Probabilistic Mapping of Dark Matter by Neural Score Matching
Benjamin Remy¹, François Lanusse¹, Zacharie Ramzi², Jia Liu³, Niall Jeffrey⁴,⁵, Jean-Luc Starck¹

1. AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France
2. CEA (Neurospin and Cosmostat), Inria (Paratil) Gif-sur-Yvette, France
3. Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA
4. Laboratoire de Physique de l’École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université Université de Paris, Paris, France
5. University College London, Gower St, London, UK

Introduction

Dark Matter present in the Large-Scale Structure of the Universe is invisible, but its presence can be inferred through the weak 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 map 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.

Our approach

Our approach allows to do the following:

1. Make full use of analytic cosmological theory to constrain the 2 point statistics of the solution.
2. Learn from cosmological simulations any differences between this analytic prior and full simulations.
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.

Weak Gravitational Lensing

Galaxies randomly distributed

Slight alignment

Shear map $\gamma$ and convergence map $\kappa$ are related through the Kaiser-Squires (1993) transformation:

$$\gamma = \text{MTPF}\kappa + n, \quad n \sim \mathcal{N}(0, \sigma^2)$$

It is an ill-posed inverse problem because of missing data and noise corruption.

We aim to provide all the possible convergence map $\kappa$ for a given observed ellipticity map $\epsilon$, thus estimate the posterior distribution:

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

The likelihood term $p(\epsilon|\kappa, \mathcal{M})$ encodes our physical understanding of the forward process that leads to the observation, given a set of cosmological parameters $\mathcal{M}$.

$$\log p(\epsilon|\kappa, \mathcal{M}) \propto -\|\mathcal{M}\gamma - \text{MTPF}\kappa\|_{L_2}^2$$

The prior term $p(\kappa)$ encodes prior knowledge on the convergence map, given by analytic cosmological theory and learned on simulations.

Prior learning with Denoising Score Matching

Prior on high dimensional images can be modeled by learned the gradient of its log probability $\nabla_\epsilon \log p(\kappa)$, which is called the score function [1].

Given a signal $x \sim p$, its noisy version $x' = x + n$, and $p_{\epsilon|x} = p \ast \mathcal{N}(0, \sigma^2)$, an optimal denoiser $r^*$ is related to the score function as [2], [3]:

$$r^*(x', \sigma) = x' + \sigma^2 \nabla_\epsilon \log p_{\epsilon|x}(x')$$

Hybrid prior

We assume that the matter density field is gaussian at large scales. Then it is fully characterised by its 2 point statistics.

$$p_\kappa(\kappa) = \frac{1}{\sqrt{\det 2\pi S}} \exp \left( -\frac{1}{2} \kappa^T S^{-1} \kappa \right)$$

with $S$ diagonal in Fourier space. Would yield only Gaussian constrained realisations or behave as a Wiener filter if the MAP is the target.

Decomposition of the score of the full prior $p(\kappa)$:

$$\nabla_\epsilon \log p(\kappa) = \nabla_\epsilon \log p_\kappa(\kappa) + \nabla_\epsilon \log p_{\epsilon|x}(\kappa)$$

Sampling from score function

MCMC procedures (Langevin Dynamics, Hamiltonian Monte Carlo) only depends on the gradient of the log distribution.

$$x_{t+1} = x_t + \frac{1}{\sqrt{\sigma^2}} \nabla_\epsilon \log p(x_t) + \sqrt{\sigma^2} w, \quad w \sim \mathcal{N}(0, I)$$

Annealing is used to avoid difficulties due to low density regions between modes. The MCMC updates are computed using a Gaussian-convolved version of the target density.

$$\sigma^2$$ gradually annealed to low temperatures and the chain progressively moves towards a point in the target distribution. $s_1 > s_2 > s_3 > s_4, p_{\epsilon|x}(\epsilon|x, s_4) \approx \hat{p}_{\epsilon|x}(\epsilon|x, s_1)$

Results

Training and testing on the MassiveVellus suite of simulations [4].

High quality mass map reconstruction from real survey [5].

References


\[\epsilon = \epsilon_1 + \gamma \text{ with } \epsilon_1 < \epsilon_2 \Rightarrow \epsilon_2 = \gamma\]