Blind Strong Gravitational Lensing Inversion: Joint Inference of Source and Lens Mass with Score-Based Models

Gabriel Missael Barco[1,2,3] **Ronan Legin**[1,2,3] **Connor Stone**[1,2,3]
Yashar Hezaveh[1,2,3,4,5,6] **Laurence Perreault-Levasseur**[1,2,3,4,5,6]
[1]Université de Montréal [2]Mila [3]Ciela Institute [4]CCA, Flatiron Institute
[5]Perimeter Institute for Theoretical Physics [6]Trottier Space Institute
{gabriel.missael.barco,ronan.legin,connor.stone,yashar.hezaveh,
laurence.perreault.levasseur}@umontreal.ca

Abstract

Score-based models can serve as expressive, data-driven priors for scientific inverse problems. In strong gravitational lensing, they enable posterior inference of a background galaxy from its distorted, multiply-imaged observation. Previous work, however, assumes that the lens mass distribution (and thus the forward operator) is known. We relax this assumption by jointly inferring the source and a parametric lens-mass profile, using a sampler based on GibbsDDRM but operating in continuous time. The resulting reconstructions yield residuals consistent with the observational noise, and the marginal posteriors of the lens parameters recover true values without systematic bias. To our knowledge, this is the first successful demonstration of joint source-and-lens inference with a score-based prior.

1 Introduction

Score-based models have been successfully applied as data-driven, expressive priors for inverse problems. For example, in the field of astrophysics, they have been used for interferometric imaging [e.g. 1, 2], strong gravitational lensing source reconstruction [e.g. 3, 4], cosmological-field inference [e.g. 5–7], deconvolution [e.g. 8, 9], and many other applications.

These applications depend on approximations or heuristics, since posterior sampling with score-based priors is, in general, intractable. Existing methods can be broadly classified into four categories [10]: guidance-based methods [e.g. 3, 11, 12], variable splitting [e.g. 13], variational Bayes [e.g. 1, 14], and sequential Monte Carlo [e.g. 15]. In this work, we focus on the first category, in which an approximate likelihood term $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t)$ is used to guide the diffusion of the prior.

Most of these approaches assume that the parameters of the forward model are known, which is often not the case in practice. Jointly inferring the operator and the underlying parameters of interest (also known as *blind inversion* in the literature [e.g. 16, 17]) is an active area of research. Several approaches approximate blind-inversion sampling with diffusion models; for example, GibbsDDRM [18], BlindDPS [19], BIRD [20], and Fast Diffusion EM [21].

Strong gravitational lensing, which describes the formation of multiple images of background sources due to the bending of their light by the mass of intervening objects, can be modeled using score-based priors for the background source [e.g. 3, 4]. Such score-based priors have not previously been applied to the problem of jointly inferring the background source and lens mass distribution. Strong lens inversion is particularly challenging in the blind scheme, as the posteriors of parametric lenses generally contain several local minima and exhibit degeneracies between the lens parameters and the

Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

source [e.g. 22, 23]. Hence, joint inference has only been possible with analytical source priors that impose Gaussian or *smoothness* assumptions [e.g. 24, 25].

Strong-lensing observations can enable many sciences, for example measuring H_0 via time-delay cosmography [e.g. 26], studying high-redshift objects [e.g. 27], and detecting dark matter subhalos [e.g. 28, 29], among other applications. Furthermore, upcoming wide-field surveys, most notably the Rubin Observatory Legacy Survey of Space and Time (LSST) and the *Euclid* space telescope, are expected to observe about 200 000 strongly lensed systems [30]. Advancing strong lens modeling is therefore crucial to extract the full scientific value of this wealth of data. In this paper, we present preliminary results on a new framework for analyzing strong lenses with score-based priors. Our contributions are:

- We explore two likelihood score approximations, CLA [3] and ΠGDM [11], for source reconstruction.
- We adapt GibbsDDRM [18] to the continuous-time regime, successfully applying it to blind strong-lensing inversion.
- We provide empirical evidence that our approach yields residuals consistent with the noise distribution and that the lens marginal posteriors are unbiased.

In Section 2 we introduce the method, assumptions, and approximations; in Section 3 we present experiments on simulated data; and we discuss limitations, outline next steps, and conclude in Section 4 and Section 5.

2 Methods

2.1 Strong gravitational lensing simulations

Strong gravitational lensing can be expressed as a linear operation $\mathbf{y} = A_{\boldsymbol{\ell}}\mathbf{x} + \boldsymbol{\eta}$ [31], where $\mathbf{y} \in \mathbb{R}^m$ is the observed (lensed) image, $\boldsymbol{\ell} \in \mathbb{R}^{n_\ell}$ is the vector of parameters of a parametric lens–mass model, $\mathbf{x} \in \mathbb{R}^n$ is a pixelated representation of the background source, $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\eta}}^2 I)$ is additive Gaussian noise, and $A_{\boldsymbol{\ell}}$ is the Jacobian of the forward model (with dependency on $\boldsymbol{\ell}$ made explicit). Our forward model also includes a point-spread function (PSF). We use `Caustics` [32] for the simulations. The lens follows an Elliptical Power-Law (EPL) profile [33] with external shear and $m = 3$ multipole. The positions of the lens and the source are also free parameters, giving a total of 12 macro parameters besides the pixelated source. See Appendix B for prior ranges.

2.2 Score-based models for solving inverse problems

A generative model for a distribution $p(\mathbf{x})$ can be constructed when we have access to the score $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ by solving the reverse-time stochastic differential equation (SDE) [34]:

$$d\mathbf{x} = \left[f(\mathbf{x}, t) - g(t)^2 \, \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] dt + g(t) \, d\bar{\mathbf{w}}, \tag{1}$$

where $p_t(\mathbf{x}_t)$ denotes the target distribution convolved with a perturbation kernel, typically a Gaussian $\mathcal{N}\big(\mu(t)\mathbf{x}, \sigma(t)^2\mathbb{I}\big)$. For the variance-exploding (VE) SDE [35] used in this work, $\mu(t) = 1$ and $\sigma(t) = \sigma_{\text{min}}\big(\sigma_{\text{max}}/\sigma_{\text{min}}\big)^t$. Given a dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$ with $\mathbf{x}_i \sim p(\mathbf{x})$, we train a neural network $\mathbf{s}_\theta(\mathbf{x}_t, t)$ to approximate the score by minimizing the denoising score-matching loss [36, 37]:

$$\mathcal{L}_\theta = \mathbb{E}_{\mathbf{x}\sim\mathcal{D}, t\sim\mathcal{U}(0,1), \mathbf{x}_t\sim p(\mathbf{x}_t|\mathbf{x})}\left[\lambda(t) \, \|\mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x})\|^2\right].$$

This learns a prior score model from the data examples. We use `score-models`[1] to train the network. Moreover, any score-based model can be turned into a zero-shot posterior sampler [e.g. 38] by replacing the prior score with the posterior score:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t). \tag{2}$$

We already have the approximation $\mathbf{s}_\theta(\mathbf{x}_t, t) \approx \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$. However, the likelihood score,

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) = \nabla_{\mathbf{x}_t} \log \int_{\mathbf{x}_0} p(\mathbf{x}_0 \mid \mathbf{x}_t) \, p(\mathbf{y} \mid \mathbf{x}_0) \, d\mathbf{x}_0, \tag{3}$$

