

COSMICRIM : RECONSTRUCTING EARLY UNIVERSE BY COMBINING DIFFERENTIABLE SIMULATIONS WITH RECURRENT INFERENCE MACHINES

Anonymous authors

Paper under double-blind review

ABSTRACT

Reconstructing the Gaussian initial conditions at the beginning of the Universe from the survey data in a forward modeling framework is a major challenge in cosmology. This requires solving a high dimensional inverse problem with an expensive, non-linear forward model: a cosmological N-body simulation. While intractable until recently, we propose to solve this inference problem using an automatically differentiable N-body solver, combined with a recurrent networks to learn the inference scheme and obtain the maximum-a-posteriori (MAP) estimate of the initial conditions of the Universe. We demonstrate using realistic cosmological observables that learnt inference is 40 times faster than traditional algorithms such as ADAM and LBFGS, which require specialized annealing schemes, and obtains solution of higher quality.

1 INTRODUCTION

Forward modeling approaches simulate the cosmological survey data, such as galaxies, in a hierarchical fashion starting from the initial conditions of the Universe. These are the most promising way of maximizing the information extraction from the next generation of cosmological surveys, because the corresponding inverse problem maps the non-Gaussian distribution of the data to a Gaussian distribution of initial conditions (Seljak et al., 2017). This enables performing the marginal integral over these Gaussian initial conditions to obtain the data likelihood as a function of parameters of interest, such as the density of dark matter and dark energy in the universe and initial conditions set by inflation. However, to achieve this we face the challenge of solving an inverse problem in a very high-dimensional space with a complex, non-linear and expensive forward model based on an N-body simulation. In this work, we demonstrate how differentiable cosmological N-body simulations, like FlowPM (Modi et al., 2020), can be combined with learnt optimization methods (Andrychowicz et al., 2016) to learn inference schemes and tackle these challenges efficiently. Specifically, we consider the problem of reconstructing the primordial initial conditions of the Universe from a sparse dark matter halo sample, which is a proxy for observed galaxies in cosmological surveys.

We approach this reconstruction problem in a hierarchical Bayesian framework wherein we combine the likelihood of the observed data with the Gaussian prior on the initial conditions to obtain a posterior on them (Jasche & Wandelt, 2013; Seljak et al., 2017). This prior depends on the initial power spectrum, which is the ultimate goal of these studies. In this work we focus on the reconstruction, which amounts to solving an optimization problem to obtain the maximum-a-posteriori (MAP) estimate of the initial conditions at a fixed fiducial power spectrum. The optimization problem is multi-million or even multi-billion in dimensionality, as we are interested in estimating the initial density field at all points in the Universe where we have observations. It is ill-conditioned because of noise and incompleteness in the data.

Over the last several years, different works have built physically motivated annealing schemes relying on the dynamics of the forward models to solve this problem (Seljak et al. (2017); Modi et al. (2018; 2019)). However this involves hundreds of iterations of the computationally expensive forward model evolving billions of particles under gravity, making the process very expensive. Instead, here we learn this inference scheme with a handful iterations using recurrent networks in the framework of recurrent inference machines (RIM) (Putzky & Welling, 2017; Morningstar et al., 2019).

RIM is an iterative map which includes the current reconstruction, a hidden memory state, and the gradient of the likelihood term that encodes the information about forward model, and proposes the next update for the current reconstruction. RIMs were originally proposed to learn an iterative inference algorithm to solve linear inverse problems without the need to explicitly specify a prior or a particular inference procedure. However our motivation is different since we already know the prior on our latent parameters. Instead, we show that with certain architectural enhancements, RIM allows us to speed up the optimization by orders of magnitude, and implicitly learn the otherwise hand-crafted annealing scheme for non-linear, ill-conditioned problems in high dimensions.

