear inverse problems into a single framework that can sample from complex, possibly highly degenerate, high-dimensional posterior distributions $p(\boldsymbol{x}|\boldsymbol{y})$ for which the forward model $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\eta}$ is non-linear. This is done by creating a diffusion model $\boldsymbol{g}_\theta$ that is itself a RIM , as shown graphically in Fig.4. To train the DiRIM, we use the loss function

$$L_{\text{DiRIM}} = \mathbb{E}_{t\sim\mathcal{U}[0,T],\boldsymbol{x}_0\sim p(\boldsymbol{x}),\boldsymbol{y}\sim p(\boldsymbol{y}|\boldsymbol{x}_0),\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_0)}\Bigg\{ W(t)\sum_{m=1}^{M} w_m \|\hat{\boldsymbol{x}}_0^{(m)} - \boldsymbol{x}_0\|_2^2 \Bigg\}, \quad (16)$$

with our choice of time weight function $W(t)$ detailed in Appendix B. For the choice of RIM iteration loss weights

$$\begin{cases} w_m = 1 & m = M \\ w_m = 0 & \text{else,} \end{cases} \quad (17)$$

the DiRIM loss function L_{DiRIM} in Eq.16 has the same minimizer as L_{CDSM} in Eq.10 (see Appendix A.2 for details). The consequence of this is that one may train the DiRIM as a denoising model via the loss function L_{DiRIM} and subsequently (following Eq.11) obtain posterior scores as

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{y}) = \frac{\boldsymbol{x}_0^{(M)} - \boldsymbol{x}_t}{\sigma^2(t)}. \quad (18)$$

This expression is exact in the limit of infinite data and model capacity, and assumes usage of the VE SDE. We give the equivalent of Eq.18 for the variance preserving (VP) SDE (Song et al., 2020) in Appendix A.3.

To generate samples from the posterior $p(\boldsymbol{x}|\boldsymbol{y})$, we proceed as described in Section 3.1.2, namely, we solve the reverse SDE Eq.7 numerically using the posterior scores obtained from Eq.18 if using the VE SDE or Eq.29 if using the VP SDE. Our choice of SDE and our numerical approach to solving the reverse SDE are detailed in Appendix C.

3.2.2. DiRIM for gravitational lensing

We apply the DiRIM framework to gravitational lensing with the goal of sampling from the posterior $p(\boldsymbol{s}, \boldsymbol{\kappa}|\boldsymbol{y})$. For this purpose we define $\boldsymbol{x}$ as the pair $(\boldsymbol{s}, \log \boldsymbol{\kappa})$, comprising a source image and a convergence map image. We work with the log convergence $\log \boldsymbol{\kappa}$ rather than the convergence $\boldsymbol{\kappa}$ because it performs better empirically. Gradients of the likelihood function, however, are still computed with respect to the convergence map itself. The full computation graph is shown in Fig.1.

Given that the noise $\boldsymbol{\eta}$ in the simulated observations is Gaussian ($\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_y^2 \boldsymbol{I})$), the likelihood function is also Gaussian and given by

$$p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\kappa}) = \mathcal{N}(\boldsymbol{f}(\boldsymbol{s}, \boldsymbol{\kappa}), \sigma_y^2 \boldsymbol{I}). \quad (19)$$

Moreover, the likelihood function is differentiable as the forward operator $\boldsymbol{f}$ is implemented in Caustics, enabling the computation of the likelihood gradients $\nabla_{\boldsymbol{s}}\mathcal{L}$ and $\nabla_{\boldsymbol{\kappa}}\mathcal{L}$.

As a modification to the framework described in the previous section, we compute, at each DiRIM iteration m, the normalized lensed image residuals

$$\frac{\delta \boldsymbol{y}^{(m)}}{\sigma_y} \equiv \frac{\boldsymbol{y} - \hat{\boldsymbol{y}}^{(m)}}{\sigma_y}, \quad (20)$$

and make $\boldsymbol{g}_\theta$ take $\delta \boldsymbol{y}^{(m)}/\sigma_y$ as input in addition to the likelihood gradients $\nabla_{\hat{\boldsymbol{s}}_0^{(m)}}\mathcal{L}$ and $\nabla_{\hat{\boldsymbol{\kappa}}_0^{(m)}}\mathcal{L}$. Additionally, we give $\log t$ as input to $\boldsymbol{g}_\theta$ as opposed to t. Finally, our

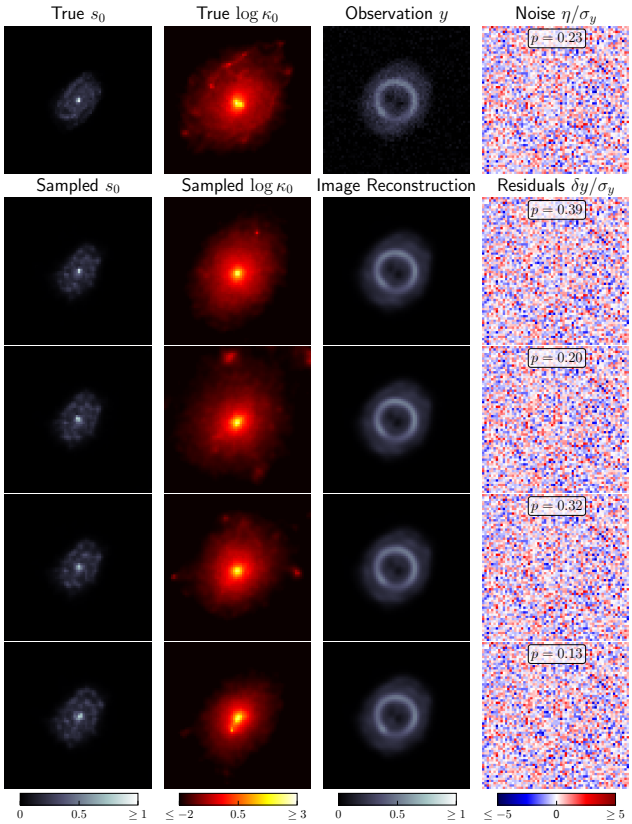

Figure 5. Results for an observation from the test set. We show four joint source and convergence map samples, and the reconstructed lensed image and residuals for each.

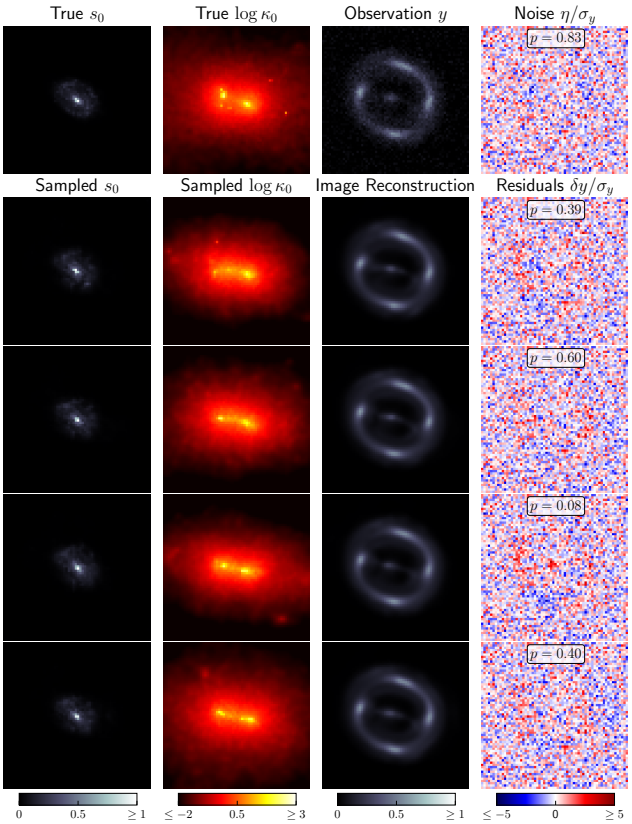

Figure 6. Same as Fig.5 for a more complex convergence map containing two main deflectors.

experiments revealed no benefit to using a hidden state $\boldsymbol{h}$ across RIM iterations and therefore we do not include any. These adaptations are featured in Fig.1.

In this work, the iteratively refined estimates of denoised sources and convergence maps ($\boldsymbol{s}_0^{(m)}$ and $\log \boldsymbol{\kappa}_0^{(m)}$) are initialized as

$$\hat{\boldsymbol{s}}_0^{(0)} = \boldsymbol{s}_t$$
$$\log \hat{\boldsymbol{\kappa}}_0^{(0)} = \log \boldsymbol{\kappa}_t, \tag{21}$$

which takes advantage of the fact that for $t \approx 0$, $\boldsymbol{s}_t$ and $\log \boldsymbol{\kappa}_t$ are close to $\boldsymbol{s}_0$ and $\log \boldsymbol{\kappa}_0$.

In this work, the DiRIM's neural network $\boldsymbol{g}_\theta$ is a U-net (Ronneberger et al., 2015). This is a common choice for diffusion models (see, e.g., (Ho et al., 2020; Song et al., 2020)). We use the U-net architecture discussed in Karras et al. (2022) which features residual blocks, skip connections, and the use of self-attention layers. The images input to $\boldsymbol{g}_\theta$ (namely, $\boldsymbol{s}_t, \log \boldsymbol{\kappa}_t, \boldsymbol{y}, \nabla_{\hat{\boldsymbol{s}}_0^{(m)}} \mathcal{L}, \nabla_{\hat{\boldsymbol{\kappa}}_0^{(m)}} \mathcal{L}, \delta \boldsymbol{y}^{(m)}/\sigma_y, \hat{\boldsymbol{s}}_0^{(m)}, \log \hat{\boldsymbol{\kappa}}_0^{(m)}$) are concatenated channel-wise, and the two output images ($\hat{\boldsymbol{s}}_0^{(m+1)}, \log \hat{\boldsymbol{\kappa}}_0^{(m+1)}$) are produced as two channels. Additional details regarding the neural network architecture and training procedure are discussed in Appendix D.

3.3. Data

In our experiments, the background source brightness distributions are taken from the SKIRT-TNG dataset (Bottrell et al., 2023), which uses dust radiative transfer post-processing with SKIRT (Camps & Baes, 2020) to make a collection of images of simulated galaxies from the IllustrisTNG cosmological magneto-hydrodynamics simulations (Nelson et al., 2019). We use cutouts with a field of view (FOV) of 51.2 kiloparsec (kpc) from the i-band frames at $z = 0$, downsampled to 64×64 pixels, to create a dataset of 42183 images. The images are normalized such that their brightest pixel is uniformly distributed between 0.9 and 1 in flux units of μJy sr^{-1}.

The convergence maps dataset consists of simulations of dark matter and baryonic matter taken from the IllustrisTNG cosmological magneto-hydrodynamics simulations (Nelson et al., 2019). Halos with a dark matter mass of at least $9 \times 10^{11} M_\odot$ are selected, and smoothed convergence maps are created with an adaptive Gaussian kernel. For more details, see (Adam et al., 2023) which used the same convergence maps. We select 42183 of those convergence maps.

The sets of 42183 background source images and convergence maps are each split into a training set (85%), a valida-

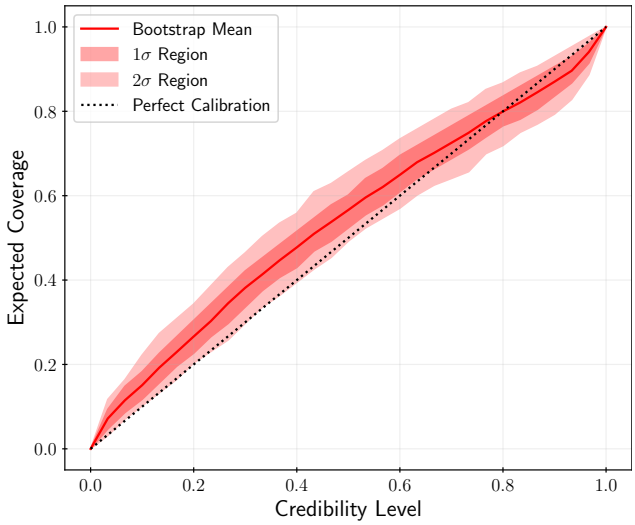

Figure 7. Tests of Accuracy with Random Points (TARP) plot computed on the test set. TARP is a necessary and sufficient condition for a posterior estimator to be unbiased. The plot demonstrates the near-perfect calibration of our posterior samples in pixel-space across the test set.