[1] github.com/AlexandreAdam/score_models

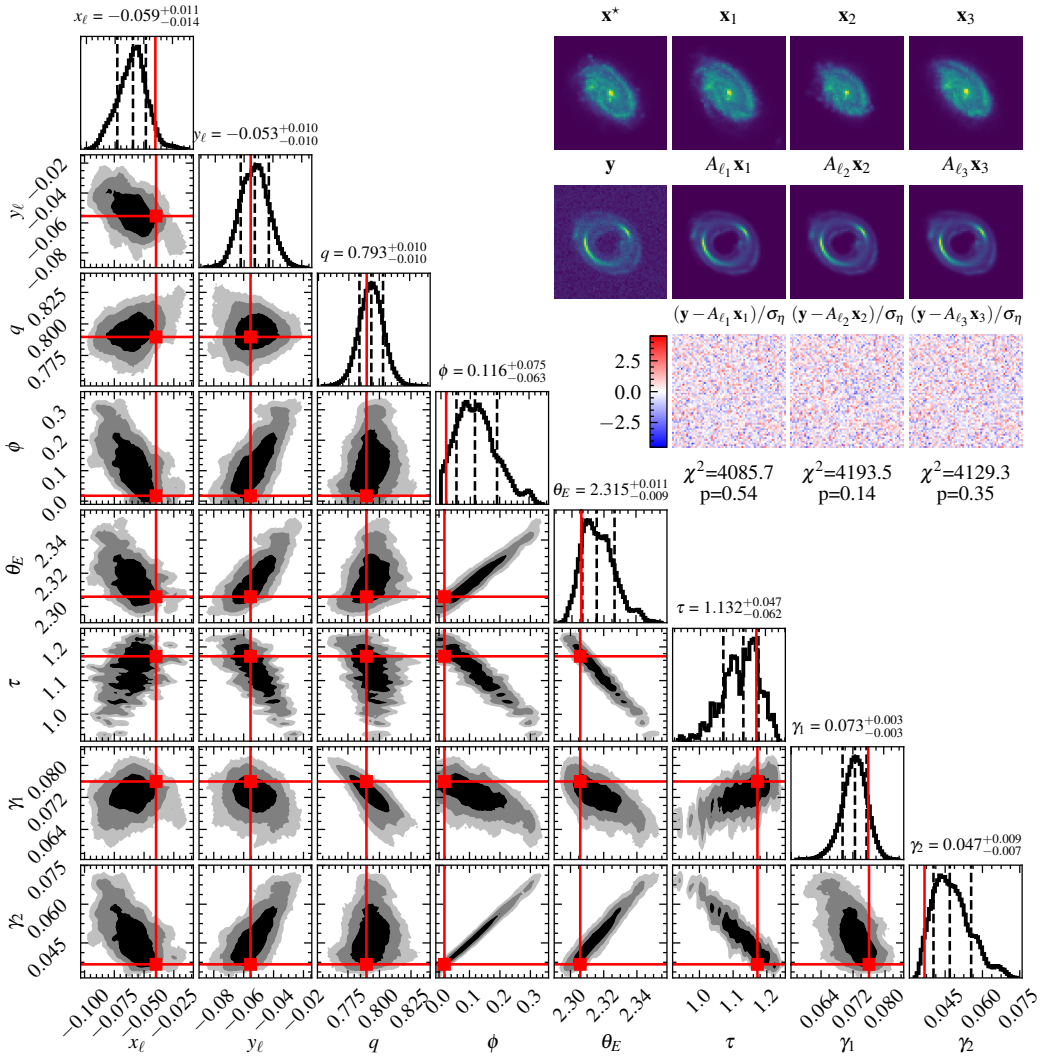

Figure 1: Simulated strong-lensing system analyzed with our joint sampler. **Top-right panel**: the observed image $\mathbf{y}$, the true source $\mathbf{x}^\star$, and three joint-posterior draws $(\mathbf{x}_i, \boldsymbol{\ell}_i)$. For each draw we show the reconstructed image $A_{\boldsymbol{\ell}_i}\mathbf{x}_i$ and the corresponding residual $(\mathbf{y} - A_{\boldsymbol{\ell}_i}\mathbf{x}_i)/\sigma_{\boldsymbol{\eta}}$, demonstrating noise-level consistency. **Bottom-left panel**: marginal lens posterior $p(\boldsymbol{\ell} \mid \mathbf{y})$ obtained from 406 joint samples, each augmented with 500 conditional lens draws as described in Appendix C.

is intractable. Here, we compare three likelihood-score approximations from the literature across different stages of our inference pipeline: Pseudoinverse-Guided Diffusion Models (ΠGDM) [11], Convolved Likelihood Approximation (CLA) [3], and Diffusion Posterior Sampling (DPS) [12]. Using Tweedie's formula, Chung et al. [12] express the posterior mean as

$$\hat{\mathbf{x}}_t := \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] = \mathbf{x}_t + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t). \tag{4}$$

With $\hat{\mathbf{x}}_t$ in hand, ΠGDM approximates the likelihood as

$$p_t(\mathbf{y} \mid \mathbf{x}_t) \approx \mathcal{N}\big(\mathbf{y} \,\big|\, A_{\boldsymbol{\ell}}\hat{\mathbf{x}}_t,\ \Sigma_{\boldsymbol{\eta}} + r_t^2 A_{\boldsymbol{\ell}} A_{\boldsymbol{\ell}}^\top\big). \tag{5}$$

CLA is similar, but it uses $\mathbf{x}_t$ directly in the mean (i.e. $A_{\boldsymbol{\ell}}\mathbf{x}_t$) instead of $\hat{\mathbf{x}}_t$. In the original formulation, CLA sets $r_t^2 = \sigma_t^2$, yet Song et al. [11] note that r_t^2 can, in general, depend on the data and inverse problem. Our choice of r_t^2 for both approximations is detailed in Appendix E.

2.3 Joint inference of source and lens

Our goal is to sample from the joint posterior

$$\mathbf{x}, \boldsymbol{\ell} \sim p(\mathbf{x}, \boldsymbol{\ell} \mid \mathbf{y}) \;\propto\; \mathcal{N}\big(\mathbf{y} \mid A_{\boldsymbol{\ell}}\mathbf{x}, \Sigma_{\boldsymbol{\eta}}\big)\, p(\mathbf{x})\, p(\boldsymbol{\ell}), \tag{6}$$

where $p(\mathbf{x})$ is implicitly represented by a score-based model, $p(\boldsymbol{\ell})$ is an analytic prior (uniform here), and the likelihood is determined by the noise distribution and the forward model.

We follow GibbsDDRM [18], a partially collapsed Gibbs sampler [39] that performs Unadjusted Langevin Algorithm (ULA) sampling [e.g. 40] of the operator parameters through the reverse diffusion process.

In the original GibbsDDRM formulation, the method builds on DDRM [41], which approximates posterior sampling for non-blind inverse problems. DDRM uses pre-trained DDPM models [42] (variance-preserving SDEs in the continuous-time view) re-parameterized in VE notation and conditioned on the observation via the singular-value decomposition of A. We instead keep the continuous-time SDE notation with a VE model, integrate with a stochastic Heun solver [43], and use an approximation to the likelihood score at time t. Aside from these differences, most procedural details remain the same.

The sampler starts from an initial lens parameter vector $\boldsymbol{\ell}_0$. We find the algorithm to be sensitive to this initialization: the macro-parameters posterior often has multiple local minima, and poor starting points cause the Langevin updates to mix slowly, an issue noted previously in lensing studies [e.g. 22]. If a given lens has been analyzed before, published values can serve as the initial guess. Alternatively, a learned estimator, such as a CNN [e.g. 44, 45], can be used.

Here, we obtain $\boldsymbol{\ell}_0$ by minimizing the negative log-likelihood under a Sérsic source model, using the Adam optimizer [46] with a multi-start strategy. Given $\boldsymbol{\ell}_0$, we solve the reverse-time SDE. At diffusion time t, the source $\mathbf{x}_t$ is updated while keeping $\boldsymbol{\ell}$ fixed, employing ΠGDM for the likelihood score with respect to $\mathbf{x}_t$ (see Appendix E for a complete justification for using ΠGDM over CLA). Next, we update the lens parameters via ULA using

$$\nabla_{\boldsymbol{\ell}} \log p(\boldsymbol{\ell} \mid \mathbf{y}, \mathbf{x}_t) \;\approx\; \nabla_{\boldsymbol{\ell}} \log\big[\mathcal{N}(\mathbf{y} \mid A_{\boldsymbol{\ell}}\hat{\mathbf{x}}_t, \Sigma_{\boldsymbol{\eta}})\, p(\boldsymbol{\ell})\big]. \tag{7}$$

In summary, we run a partially collapsed Gibbs sampler through the reverse diffusion process, alternating between updating the lens parameters with ULA and the source with ΠGDM. Finally, we perform a full Gibbs sweep starting from the joint sample $(\mathbf{x}, \boldsymbol{\ell}) \sim p(\mathbf{x}, \boldsymbol{\ell} \mid \mathbf{y})$ updating the source with CLA, which has been more tested for non-blind lensing inversion, and the lens with NUTS [47].

3 Experiments and results

3.1 Dataset

We use $60\,774$ images from the SKIRT–TNG dataset [48], produced by applying dust–radiative-transfer post-processing [49] to galaxies in the TNG cosmological magneto-hydrodynamical simulations[2] [50]. The i-band frames are downsampled to 64×64 pixels and converted to flux units of $\mu\mathrm{Jy\,sr}^{-1}$ to train the SBM. A further $4\,293$ images from the same dataset are reserved as a validation set, used only for inference.

