Probabilistic Mapping of Dark Matter by Neural Score Matching

Benjamin Remy
AIM, CEA, CNRS
Université Paris-Saclay
Université Paris Diderot
Sorbonne Paris Cité
F-91191 Gif-sur-Yvette, France
benjamin.remy@cea.fr

François Lanusse
AIM, CEA, CNRS
Université Paris-Saclay
Université Paris Diderot
Sorbonne Paris Cité
F-91191 Gif-sur-Yvette, France
francois.lanusse@cea.fr

Zaccharie Ramzi
CEA (Neurospin and Cosmostat)
Inria (Parietal)
Gif-sur-Yvette, France
zaccharie.ramzi@inria.fr

Jia Liu
Berkeley Center for Cosmological Physics
University of California
Berkeley, CA 94720, USA
jialiu@berkeley.edu

Niall Jeffrey[a,b]
[a]Laboratoire de Physique de
l'Ecole Normale Supérieure, ENS
Université PSL, CNRS
Sorbonne Université
Université de Paris, Paris, France
[b]University College London
Gower St, London, UK
niall.jeffrey@phys.ens.fr

Jean-Luc Starck
AIM, CEA, CNRS
Université Paris-Saclay
Université Paris Diderot
Sorbonne Paris Cité
F-91191 Gif-sur-Yvette, France
jstarck@cea.fr

Abstract

The Dark Matter present in the Large-Scale Structure of the Universe is invisible, but its presence can be inferred through the small gravitational lensing effect it has on the images of far away galaxies. By measuring this lensing effect on a large number of galaxies it is possible to reconstruct maps of the Dark Matter distribution on the sky. This, however, represents an extremely challenging inverse problem due to missing data and noise dominated measurements. In this work, we present a novel methodology for addressing such inverse problems by combining elements of Bayesian statistics, analytic physical theory, and a recent class of Deep Generative Models based on Neural Score Matching. This approach allows to do the following: (1) make full use of analytic cosmological theory to constrain the 2pt statistics of the solution, (2) learn from cosmological simulations any differences between this analytic prior and full simulations, and (3) obtain samples from the full Bayesian posterior of the problem for robust Uncertainty Quantification. We present an application of this methodology on the first deep-learning-assisted Dark Matter map reconstruction of the Hubble Space Telescope COSMOS field.

1 Introduction

Deep Learning methods have proven extremely efficient at solving a wide range of inverse problems such as deconvolution, denoising, super-resolution, inpainting, etc. As a concrete example, deep learning models are now state-of-the-art in MRI [1]. Yet, to apply these methods in the Physical Sciences, two particular preoccupations always remain: How do we robustly quantify uncertainties on a deep learning solution? How do we combine deep learning with existing and trusted physical models? In this work, we present a methodology allowing us to address both of these questions in a consistent Bayesian framework and apply it to a challenging instance of inverse problems from the field of Cosmology: mapping the distribution of Dark Matter in the Large-Scale Structure of the Universe from its weak gravitational lensing effect. Several factors make this problem a perfect example:

- The mapping problem is ill-posed, there is *no unique solution* and it cannot be solved without making some assumptions on the solution: we want a method that doesn't return a single solution, but a range of solutions covering all possible Dark Matter maps compatible with the observations and our a priori knowledge.

- Analytic cosmological *theory can predict some properties of Dark Matter maps*, in particular their two-point functions: we want to use as much as possible analytic modeling to constrain these solutions, i.e. we do not want to use deep learning in the regimes where theory applies.

- Numerical cosmological *simulations can sample realistic realisations of Dark Matter maps*, in all of their complexity: we want to restrict the use of deep learning to only modeling *residuals* between analytic theory and full simulations.