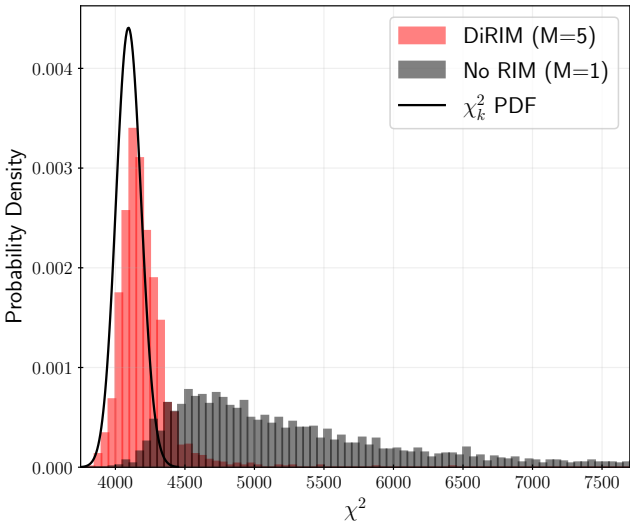

Figure 8. Distribution of χ^2 values computed using simulated observations from the test set and the image reconstructions obtained with posterior samples. The χ^2 value is computed for one posterior sample for each observation in the test set, and the resulting distribution is compared to the χ^2_k probability density function for $k = 64^2$.

tion set (10%) and a test set (5%).

To create the simulated observations, the lens and source redshifts are fixed at $z_l = 0.5$ and $z_s = 1$, respectively, and a flat ΛCDM cosmology with $H_0 = 67.66$ km s^{-1} Mpc^{-1} and $\Omega_m = 0.3097$ is assumed. A Gaussian PSF with a standard deviation of $0.12''$ is used. The lensed images are rendered at a size of 64×64 with a field of view of $7.68''$. We add Gaussian additive noise $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_y^2 \boldsymbol{I})$ with $\sigma_y = 0.03$ to the observations, resulting in a peak SNR uniformly distributed between $0.9/0.03 = 30$ and $1/0.03 = 33.33$.

4. Results and discussion

Fig.5 shows an example of a lens model obtained for an observation from the test set using the methods presented above. The convergence map consists of a single main deflector, making this system easier to model. We show four joint source and convergence map posterior samples. By displaying the normalized residuals between the simulated observation and the image reconstructions, we show that the system is modeled to the noise level. The difference in the size of the source between samples is a demonstration of the ability of DiRIM to sample posteriors with degeneracies, in this case, the degeneracy between the size of the source and the radial slope of the convergence map.

Fig.6 shows another example of a lens model. The convergence map consists of two main deflectors, making this system harder to model. In contrast to our methods, modeling this system with analytical profiles would entail a significant computational cost and the need for heavy user input. The normalized residuals show some structure and the corresponding p-values are reduced in comparison to the observational noise. This indicates that small biases remain in our results, particularly when modeling systems with multiple main deflectors.

In Appendix E, we show ten additional examples of lens models, with observations taken randomly from the test set. The results are consistent with the previous discussion.

To assess the calibration of posterior samples generated with our methods, we show in Fig.7 a TARP (Tests of Accuracy with Random Points) plot. Calibrated TARP is a necessary and sufficient condition for a posterior estimator to be unbiased (Lemos et al., 2023). The test reveals only a small miscalibration, consistent with results discussed previously.

To assess the goodness of fit of the posterior samples generated with our methods, we show in Fig.8 the distribution of χ^2 values computed using simulated observations from the test set and the image reconstructions obtained with posterior samples. We compute the χ^2 value of one posterior sample per observation for all 2110 observations in the test set, and compare the resulting distribution to the theoretical distribution, which is the χ^2_k distribution with $k = 64^2$ degrees of freedom. The small extended tail of the distribution is consistent with the results discussed previously.

We now demonstrate the importance of the iterative refinement of posterior scores via the RIM framework in our methods. We do this by training an additional neural net-

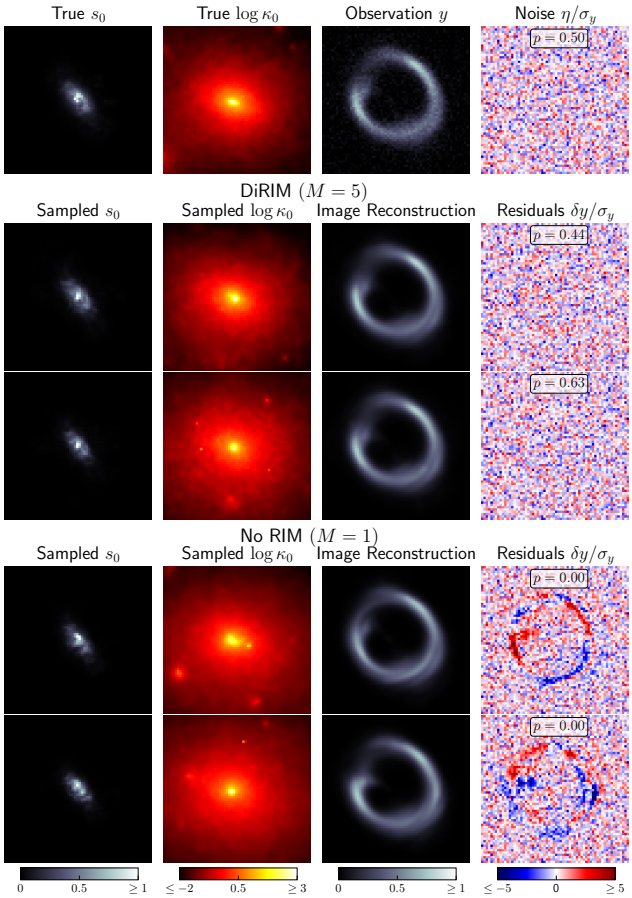

Figure 9. Comparison of results for an observation from the test set with the number of RIM iterations varied between $M = 5$ and $M = 1$. The latter disables the iterative refinement of posterior scores via the RIM framework, which causes the inference to be biased. This demonstrates that the RIM is an essential component of our method.

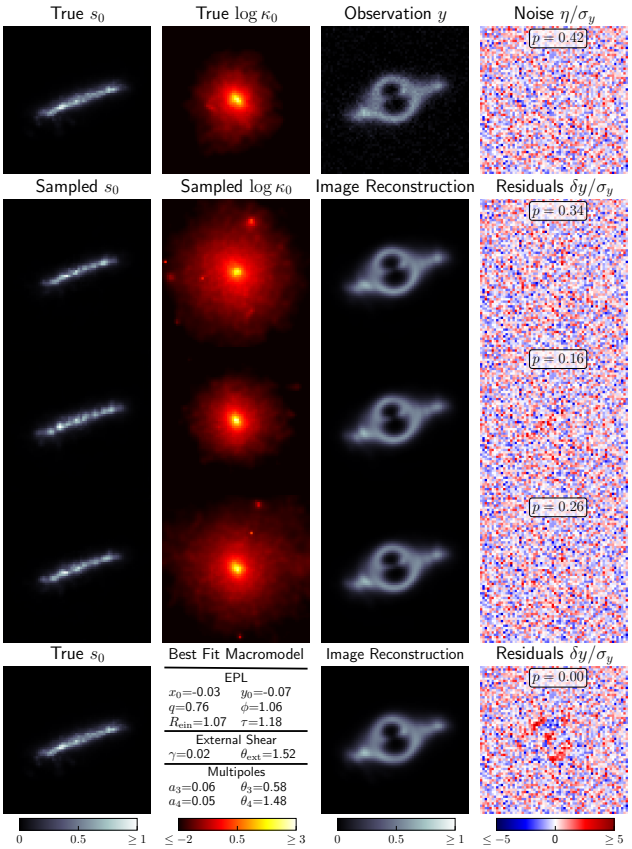

Figure 10. Comparison between our method and a traditional method for an observation from the test set. The traditional method consists of modeling the foreground mass with a low-dimensional parametric macromodel. We fix the source to the ground truth and determine the maximum likelihood macromodel parameters. Even in this idealized case, the lack of flexibility in the foreground mass model prevents modeling the observation down to the noise level.

work with the number of RIM iterations set to $M = 1$, while keeping everything else fixed. This effectively disables the iterative refinement of posterior scores via the RIM framework, making it closely mimic the behavior of baseline conditional denoising score matching (see Sec.3.1.2 and particularly Fig.2). In Fig.9, we model an observation from the test set using this model. The resulting lensed images residuals show significant structure and multiple $\geq 5\sigma_y$ pixels, indicating that the system was not modeled to the noise level. This, in particular, means that baseline conditional denoising score matching fails to accurately model strong gravitational lenses. To further demonstrate this, we add to Fig.8 an additional distribution of χ^2 values obtained from posterior samples of the $M = 1$ model. The distribution is shifted to the right and has a heavily extended tail.

In comparison to methods modeling the foreground mass distribution using parametric models, our method has additional flexibility, which translates into improved modeling of the observations. As an illustrative example, we show in Fig.10 an example of a system for which the convergence map consists of a single deflector, but for which a parametric fit fails to model the observation to the noise level[2]. To demonstrate this, the source image is fixed to the ground truth, and a parametric model comprising an Elliptical Power-Law (EPL) perturbed with external shear and $m = 3, 4$ multipoles is used to model the foreground mass. This gives a total of 12 macromodel parameters, which are described in Appendix F. The parameters of the macromodel are determined by maximizing the likelihood of the reconstructed image given the data. The residuals between the observation and image reconstruction indicate a bias caused by the lack of flexibility of the parametric model. This is in contrast to our method, which models the

[2]We find that a large fraction of systems with a single deflector can be modeled down to the noise level using the parametric model described for the deflector, however we focus here on an example where this is not the case.

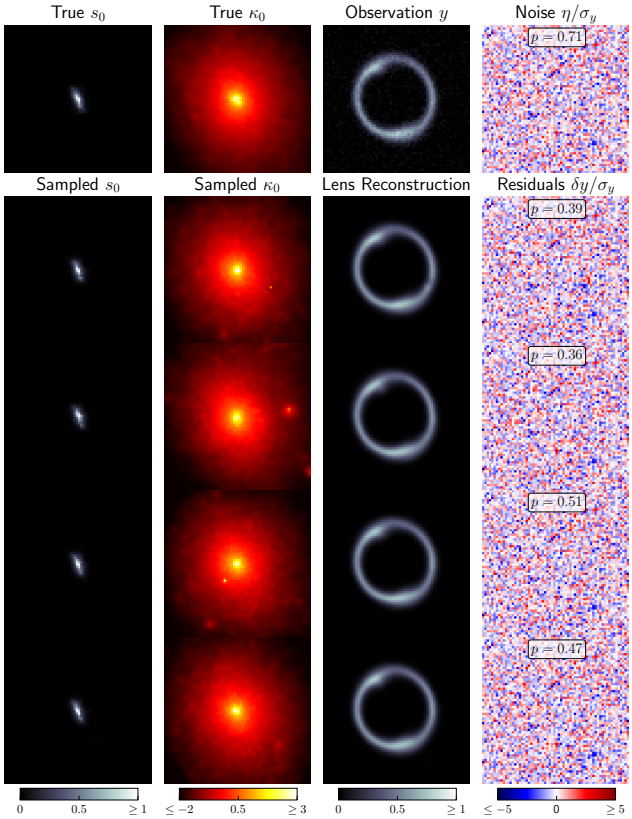

Figure 11. Test of our methods on out-of-distribution (OOD) data. The convergence map is the sum of two randomly selected convergence maps from the test set, assessing the robustness of the model to changes in the physical properties of the convergence maps. We show four joint posterior samples and the corresponding lensed image reconstruction and residuals.

observation down to the noise level.

Finally, to assess the ability of DiRIM to generalize to out-of-distribution (OOD) data, we show in Fig.11 an example of a lens model for which the convergence map is out-of-distribution with respect to the training set. We take the convergence map to be the sum of two randomly selected convergence maps from the test set. This results in a convergence map with two co-located halos that have not interacted with each other. The ability of DiRIM to model this system shows robustness to changes in the physical properties of the convergence maps.

Overall, our results demonstrate that DiRIMs can leverage highly flexible representations of the source and foreground mass to model, in a Bayesian framework, realistic gravitational lensing observations down to the noise level. The small remaining biases present in our results will be the subject of future work.