3.2 Joint–inference experiment

Figure 1 illustrates a representative run of our joint sampler. The observation $\mathbf{y}$ is generated from a ground-truth source $\mathbf{x}^\star$ and lens parameters $\boldsymbol{\ell}^\star$. In the top right panel, we display three joint-posterior draws $(\mathbf{x}_i, \boldsymbol{\ell}_i) \sim p(\mathbf{x}, \boldsymbol{\ell} \mid \mathbf{y})$. Each sampled source recovers the overall morphology of $\mathbf{x}^\star$ while exhibiting natural variability in size and pixel-scale structure. The corresponding reconstructions $A_{\boldsymbol{\ell}_i}\mathbf{x}_i$ closely match the observation, and the normalized residuals are consistent with the noise model. The bottom-left panel shows the marginal lens posterior $p(\boldsymbol{\ell} \mid \mathbf{y})$, estimated from 406 joint samples as detailed in Appendix C. Known degeneracies among strong-lensing parameters are properly explored, and the true lens parameters $\boldsymbol{\ell}^\star$ lie well within the high-probability region. A full corner plot of all 12 lens parameters, additional source–reconstruction pairs, and three further lensing systems are provided in Appendix F.

[2] www.tng-project.org

4 Future work and limitations

Although the results are promising, we still need to test the robustness and generality of the method. In future iterations, we plan to run a coverage test with TARP [51], a sample-based diagnostic that is necessary and sufficient for posterior coverage. Furthermore, the overall pipeline contains several hyperparameters and design choices that remain unexplored. For example, we intend to perform an ablation study on the r_t^2 parameter in both CLA and ΠGDM for this specific problem. We also plan to benchmark non-blind inverse approaches in lensing, following the protocol of Zheng et al. [10].

A key limitation of our approach is its strong dependence on the initialization $\boldsymbol{\ell}_0$. Because ULA explores a single mode and implicitly assumes a log-concave target, multi-modal or strongly non-log-concave posteriors are difficult to sample. Preliminary experiments show that mode switching (or recovery from a poor initialization) is possible when using a *swarm* of walkers, or by performing expectation–maximization through the diffusion process instead of sampling, as in Laroche et al. [21].

Another limitation of the current framework is its assumption of a purely additive Gaussian noise model with a known likelihood. This may not fully capture the complex noise properties of real observations. Future work will explore relaxing this assumption. One direction is to adopt methods that can handle complex or non-Gaussian noise, for instance, by building a data-driven noise model using a score-based likelihood characterization [e.g. 52]. A second, complementary approach is to treat observational uncertainty as an intrinsic part of the diffusion process itself, using deconvolution or mixed-noise learning techniques [e.g. 53, 54].

Finally, it would be valuable to compare analytical priors with learned priors when the score is available. A natural baseline is a Gaussian prior, which has a long history in lensing modeling [e.g. 24]. The full source code implementing our sampler will be released publicly upon journal publication.

5 Conclusion

We have presented the first score-based framework for *blind* strong-lensing inversion, jointly sampling the pixelated source and the parametric lens mass. Our approach couples a VE score model with a GibbsDDRM-like sampler: the source is updated with ΠGDM, and the lens parameters with an Unadjusted Langevin step. On simulated data, the sampler yields residuals that match the noise distribution and recovers all lens parameters without systematic bias, showing that it can handle the nonlinearities and parameter degeneracies characteristic of strong lensing. In future work, we plan to improve and extend the framework and apply it to the large samples of lenses expected from LSST and *Euclid*, enabling fully Bayesian analyses at next-generation survey scale.

Acknowledgments and Disclosure of Funding

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). The work is in part supported by computational resources provided by Calcul Québec and the Digital Research Alliance of Canada. G.M.B. thanks Sacha Perry-Fagant, Nicolas Payot, and Alexandre Adam for useful discussions and valuable input while this work was being completed. Y.H. and L.P.-L. acknowledge support from the Canada Research Chairs Program, the National Sciences and Engineering Council of Canada (NSERC) through grants RGPIN-2020-05073 and 05102, and the Fonds de recherche du Québec through grant CFQCU-2024-348060. R.L. acknowledges support from the NSERC CGS D scholarship. C.S. acknowledges the support of an NSERC Postdoctoral Fellowship and a CITA National Fellowship. 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).

References

[1] Berthy T. Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L. Bouman, and William T. Freeman. Score-based diffusion models as principled priors for inverse imaging.

In *International Conference on Computer Vision (ICCV)*, pages 10486–10497, 2023. URL https://doi.org/10.1109/ICCV51070.2023.00965.

[2] Noé Dia, M. J. Yantovski-Barth, Alexandre Adam, Micah Bowles, Laurence Perreault-Levasseur, Yashar Hezaveh, and Anna Scaife. Iris: A bayesian approach for image reconstruction in radio interferometry with expressive score-based priors, 2025. URL https://arxiv.org/abs/2501.02473.

[3] Alexandre Adam, Adam Coogan, Nikolay Malkin, Ronan Legin, Laurence Perreault-Levasseur, Yashar Hezaveh, and Yoshua Bengio. Posterior samples of source galaxies in strong gravitational lenses with score-based priors. In *Machine Learning and the Physical Sciences Workshop, NeurIPS 2022*, January 2022. URL https://arxiv.org/pdf/2211.03812.

[4] Konstantin Karchev, Noemi Anau Montel, Adam Coogan, and Christoph Weniger. Strong-Lensing Source Reconstruction with Denoising Diffusion Restoration Models. In *Machine Learning and the Physical Sciences Workshop, NeurIPS 2022*, 11 2022. URL https://arxiv.org/pdf/2211.04365.

[5] Ronan Legin, Matthew Ho, Pablo Lemos, Laurence Perreault-Levasseur, Shirley Ho, Yashar Hezaveh, and Benjamin Wandelt. 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, 10 2023. ISSN 1745-3925. doi:10.1093/mnrasl/slad152.

[6] Victoria Ono, Core Francisco Park, Nayantara Mudur, Yueying Ni, Carolina Cuesta-Lazaro, and Francisco Villaescusa-Navarro. Debiasing with Diffusion: Probabilistic Reconstruction of Dark Matter Fields from Galaxies with CAMELS. *The Astrophysical Journal*, 970(2):174, August 2024. doi:10.3847/1538-4357/ad5957.

[7] Thomas Flöss, William R. Coulton, Adriaan J. Duivenvoorden, Francisco Villaescusa-Navarro, and Benjamin D. Wandelt. Denoising diffusion delensing: reconstructing the non-Gaussian CMB lensing potential with diffusion models. *Monthly Notices of the Royal Astronomical Society*, 533(1):423–432, September 2024. doi:10.1093/mnras/stae1818.

[8] Zhiwei Xue, Yuhang Li, Yash Patel, and Jeffrey Regier. Diffusion models for probabilistic deconvolution of galaxy images. In *Machine Learning for Astrophysics Workshop, ICML 2023*, 2023. URL https://arxiv.org/abs/2307.11122.

[9] Alexandre Adam, Connor Stone, Connor Bottrell, Ronan Legin, Yashar Hezaveh, and Laurence Perreaul-Levasseur. 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. doi:10.3847/1538-3881/adb039.

[10] Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy Feng, Caifeng Zou, Yu Sun, Nikola Borislavov Kovachki, Zachary E Ross, Katherine Bouman, and Yisong Yue. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=U3PBITXNG6.

[11] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

[12] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

[13] Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 118389–118427. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/d65c4ce22241138c1784ff753d4c746c-Paper-Conference.pdf.

[14] Berthy Feng and Katherine Bouman. Variational bayesian imaging with an efficient surrogate score-based prior. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=db2pFKVcm1.

[15] Zehao Dou and Yang Song. Diffusion Posterior Sampling for Linear Inverse Problem Solving: A Filtering Perspective. In *The Twelfth International Conference on Learning Representations*, October 2023. URL https://openreview.net/forum?id=tplXNcHZs1.

[16] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971, 2009. doi:10.1109/CVPR.2009.5206815.

[17] Angela Gao, Jorge Castellanos, Yisong Yue, Zachary Ross, and Katherine Bouman. Deepgem: Generalized expectation-maximization for blind inversion. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 11592–11603. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/606c90a06173d69682feb83037a68fec-Paper.pdf.