2 COSMICRIM

We are interested in reconstructing the initial density field \mathbf{x} given a cosmological observable \mathbf{y} , which can be modeled with a non-linear differentiable forward model \mathcal{F} . Let $p(\mathbf{y}|\mathbf{x}) = N(\mathcal{F}(\mathbf{x}), \sigma^2)$ be the likelihood of the observed data, which we model as a Gaussian noise with variance σ^2 that is given by the experiment. Let $p_\theta(\mathbf{x}) = N(0, \theta)$ be the Gaussian prior on the initial conditions parametrized with the initial power spectrum θ , which we assume to be fixed at some fiducial value in this work. Then our reconstruction of the initial conditions \mathbf{x} corresponds to the maximum-a-posteriori estimate, which maximized the joint $p(\mathbf{x}, \mathbf{y})$

$$\max_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{x}} [\ln p(\mathbf{y}|\mathbf{x}) + \ln p_\theta(\mathbf{x})] \quad (1)$$

Taking inspiration from RIM (Putzky & Welling, 2017), we optimize this posterior by training a recurrent neural network that follows the following update equation at every iteration t ,

$$\begin{aligned} \mathbf{s}_{t+1} &= \mathbf{s}_t + h_\phi(\mathbf{s}_t, \nabla_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{y}), \mathbf{x}_t), \\ \mathbf{x}_{t+1} &= \mathbf{x}_t + g_\phi(\mathbf{x}_t, \nabla_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{y}), \mathbf{s}_{t+1}). \end{aligned} \quad (2)$$

The update is informed by the gradient of the joint at the current position $\nabla_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{y})$, as well as by the state vector \mathbf{s}_t that keeps track of previous information in the style of Recurrent Neural Networks. Here g_ϕ and h_ϕ are the neural network functions parametrized by learnt weights ϕ .

We construct a 10-step RIM i.e. $t = 0, 1 \dots 10$. During training, RIM simulates all these steps of the inference and at each step the model will produce a prediction. The total loss to train parameters ϕ combines all these steps to produce predictions that improve over time. For CosmicRIM, we use a simple mean squared loss that combines every step such that $\mathcal{L} = \sum_{t=0}^{10} (\mathbf{x}_t(\phi) - \mathbf{x}_{\text{true}})^2$. For training, we use ADAM optimization (Kingma & Ba, 2014) with starting learning rate of 10^{-4} .

We show architecture of CosmicRIM in Figure 1. Every convolution and LSTM kernel is of size 5 and has 16 channels. There are two more notable features of CosmicRIM -

1. ADAM optimization significantly outperforms simple gradient descent for our optimization problem, with the latter failing completely. Thus instead of using the gradients $\nabla_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{y})$ directly at the current position in the update function g_ϕ , we use the update predicted by the ADAM algorithm \mathbf{a}_t , which itself is constructed by using the gradients from the previous RIM iterations.

2. Our non-linear forward model couples different scales, and the information primarily flows from large to small scales due to gravitational evolution. Thus we use U-Net (Ronneberger et al., 2015) like architecture with different LSTM cells (Hochreiter & Schmidhuber, 1997) in parallel to learn updates for large and small scales separately before merging them again.

3 RESULTS

We consider dark matter halo field with number density $\bar{n} = 10^{-3} \text{ (h/Mpc)}^3$ as our observable. This will be typical of upcoming cosmological surveys like Dark Energy Spectroscopic Instrument (DESI) (DESI Collaboration et al., 2016). Our data is generated with 512^3 particles evolved in 400 Mpc/h box with 40 steps of FastPM scheme (Feng et al., 2016)) and force resolution $B=2$.

For our forward model, we evolve 64^3 dark matter particles on a uniform periodic mesh using second order Lagrangian perturbation theory (2-LPT) and combine it with a quadratic bias model to mimic halos. The two bias parameters, b_1 and b_2 , are fit by matching the model and the data at the level of the fields in Fourier space. This is then also used as our data likelihood in Eq. 1 with the noise variance σ^2 given by the error power spectrum of the residual field after fitting the bias

function. We refer the reader to Schmittfull et al. (2018); Modi et al. (2019), for further details on the forward model and the corresponding likelihood function. Figure 2 shows the projection of true initial conditions and the halo data as well as the reconstructed fields with CosmicRIM.