5. Conclusion

We have presented DiRIM, a first-of-its-kind method for pixel-space joint posterior sampling of the source and foreground mass distribution in strong gravitational lensing. Using simulated data, we showed that it can model realistic observations down to the noise level.

In the future, we plan to apply this framework to other high-dimensional non-linear inverse problems. It is potentially applicable to any inverse problem for which the forward model can be made differentiable and computationally inexpensive. Additionally, motivated by the success of latent diffusion models, we plan to investigate a latent-space formulation of DiRIM, which would allow higher resolution generation at a reduced computational cost.

Software and Data

The source code and scripts used in this work are publicly available as a Python package under the name DiRIM_Lensing[3]. The source and convergence map images used during training and testing are available and hosted on Zenodo[4].

This work made use of the packages CAUSTICS (Stone et al., 2024), PYTORCH (Paszke et al., 2019), NUMPY (Harris et al., 2020), TARP (Lemos et al., 2023), MATPLOTLIB (Hunter, 2007) and PYDANTIC (Colvin et al., 2026).

Acknowledgments

The authors thank Alexandre Adam, Nicolas Payot, Ronan Legin and Huilin Tai for helpful discussions and comments. This work is supported in part by computational resources provided by Calcul Quebec and the Digital Research Alliance of Canada. This work is partially supported by Schmidt Futures, a philanthropic initiative founded by Eric and Wendy Schmidt as part of the Virtual Institute for Astrophysics (VIA). G.P. acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number 601521]. G.M.B. acknowledges support from the Fonds de recherche du Québec – Nature et technologies (FRQNT) under a Doctoral Research Scholarship (doi:10.69777/368273). Y.H. and L.P. acknowledge support from the Canada Research Chairs Program, the National Sciences and Engineering Council of Canada through grants RGPIN-2020-05073 and 05102.

[3] https://github.com/GuillaumePayeur/DiRIM_Lensing
[4] https://zenodo.org/records/19668019

Impact statement

This paper presents work whose goal is to advance the fields of Physics and Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Adam, A., Coogan, A., Malkin, N., Legin, R., Perreault-Levasseur, L., Hezaveh, Y., and Bengio, Y. Posterior samples of source galaxies in strong gravitational lenses with score-based priors. 2022. doi:10.48550/ARXIV.2211.03812. URL https://arxiv.org/abs/2211.03812.

Adam, A., Perreault-Levasseur, L., Hezaveh, Y., and Welling, M. Pixelated reconstruction of foreground density and background surface brightness in gravitational lensing systems using recurrent inference machines. *The Astrophysical Journal*, 951(1):6, 2023. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/accf84. URL https://iopscience.iop.org/article/10.3847/1538-4357/accf84.

Adam, A., Stone, C., Bottrell, C., Legin, R., Hezaveh, Y., and Perreault-Levasseur, L. Echoes in the noise: Posterior samples of faint galaxy surface brightness profiles with score-based likelihoods and priors. *The Astronomical Journal*, 169(5):254, May 2025. ISSN 0004-6256, 1538-3881. doi:10.3847/1538-3881/adb039. URL https://iopscience.iop.org/article/10.3847/1538-3881/adb039.

Anau Montel, N., Coogan, A., Correa, C., Karchev, K., and Weniger, C. Estimating the warm dark matter mass from strong lensing images with truncated marginal neural ratio estimation. *Monthly Notices of the Royal Astronomical Society*, 518(2):2746–2760, November 2022. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stac3215. URL https://academic.oup.com/mnras/article/518/2/2746/6821314.

Anderson, B. D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3): 313–326, May 1982. ISSN 03044149. doi:10.1016/0304-4149(82)90051-5. URL https://linkinghub.elsevier.com/retrieve/pii/0304414982900515.

Andry, G., Lewin, S., Rozet, F., Rochman, O., Mangeleer, V., Pirlet, M., Faulx, E., Grégoire, M., and Louppe, G. Appa: Bending weather dynamics with latent diffusion models for global data assimilation. (arXiv:2504.18720), November 2025. doi:10.48550/arXiv.2504.18720. URL http://arxiv.org/abs/2504.18720. arXiv:2504.18720.

Ballard, D. J., Enzi, W. J. R., Collett, T. E., Turner, H. C., and Smith, R. J. Gravitational imaging through a triple source plane lens: revisiting the λcdm-defying dark subhalo in sdssj0946+1006. *Monthly Notices of the Royal Astronomical Society*, 528(4):7564–7586, February 2024. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stae514. URL https://academic.oup.com/mnras/article/528/4/7564/7611708.

Barco, G. M., Adam, A., Stone, C., Hezaveh, Y., and Perreault-Levasseur, L. Tackling the problem of distributional shifts: Correcting misspecified, high-dimensional data-driven priors for inverse problems. *The Astrophysical Journal*, 980(1):108, February 2025a. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ad9b92. URL https://iopscience.iop.org/article/10.3847/1538-4357/ad9b92.

Barco, G. M., Legin, R., Stone, C., Hezaveh, Y., and Perreault-Levasseur, L. Blind strong gravitational lensing inversion: Joint inference of source and lens mass with score-based models. 2025b. doi:10.48550/ARXIV.2511.04792. URL https://arxiv.org/abs/2511.04792.

Barkana, R. Fast calculation of a family of elliptical gravitational lens models. *The Astrophysical Journal*, 502(2): 531–537, August 1998. ISSN 0004-637X, 1538-4357. doi:10.1086/305950. URL https://iopscience.iop.org/article/10.1086/305950.

Batzolis, G., Stanczuk, J., Schönlieb, C.-B., and Etmann, C. Conditional image generation with score-based diffusion models. (arXiv:2111.13606), November 2021. doi:10.48550/arXiv.2111.13606. URL http://arxiv.org/abs/2111.13606. arXiv:2111.13606.

Bayer, D., Chatterjee, S., Koopmans, L. V. E., Vegetti, S., McKean, J. P., Treu, T., Fassnacht, C. D., and Glazebrook, K. Probing sub-galactic mass structure with the power spectrum of surface-brightness anomalies in high-resolution observations of galaxy–galaxy strong gravitational lenses. ii. observational constraints on the subgalactic matter power spectrum. *Monthly Notices of the Royal Astronomical Society*, 523(1): 1310–1325, May 2023a. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stad1402. URL https://academic.oup.com/mnras/article/523/1/1310/7160470.

Bayer, D., Koopmans, L. V. E., McKean, J. P., Vegetti, S., Treu, T., Fassnacht, C. D., and Glazebrook, K. Probing sub-galactic mass structure with the power spectrum of surface-brightness anomalies in high-resolution observations of galaxy-galaxy strong gravitational lenses – i. power-spectrum measurement and feasibility study.

Monthly Notices of the Royal Astronomical Society, 523 (1):1326–1345, May 2023b. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stad1403. URL https://academic.oup.com/mnras/article/523/1/1326/7160466.

Birrer, S., Millon, M., Sluse, D., Shajib, A. J., Courbin, F., Erickson, S., Koopmans, L. V. E., Suyu, S. H., and Treu, T. Time-delay cosmography: Measuring the hubble constant and other cosmological parameters with strong gravitational lensing. *Space Science Reviews*, 220(5):48, August 2024. ISSN 0038-6308, 1572-9672. doi:10.1007/s11214-024-01079-w. URL https://link.springer.com/10.1007/s11214-024-01079-w.

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., Gavazzi, R., Moustakas, L. A., Wayth, R., and Schlegel, D. J. The sloan lens acs survey. v. the full acs strong-lens sample. *The Astrophysical Journal*, 682 (2):964–984, August 2008. ISSN 0004-637X, 1538-4357. doi:10.1086/589327. URL https://iopscience.iop.org/article/10.1086/589327.

Boruah, S. S., Jacob, M., and Jain, B. Diffusion-based mass map reconstruction from weak lensing data. *Physical Review D*, 111(8):083542, April 2025. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.111.083542. URL https://link.aps.org/doi/10.1103/PhysRevD.111.083542.

Bottrell, C., Yesuf, H. M., Popping, G., Omori, K. C., Tang, S., Ding, X., Pillepich, A., Nelson, D., Eisert, L., Gao, H., Goulding, A. D., Kalita, B. S., Luo, W., Greene, J. E., Shi, J., and Silverman, J. D. IllustrisTNG in the HSC-SSP: image data release and the major role of mini mergers as drivers of asymmetry and star formation. *Monthly Notices of the Royal Astronomical Society*, 527(3):6506–6539, November 2023. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stad2971. URL https://academic.oup.com/mnras/article/527/3/6506/7286659.

Bourdin, A., Legin, R., Ho, M., Adam, A., Hezaveh, Y., and Perreault-Levasseur, L. Inpainting galaxy counts onto n-body simulations over multiple cosmologies and astrophysics. (arXiv:2408.00839), August 2024. doi:10.48550/arXiv.2408.00839. URL http://arxiv.org/abs/2408.00839. arXiv:2408.00839.

Camps, P. and Baes, M. Skirt 9: Redesigning an advanced dust radiative transfer code to allow kinematics, line transfer and polarization by aligned dust grains. *Astronomy and Computing*, 31:100381, April 2020. ISSN 22131337. doi:10.1016/j.ascom.2020.100381. URL https://linkinghub.elsevier.com/retrieve/pii/S2213133720300354.

Chen, N., Zhang, Y., Zen, H., Weiss, R. J., Norouzi, M., and Chan, W. Wavegrad: Estimating gradients for waveform generation. (arXiv:2009.00713), October 2020. doi:10.48550/arXiv.2009.00713. URL http://arxiv.org/abs/2009.00713. arXiv:2009.00713.

Colaço, L. R., Holanda, R. F. L., Santana, Z. C., and Silva, R. A joint analysis of strong lensing and type ia supernovae to determine the hubble constant. *The European Physical Journal C*, 85(5):577, May 2025. ISSN 1434-6052. doi:10.1140/epjc/s10052-025-14315-6. URL https://doi.org/10.1140/epjc/s10052-025-14315-6.

Colvin, S., Jolibois, E., Ramezani, H., Garcia Badaracco, A., Dorsey, T., Montague, D., Matveenko, S., Trylesinski, M., Runkle, S., Hewitt, D., Hall, A., and Plot, V. Pydantic Validation, March 2026. URL https://github.com/pydantic/pydantic.

Cuesta-Lazaro, C. and Mishra-Sharma, S. A point cloud approach to generative modeling for galaxy surveys at the field level. (arXiv:2311.17141), December 2024. doi:10.48550/arXiv.2311.17141. URL http://arxiv.org/abs/2311.17141. arXiv:2311.17141.

Dia, N., Yantovski-Barth, M. J., Adam, A., Bowles, M., Lemos, P., Scaife, A. M. M., Hezaveh, Y., and Perreault-Levasseur, L. Bayesian imaging for radio interferometry with score-based priors. (arXiv:2311.18012), November 2023. doi:10.48550/arXiv.2311.18012. URL http://arxiv.org/abs/2311.18012. arXiv:2311.18012.

Fan, X., Wang, F., Yang, J., Keeton, C. R., Yue, M., Zabludoff, A., Bian, F., Bonaglia, M., Georgiev, I. Y., Hennawi, J. F., Li, J., McGreer, I. D., Naidu, R., Pacucci, F., Rabien, S., Thompson, D., Venemans, B., Walter, F., Wang, R., and Wu, X.-B. The discovery of a gravitationally lensed quasar at z = 6.51. *The Astrophysical Journal Letters*, 870(2):L11, January 2019. ISSN 2041-8205, 2041-8213. doi:10.3847/2041-8213/aaeffe. URL https://iopscience.iop.org/article/10.3847/2041-8213/aaeffe.

Galan, A., Vernardos, G., Minor, Q., Sluse, D., Van De Vyvere, L., and Gomer, M. Exploiting the diversity of modeling methods to probe systematic biases in strong lensing analyses. *Astronomy & Astrophysics*, 692:A87, December 2024. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202451095. URL https://www.aanda.org/10.1051/0004-6361/202451095.

Gilman, D., Birrer, S., Nierenberg, A., Treu, T., Du, X., and Benson, A. Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses. *Monthly Notices of the Royal Astronomical*

Society, 491(4):6077–6101, February 2020. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stz3480. URL https://academic.oup.com/mnras/article/491/4/6077/5673494.