[18] Naoki Murata, Koichi Saito, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano Ermon. GibbsDDRM: A partially collapsed Gibbs sampler for solving blind inverse problems with denoising diffusion restoration. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 25501–25522. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/murata23a.html.

[19] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel Diffusion Models of Operator and Image for Blind Inverse Problems. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6059–6069, Vancouver, BC, Canada, June 2023. IEEE. ISBN 979-8-3503-0129-8. doi:10.1109/CVPR52729.2023.00587.

[20] Hamadi Chihaoui, Abdelhak Lemkhenter, and Paolo Favaro. Blind image restoration via fast diffusion inversion. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=HfSJlBRkKJ.

[21] Charles Laroche, Andrés Almansa, and Eva Coupete. Fast Diffusion EM: a diffusion model for blind inverse problems with application to deconvolution. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5259–5269, Waikoloa, HI, USA, January 2024. IEEE. ISBN 979-8-3503-1892-0. doi:10.1109/WACV57701.2024.00519.

[22] Brendon J. Brewer and Geraint F. Lewis. Strong Gravitational Lens Inversion: A Bayesian Approach. *The Astrophysical Journal*, 637(2):608–619, February 2006. doi:10.1086/498409.

[23] Peter Schneider and Dominique Sluse. Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing. *Astronomy and Astrophysics*, 559:A37, November 2013. doi:10.1051/0004-6361/201321882.

[24] S. H. Suyu, P. J. Marshall, M. P. Hobson, and R. D. Blandford. A Bayesian analysis of regularized source inversions in gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 371(2):983–998, September 2006. doi:10.1111/j.1365-2966.2006.10733.x.

[25] S. Vegetti and L. V. E. Koopmans. Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in Galaxies. *Monthly Notices of the Royal Astronomical Society*, 392(3):945–963, January 2009. doi:10.1111/j.1365-2966.2008.14005.x.

[26] Kenneth C. Wong, Sherry H. Suyu, Geoff C. F. Chen, Cristian E. Rusu, Martin Million, Dominique Sluse, Vivien Bonvin, Christopher D. Fassnacht, Stefan Taubenberger, Matthew W. Auger, Simon Birrer, James H. H. Chan, Frederic Courbin, Stefan Hilbert, Olga Tihhonova, Tommaso Treu, Adriano Agnello, Xuheng Ding, In Jee, Eiichiro Komatsu, Anowar J. Shajib, Alessandro Sonnenfeld, Roger D. Blandford, Léon V. E. Koopmans, Philip J. Marshall, and Georges Meylan. H0LiCOW - XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes. *MNRAS*, 498(1):1420–1439, October 2020. doi:10.1093/mnras/stz3094.

[27] Chien Y. Peng, Chris D. Impey, Hans-Walter Rix, Christopher S. Kochanek, Charles R. Keeton, Emilio E. Falco, Joseph Lehár, and Brian A. McLeod. Probing the Coevolution of Supermassive Black Holes and Galaxies Using Gravitationally Lensed Quasar Hosts. *ApJ*, 649(2):616–634, October 2006. doi:10.1086/506266.

[28] S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans. Gravitational detection of a low-mass dark satellite galaxy at cosmological distance. *Nature*, 481(7381):341–343, January 2012. doi:10.1038/nature10669.

[29] Yashar D. Hezaveh, Neal Dalal, Daniel P. Marrone, Yao-Yuan Mao, Warren Morningstar, Di Wen, Roger D. Blandford, John E. Carlstrom, Christopher D. Fassnacht, Gilbert P. Holder, Athol Kemball, Philip J. Marshall, Norman Murray, Laurence Perreault Levasseur, Joaquin D. Vieira, and Risa H. Wechsler. Detection of Lensing Substructure Using ALMA Observations of the Dusty Galaxy SDP.81. *The Astrophysical Journal*, 823(1):37, May 2016. doi:10.3847/0004-637X/823/1/37.

[30] Thomas E. Collett. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *The Astrophysical Journal*, 811(1):20, September 2015. doi:10.1088/0004-637X/811/1/20.

[31] S. J. Warren and S. Dye. Semilinear Gravitational Lens Inversion. *The Astrophysical Journal*, 590(2):673–682, June 2003. doi:10.1086/375132.

[32] Connor Stone, Alexandre Adam, Adam Coogan, M. Yantovski-Barth, Andreas Filipp, Landung Setiawan, Cordero Core, Ronan Legin, Charles Wilson, Gabriel Barco, Yashar Hezaveh, and Laurence Perreault-Levasseur. Caustics: A Python Package for Accelerated Strong Gravitational Lensing Simulations. *The Journal of Open Source Software*, 9(103):7081, November 2024. doi:10.21105/joss.07081.

[33] Rennan Barkana. Fast Calculation of a Family of Elliptical Mass Gravitational Lens Models. *The Astrophysical Journal*, 502(2):531–537, August 1998. doi:10.1086/305950.

[34] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *The Ninth International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

[35] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf.

[36] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005. URL http://jmlr.org/papers/v6/hyvarinen05a.html.

[37] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, 2011. doi:10.1162/NECO_a_00142.

[38] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 14715–14728. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/5e6cec2a9520708381fe520246018e8b-Paper-Conference.pdf.

[39] David A Van Dyk and Taeyoung Park. Partially collapsed gibbs samplers. *Journal of the American Statistical Association*, 103(482):790–796, 2008. doi:10.1198/016214508000000409.

[40] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, December 1996. ISSN 1350-7265. URL https://projecteuclid.org/journals/bernoulli/volume-2/issue-4/Exponential-convergence-of-Langevin-distributions-and-their-discrete-appro

ximations/bj/1178291835.full. Publisher: Bernoulli Society for Mathematical Statistics and Probability.

[41] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 23593–23606. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/95504595b6169131b6ed6cd72eb05616-Paper-Conference.pdf.

[42] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[43] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, Heidelberg, 1992. ISBN 978-3-642-08107-1 978-3-662-12616-5. doi:10.1007/978-3-662-12616-5.

[44] Yashar D. Hezaveh, Laurence Perreault Levasseur, and Philip J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669): 555–557, August 2017. doi:10.1038/nature23463.

[45] Laurence Perreault Levasseur, Yashar D. Hezaveh, and Risa H. Wechsler. Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. *The Astrophysical Journal Letter*, 850(1):L7, November 2017. doi:10.3847/2041-8213/aa9704.

[46] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

[47] Matthew D. Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014. URL http://jmlr.org/papers/v15/hoffman14a.html.

[48] Connor Bottrell, Hassen M. Yesuf, Gergö Popping, Kiyoaki Christopher Omori, Shenli Tang, Xuheng Ding, Annalisa Pillepich, Dylan Nelson, Lukas Eisert, Hua Gao, Andy D. Goulding, Boris S. Kalita, Wentao Luo, Jenny E. Greene, Jingjing Shi, and John D. Silverman. IllustrisTNG in the HSC-SSP: image data release and the major role of mini mergers as drivers of asymmetry and star formation. *MNRAS*, 527(3):6506–6539, January 2024. doi:10.1093/mnras/stad2971.

[49] P. Camps and M. Baes. 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, 2020. ISSN 2213-1337. doi:10.1016/j.ascom.2020.100381.

[50] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi:10.1186/s40668-019-0028-x.

[51] Pablo Lemos, Adam Coogan, Yashar Hezaveh, and Laurence Perreault-Levasseur. Sampling-based accuracy testing of posterior estimators for general inference. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 19256–19273. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/lemos23a.html.

[52] Ronan Legin, Alexandre Adam, Yashar Hezaveh, and Laurence Perreault-Levasseur. Beyond Gaussian Noise: A Generalized Approach to Likelihood Analysis with Non-Gaussian Noise. *The Astrophysical Journal Letters*, 949(2):L41, June 2023. doi:10.3847/2041-8213/acd645.

[53] Haoye Lu, Qifan Wu, and Yaoliang Yu. Stochastic forward–backward deconvolution: Training diffusion models with finite noisy datasets. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 40741–40768. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/lu25n.html.

[54] Paul Hagemann, Robert Gruhlke, Bernhard Stankewitz, Claudia Schillings, and Gabriele Steidl. Provable mixed-noise learning with flow-matching, 2025. URL https://arxiv.org/abs/2508.18122.

[55] Conor M O'Riordan and Simona Vegetti. Angular complexity in strong lens substructure detection. *Monthly Notices of the Royal Astronomical Society*, 528(2):1757–1768, 01 2024. ISSN 0035-8711. doi:10.1093/mnras/stae153.

