De-noising non-Gaussian fields in cosmology with normalizing flows

Adam Rouhiainen
Department of Physics
University of Wisconsin-Madison
Madison, WI 53706
rouhiainen@wisc.edu

Moritz Münchmeyer
Department of Physics
University of Wisconsin-Madison
Madison, WI 53706
muenchmeyer@wisc.edu

Abstract

Fields in cosmology, such as the matter distribution, are observed by experiments up to experimental noise. The first step in cosmological data analysis is usually to de-noise the observed field using an analytic or simulation driven prior. On large enough scales, such fields are Gaussian, and the de-noising step is known as Wiener filtering. However, on smaller scales probed by upcoming experiments, a Gaussian prior is substantially sub-optimal because the true field distribution is very non-Gaussian. Using normalizing flows, it is possible to learn the non-Gaussian prior from simulations (or from more high-resolution observations), and use this knowledge to de-noise the data more effectively. We show that we can train a flow to represent the matter distribution of the universe, and evaluate how much signal-to-noise can be gained in idealized conditions, as a function of the experimental noise. We also introduce a patching method to reconstructing information on arbitrarily large images by dividing them up into small maps (where we reconstruct non-Gaussian features), and patching the small posterior maps together on large scales (where the field is Gaussian).

1 Introduction

Normalizing flows [1] have been shown to be very effective at learning high-dimensional probability distribution functions (PDFs), in particular when the random variables are spatially organized as in an image. This has led to a lot of recent work where PDFs in physics have been parametrized with flows, in particular in the domain of lattice QCD [2]. In our precursor work [3], we evaluated how well various flows can learn sample generation and density estimation of cosmological fields. In the present work, we use the learned flow for a practical application, finding the maximum a posterior (MAP) value of a noisy observation. The potential gain of the method is that observations from galaxy survey telescopes could be de-noised using a normalizing flow to ultimately reach better cosmological constraints. Here we do not yet go all the way to real data application but make some simplifications to focus on the machine learning problem.

In cosmology, apart from our previous work [3], flows have recently been used to represent the matter distribution of the universe in [4]. This paper designed a rotation invariant flow, TRENF, specifically for cosmology, while here we used a classical real NVP flow [5] which is only translation invariant. The TRENF paper is not solving the posterior reconstruction task discussed here but is concerned with the parameter dependence of the flow on cosmological parameters. The closest existing works which we are aware of are [6] and [7], which also aim to improve the posterior of a noisy observation of a Gaussian field, by using a learned prior. However in their case, a score matching approach was used which learns only gradients, rather than a normalizing flow that gives the complete normalized PDF. Further, these works did not systematically study how much signal to noise can be gained

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

depending on the noise in the experiment, and considered the case of the lensing convergence map, rather than the matter distribution. In the following we will use the simulated matter distribution as a proxy for observable non-Gaussian fields that depend on the matter distribution, including the smoothed galaxy field in a galaxy survey, the convergence map of galaxy surveys, or the secondary anisotropies of a a high-resolution CMB survey (such as kSZ and CMB lensing) [8] [9].

The real NVP flow, used in this work to learn the matter distribution from simulations, is expressive and is fast for both sampling and inference. Of particular use for us is that real NVP flow can be trained on either periodic or non-periodic data simply by setting the padding mode of the network convolutions to either periodic or zero padding. We use the same flow architecture for periodic data (Section 3) and non-periodic data (Section 4) in this work, only changing the convolution padding mode. The details of our network resemble [2], which we have modified for this work under the CC BY license. Our flow stacks 16 affine coupling layers, each with their own CNN. We use 3 convolutional layers with kernel size 3 and leaky ReLU activation functions, and we use 12 hidden feature maps after each of the first and second convolutions. This architecture has a receptive field of 97 px, which on our data corresponds to a Fourier mode of $k = 1.6 \times 10^{-2}\,h/\mathrm{Mpc}$. In practice however, the real NVP network is learning the higher k modes more accurately [3]. This is acceptable for our purposes, as the small k modes will be reconstructed with Wiener filtering. Testing different sets of hyperparameters, we found improved generalization to OOD data with this setup of a relatively small 26,336 parameters, at the cost of a larger receptive field.