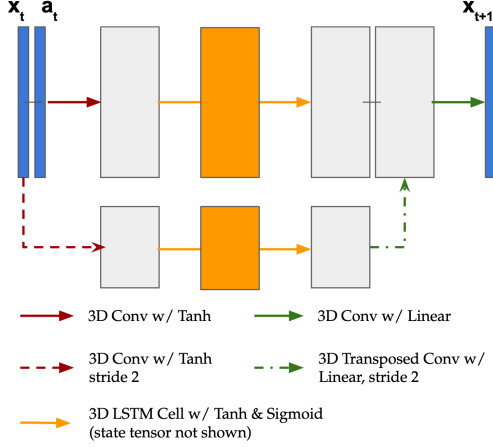


Figure 1: Architecture of CosmicRIM: at every iteration t , RIM takes the current position (\mathbf{x}_t) and proposed ADAM update (\mathbf{a}_t) to predict an update for the next iteration (\mathbf{x}_{t+1})

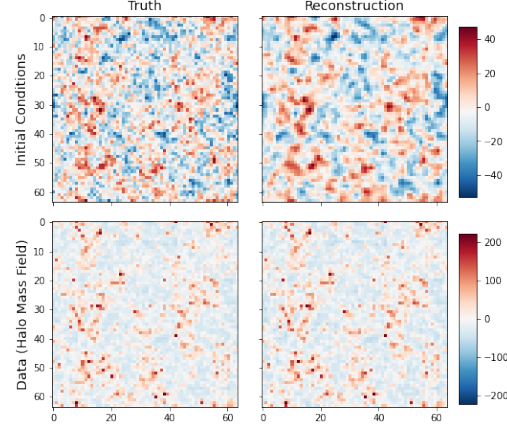


Figure 2: Projection of the initial conditions and the halo mass data (top & bottom row) for true fields and reconstruction with CosmicRIM (first & second column).

To statistically quantify how well we reconstruct the initial field, we compare it to the true initial field by measuring their transfer function t_f , the square root of the ratio of the power spectra of two fields, and their cross correlation r_c

$$t_f(k) = \sqrt{\frac{P_a(k)}{P_b(k)}}; \quad r_c(k) = \frac{P_{ab}(k)}{\sqrt{P_a(k) \times P_b(k)}} \quad (3)$$

with P_{ab} being the cross power spectra of the two fields and a , b correspond to reconstructed and true initial fields in our comparison.

In Figure 3, we compare the reconstruction of CosmicRIM with other optimization schemes when starting from an un-informative white noise initial position with unit variance. After training, CosmicRIM takes 10 steps for reconstruction. We compare it to the ADAM optimization for 10 and 100 steps starting from the same position and find that CosmicRIM clearly outperforms both of these.

We also implement an annealing scheme developed in Modi et al. (2018); Feng et al. (2018), motivated by the knowledge of cosmological dynamics. The dynamics are linear on large scales and hence these should converge quickly to correct solution while information is saturated on small scales due to noise and shell-crossing. Hence the annealing scheme smooths the loss function on small scales to reconstruct the large scales first, and then successively decreases the smoothing scale. We implement this with ADAM and LBFGS algorithm, which is a pseudo-second order scheme, in 400 steps. CosmicRIM is on par with these results for cross-correlation but clearly improves upon the transfer function on the large scales. However most importantly, it takes 40 times fewer iterations than these algorithms.

Performance of RIM can be significantly improved on small scales by using a more informative initial position than white noise with unit power. We construct this initial position by doing a single step standard reconstruction (Eisenstein et al., 2007) with the observed data, since it takes us closest to the correct solution on the largest scales. Additionally, though not shown here, we verify that using ADAM or LBFGS starting from the final output of CosmicRIM does not improve our reconstruction with respect to any metric of interest, suggesting RIM already finds a good MAP for the problem.

Finally, to understand the optimization route learnt by CosmicRIM, we also look at the metrics for all the 10 inference steps made by RIM in Figure 4. We find that implicitly, RIM learns a scheme similar to the one that annealing imposes explicitly, by learning to reconstruct the more informative large scales first which are to converge faster, and then it pushes on smaller scales.