Gilman, D., Bovy, J., Treu, T., Nierenberg, A., Birrer, S., Benson, A., and Sameie, O. Strong lensing signatures of self-interacting dark matter in low-mass haloes. *Monthly Notices of the Royal Astronomical Society*, 507 (2):2432–2447, August 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab2335. URL https://academic.oup.com/mnras/article/507/2/2432/6350574.

Grillo, C., Suyu, S. H., Rosati, P., Mercurio, A., Balestra, I., Munari, E., Nonino, M., Caminha, G. B., Lombardi, M., De Lucia, G., Borgani, S., Gobat, R., Biviano, A., Girardi, M., Umetsu, K., Coe, D., Koekemoer, A. M., Postman, M., Zitrin, A., Halkola, A., Broadhurst, T., Sartoris, B., Presotto, V., Annunziatella, M., Maier, C., Fritz, A., Vanzella, E., and Frye, B. Clash-vlt: Insights on the mass substructures in the frontier fields cluster macs j0416.1–2403 through accurate strong lens modeling. *The Astrophysical Journal*, 800(1):38, February 2015. ISSN 1538-4357. doi:10.1088/0004-637X/800/1/38. URL https://iopscience.iop.org/article/10.1088/0004-637X/800/1/38.

Harris, C. R., Millman, K. J., Walt, S. J. v. d., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. v., Brett, M., Haldane, A., Río, J. F. d., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with numpy. (arXiv:2006.10256), 2020. doi:10.48550/arXiv.2006.10256. URL http://arxiv.org/abs/2006.10256. arXiv:2006.10256.

Hezaveh, Y. D., Dalal, N., Marrone, D. P., Mao, Y.-Y., Morningstar, W., Wen, D., Blandford, R. D., Carlstrom, J. E., Fassnacht, C. D., Holder, G. P., Kemball, A., Marshall, P. J., Murray, N., Levasseur, L. P., Vieira, J. D., and Wechsler, R. H. Detection of lensing substructure using alma observations of the dusty galaxy sdp.81. *The Astrophysical Journal*, 823(1):37, May 2016. ISSN 0004-637X, 1538-4357. doi:10.3847/0004-637X/823/1/37. URL https://iopscience.iop.org/article/10.3847/0004-637X/823/1/37.

Hezaveh, Y. D., Levasseur, L. P., and Marshall, P. J. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669): 555–557, August 2017. ISSN 0028-0836, 1476-4687. doi:10.1038/nature23463. URL https://www.nature.com/articles/nature23463.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. 2020. doi:10.48550/ARXIV.2006.11239. URL https://arxiv.org/abs/2006.11239.

Hsueh, J.-W., Enzi, W., Vegetti, S., Auger, M. W., Fassnacht, C. D., Despali, G., Koopmans, L. V. E., and McKean, J. P. Sharp – vii. new constraints on the dark matter free-streaming properties and substructure abundance from gravitationally lensed quasars. *Monthly Notices of the Royal Astronomical Society*, 492(2): 3047–3059, February 2020. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stz3177. URL https://academic.oup.com/mnras/article/492/2/3047/5626368.

Huang, X., Baltasar, S., Ratier-Werbin, N., Storfer, C., Sheu, W., Agarwal, S., Tamargo-Arizmendi, M., Schlegel, D., Aguilar, J., Ahlen, S., Aldering, G., Banka, S., BenZvi, S., Bianchi, D., Bolton, A., Brooks, D., Cikota, A., Claybaugh, T., De La Macorra, A., Dey, A., Doel, P., Edelstein, J., Filipp, A., Forero-Romero, J. E., Gaztañaga, E., Gontcho A Gontcho, S., Gu, A., Gutierrez, G., Honscheid, K., Jullo, E., Juneau, S., Kehoe, R., Kirkby, D., Kisner, T., Kremin, A., Kwon, K., Lambert, A., Landriau, M., Lang, D., Le Guillou, L., Liu, J., Meisner, A., Miquel, R., Moustakas, J., Myers, A. D., Perlmutter, S., Pérez-Ràfols, I., Prada, F., Rossi, G., Rubin, D., Sanchez, E., Schubnell, M., Shu, Y., Silver, E., Sprayberry, D., Suzuki, N., Tarlé, G., Weaver, B., and Zou, H. Desi strong lens foundry. i. hst observations and modeling with giga-lens. *The Astrophysical Journal*, 998(1):69, February 2026. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ae22d2. URL https://iopscience.iop.org/article/10.3847/1538-4357/ae22d2.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. ISSN 1521-9615. doi:10.1109/MCSE.2007.55. URL http://ieeexplore.ieee.org/document/4160265/.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005. URL https://api.semanticscholar.org/CorpusID:1152227.

Jacobs, C., Collett, T., Glazebrook, K., McCarthy, C., Qin, A. K., Abbott, T. M. C., Abdalla, F. B., Annis, J., Avila, S., Bechtol, K., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., da Costa, L. N., Davis, C., De Vicente, J., Desai, S., Diehl, H. T., Doel, P., Eifler, T. F., Flaugher, B., Frieman, J., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Goldstein, D. A., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hartley, W. G., Hollowood, D. L., Honscheid, K., Hoyle, B., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., Li, T. S.,

Lima, M., Lin, H., Maia, M. A. G., Martini, P., Miller, C. J., Miquel, R., Nord, B., Plazas, A. A., Sanchez, E., Scarpine, V., Schubnell, M., Serrano, S., Sevilla-Noarbe, I., Smith, M., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Vikram, V., Walker, A. R., Zhang, Y., Zuntz, J., and Collaboration, D. Finding high-redshift strong lenses in des using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 484(4):5330–5349, April 2019. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stz272. URL https://academic.oup.com/mnras/article/484/4/5330/5301418.

Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models. (arXiv:2206.00364), October 2022. doi:10.48550/arXiv.2206.00364. URL http://arxiv.org/abs/2206.00364. arXiv:2206.00364.

Kazerouni, A., Aghdam, E. K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., and Merhof, D. Diffusion models for medical image analysis: A comprehensive survey. 2022. doi:10.48550/ARXIV.2211.07804. URL https://arxiv.org/abs/2211.07804.

Kloeden, P. E. and Platen, E. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Springer, Berlin, Heidelberg, 1992. doi:10.1007/978-3-662-12616-5.

Knabel, S., Holwerda, B. W., Nightingale, J., Treu, T., Bilicki, M., Brough, S., Driver, S., Finnerty, L., Haberzettl, L., Hegde, S., Hopkins, A. M., Kuijken, K., Liske, J., Pimbblet, K. A., Steele, R. C., and Wright, A. H. Modelling strong lenses from wide-field ground-based observations in kids and gama. *Monthly Notices of the Royal Astronomical Society*, 520(1): 804–827, January 2023. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stad133. URL https://academic.oup.com/mnras/article/520/1/804/6987696.

Koopmans, L. V. E. Gravitational imaging of cold dark matter substructures: Gravitational imaging of cdm substructures. *Monthly Notices of the Royal Astronomical Society*, 363(4):1136–1144, October 2005. ISSN 00358711. doi:10.1111/j.1365-2966.2005.09523.x. URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2005.09523.x.

Lai, C.-H., Song, Y., Kim, D., Mitsufuji, Y., and Ermon, S. The principles of diffusion models. (arXiv:2510.21890), October 2025. doi:10.48550/arXiv.2510.21890. URL http://arxiv.org/abs/2510.21890. arXiv:2510.21890.

Legin, R., Ho, M., Lemos, P., Perreault-Levasseur, L., Ho, S., Hezaveh, Y., and Wandelt, B. Posterior sampling of the initial conditions of the universe from non-linear large scale structures using score-based generative models. *Monthly Notices of the Royal Astronomical Society: Letters*, 527(1):L173–L178, October 2023. ISSN 1745-3925, 1745-3933. doi:10.1093/mnrasl/slad152. URL https://academic.oup.com/mnrasl/article/527/1/L173/7313644.

Legin, R., Stone, C., Adam, A., Barco, G. M., Coogan, A., Malkin, N., Perreault-Levasseur, L., and Hezaveh, Y. Mind the information gap: Unveiling detailed morphologies of z 0.5-1.0 galaxies with slacs strong lenses and data-driven analysis. 2025. doi:10.48550/ARXIV.2511.19595. URL https://arxiv.org/abs/2511.19595.

Lemos, P., Coogan, A., Hezaveh, Y., and Perreault-Levasseur, L. Sampling-based accuracy testing of posterior estimators for general inference. (arXiv:2302.03026), 2023. doi:10.48550/arXiv.2302.03026. URL http://arxiv.org/abs/2302.03026. arXiv:2302.03026.

Lines, N. E. P., Li, T., Collett, T. E., Holloway, P., Nightingale, J. W., Rojas, K., Verma, A., and Walmsley, M. The revolution in strong lensing discoveries from euclid. *Nature Astronomy*, 9(8):1116–1122, August 2025. ISSN 2397-3366. doi:10.1038/s41550-025-02616-5. URL https://www.nature.com/articles/s41550-025-02616-5.

Lønning, K., Putzky, P., Sonke, J.-J., Reneman, L., Caan, M. W. A., and Welling, M. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis*, 53:64–78, April 2019. ISSN 1361-8423. doi:10.1016/j.media.2019.01.005.

Marques-Chaves, R., Pérez-Fournon, I., Shu, Y., Martínez-Navajas, P. I., Bolton, A. S., Kochanek, C. S., Oguri, M., Zheng, Z., Mao, S., Montero-Dorta, A. D., Cornachione, M. A., and Brownstein, J. R. Discovery of a very bright and intrinsically very luminous, strongly lensed lyα emitting galaxy at z = 2.82 in the boss emission-line lens survey*. *The Astrophysical Journal Letters*, 834(2):L18, January 2017. ISSN 2041-8205, 2041-8213. doi:10.3847/2041-8213/834/2/L18. URL https://iopscience.iop.org/article/10.3847/2041-8213/834/2/L18.

Marques-Chaves, R., Pérez-Fournon, I., Gavazzi, R., Martínez-Navajas, P. I., Riechers, D., Rigopoulou, D., Cabrera-Lavers, A., Clements, D. L., Cooray, A., Farrah, D., Ivison, R. J., Jiménez-Ángel, C. E., Nayyeri, H., Oliver, S., Omont, A., Scott, D., Shu, Y., and Wardlow, J. The strong gravitationally lensed herschel galaxy hlock01: Optical spectroscopy reveals a close galaxy

merger with evidence of inflowing gas. *The Astrophysical Journal*, 854(2):151, February 2018. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/aaabb7. URL https://iopscience.iop.org/article/10.3847/1538-4357/aaabb7.

Marques-Chaves, R., Pérez-Fournon, I., Shu, Y., Colina, L., Bolton, A., Álvarez Márquez, J., Brownstein, J., Cornachione, M., Geier, S., Jiménez-Ángel, C., Kojima, T., Mao, S., Montero-Dorta, A., Oguri, M., Ouchi, M., Poidevin, F., Shirley, R., and Zheng, Z. Rest-frame uv properties of luminous strong gravitationally lensed lyα emitters from the bells gallery survey. *Monthly Notices of the Royal Astronomical Society*, 492(1):1257–1278, February 2020. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stz3500. URL https://academic.oup.com/mnras/article/492/1/1257/5675639.

Maruyama, G. Continuous markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48–90, 1955. doi:10.1007/BF02846028.

Meneghetti, M. Introduction to gravitational lensing: With python examples. *Introduction to Gravitational Lensing: With Python Examples*, 2022. URL https://ui.adsabs.harvard.edu/abs/2022iglp.book.....M/abstract.

Minor, Q., Gad-Nasr, S., Kaplinghat, M., and Vegetti, S. An unexpected high concentration for the dark substructure in the gravitational lens sdssj0946+1006. *Monthly Notices of the Royal Astronomical Society*, 507 (2):1662–1683, August 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab2247. URL https://academic.oup.com/mnras/article/507/2/1662/6342121.

Modi, C., Lanusse, F., Seljak, U., Spergel, D. N., and Perreault-Levasseur, L. Cosmicrim: Reconstructing early universe by combining differentiable simulations with recurrent inference machines. 2021. doi:10.48550/ARXIV.2104.12864. URL https://arxiv.org/abs/2104.12864.