[56] Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 2018. URL http://jmlr.org/papers/v20/18-403.html.

[57] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4. doi:10.1007/978-3-319-24574-4_28.

[58] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12438–12448. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/92c3b916311a5517d9290576e3ea37ad-Paper.pdf.

[59] Pablo Lemos, Sammy Nasser Sharief, Nikolay Malkin, Salma Salhi, Connor Stone, Laurence Perreault-Levasseur, and Yashar Hezaveh. PQMass: Probabilistic Assessment of the Quality of Generative Models using Probability Mass Estimation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=n7qGCmluZr.

[60] Gabriel Missael Barco, Alexandre Adam, Connor Stone, Yashar Hezaveh, and Laurence Perreault-Levasseur. Tackling the problem of distributional shifts: Correcting misspecified, high-dimensional data-driven priors for inverse problems. *The Astrophysical Journal*, 980(1): 108, feb 2025. doi:10.3847/1538-4357/ad9b92.

A Strong gravitational-lensing simulator

We use `Caustics` [32] to simulate strong gravitational lensing because it is fully differentiable, allowing us to compute the gradients with respect to both the source and lens parameters required for inference. `Caustics` supports pixelated and parametric descriptions of the source and the lens, and it provides several predefined parametric mass models. In our experiments, the images $\mathbf{x}$, $\mathbf{y}$, and $\boldsymbol{\eta}$ are elements of $\mathbb{R}^{64\times 64}$. Consequently, the Jacobian of the forward model is

$$A_{\boldsymbol{\ell}} = \nabla_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\ell}) \in \mathbb{R}^{4096\times 4096}. \tag{8}$$

The lens mass is modeled as an Elliptical Power-Law (EPL) profile with external shear and $m = 3$ multipole.

Our forward model also includes a Gaussian point-spread function (PSF) with $\mathrm{FWHM} = 0.375''$. In principle, the simulator first applies the lensing transformation and then convolves the result with the PSF. Because convolution with a fixed kernel is a linear operation (represented by a circulant matrix) and the PSF parameters are held constant, we denote the combined lensing-plus-PSF operator simply by $A_{\boldsymbol{\ell}}$.

The lens and source redshifts are assumed to be known and fixed at $z_\ell = 0.5$ and $z_s = 1.0$, respectively. The source plane has a field of view (FOV) of $6.24''$, while the observation plane spans $12''$. The $6.24''$ source FOV corresponds to $\simeq 50\,\mathrm{kpc}$ at $z = 1$, matching the window used when training the galaxy prior, whereas the $12''$ observation window is wide enough to encompass all simulated lenses and is consistent with the typical angular extent of real strong-lensing systems. Finally, we add Gaussian additive noise to the simulations, with $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_\eta^2 \mathbb{I})$ and $\sigma_\eta = 0.35$.

B Uniform prior ranges

Table 1 lists the uniform priors used for both the simulations and the inference runs. The intervals are chosen to encompass the bulk of galaxy–scale strong lenses reported in the literature. All distance-related quantities are expressed in arcseconds, and we follow the `Caustics` convention for angles (measured counter-clockwise from the positive x-axis) and for all other parameters.

The m=3 multipole amplitude is kept small, consistent with previously reported values [e.g. 55]. Although a_m is weakly constrained in our experiments, and a wider inference prior would explore the parameter space more thoroughly, we retain the same interval for simulation and inference so that a future coverage test remains well defined.

Plane	Component	Parameter	Uniform prior
Lens	Lens center	x_ℓ	$[-0.25,\ 0.25]$
		y_ℓ	$[-0.25,\ 0.25]$
	EPL profile	q	$[0.70,\ 1.00]$
		ϕ	$[0,\ \pi]$
		θ_{E}	$[1,\ 3]$
		τ	$[0.75,\ 1.25]$
	External shear	γ_1	$[-0.25,\ 0.25]$
		γ_2	$[-0.25,\ 0.25]$
	Multipole m=3	a_m	$[0.00,\ 0.015]$
		ϕ_m	$[0,\ 2\pi]$
Source	Source center	x_s	$[-0.35,\ 0.35]$
		y_s	$[-0.35,\ 0.35]$
	Sérsic light†	q_s	$[0.05,\ 1.00]$
		ϕ_s	$[0,\ \pi]$
		n_s	$[0.3,\ 10]$
		R_s	$[0.1,\ 3]$
		I_s	$[0.6,\ 100]$

Table 1: Uniform priors for all parameters used in the simulations. †The Sérsic source parameters are used only to initialize $\boldsymbol{\ell}_0$.

C Marginal lens posterior

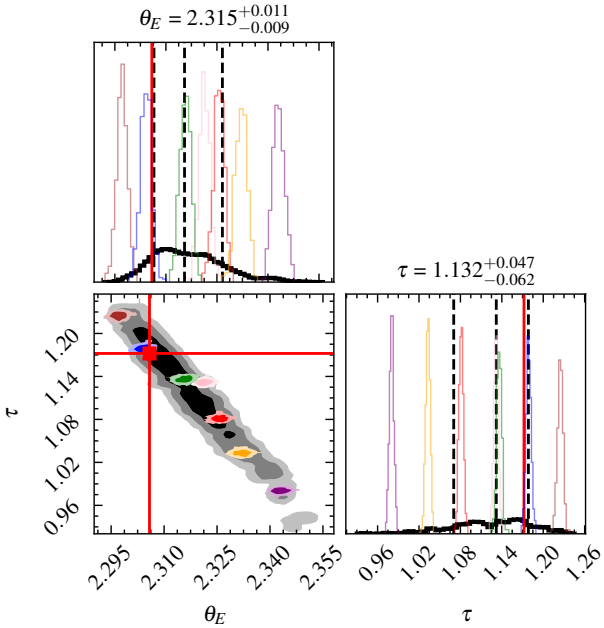

Figure 2: Marginal lens posterior $p(\boldsymbol{\ell} \mid \mathbf{y})$ (black contours and histograms) compared with six conditional posteriors $p(\boldsymbol{\ell} \mid \mathbf{y}, \mathbf{x})$ (colored curves), where each $\mathbf{x}$ is drawn from the joint sampler. Only the parameters θ_{E} and τ are shown for clarity.

To estimate the marginal lens posterior, we write

$$
\begin{aligned}
p(\boldsymbol{\ell} \mid \mathbf{y}) &= \int_{\mathbf{x}_0} p(\boldsymbol{\ell}, \mathbf{x}_0 \mid \mathbf{y})\, d\mathbf{x}_0 = \int_{\mathbf{x}_0} p(\boldsymbol{\ell} \mid \mathbf{x}_0, \mathbf{y})\, p(\mathbf{x}_0 \mid \mathbf{y})\, d\mathbf{x}_0 \\
&= \int_{\mathbf{x}_0} \int_{\bar{\boldsymbol{\ell}}} p(\boldsymbol{\ell} \mid \mathbf{x}_0, \mathbf{y})\, p(\mathbf{x}_0, \bar{\boldsymbol{\ell}} \mid \mathbf{y})\, d\bar{\boldsymbol{\ell}}\, d\mathbf{x}_0 \qquad (9) \\
&= \mathbb{E}_{(\mathbf{x}_0, \bar{\boldsymbol{\ell}}) \sim p(\mathbf{x}_0, \bar{\boldsymbol{\ell}} \mid \mathbf{y})} \big[p(\boldsymbol{\ell} \mid \mathbf{x}_0, \mathbf{y}) \big] \qquad (10) \\
&\approx \frac{1}{n} \sum_{i=1}^{n} p\big(\boldsymbol{\ell} \mid \mathbf{x}_0^{(i)}, \mathbf{y}\big), \qquad (11)
\end{aligned}
$$

where $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0, \bar{\boldsymbol{\ell}} \mid \mathbf{y})$ are samples from the joint sampler. We sample the conditional density with the NUTS sampler [47] as implemented in `Pyro` [56].

This strategy is more efficient than drawing a very large number of joint samples: we obtain a dense estimate of the marginal by running short conditional chains for several fixed sources. In addition, the conditionals $p(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\ell})$ and $p(\boldsymbol{\ell} \mid \mathbf{y}, \mathbf{x})$ are better understood than the approximate joint sampler, which still awaits a formal coverage study.