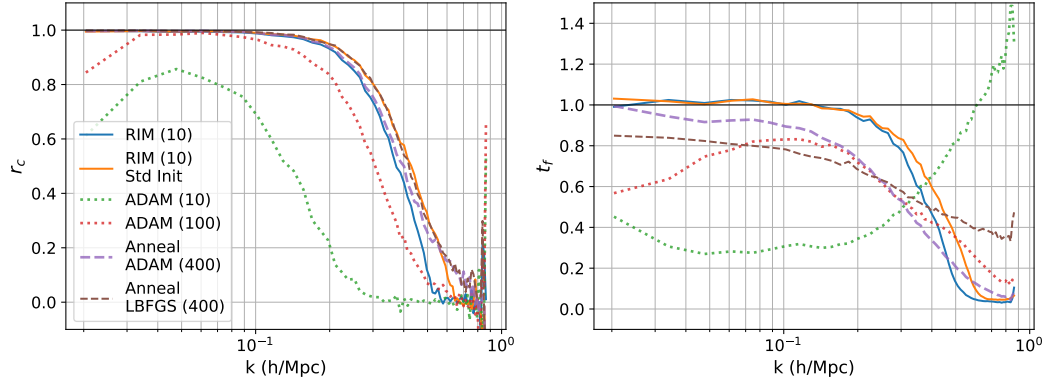


Figure 3: Cross correlation and transfer function of reconstructed initial field with the truth for different optimization schemes: comparing i) 10 step RIM (white noise initial position and standard reconstruction initial position) with ii) 10, 100 step ADAM and iii) annealed optimization (requires 400 steps) using ADAM and LBFGS algorithms

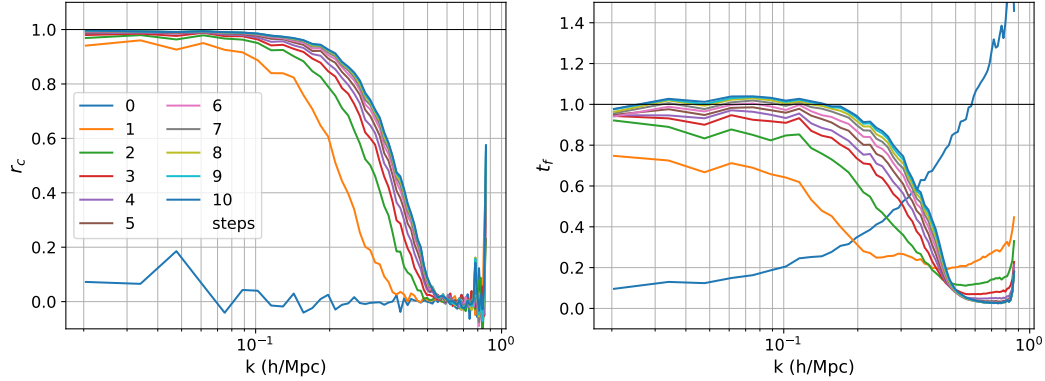


Figure 4: Iterations of RIM : evolution of cross correlation and transfer function over 10 steps of optimization when starting from white noise initial position

4 CONCLUSION

We combine differentiable forward models in cosmology, such as FlowPM, with learnt inference schemes such as recurrent inference machines for learning inference in high-dimensional cosmological problems with complex, non-linear forward model. We develop CosmicRIM to reconstruct the initial conditions of the Universe from a dark matter halo mass survey data and show that it outperforms traditional optimization schemes, including domain-specialized annealing schemes, and requires 40 times fewer iterations in our tests. Moreover, analyzing the RIM iterations, we find that it follows an optimization path similar to the one that the specialized annealing scheme is constructed to impose based on knowledge of cosmological dynamics.

Fast optimization opens doors to many applications that require multiple rounds of inference on a similar class of problems. In cosmology, this can be used to reconstruct Baryon Acoustic Oscillations that involves reconstruction on a large number of mock catalogs to estimate covariance matrices (Beutler et al., 2017), or for using simulations to estimate the posterior of the band powers of the reconstructed field (Seljak et al., 2017). One of the drawbacks of RIM is the high memory requirements for training. However in preliminary studies, we find that these can be overcome by breaking down the optimization path in steps learnt with successive networks, each trained sequentially on the output of previous networks. This will allow to scale up the learned optimization method proposed here, as is necessary for tackling the largest of the next generation of cosmological surveys.