Morningstar, W. R., Hezaveh, Y. D., Levasseur, L. P., Blandford, R. D., Marshall, P. J., Putzky, P., and Wechsler, R. H. Analyzing interferometric observations of strong gravitational lenses with recurrent and convolutional neural networks, July 2018. URL http://arxiv.org/abs/1808.00011. arXiv:1808.00011.

Morningstar, W. R., Levasseur, L. P., Hezaveh, Y. D., Blandford, R., Marshall, P., Putzky, P., Rueter, T. D., Wechsler, R., and Welling, M. Data-driven Reconstruction of Gravitationally Lensed Galaxies Using Recurrent Inference Machines. *The Astrophysical Journal*, 883(1):14, September 2019. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ab35d7. URL https://iopscience.iop.org/article/10.3847/1538-4357/ab35d7.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B., and Hernquist, L. The illustristng simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, December 2019. ISSN 2197-7909. doi:10.1186/s40668-019-0028-x. URL https://link.springer.com/10.1186/s40668-019-0028-x.

Nightingale, J., Mahler, G., McCleary, J., He, Q., Hogg, N. B., Amvrosiadis, A., Gozaliasl, G., Mercier, W., Scognamiglio, D., Berman, E., Leroy, G., Liu, D., Massey, R. J., Shuntov, M., von Wietersheim-Kramsta, M., Franco, M., Paquereau, L., Ilbert, O., Allen, N., Toft, S., Akins, H. B., Casey, C. M., Kartaltepe, J. S., Koekemoer, A. M., McCracken, H. J., Rhodes, J. D., Robertson, B. E., Drakos, N. E., Faisst, A. L., and Jin, S. The cosmos-web lens survey (cowls) i: Discovery of >100 high redshift strong lenses in contiguous jwst imaging. 2025. doi:10.48550/ARXIV.2503.08777. URL https://arxiv.org/abs/2503.08777.

Nightingale, J. W. and Dye, S. Adaptive semi-linear inversion of strong gravitational lens imaging. *Monthly Notices of the Royal Astronomical Society*, 452(3): 2940–2959, 2015. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stv1455. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stv1455.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. (arXiv:1912.01703), December 2019. doi:10.48550/arXiv.1912.01703. URL http://arxiv.org/abs/1912.01703. arXiv:1912.01703.

Powell, D. M., Vegetti, S., McKean, J. P., White, S. D. M., Ferreira, E. G. M., May, S., and Spingola, C. A lensed radio jet at milli-arcsecond resolution – ii. constraints on fuzzy dark matter from an extended gravitational arc. *Monthly Notices of the Royal Astronomical Society: Letters*, 524(1):L84–L88, 2023. ISSN 1745-3925, 1745-3933. doi:10.1093/mnrasl/slad074. URL https://academic.oup.com/mnrasl/article/524/1/L84/7197445.

Putzky, P. and Welling, M. Recurrent inference machines for solving inverse problems. (arXiv:1706.04008), 2017.

doi:10.48550/arXiv.1706.04008. URL http://arxiv.org/abs/1706.04008. arXiv:1706.04008.

Qi, J.-Z., Zhao, J.-W., Cao, S., Biesiada, M., and Liu, Y. Measurements of the hubble constant and cosmic curvature with quasars: ultracompact radio structure and strong gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 503(2):2179–2186, March 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab638. URL https://academic.oup.com/mnras/article/503/2/2179/6159476.

Remy, B., Lanusse, F., Jeffrey, N., Liu, J., Starck, J.-L., Osato, K., and Schrabback, T. Probabilistic mass-mapping with neural score estimation. *Astronomy & Astrophysics*, 672:A51, April 2023. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202243054. URL https://www.aanda.org/10.1051/0004-6361/202243054.

Rhea, C., Hlavacek-Larrondo, J., Adam, A., Kraft, R., Bogdan, A., Perreault-Levasseur, L., and Prunier, M. Deconvolving X-ray Galaxy Cluster Spectra Using a Recurrent Inference Machine, September 2024. URL http://arxiv.org/abs/2409.10711. arXiv:2409.10711.

Riveros, J. K., Saavedra, P. A., Hortúa, H. J., García-Farieta, J. E., and Olier, I. Conditional diffusion-flow models for generating 3d cosmic density fields: applications to f (r) cosmologies. *Machine Learning: Science and Technology*, 6(3):035031, 2025. ISSN 2632-2153. doi:10.1088/2632-2153/adf8b1. URL https://iopscience.iop.org/article/10.1088/2632-2153/adf8b1.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. (arXiv:1505.04597), May 2015. doi:10.48550/arXiv.1505.04597. URL http://arxiv.org/abs/1505.04597. arXiv:1505.04597.

Rozet, F., Andry, G., Lanusse, F., and Louppe, G. Learning diffusion priors from observations by expectation maximization. (arXiv:2405.13712), October 2025. doi:10.48550/arXiv.2405.13712. URL http://arxiv.org/abs/2405.13712. arXiv:2405.13712.

Sabidussi, E., Klein, S., Caan, M., Bazrafkan, S., Den Dekker, A., Sijbers, J., Niessen, W., and Poot, D. Recurrent inference machines as inverse problem solvers for MR relaxometry. *Medical Image Analysis*, 74:102220, December 2021. ISSN 13618415. doi:10.1016/j.media.2021.102220. URL https://linkinghub.elsevier.com/retrieve/pii/S1361841521002656.

Saha, P., Coles, J., Macciò, A. V., and Williams, L. L. R. The hubble time inferred from 10 time delay lenses. *The Astrophysical Journal*, 650(1):L17–L20, October 2006. ISSN 0004-637X, 1538-4357. doi:10.1086/507583. URL https://iopscience.iop.org/article/10.1086/507583.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. Image super-resolution via iterative refinement. (arXiv:2104.07636), 2021. doi:10.48550/arXiv.2104.07636. URL http://arxiv.org/abs/2104.07636. arXiv:2104.07636.

Saharia, C., Chan, W., Chang, H., Lee, C. A., Ho, J., Salimans, T., Fleet, D. J., and Norouzi, M. Palette: Image-to-image diffusion models. (arXiv:2111.05826), May 2022. doi:10.48550/arXiv.2111.05826. URL http://arxiv.org/abs/2111.05826. arXiv:2111.05826.

Savary, E., Rojas, K., Maus, M., Clément, B., Courbin, F., Gavazzi, R., Chan, J. H. H., Lemon, C., Vernardos, G., Cañameras, R., Schuldt, S., Suyu, S. H., Cuillandre, J.-C., Fabbro, S., Gwyn, S., Hudson, M. J., Kilbinger, M., Scott, D., and Stone, C. Strong lensing in unions: Toward a pipeline from discovery to modeling. *Astronomy & Astrophysics*, 666:A1, October 2022. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202142505. URL https://www.aanda.org/10.1051/0004-6361/202142505.

Shajib, A. J., Birrer, S., Treu, T., Agnello, A., Buckley-Geer, E. J., Chan, J. H. H., Christensen, L., Lemon, C., Lin, H., Millon, M., Poh, J., Rusu, C. E., Sluse, D., Spiniello, C., Chen, G. C.-F., Collett, T., Courbin, F., Fassnacht, C. D., Frieman, J., Galan, A., Gilman, D., More, A., Anguita, T., Auger, M. W., Bonvin, V., McMahon, R., Meylan, G., Wong, K. C., Abbott, T. M. C., Annis, J., Avila, S., Bechtol, K., Brooks, D., Brout, D., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Costanzi, M., da Costa, L. N., De Vicente, J., Desai, S., Dietrich, J. P., Doel, P., Drlica-Wagner, A., Evrard, A. E., Finley, D. A., Flaugher, B., Fosalba, P., García-Bellido, J., Gerdes, D. W., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hollowood, D. L., Honscheid, K., Huterer, D., James, D. J., Jeltema, T., Krause, E., Kuropatkin, N., Li, T. S., Lima, M., MacCrann, N., Maia, M. A. G., Marshall, J. L., Melchior, P., Miquel, R., Ogando, R. L. C., Palmese, A., Paz-Chinchón, F., Plazas, A. A., Romer, A. K., Roodman, A., Sako, M., Sanchez, E., Santiago, B., Scarpine, V., Schubnell, M., Scolnic, D., Serrano, S., Sevilla-Noarbe, I., Smith, M., Soares-Santos, M., Suchyta, E., Tarle, G., Thomas, D., Walker, A. R., and Zhang, Y. Strides: a 3.9 per cent measurement of the hubble constant from the strong lens system des j0408-5354. *Monthly Notices of the Royal Astronomical Society*, 494(4):6072–6102, 2020. ISSN

0035-8711, 1365-2966. doi:10.1093/mnras/staa828. URL https://academic.oup.com/mnras/article/494/4/6072/5813265.

Shajib, A. J., Mozumdar, P., Chen, G. C.-F., Treu, T., Cappellari, M., Knabel, S., Suyu, S. H., Bennert, V. N., Frieman, J. A., Sluse, D., Birrer, S., Courbin, F., Fassnacht, C. D., Villafaña, L., and Williams, P. R. Tdcosmo: Xii. improved hubble constant measurement from lensing time delays using spatially resolved stellar kinematics of the lens galaxy. *Astronomy & Astrophysics*, 673:A9, May 2023. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202345878. URL https://www.aanda.org/10.1051/0004-6361/202345878.

Shajib, A. J., Vernardos, G., Collett, T. E., Motta, V., Sluse, D., Williams, L. L. R., Saha, P., Birrer, S., Spiniello, C., and Treu, T. Strong lensing by galaxies. *Space Science Reviews*, 220(8):87, November 2024. ISSN 1572-9672. doi:10.1007/s11214-024-01105-x. URL https://doi.org/10.1007/s11214-024-01105-x.

Shajib, A. J., Smith, G. P., Birrer, S., Verma, A., Arendse, N., Collett, T., Daylan, T., Serjeant, S., and Collaboration, L. S. L. S. Strong gravitational lenses from the vera c. rubin observatory. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 383(2295):20240117, May 2025. ISSN 1364-503X, 1471-2962. doi:10.1098/rsta.2024.0117. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2024.0117.

Shang, S., Shan, Z., Liu, G., Wang, L., Wang, X., Zhang, Z., and Zhang, J. Resdiff: Combining cnn and diffusion model for image super-resolution. 2023. doi:10.48550/ARXIV.2303.08714. URL https://arxiv.org/abs/2303.08714.

Sharief, S., Zeghal, J., Barco, G. M., Lemos, P., Hezaveh, Y., and Perreault-Levasseur, L. Mira: A score for conditional distribution accuracy and model comparison. (arXiv:2605.02014), May 2026. doi:10.48550/arXiv.2605.02014. URL http://arxiv.org/abs/2605.02014. arXiv:2605.02014.

Shu, X., Yang, L., Liu, D., Wang, W.-H., Wang, T., Han, Y., Huang, X., Lim, C.-F., Chang, Y.-Y., Zheng, W., Zheng, X., Wang, J., and Kong, X. A census of optically dark massive galaxies in the early universe from magnification by lensing galaxy clusters. *The Astrophysical Journal*, 926(2):155, February 2022a. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac3de5. URL https://iopscience.iop.org/article/10.3847/1538-4357/ac3de5.

Shu, Y., Cañameras, R., Schuldt, S., Suyu, S. H., Taubenberger, S., Inoue, K. T., and Jaelani, A. T. Holismokes: Viii. high-redshift, strong-lens search in the hyper suprime-cam subaru strategic program. *Astronomy & Astrophysics*, 662:A4, 2022b. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202243203. URL https://www.aanda.org/10.1051/0004-6361/202243203.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. 2019. doi:10.48550/ARXIV.1907.05600. URL https://arxiv.org/abs/1907.05600.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv e-prints*, pp. arXiv:2011.13456, November 2020. doi:10.48550/arXiv.2011.13456. URL https://ui.adsabs.harvard.edu/abs/2020arXiv201113456S/abstract.

Song, Y., Shen, L., Xing, L., and Ermon, S. Solving inverse problems in medical imaging with score-based generative models. 2021. doi:10.48550/ARXIV.2111.08005. URL https://arxiv.org/abs/2111.08005.

Sonnenfeld, A. On the choice of lens density profile in time delay cosmography. *Monthly Notices of the Royal Astronomical Society*, 474(4):4648–4659, March 2018. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stx3105. URL http://academic.oup.com/mnras/article/474/4/4648/4683265.