Figure 2 shows the resulting marginal posterior (black) together with six conditional posteriors (colors) from different source draws. Each conditional distribution is much narrower and nested within the marginal, illustrating how source–lens coupling broadens the overall posterior. Relying on a single point estimate for $\boldsymbol{\ell}$ without fully exploring the joint posterior can therefore lead to biased scientific conclusions. For the marginal lens posteriors displayed in this work, we get 406 joint samples, and 500 conditional lens posterior samples per source.

D Score-based model: architecture and training

We train a variance-exploding (VE) score model on the galaxy dataset described in Section 3, which contains simulated galaxy images $\mathbf{x} \in \mathbb{R}^{64 \times 64}$ covering a $50\,\mathrm{kpc}$ window. Model definition

and training are implemented with the `score-models`[3] package. The network $\mathbf{s}_\theta(\mathbf{x}, t)$ is a noise-conditional score network [35] with a U-Net architecture [57].

Network hyperparameters. Within `score-models` we adopt "nf": 64, "ch_mult": [1, 2, 2, 3], and "num_blocks": 3, and leave all other settings at their defaults.

Optimization. The model is trained with Adam [46] using a learning rate 5×10^{-5}, an EMA decay of 0.999, a batch size of 256, and 1000 epochs ($\approx 230\,000$ optimization steps). Training required 45 h on a single A100 40GB GPU.

SDE parameters and data normalization. We parametrize time as $t \in [0, 1]$ with $\sigma_{\min} = 0.001$ and $\sigma_{\max} = 50$. The data are normalized as $(\mathbf{x} - M)/C$ with $M = 0.125$ and $C = 10.115$, so that most pixel values lie below 1. Following Song and Ermon [58], the maximum noise level is set from pair-wise distances; we use the 95th percentile (rather than the maximum) to avoid the undue influence of a few very bright galaxies.

Statistical fit. We evaluate the learned prior with PQMASS [59], a sample-based χ^2 test that divides the data space into n regions and estimates the densities of both sample sets in each region. The regions are defined by randomly selecting n reference points from one sample set and constructing the corresponding Voronoi tessellation. The resulting statistic, χ^2_{PQM}, follows a χ^2 distribution with $n - 1$ degrees of freedom; we report its value averaged over m random re-tessellations. Using 4000 prior samples, $n = 100$ regions, $m = 2000$ re-tessellations, and 4000 training images, we obtain $\chi^2_{\mathrm{PQM}} = 104.39 \pm 13.19$; comparing to 4000 validation images yields $\chi^2_{\mathrm{PQM}} = 116.2 \pm 14.1$. Both values are close to the expected mean (99) of the $\chi^2_{(99)}$ distribution, indicating that the learned score model faithfully represents the galaxy distribution.

E Inference hyperparameters and implementation details

Source-only inference (non-blind inversion). For source-only inference, we adopt the Convolved Likelihood Approximation because it has been shown to satisfy the TARP coverage test [51] under appropriate noise levels, solver accuracy, and prior choice [60]. In our experiments, CLA yields residual χ^2 values clustered around the ground-truth noise realization (to which we have access). We integrate the SDE with 1 000 steps.

Schedule for r_t^2. For both CLA and ΠGDM we set

$$r_t^2 = \sigma(t)^2\big(C^2 t^4 + 1\big), \tag{12}$$

instead of the original choices $r_t^2 = \sigma(t)^2$ (CLA) or $r_t^2 = \sigma(t)^2/\big[\sigma(t)^2 + 1\big]$ (ΠGDM). The original schedules often lead to unstable diffusion, especially for bright, high-S/N galaxies, whereas the schedule in Equation 12 performs robustly in all our tests. The boundary condition at $t \to 0$ remains exact, so the score approximation is unchanged at the data end of the SDE. A formal ablation study of r_t^2, ideally using a coverage metric such as TARP, is left for future work.

Joint inference. For blind inversion, we prefer ΠGDM, because the modifications CLA makes to $\mathbf{x}_t$ degrade the Tweedie estimate $\hat{\mathbf{x}}_t$ required for the lens update. Figure 3 and Figure 4 compare the two methods at $t = 0.5$: the ΠGDM estimate of $\hat{\mathbf{x}}_t$ resembles a realistic galaxy and lies close to the ground truth, whereas the CLA estimate is noticeably degraded. With CLA, the joint sampler often diverges; with ΠGDM, it converges reliably.

We run 400 reverse-diffusion steps and, at each, perform 500 ULA updates of the lens parameters with a step size 10^{-7} to limit discretization bias. The lens is updated only for $t \in [0.7,\ 0.2]$: at high t the Tweedie estimate is inaccurate, and for $t < 0.2$ the source has essentially converged [see also 18]. Figure 5 illustrates the evolution of $\mathbf{x}_t$ and $\hat{\mathbf{x}}_t$ throughout the diffusion process for an experiment doing source-only inference.

[3] github.com/AlexandreAdam/score_models

Lens initialization and sampling. We obtain $\boldsymbol{\ell}_0$ by minimizing the negative log-likelihood under a Sérsic source model. Adam [46] is run from 1 250 random starts drawn from the priors in Appendix B, using a learning rate of 0.25 and 8 000 optimization steps. For conditional lens sampling, we use NUTS with an initial step size 10^{-3}, 750 warm-up steps, and a maximum tree depth of 9.

Computation time. All experiments were performed on a single A100 GPU (40 GB). Approximate wall-times are:

- Lens initialization: $\sim$30 min;
- Joint sampling of seven $(\mathbf{x}_i, \boldsymbol{\ell}_i)$ pairs (parallel): $\sim$40 min;
- Source-only sampling (seven sources): $\sim$5 min;
- Conditional lens sampling with NUTS for seven lens conditional posteriors: $\sim$1 h (varies with posterior complexity).

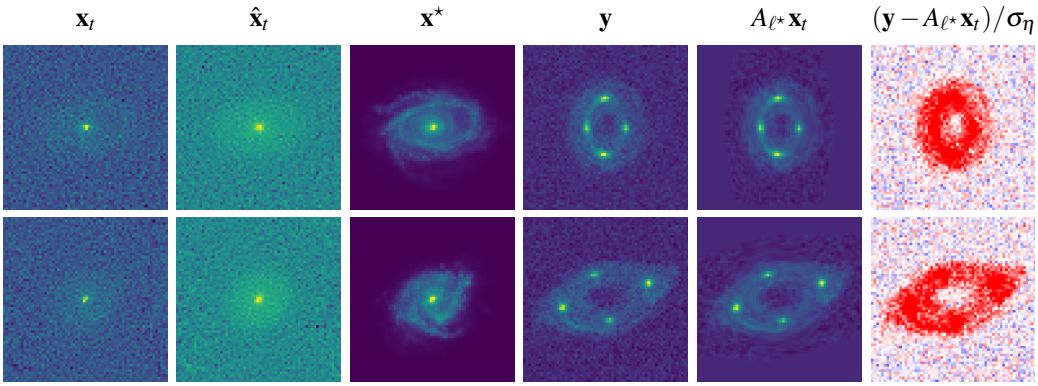

Figure 3: Source-only inference variables with CLA at $t = 0.5$.

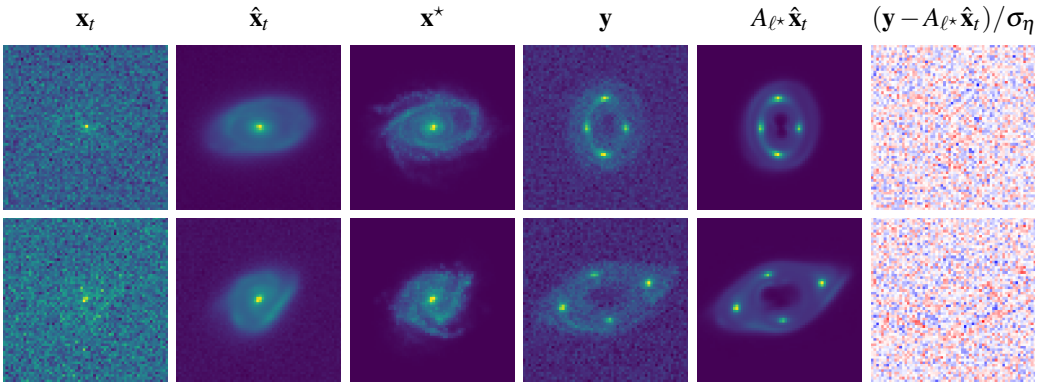

Figure 4: Source-only inference variables with ΠGDM at $t = 0.5$.