2 Mapping the Invisible Dark Matter through its Gravitational Effects

The field of cosmology is investing multiple-billion dollars in space- and ground-based telescopes, with the aim to unveil the fundamental nature of our Universe: dark energy, Dark Matter, cosmic inflation, and Einstein's theory of gravity. The matter distribution—where and how all the matter is positioned throughout space—holds the key to these puzzles. The main difficulty stems from the fact that visible matter, the stars and galaxies, only accounts for 15% of total matter, and the remaining 85% is in the form of invisible Dark Matter. Under Einstein's theory of gravity, the path of a light ray is bent when it passes by massive objects, an effect similar to that of magnifying glasses (and hence the name "gravitational lensing"). Using distant galaxies as our background image, by measuring the distortion of the galaxy shape, we can infer the total matter distribution.

In practice, our task is to infer the matter distribution map, known as a convergence κ map or simply as a mass map, from measurements of individual galaxy ellipticities ϵ. These ellipticities are affected by gravitational lensing and can be used as estimators of the lensing signal. From a Bayesian perspective, the shape-to-mass $\epsilon \rightarrow \kappa$ inference problem can be quantified as,

$$p(\kappa|\epsilon, \mathcal{M}) \propto p(\epsilon|\kappa, \mathcal{M})\, p(\kappa|\mathcal{M}), \qquad (1)$$

where $\mathcal{M}$ is the cosmological model that encapsulates all the fundamental physics of interest. The likelihood factor $p(\epsilon|\kappa, \mathcal{M})$, which describes how light-rays are bent by gravity, how measurements are affected by noise, and accounts for missing observational data. All these factors contribute to making the problem non-invertible, but are all well-understood. The prior distribution however, $p(\kappa|\mathcal{M})$, which describes the detailed matter distribution under any model $\mathcal{M}$, lacks a full analytic expression. Cosmological theory can provide an analytic Gaussian prior which is accurate but only on large scales.

While a number of approximate methods have been adopted to quantify this prior term on smaller scales [2, 3, 4, 5, 6, 7], at present the only accurate physical model of convergence maps for Dark Matter mapping rely on large-scale numerical simulations (presented in [8, 9]), such as the `MassiveNuS` simulations used in this work [10]. As described in [9], forward simulation lets us draw samples from $p(\kappa|\mathcal{M})$. Our goal in this work is to sample the full posterior distribution $p(\kappa|\epsilon, \mathcal{M})$.

3 Neural Score Matching for Bayesian Inverse Problems

With the fast development of Deep Generative Models, it has recently become possible to learn high dimensional probability distributions from data, opening the prospect of data-driven deep priors. In particular, generative models with explicit likelihoods, like PixelCNNs [11] and Normalizing Flow [12], can directly learn, evaluate, and sample any $p(\boldsymbol{x})$ given enough training examples. However, a significant realization is that in the perspective of using such generative models as part of a Bayesian inverse problem such as Equation 1, i.e. alongside a likelihood term, there is no need for a full normalized probability distribution function $p(\boldsymbol{x})$. Instead, all we need to sample the posterior with modern gradient-based inference techniques such as Hamiltonian Monte-Carlo [HMC, 13] or Black-Box Variational Inference [14], is to have access to the *score* $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$. Very recent work [15, 16] has demonstrated that generative models based on learning this score can easily reach GAN quality samples and are both more scalable and easier to train.

Denoising Score Matching As originally identified in [17] and [18], the score of a given target distribution P can be modeled using a Denoising Auto-Encoder (DAE), i.e. by introducing an auto-encoding function $\boldsymbol{r}_\theta : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}^n$ trained to reconstruct a true $\boldsymbol{x} \sim P$ given a noisy version $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{n}$ with $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ under an ℓ_2 loss. An optimal denoiser $\boldsymbol{r}^\star$ would then be achieved for:

$$\boldsymbol{r}^\star(\boldsymbol{x}, \sigma) = \boldsymbol{x} + \sigma^2 \nabla_{\boldsymbol{x}} \log p_{\sigma^2}(\boldsymbol{x}) \tag{2}$$

where $p_{\sigma^2} = p * \mathcal{N}(0, \sigma^2)$. In other words, the optimal denoiser is closely related to the score we wish to learn and when the noise variance σ^2 tends to zero, should exactly match the score of the target density. In practice to learn this score efficiently, we adopt the noise-conditional Denoising Score Matching (DSM) technique proposed by [19], and train a model to minimize the loss:

$$\mathcal{L}_{DSM} = \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim \mathcal{N}(0,I) \\ \sigma \sim \mathcal{N}(0,s^2)}} \| \boldsymbol{u} + \sigma \boldsymbol{r}_\theta(\boldsymbol{x} + \sigma \boldsymbol{u}, \sigma) \|_2^2 \tag{3}$$