Stone, C., Adam, A., Coogan, A., Yantovski-Barth, M. J., Filipp, A., Setiawan, L., Core, C., Legin, R., Wilson, C., Barco, G. M., Hezaveh, Y., and Perreault-Levasseur, L. Caustics: A python package for accelerated strong gravitational lensing simulations. *Journal of Open Source Software*, 9(103):7081, November 2024. ISSN 2475-9066. doi:10.21105/joss.07081. URL https://joss.theoj.org/papers/10.21105/joss.07081.

Stone, C., Legin, R., Adam, A., Malkin, N., Barco, G. M., Perreault-Levasseur, L., and Hezaveh, Y. Pixellated posterior sampling of point spread functions in astronomical images. 2025. doi:10.48550/ARXIV.2511.19594. URL https://arxiv.org/abs/2511.19594.

Suyu, S. H., Marshall, P. J., Hobson, M. P., and Blandford, R. D. A bayesian analysis of regularized source inversions in gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 371(2):983–998, 2006. ISSN 0035-8711, 1365-2966. doi:10.1111/j.1365-2966.2006.10733.x. URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2006.10733.x.

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., and Gavazzi, R. Detection of a dark substructure through gravitational imaging. 2009. doi:10.48550/ARXIV.0910.0760. URL https://arxiv.org/abs/0910.0760.

Vegetti, S., Birrer, S., Despali, G., Fassnacht, C. D., Gilman, D., Hezaveh, Y., Perreault Levasseur, L., McKean, J. P., Powell, D. M., O'Riordan, C. M., and Vernardos, G. Strong gravitational lensing as a probe of dark matter. *Space Science Reviews*, 220(5):58, August 2024. ISSN 0038-6308, 1572-9672. doi:10.1007/s11214-024-01087-w. URL https://link.springer.com/10.1007/s11214-024-01087-w.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Computation*, 23 (7):1661–1674, 2011. ISSN 0899-7667, 1530-888X. doi:10.1162/NECO_a_00142. URL https://direct.mit.edu/neco/article/23/7/1661-1674/7677.

Warren, S. J. and Dye, S. Semilinear gravitational lens inversion. *The Astrophysical Journal*, 590(2):673–682, 2003. ISSN 0004-637X, 1538-4357. doi:10.1086/375132. URL https://iopscience.iop.org/article/10.1086/375132.

Wong, K. C., Suyu, S. H., Chen, G. C.-F., Rusu, C. E., Millon, M., Sluse, D., Bonvin, V., Fassnacht, C. D., Taubenberger, S., Auger, M. W., Birrer, S., Chan, J. H. H., Courbin, F., Hilbert, S., Tihhonova, O., Treu, T., Agnello, A., Ding, X., Jee, I., Komatsu, E., Shajib, A. J., Sonnenfeld, A., Blandford, R. D., Koopmans, L. V. E., Marshall, P. J., and Meylan, G. H0licow – xiii. a 2.4 per cent measurement of h0 from lensed quasars: 5.3σ tension between early- and late-universe probes. *Monthly Notices of the Royal Astronomical Society*, 498(1):1420–1439, October 2020. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stz3094. URL https://academic.oup.com/mnras/article/498/1/1420/5849454.

Wu, Z., Sun, Y., Chen, Y., Zhang, B., Yue, Y., and Bouman, K. L. Principled probabilistic imaging using diffusion models as plug-and-play priors. (arXiv:2405.18782), November 2024. doi:10.48550/arXiv.2405.18782. URL http://arxiv.org/abs/2405.18782. arXiv:2405.18782.

Xu, D., Sluse, D., Gao, L., Wang, J., Frenk, C., Mao, S., Schneider, P., and Springel, V. How well can cold dark matter substructures account for the observed radio flux-ratio anomalies. *Monthly Notices of the Royal Astronomical Society*, 447(4):3189–3206, March 2015. ISSN 1365-2966, 0035-8711. doi:10.1093/mnras/stu2673. URL http://academic.oup.com/mnras/article/447/4/3189/1749822/How-well-can-cold-dark-matter-substructures.

Xue, Z., Li, Y., Patel, Y., and Regier, J. Diffusion models for probabilistic deconvolution of galaxy images. 2023. doi:10.48550/ARXIV.2307.11122. URL https://arxiv.org/abs/2307.11122.

A. derivations concerning diffusion

A.1. Relation between score matching and denoising

For the VE SDE Eq.9, one can show (Lai et al., 2025) that the perturbation kernel $p(\boldsymbol{x}_t|\boldsymbol{x}_0)$ is given by

$$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{x}_0, \sigma^2(t)I), \tag{22}$$

and thus

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) = \frac{\boldsymbol{x}_0 - \boldsymbol{x}_t}{\sigma^2(t)}. \tag{23}$$

Denoting

$$\mathbb{E} \equiv \mathbb{E}_{t\sim\mathcal{U}[0,T],\boldsymbol{x}_0\sim p(\boldsymbol{x}),\boldsymbol{y}\sim p(\boldsymbol{y}|\boldsymbol{x}_0),\boldsymbol{x}_t\sim p(\boldsymbol{x}_t|\boldsymbol{x}_0)}, \tag{24}$$

we can thus write the loss L_{CDSM} in Eq.8 as

$$\begin{aligned} L_{\text{CDSM}} &= \mathbb{E}\left\{\lambda(t)\left\|\boldsymbol{s}_\theta(\boldsymbol{x}_t, \boldsymbol{y}, t) - \frac{\boldsymbol{x}_0 - \boldsymbol{x}_t}{\sigma^2(t)}\right\|_2^2\right\} \\ &= \mathbb{E}\left\{\frac{\lambda(t)}{\sigma^4(t)}\|\boldsymbol{s}_\theta(\boldsymbol{x}_t, \boldsymbol{y}, t)\sigma^2(t) + \boldsymbol{x}_t - \boldsymbol{x}_0\|_2^2\right\} \\ &= \mathbb{E}\left\{\frac{\lambda(t)}{\sigma^4(t)}\|\boldsymbol{g}_\theta(\boldsymbol{x}_t, \boldsymbol{y}, t) - \boldsymbol{x}_0\|_2^2\right\}, \end{aligned} \tag{25}$$

which is Eq.10.

A.2. Relation between DiRIM loss and CDSM loss

The minimizer of the DiRIM loss Eq.16, using again the notation of Eq.24, is

$$\operatorname*{argmin}_\theta L_{\text{DiRIM}} = \operatorname*{argmin}_\theta \mathbb{E}\left\{W(t)\sum_{m=1}^{M} w_m \|\hat{\boldsymbol{x}}_0^{(m)} - \boldsymbol{x}_0\|_2^2\right\}. \tag{26}$$

For the choice of weights w_m in Eq.17, this reads

$$\begin{aligned} \operatorname*{argmin}_\theta L_{\text{DiRIM}} &= \operatorname*{argmin}_\theta \mathbb{E}\{W(t)\|\hat{\boldsymbol{x}}_0^{(M)} - \boldsymbol{x}_0\|_2^2\} \\ &= \operatorname*{argmin}_\theta \mathbb{E}\left\{\frac{\lambda(t)}{\sigma^4(t)}\|\hat{\boldsymbol{x}}_0^{(M)} - \boldsymbol{x}_0\|_2^2\right\} \\ &= \operatorname*{argmin}_\theta L_{\text{CDSM}}, \end{aligned} \tag{27}$$

where the last line follows from identifying $\boldsymbol{g}_\theta(\boldsymbol{x}_t, \boldsymbol{y}, t) \equiv \hat{\boldsymbol{x}}_0^{(M)}$.

A.3. DiRIM with the VP SDE

For the variance preserving (VP) SDE (Song et al., 2020) defined by

$$d\boldsymbol{x} = -\frac{1}{2}\beta(t)\boldsymbol{x}dt + \sqrt{\beta(t)}d\boldsymbol{w}, \tag{28}$$

the equivalent of Eq.18 is

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{y}) = \frac{\boldsymbol{x}_0^{(M)}\alpha(t) - \boldsymbol{x}_t}{\sigma^2(t)}, \tag{29}$$

where $\alpha(t)$ and $\sigma(t)$ are defined as

$$\begin{aligned}\alpha(t) &= \exp\Big(-\frac{1}{2}\int_0^t \beta(s)ds\Big)\\ \sigma(t) &= \sqrt{1-\alpha^2(t)}.\end{aligned} \tag{30}$$

To see this, consider that for the VP SDE Eq.28, the perturbation kernel $p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$ is given by (Lai et al., 2025)

$$p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{x}_0\alpha(t), \sigma^2(t)I), \tag{31}$$

with $\alpha(t)$ and $\sigma(t)$ defined in Eq.30. Thus,

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{x}_0) = \frac{\boldsymbol{x}_0\alpha(t) - \boldsymbol{x}_t}{\sigma^2(t)}. \tag{32}$$

Using again the notation of Eq.24, we can thus write the loss L_{CDSM} in Eq.8 as

$$\begin{aligned}L_{\text{CDSM}} &= \mathbb{E}\Big\{\lambda(t)\Big\|\boldsymbol{s}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) - \frac{\boldsymbol{x}_0\alpha(t)-\boldsymbol{x}_t}{\sigma^2(t)}\Big\|_2^2\Big\}\\ &= \mathbb{E}\Big\{\frac{\lambda(t)\alpha^2(t)}{\sigma^4(t)}\Big\|\frac{\boldsymbol{s}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t)\sigma^2(t)+\boldsymbol{x}_t}{\alpha(t)} - \boldsymbol{x}_0\Big\|_2^2\Big\}\\ &= \mathbb{E}\Big\{\frac{\lambda(t)\alpha^2(t)}{\sigma^4(t)}\Big\|\boldsymbol{g}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) - \boldsymbol{x}_0\Big\|_2^2\Big\},\end{aligned} \tag{33}$$

where $\boldsymbol{g}_\theta$ and $\boldsymbol{s}_\theta$ are related by

$$\boldsymbol{s}_\theta = \frac{\boldsymbol{g}_\theta\alpha(t) - \boldsymbol{x}_t}{\sigma^2(t)}. \tag{34}$$

As in the VE case, identifying $\boldsymbol{g}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) \equiv \hat{\boldsymbol{x}}_0^{(M)}$, we conclude that we can obtain posterior scores via

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{y}) = \frac{\boldsymbol{x}_0^{(M)}\alpha(t) - \boldsymbol{x}_t}{\sigma^2(t)}, \tag{35}$$

which is Eq.29.

B. Choice of diffusion time weights in the loss function

In this work, the function $W(t)$ in the loss function Eq.16 is chosen by considering two factors:

1. For $t \approx 1$, the denoising neural network $g_\theta(\boldsymbol{x}_t, \boldsymbol{y}, t)$ is mostly informed by the observation $\boldsymbol{y}$ over the noisy source and convergence map $\boldsymbol{x_t}$, since the latter is almost pure noise. Hence, the function $W(t)$ is chosen to be constant near $t = 1$.

2. For $t \approx 0$, the denoising neural network $g_\theta(\boldsymbol{x}_t, \boldsymbol{y}, t)$ is mostly informed by the noisy source and convergence map $\boldsymbol{x}_t$ over the observation $\boldsymbol{y}$, since the former are almost noiseless. Hence, the function $W(t)$ is chosen to satisfy $W(t) \propto \alpha(t)^2/\sigma(t)^2$, which is standard for unconditional denoising models (see e.g., Song et al. 2020).

$W(t)$ is therefore set to

$$W(t) = \frac{\bar{\sigma}^2}{\min\{\bar{\sigma}^2, \sigma(t)^2/\alpha^2(t)\}}, \tag{36}$$

where $\bar{\sigma}$ is a free parameter which governs at which point in time $W(t)$ transitions between its large t constant phase and its low t phase. We set $\bar{\sigma} = 2 \times 10^{-2}$ since we found that it performs best empirically.

C. SDE implementation details

In this work, we use the VP SDE Eq.28 with the noise schedule $\beta(t)$ given by

$$\beta(t) = \beta_{\text{min}}\Big(\frac{\beta_{\text{max}}}{\beta_{\text{min}}}\Big)^t, \tag{37}$$

with $\beta_{\text{min}} = 2 \times 10^{-3}$ and $\beta_{\text{max}} = 250$. To avoid numerical instabilities at $t = 0$ during training and sampling (Song et al., 2020), we train the DiRIM only on the time interval $[\epsilon, 1]$ with $\epsilon = 10^{-3}$.