F Complete lens posteriors and additional experiments

Figure 6 presents the full marginal posterior for all 12 macro parameters in the main experiment, together with additional source samples, their lens reconstructions, and residual maps. To demonstrate robustness, we include three further simulated systems in Figure 7, Figure 8, and Figure 9. Each figure follows the same layout: the corner plot of the marginal lens posterior, ground-truth lens values (red markers), and several joint posterior draws with their corresponding residuals.

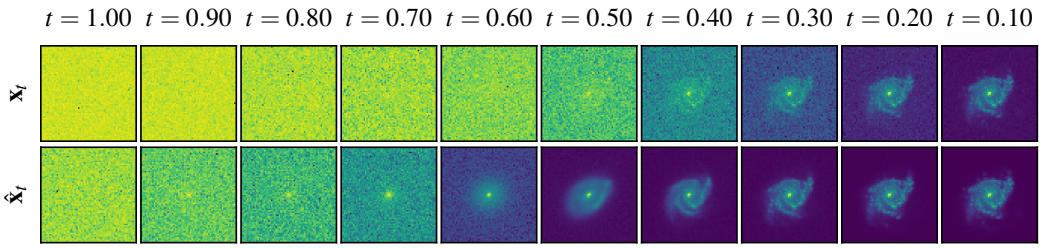

Figure 5: Evolution of the Tweedie posterior mean $\hat{\mathbf{x}}_t$ during ΠGDM source-only inference.

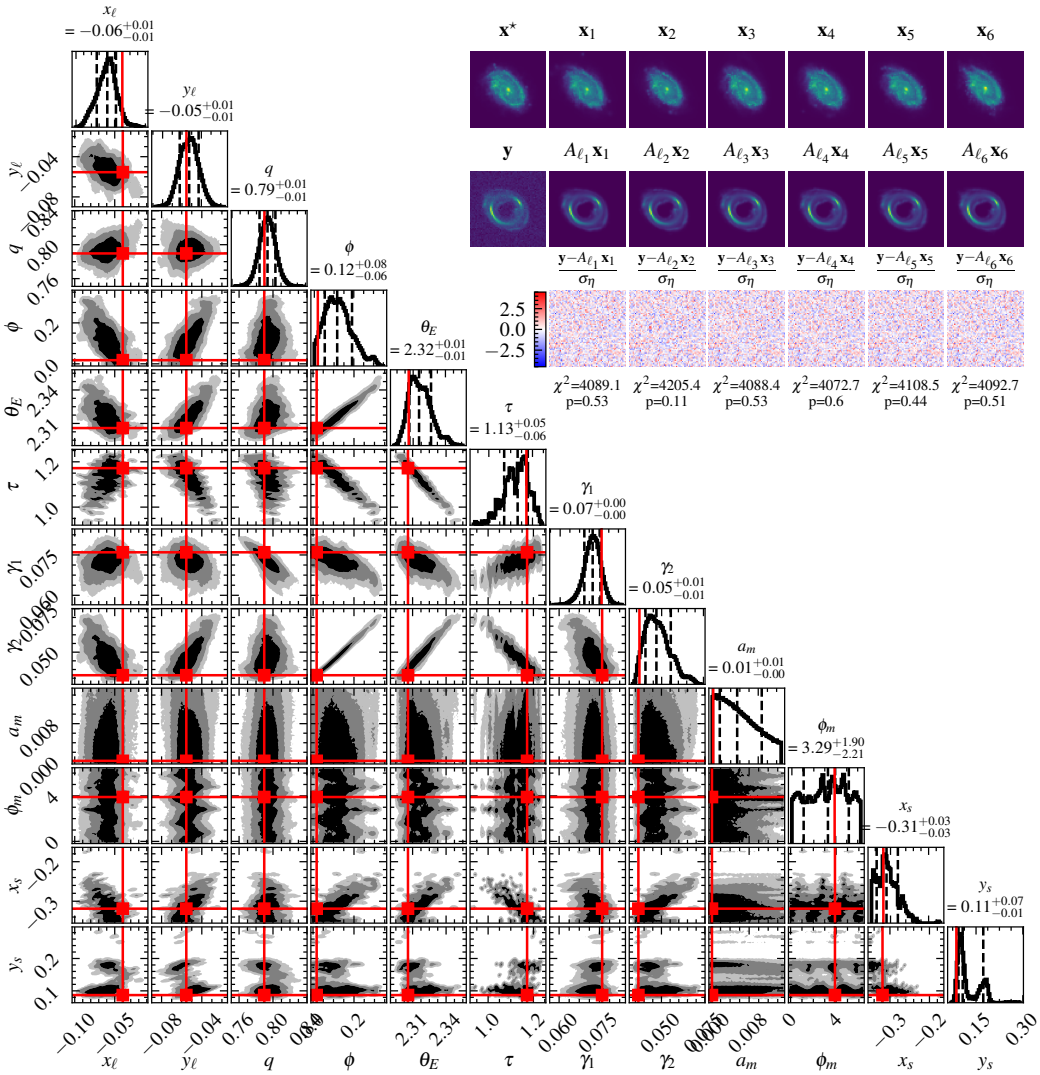

Figure 6: Full marginal posterior for the 12 lens parameters in the main experiment, along with several source–lens reconstructions and residuals.

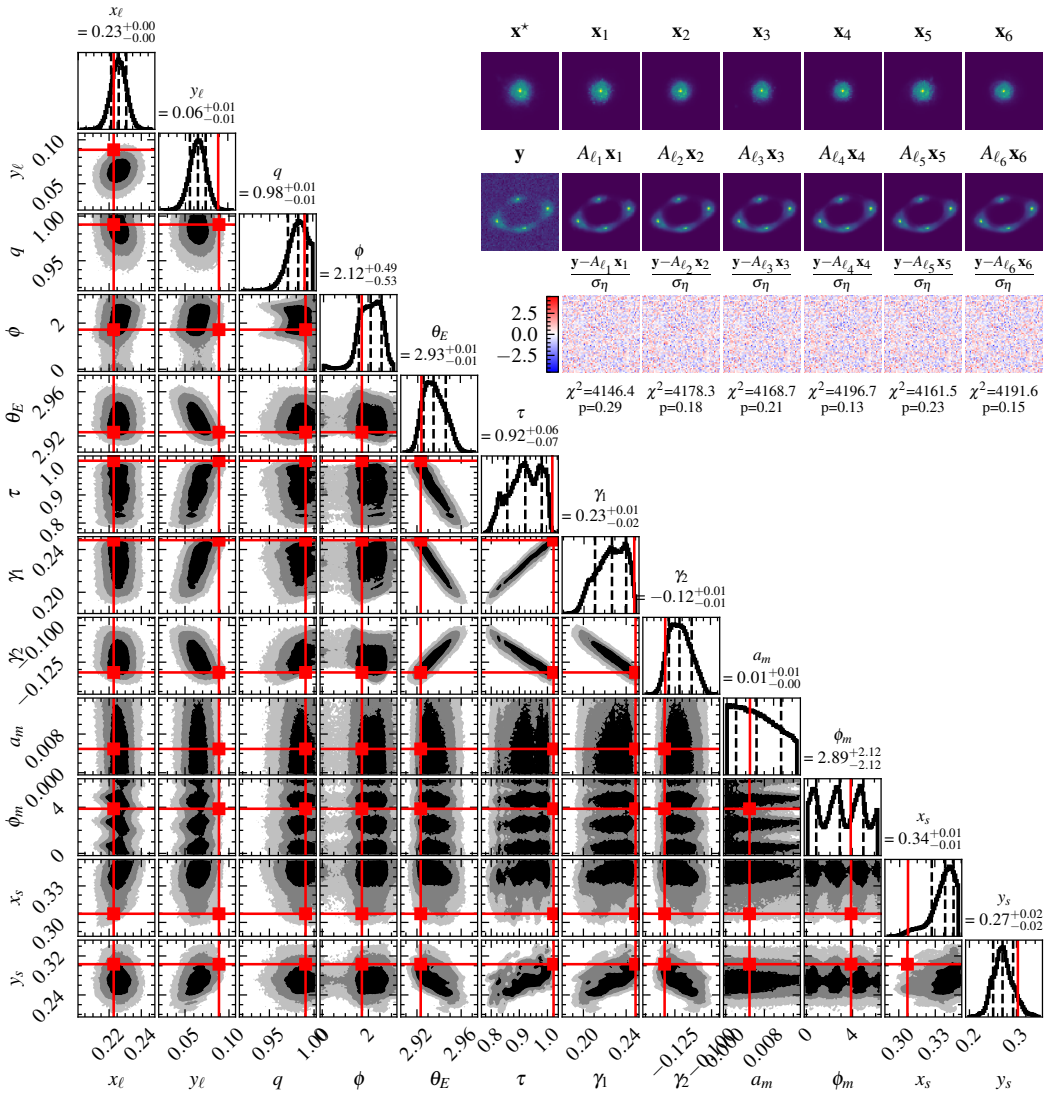

Figure 7: Experiment 2: marginal lens posterior and joint samples for a second simulated system.