In this expression, note that $\boldsymbol{r}_\theta(\boldsymbol{x} + \sigma \boldsymbol{u}, \sigma)$ directly models the score $\nabla_{\boldsymbol{x}} \log p_{\sigma^2}$ and is conditioned on the level of noise σ used to corrupt the input. In practice, the network is a 3-scale U-net with residual blocks composed of convolutions followed by a batch normalisation.

Annealed Hamiltonian-Monte Carlo Sampling Given the noise-conditional neural scores learned with the procedure described above, it is now possible to use a variety of inference methods to access the Bayesian posterior. In this work, we adopt an annealed HMC procedure which provides an efficient way to obtain independent samples from the target posterior despite the high dimensionality of the problem. We consider a tempered version of a target posterior density:

$$\log p_{\sigma^2}(\boldsymbol{x}|\boldsymbol{y}) = \log p_{\sigma^2}(\boldsymbol{y}|\boldsymbol{x}) + \log p_{\sigma^2}(\boldsymbol{x}) \tag{4}$$

where σ^2 plays the role of the inverse temperature found in thermodynamic annealing methods, and we recognise the $\log p_{\sigma^2}(\boldsymbol{x})$ term modeled by the score network $\boldsymbol{r}_\theta(., \sigma)$. At high temperatures, i.e. large σ, the tempered distribution tends to a wide Gaussian, all modes merge together, and an HMC with a diagonal mass matrix can efficiently navigate the posterior distribution. This distribution is then gradually annealed to low temperatures and parallel independent chains progressively moves towards different points in the target distribution. Once we have reached low temperature, we retrieve the last sample from each parallel chain, yielding a batch of independent samples of the posterior distribution. This procedure is similar to the Annealed Langevin diffusion proposed in [15].

4 Hybrid Analytic and Neural Score Modeling for Mass-Mapping

The method presented in the previous section was based on learning the full score from simulations. However, some properties of the Dark Matter maps are known from analytic cosmology models. We now present how we can directly combine analytic modeling and deep Neural Score Matching.
The two-point statistics of weak lensing fields can be accurately predicted from theory over a range of scales, and this forms the basis of most lensing-based cosmological constraints to date. Only on small scales do the theoretical models start to deviate from reality. Since these two-point properties are so well-known, we do not wish to leave it up to the neural network to learn them, and instead
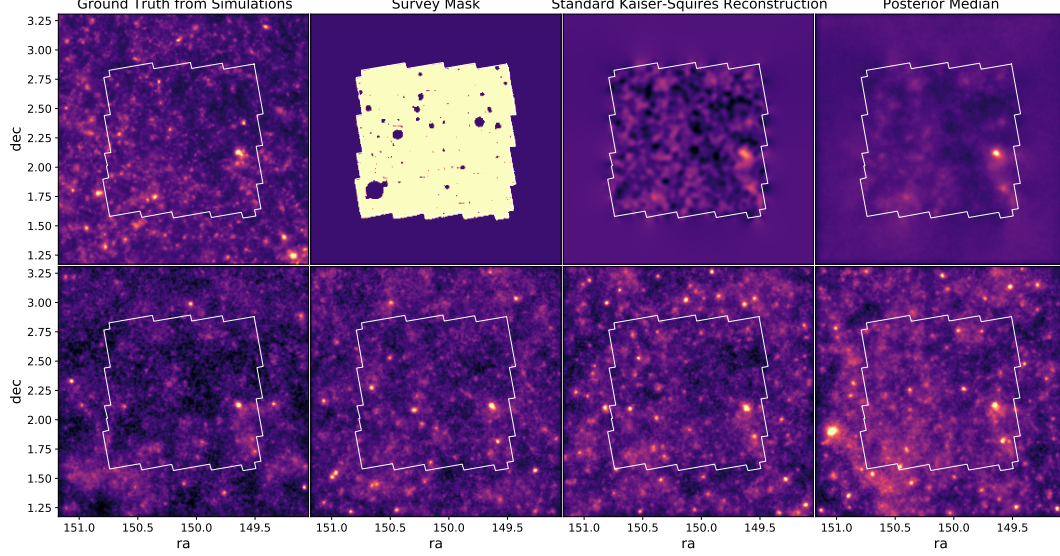
Figure 1: Reconstruction of Dark Matter maps on simulated Hubble Space Telescope lensing measurements. The top row shows from left to right, true map from the `MassiveNuS` simulations, the mask showing the extent of the COSMOS survey, a reconstructed map using a conventional Kaiser-Squires method, and the median of our posterior samples. The bottom row shows multiple samples from the posterior obtained with our procedure.