REFERENCES

- Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. *arXiv e-prints*, art. arXiv:1606.04474, June 2016.
- Florian Beutler, Hee-Jong Seo, Ashley J. Ross, Patrick McDonald, Shun Saito, Adam S. Bolton, Joel R. Brownstein, Chia-Hsun Chuang, Antonio J. Cuesta, Daniel J. Eisenstein, Andreu Font-Ribera, Jan Niklas Grieb, Nick Hand, Francisco-Shu Kitaura, Chirag Modi, Robert C. Nichol, Will J. Percival, Francisco Prada, Sergio Rodriguez-Torres, Natalie A. Roe, Nicholas P. Ross, Salvador Salazar-Albornoz, Ariel G. Sánchez, Donald P. Schneider, Anže Slosar, Jeremy Tinker, Rita Tojeiro, Mariana Vargas-Magaña, and Jose A. Vazquez. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space. *MNRAS*, 464(3):3409–3430, January 2017. doi: 10.1093/mnras/stw2373.
- DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Bolland, and et al. The DESI Experiment Part I: Science, Targeting, and Survey Design. *ArXiv e-prints*, October 2016.
- Daniel J. Eisenstein, Hee-Jong Seo, Edwin Sirko, and David N. Spergel. Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak. *ApJ*, 664(2):675–679, August 2007. doi: 10.1086/518712.
- Yu Feng, Man-Yat Chu, Uroš Seljak, and Patrick McDonald. FASTPM: a new scheme for fast simulations of dark matter and haloes. *MNRAS*, 463(3):2273–2286, Dec 2016. doi: 10.1093/mnras/stw2123.
- Yu Feng, Uroš Seljak, and Matias Zaldarriaga. Exploring the posterior surface of the large scale structure reconstruction. *Journal of Cosmology and Astro-Particle Physics*, 2018(7):043, Jul 2018. doi: 10.1088/1475-7516/2018/07/043.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Jens Jasche and Benjamin D. Wandelt. Bayesian physical reconstruction of initial conditions from large-scale structure surveys. *MNRAS*, 432(2):894–913, June 2013. doi: 10.1093/mnras/stt449.
- D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, December 2014.
- C. Modi, Y. Feng, and U. Seljak. Cosmological reconstruction from galaxy light: neural network based light-matter connection. *Journal of Cosmology and Astro-Particle Physics*, 10:028, October 2018. doi: 10.1088/1475-7516/2018/10/028.
- C. Modi, M. White, A. Slosar, and E. Castorina. Reconstructing large-scale structure with neutral hydrogen surveys. *arXiv e-prints*, art. arXiv:1907.02330, Jul 2019.
- Chirag Modi, Francois Lanusse, and Uros Seljak. FlowPM: Distributed TensorFlow Implementation of the FastPM Cosmological N-body Solver. *arXiv e-prints*, art. arXiv:2010.11847, October 2020.
- Warren R. Morningstar, Laurence Perreault Levasseur, Yashar D. Hezaveh, Roger Blandford, Phil Marshall, Patrick Putzky, Thomas D. Rueter, Risa Wechsler, and Max Welling. Data-driven reconstruction of gravitationally lensed galaxies using recurrent inference machines. *The Astrophysical Journal*, 883(1):14, sep 2019. doi: 10.3847/1538-4357/ab35d7. URL <https://doi.org/10.3847/1538-4357/ab35d7>.
- Patrick Putzky and Max Welling. Recurrent Inference Machines for Solving Inverse Problems. *arXiv e-prints*, art. arXiv:1706.04008, June 2017.
- O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pp. 234–241. Springer, 2015. URL <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a>. (available on arXiv:1505.04597 [cs.CV]).

- M. Schmittfull, M. Simonović, V. Assassi, and M. Zaldarriaga. Modeling Biased Tracers at the Field Level. *arXiv e-prints*, November 2018.
- U. Seljak, G. Aslanyan, Y. Feng, and C. Modi. Towards optimal extraction of cosmological information from nonlinear data. *Journal of Cosmology and Astro-Particle Physics*, 2017(12):009, Dec 2017. doi: 10.1088/1475-7516/2017/12/009.