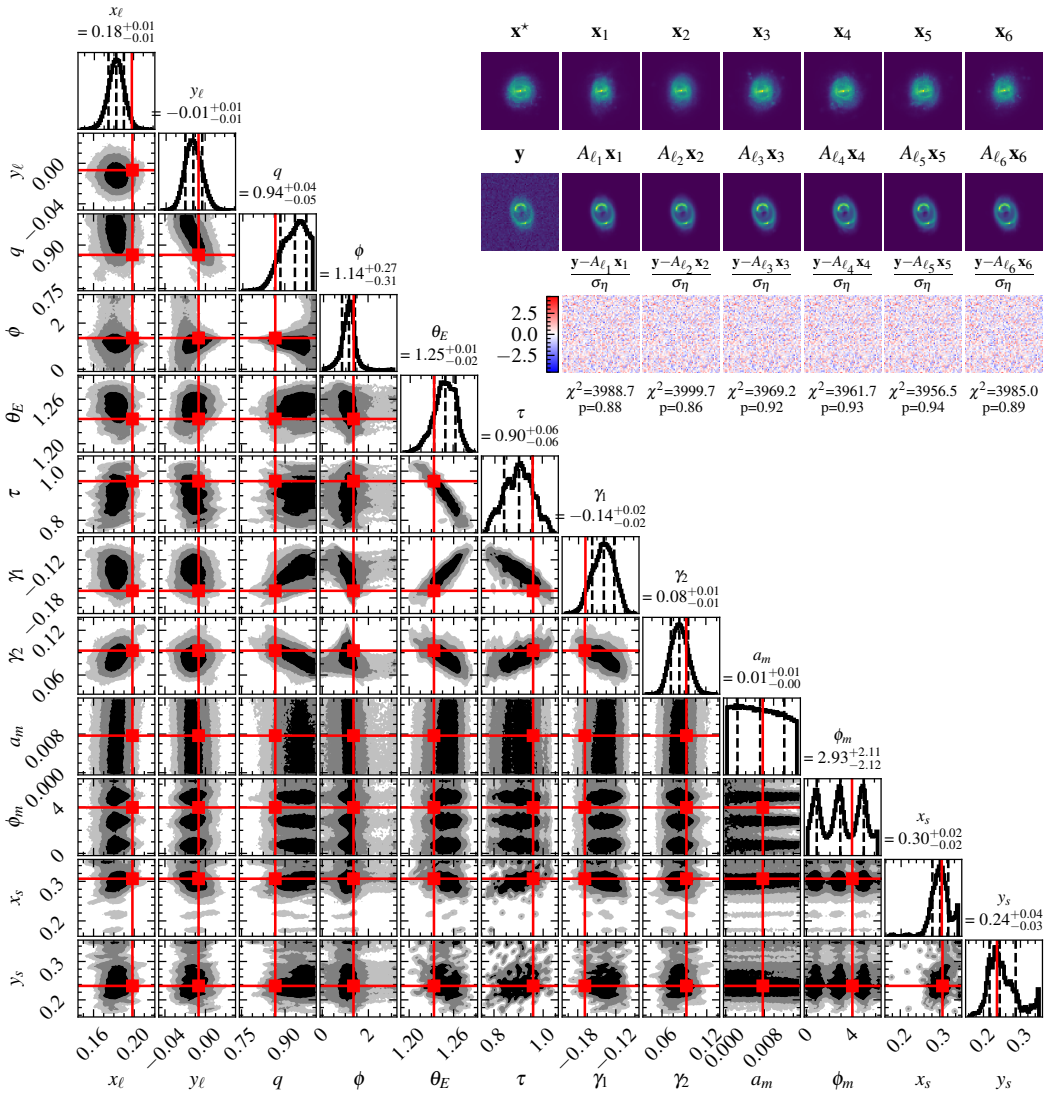

Figure 8: Experiment 3: marginal lens posterior and joint samples for a second simulated system.

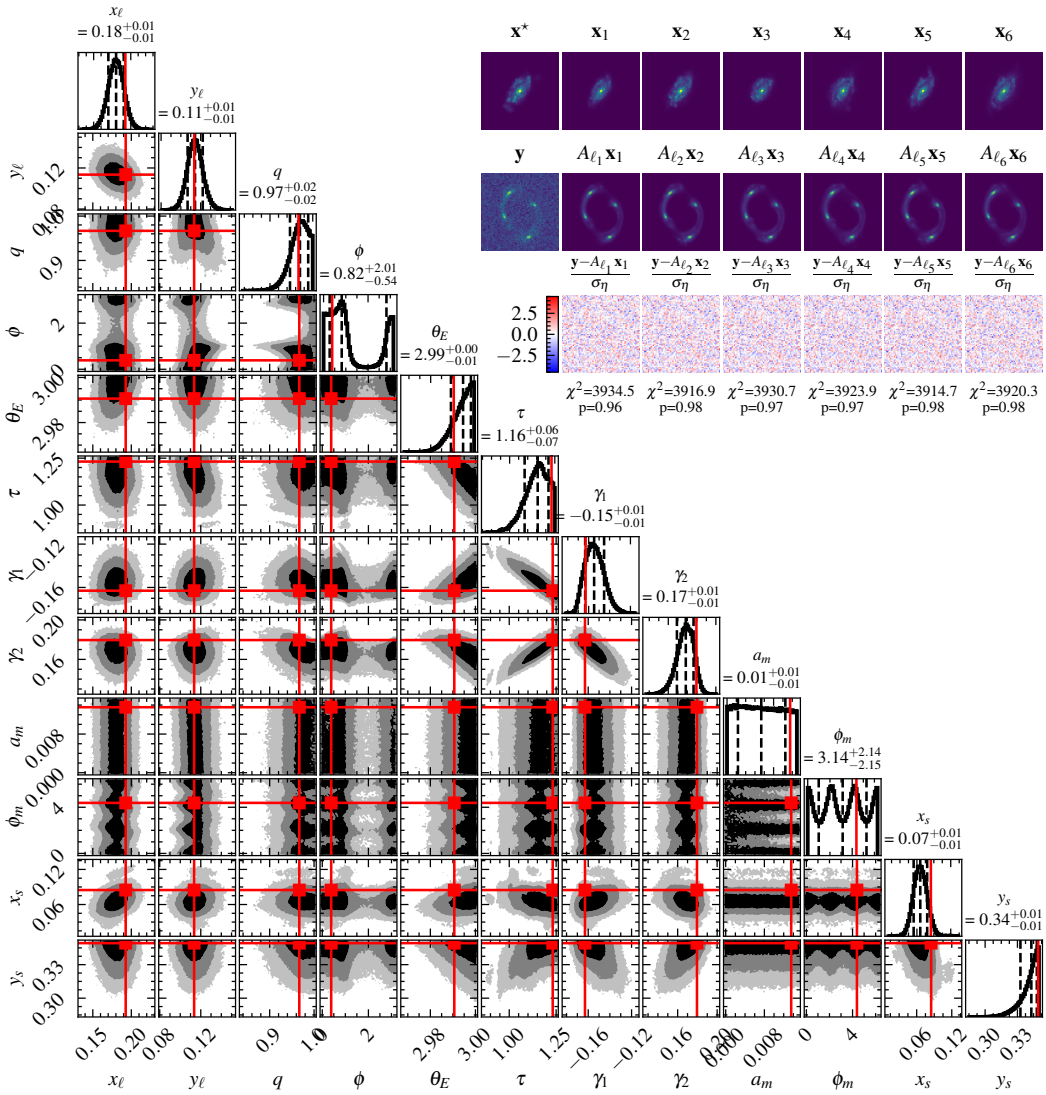

Figure 9: Experiment 4: marginal lens posterior and joint samples for a third simulated system.