want to directly impose this prior knowledge on our solution. This prior on $\boldsymbol{\kappa}$ can be expressed as a Gaussian with a covariance $\mathbf{S}$ which becomes diagonal in Fourier space:

$$p_{th}(\boldsymbol{\kappa}) = \frac{1}{\sqrt{\det 2\pi\mathbf{S}}} \exp\left(-\frac{1}{2}\boldsymbol{\kappa}^{\dagger}\mathbf{S}^{-1}\boldsymbol{\kappa}\right) \qquad (5)$$

Using this Gaussian prior alone in a reconstruction would yield Gaussian constrained realisations [20] or the conventional Wiener filter if the maximum a posteriori solution is the target [21]. But we know the convergence field is *not* Gaussian, so we aim to learn the difference between the true distribution of convergence maps and a Gaussian model. In our Score Matching framework, we propose the following decomposition of the score of the full prior $p(\boldsymbol{\kappa})$:

$$\nabla_{\boldsymbol{\kappa}} \log p(\boldsymbol{\kappa}) = \nabla_{\boldsymbol{\kappa}} \log p_{th}(\boldsymbol{\kappa}) + \boldsymbol{r}_{\theta}(\boldsymbol{\kappa}, \nabla_{\boldsymbol{\kappa}} \log p_{th}(\boldsymbol{\kappa})) \qquad (6)$$

where $\boldsymbol{r}_{\theta}$ is a score network, tasked with modeling the difference between true and Gaussian scores. This network is trained by Denoising Score Matching, simply by adapting Equation 3 to add the theory score to the network output. The main advantage of this approach is that we decrease our reliance on the neural score model to a minimum, lowering the demand on the model complexity.

5   Results

To demonstrate our methodology, we present the first deep-learning assisted, probabilistic, Dark Matter map reconstruction of the Hubble Space Telescope (HST) Cosmic Evolution Survey (COSMOS) field [22], using data described in [23], and basing our prior on the `MassiveNuS` suite of simulations [10]. The simulations contain 1024^3 particles in a $(512 \ h^{-1}\text{Mpc})^3$ box with periodic boundary conditions. 10,000 mass maps are generated by ray-tracing through the simulation boxes, following the physics of gravitational lensing. We make use of the maps for source galaxies at $z = 1$, which corresponds to the mean redshift of our COSMOS lensing catalog. Along with these maps, we use the corresponding theoretical power spectrum to build the analytic prior $p_{th}(\boldsymbol{\kappa})$ for training the score network.