To solve the reverse SDE from $t = 1$ to $t = \epsilon$, we use the Euler-Maruyama (EM) method (Maruyama, 1955; Kloeden & Platen, 1992) with 1000 steps. After these steps, we pass the sampled source and convergence map through the denoising model to obtain the finalized samples. This effectively projects the samples from $t = \epsilon$ to $t = 0$ in one step.

D. U-net architecture and training procedure

In this work, the architecture of the U-net $\boldsymbol{g}_\theta$ is the one discussed in Karras et al. 2022. We adapt its implementation found in the EDM[5] Python package. Our choice of hyperparameters when training the model is detailed in Tab.1.

During training and at inference time, the likelihood gradients $\nabla_{\hat{\boldsymbol{s}}_0^{(m)}}\mathcal{L}$ and $\nabla_{\hat{\boldsymbol{\kappa}}_0^{(m)}}\mathcal{L}$ ($\nabla_s\mathcal{L}$ and $\nabla_\kappa\mathcal{L}$ for short) are

[5] https://github.com/nvlabs/edm

Table 1. U-net hyperparameters used in this work.

Hyperparameter	Value
Levels in encoder & decoder	5
Channels per level	64, 128, 256, 512, 512
Residual blocks per level	2
Resolutions with attention	4, 8
Number of weights	112 264 706

Table 2. RIM hyperparameters used in this work.

Hyperparameter	Value
Number of RIM iterations (M)	5
Iteration loss weights (w_m)	$\{0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$
Gradients normalization (c_s, c_κ)	$(100, 100)$

pre-processed with a $\tanh$ function as

$$\nabla_s \mathcal{L} \to \tanh(\nabla_s \mathcal{L}/c_s) \tag{38}$$

$$\nabla_\kappa \mathcal{L} \to \tanh(\nabla_\kappa \mathcal{L}/c_\kappa), \tag{39}$$

where c_s and c_κ are normalization factors treated as hyperparameters. Their value and other RIM hyperparameters are detailed in Tab.2

During training, the weights of the neural network $\boldsymbol{g}_\theta$ are updated using backpropagation through time (BPTT). Some sections of the computation graph are not recorded for BPTT to avoid instabilities from log-domain backpropagation. This is shown in the full unrolled computation graph in Fig.12. In addition, we use a reduced number of RIM iterations ($M = 2$) for the first few epochs of training, since failure to do so resulted in instabilities during some of our experiments.

During training, the source images and convergence maps from the training sets are not paired; instead, they are shuffled at every epoch. Additionally, both source and convergence map images are randomly rotated and mirrored during training. Together, these two choices prevent overfitting. Once the learning rate has decayed to 1×10^{-6}, we fix the learning rate and turn on exponential moving average (EMA) of the weights, and continue training until convergence of the validation loss. Hyperparameters relating to the training of the model are detailed in Tab. 3.

The training of the model required 72 hours on a single A100 40GB GPU. At inference time, the generation of joint source and kappa map samples required approximately 5 seconds per sample on the same GPU.

E. Additional Reconstructions

Figs.13-22 show ten examples of lens models with observations taken randomly from the test set.

Table 3. Training hyperparameters used in this work.

Hyperparameter	Value
Optimizer	Adam
Initial Learning rate	2×10^{-4}
Learning rate schedule	Exponential decay
Decay rate per epoch	0.96
Batch size	16
Gradient norm clipping	10
Weight EMA decay rate per epoch	0.9999

F. Foreground mass macromodel

The macromodel for the foreground mass distribution used for comparison purposes in Fig.10 consists of an elliptical power law (EPL) (Barkana, 1998) main halo with convergence

$$\kappa_{\text{EPL}}(\theta_x, \theta_y) = \frac{2-\tau}{2}\left(\frac{R_E\sqrt{q}}{\sqrt{q^2\theta_x^2 + \theta_y^2}}\right)^\tau, \tag{40}$$

parameterized by the Einstein radius $R_E['']$, the density slope τ and the axis ratio q. In addition, the EPL component is rotated by an angle ϕ. The EPL main halo is augmented with an external shear, which has zero convergence and is implemented by its reduced deflection angle

$$\boldsymbol{\alpha}(\theta_x, \theta_y) = \gamma \begin{pmatrix} \theta_x \cos(2\theta_{\text{ext}}) + \theta_y \sin(2\theta_{\text{ext}}) \\ \theta_x \sin(2\theta_{\text{ext}}) - \theta_y \cos(2\theta_{\text{ext}}) \end{pmatrix}, \tag{41}$$

parameterized by the external shear amplitude γ and its orientation θ_{ext}. The EPL main halo is also augmented with $m = 3$ and $m = 4$ multipoles with convergence

$$\kappa_{\text{multipoles}}(\theta_x, \theta_y) = \sum_{m \in \{3,4\}} \frac{a_m}{2r} \cos(m(\psi - \theta_m)), \tag{42}$$

where $r = \sqrt{\theta_x^2 + \theta_y^2}$ and $\psi = \arctan(\theta_y/\theta_x)$. The multipoles are parameterized by the amplitudes $a_m[('')^{-1}]$ and the orientations θ_m. The EPL, external shear, and multipole components are given common center coordinates parameterized by $(\boldsymbol{x}_0['']$, $\boldsymbol{y}_0['''])$. This represents a total of 12 macromodel parameters.

When using this macromodel, the convergence map is not rendered on a grid; instead, the deflection angles are evaluated analytically at the required angular positions during ray tracing. We have, however, also conducted experiments where the convergence map was rendered on a 64×64 grid with various levels of upsampling beforehand. This did not improve the fit to the data in Fig.10. This confirms that the poor fit is caused by the lack of flexibility in the foreground mass model.

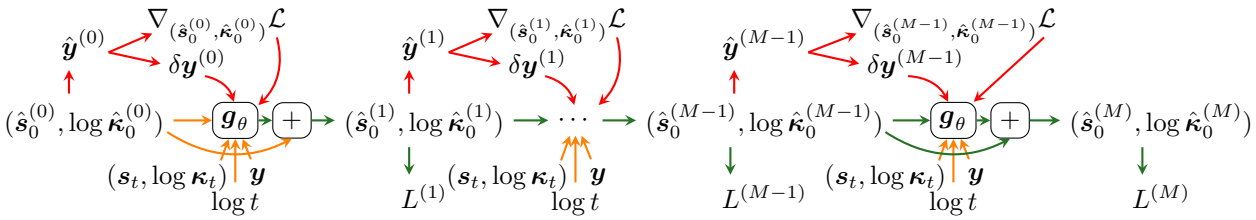

Figure 12. Unrolled computation graph of the DiRIM. During training, loss gradients with respect to the weights are backpropagated through the green edges only. The orange edges do not contribute to the gradient of the loss function with respect to model parameters. Gradients are not backpropagated through the red edges to avoid instabilities from log-domain backpropagation.

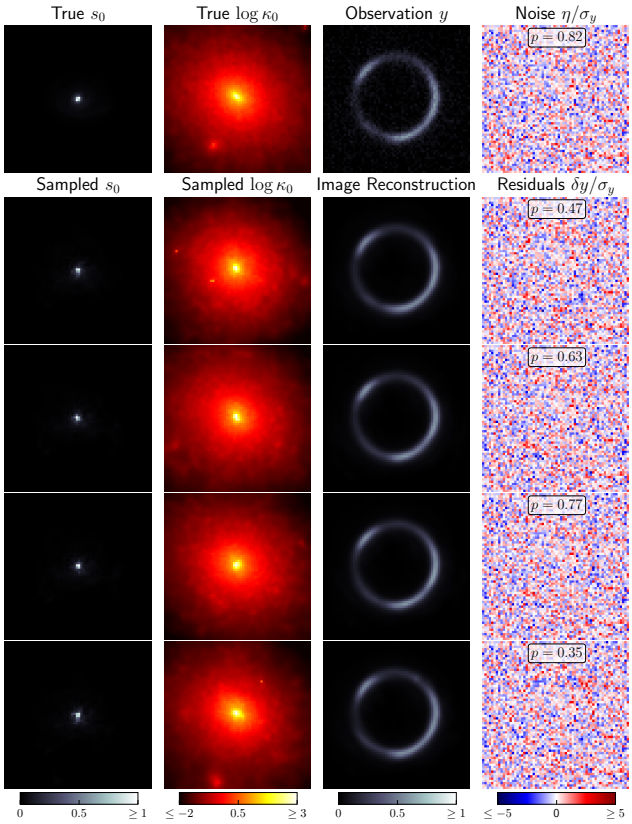

Figure 13. Same as Fig.5 for a randomly selected test set example.

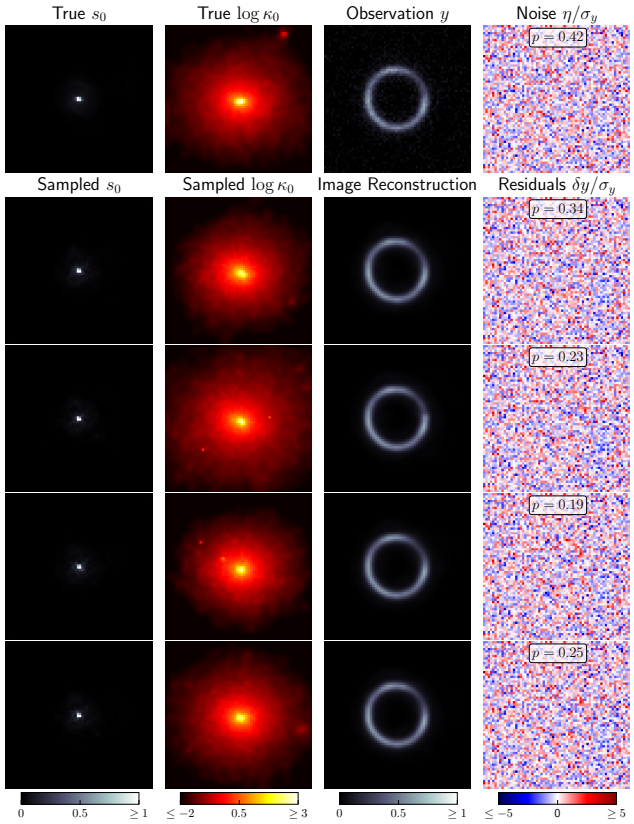

Figure 14. Same as Fig.5 for a randomly selected test set example.

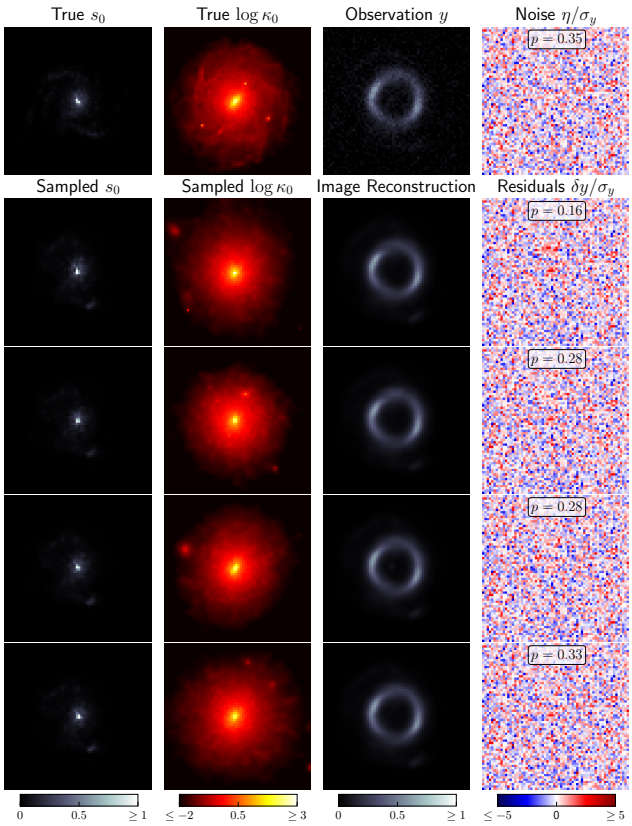

Figure 15. Same as Fig.5 for a randomly selected test set example.