Validation on simulation   We first emulate from `MassiveNuS` simulations a mock COSMOS lensing catalog. We use the actual distribution of galaxies in that survey to create a binary mask representing where lensing is actually measured, and we mock observations by adding realistic levels of noise matching the COSMOS bright galaxy sample of [23]. Figure 1 illustrates the results of our
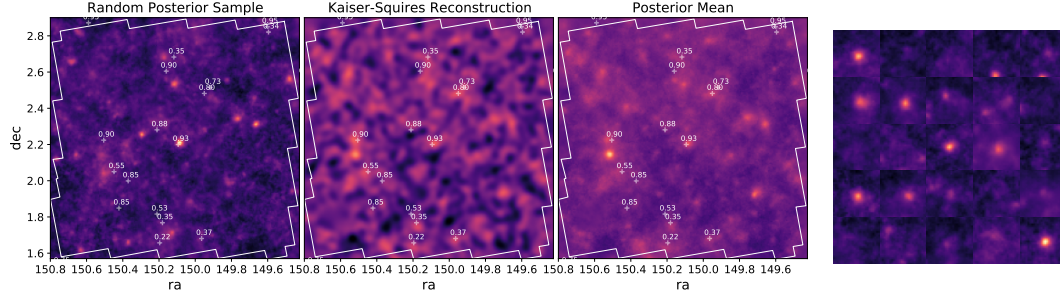
Figure 2: Dark Matter map reconstruction of HST COSMOS survey. We provide both a random sample of the posterior and the posterior mean, to be compared to a conventional Kaiser-Squires reconstruction. Positions of known X-ray clusters and their redshifts are indicated in white. The rightmost plot shows cutouts from independent posterior samples of the central region of the map where a massive cluster is known to be present. As can be seen, the posterior is bi-modal in this region, with the cluster present only in some of the samples.

method on this simulated dataset. The bottom row shows 4 independent samples from the posterior. Note that because the data is very noisy, independent posterior samples can exhibit significant variability, except where a strong signal is present. This is for instance the case of the massive cluster visible in the bottom right corner of the field, visible even in the map recovered by a traditional Kaiser-Squires [24] reconstruction. This structure being highly significant, it therefore appears virtually in all posterior samples. Also note how realistic each posterior sample looks compared to the simulation, demonstrating the quality of our score matching model on these 320x320 pixel images.

COSMOS field reconstruction   Finally, we present the application of our method to the real COSMOS field, based on the bright galaxy sample described in [23]. Not only does our posterior mean map represents the highest quality map ever made of the COSMOS field, we have access to the full posterior for interpreting these results. As an example, Figure 2 on the right shows cutouts from independent posterior samples of the central region of the map where a massive cluster is known to be present. We see that the posterior is bi-modal in this region, with the cluster present in some maps and not in others. Measuring the frequency of the cluster appearing in posterior samples would yield a robust quantification of the significance of this structure.

These results demonstrate the benefits of our proposed methodology, which optimally makes use of our physical models and knowledge (through the known likelihood term, theoretical prior on large scales, numerical simulations for small scale non-gaussian prior) and allows us to access a full posterior distribution in high dimension.

Broader Impact

The methodology described here finds many useful applications outside of cosmology. Of particular interest, the same method can be applied to medical imaging (e.g. MRI) and help physicians assess the significance of particular parts of an image before making a diagnostic. We believe this work does not entail any negative consequences or ethical issues.

References

[1] Zaccharie Ramzi, Philippe Ciuciu, and Jean Luc Starck. Benchmarking MRI reconstruction neural networks on large public datasets. *Applied Sciences (Switzerland)*, 10(5), 2020.

[2] P. J. Marshall, M. P. Hobson, S. F. Gull, and S. L. Bridle. Maximum-entropy weak lens reconstruction: improved methods and application to data. MNRAS, 335(4):1037–1048, October 2002.

[3] F. Lanusse, J.-L. Starck, A. Leonard, and S. Pires. High resolution weak lensing mass mapping combining shear and flexion. A&A, 591,A2, Jun 2016.

[4] Justin Alsing, Alan Heavens, Andrew H Jaffe, Alina Kiessling, Benjamin Wandelt, and Till Hoffmann. Hierarchical cosmic shear power spectrum inference. *Monthly Notices of the Royal Astronomical Society*, 455(4):4452–4466, 2016.

[5] Justin Alsing, Alan Heavens, and Andrew H Jaffe. Cosmological parameters, shear maps and power spectra from cfhtlens using bayesian hierarchical inference. *Monthly Notices of the Royal Astronomical Society*, 466(3):3272–3292, 2017.

[6] N. Jeffrey, F. B. Abdalla, O. Lahav, F. Lanusse, J. L. Starck, A. Leonard, D. Kirk, C. Chang, E. Baxter, and T. Kacprzak. Improving weak lensing mass map reconstructions using Gaussian and sparsity priors: application to DES SV. MNRAS, 479(3):2871–2888, Sep 2018.

[7] M. A. Price, X. Cai, J. D. McEwen, M. Pereyra, and T. D. Kitching. Sparse bayesian mass-mapping with uncertainties: local credible intervals. *Mon. Not. Roy. Astron. Soc.*, 492(1):394–404, 2019.

[8] Masato Shirasaki, Naoki Yoshida, and Shiro Ikeda. Denoising weak lensing mass maps with deep learning. Phys. Rev. D, 100(4):043527, August 2019.

[9] Niall Jeffrey, François Lanusse, Ofer Lahav, and Jean-Luc Starck. Deep learning dark matter map reconstructions from DES SV weak lensing data. MNRAS, 492(4):5023–5029, March 2020.

[10] Jia Liu, Simeon Bird, José Manuel Zorrilla Matilla, J Colin Hill, Zoltán Haiman, Mathew S Madhavacheril, David N Spergel, and Andrea Petri. MassiveNuS: Cosmological massive neutrino simulations. *Journal of Cosmology and Astroparticle Physics*, 2018(3), 2018.

[11] Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P. Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *CoRR*, abs/1701.05517, 2017.

[12] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *32nd International Conference on Machine Learning, ICML 2015*, 2:1530–1538, 2015.

[13] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo, 2018.

[14] Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14(4):1303–1347, 2013.

[15] Yang Song and Stefano Ermon. Generative Modeling by Estimating Gradients of the Data Distribution. (NeurIPS), 2019.

[16] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models, 2020.

[17] Pascal Vincent. A connection between scorematching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

[18] Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data generating distribution. *1st International Conference on Learning Representations, ICLR 2013 - Conference Track Proceedings*, 15:3743–3773, 2013.

[19] Jae Hyun Lim, Aaron Courville, Christopher Pal, and Chin-Wei Huang. AR-DAE: Towards Unbiased Neural Entropy Gradient Estimation. 2020.

[20] Yehuda Hoffman and Erez Ribak. Constrained Realizations of Gaussian Fields: A Simple Algorithm. *apjl*, 380:L5, October 1991.

[21] O. Lahav, K. B. Fisher, Y. Hoffman, C. A. Scharf, and S. Zaroubi. Wiener Reconstruction of All-Sky Galaxy Surveys in Spherical Harmonics. *apjl*, 423:L93, March 1994.

[22] N. Scoville, R. G. Abraham, H. Aussel, J. E. Barnes, A. Benson, A. W. Blain, D. Calzetti, A. Comastri, P. Capak, C. Carilli, and et al. Cosmos: Hubble space telescope observations. *The Astrophysical Journal Supplement Series*, 172(1):38–45, Sep 2007.

[23] T. Schrabback, J. Hartlap, B. Joachimi, M. Kilbinger, P. Simon, K. Benabed, M. Bradač, T. Eifler, T. Erben, C. D. Fassnacht, and et al. Evidence of the accelerated expansion of the universe from weak lensing tomography with cosmos. *Astronomy and Astrophysics*, 516:A63, Jun 2010.

[24] Nick Kaiser and Gordon Squires. Mapping the Dark Matter with Weak Gravitational Lensing. ApJ, 404:441, Feb 1993.