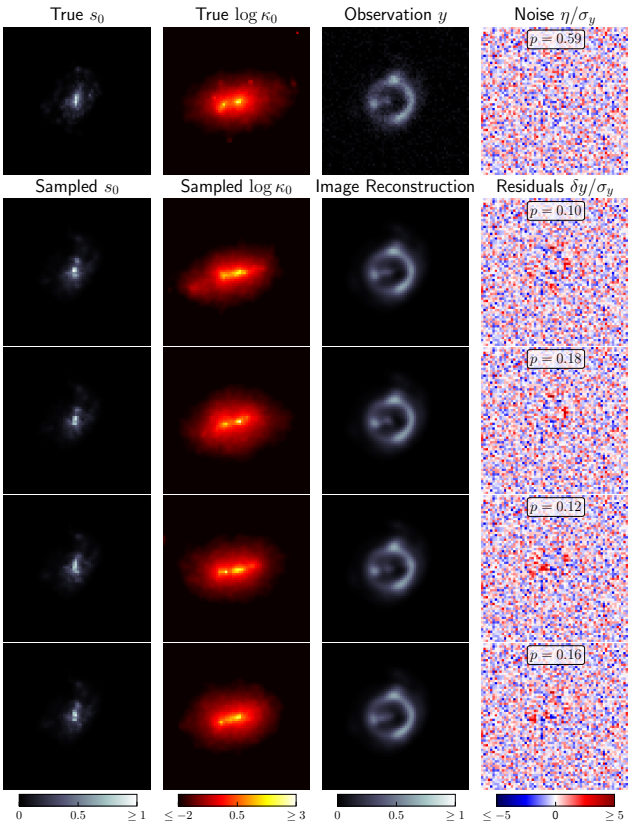

Figure 16. Same as Fig.5 for a randomly selected test set example.

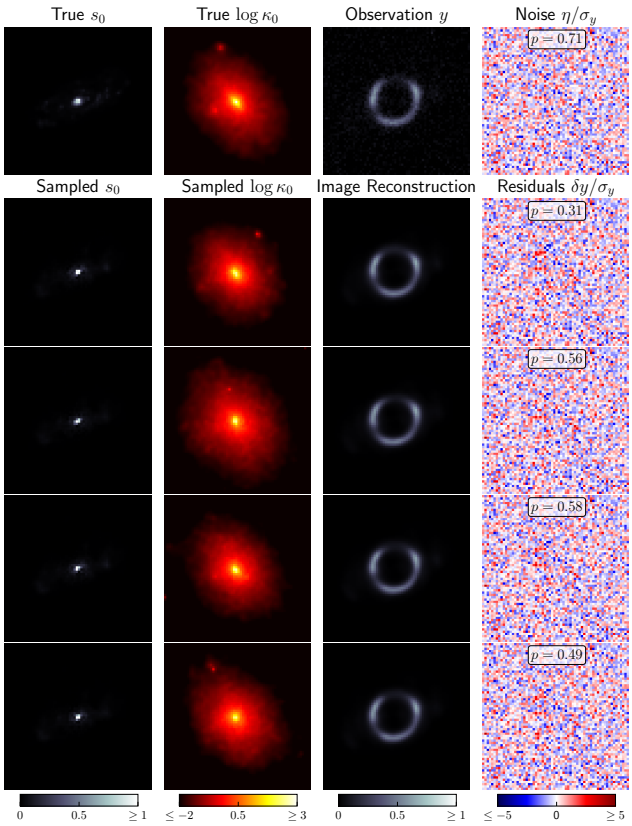

Figure 17. Same as Fig.5 for a randomly selected test set example.

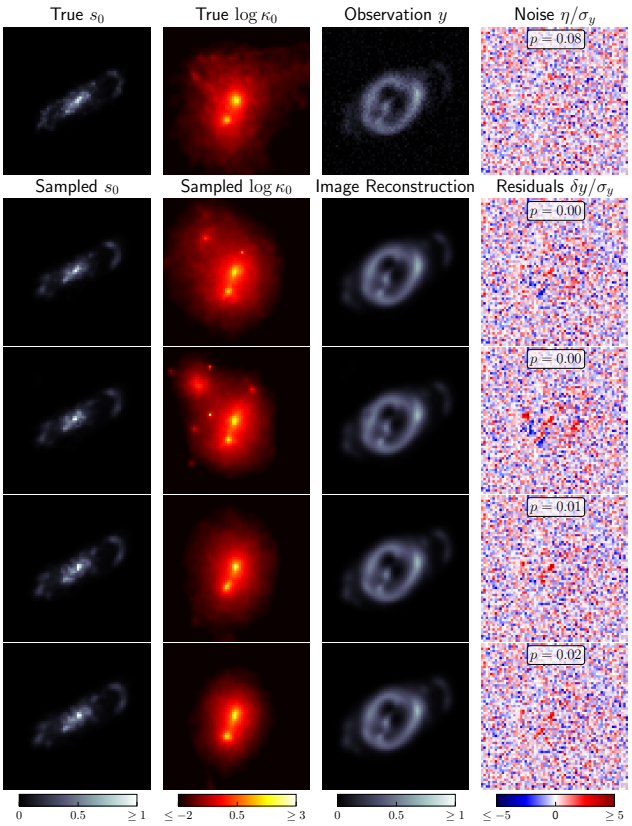

Figure 18. Same as Fig.5 for a randomly selected test set example.

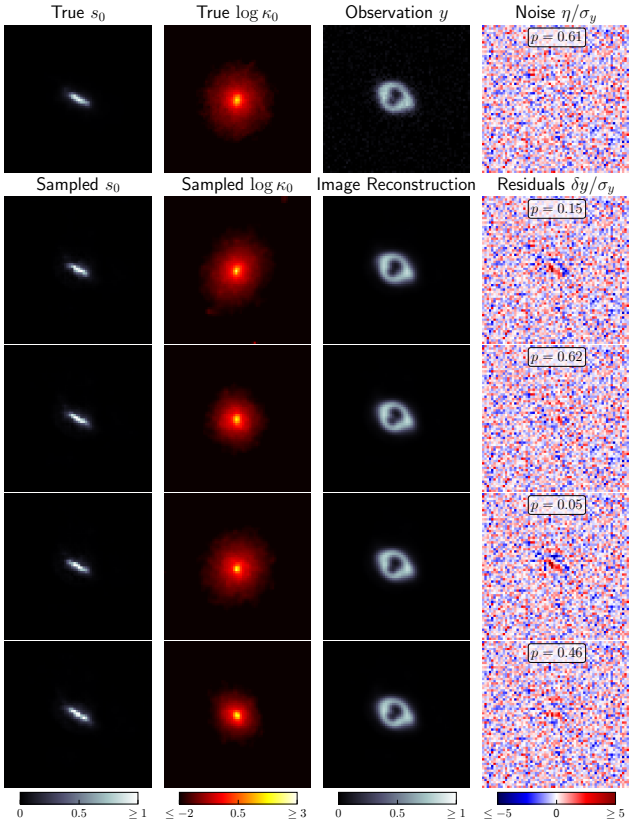

Figure 19. Same as Fig.5 for a randomly selected test set example.

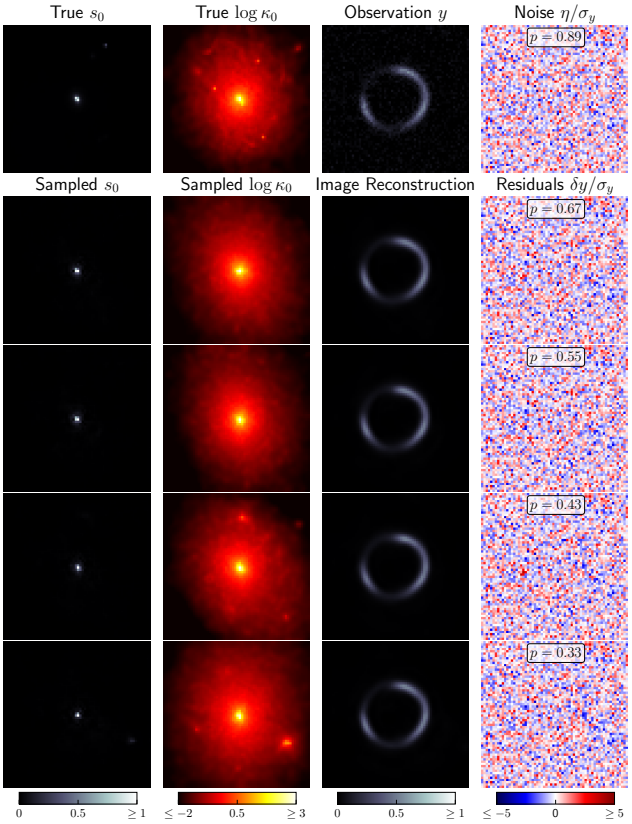

Figure 20. Same as Fig.5 for a randomly selected test set example.

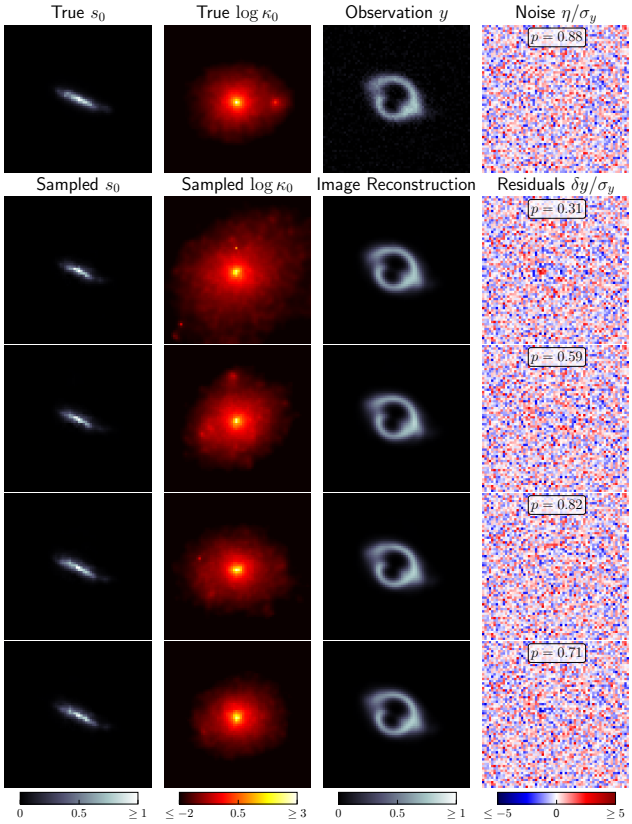

Figure 21. Same as Fig.5 for a randomly selected test set example.

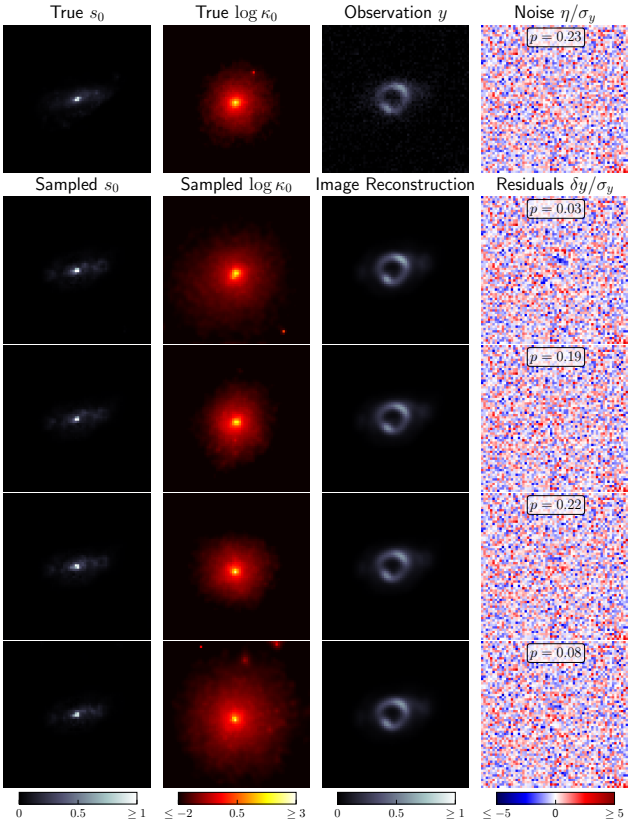

Figure 22. Same as Fig.5 for a randomly selected test set example.