2 Method - Optimizing the posterior

In a cosmological experiment, such as a survey of the galaxy distribution, one can assume that the vector of data $d = s + n$ is a sum of mutually uncorrelated signal s and noise n, with covariances $S = \langle ss^{\mathrm{T}} \rangle$ and $N = \langle nn^{\mathrm{T}} \rangle$. The first step in cosmological data analysis is often to find the MAP $\hat{s}$ of the signal given the data d, assuming that S and N are known. The MAP is given by maximizing the posterior

$$\ln P(s|d) \propto \ln P(d|s) + \ln P(s) \tag{1}$$

$$= -\frac{1}{2}(s - d)^{\mathrm{T}} N^{-1}(s - d) + \ln P(s) \tag{2}$$

with respect to the signal s to find the MAP $\hat{s}$. Here we assumed that the noise of the experiment, which appears in the likelihood, is Gaussian, which is usually the case in practice. If we also assume that the signal is a Gaussian field, i.e. that the prior is

$$\ln P(s) \propto -\frac{1}{2} s^{\mathrm{T}} S^{-1} s, \tag{3}$$

then there is an analytic solution to the maximization, called Wiener filtering, given by

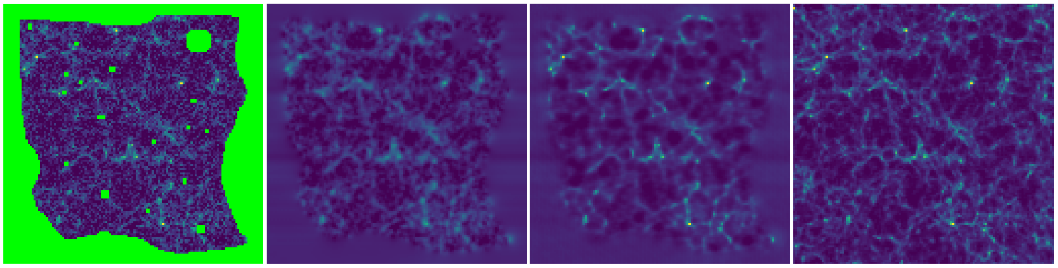
$$\hat{s}_{\mathrm{WF}} = S(S + N)^{-1} d. \tag{4}$$

Wiener filtering is very common in cosmology, see for example [10], [11]. However, upcoming surveys in cosmology such as Rubin Observatory [12] or Simons Observatory [13] probe the matter distribution with such high resolution, that scales are being measured where the Gaussianity assumption of the signal prior does not hold at all. Until recently, it would have been difficult to improve upon this assumption, because no good analytic expressions for the matter distribution $P(s)$ at non-Gaussian scales exist. In this work, we introduce posterior reconstruction of non-Gaussian signal maps s where the prior is learned with a normalizing flow.

Based on the learned differentiable prior we can either find the MAP solution to $\ln P(s|d)$, or perform Hamiltonian Monte-Carlo to make probabilistic instances of solutions to $\ln P(s|d)$. This work will make use of MAP solutions, while HMC results will be presented in the near future. A benefit of using a learned prior over directly training a neural network for de-noising is that the noise matrix N only appears in the likelihood term $\ln P(d|s)$, which is easy to compute for Gaussian noise. Therefore a single trained flow used as the prior $\ln P(s)$ may be used to de-noise any amount of noise N.

3 Results - De-noising applied to simulations

We use the particle mesh code `FastPM` [14] to generate an ensemble of simulations of the matter distribution of patches of the universe, and project them to 2 dimensions for computational simplicity.



(a) Observed (noisy, mask) (b) Wiener filtered (c) Flow posterior (d) Truth

Figure 1: Observed, reconstructed posterior, and truth maps for $1.0\tilde{\sigma}$ noise and a mask. Wiener filtering reduces noise on large and moderate length scales at the cost of over-smoothing small length scales. The flow posterior maps correctly retain the high frequency modes. The maps have length 128 px, with physical length $512\ \mathrm{Mpc}/h$.

Using cosmological parameters $\Omega_M = 0.315$ and $\sigma_8 = 0.811$, we simulate 128^3 particles in a 512 Mpc$/h$ side-length periodic box, and run 10 steps from scale factor $a = 10$ to $a = 1$. The particles are then fitted to mesh, creating 3D arrays of 128^3 px. We make four 2D projections per 3D box by projecting two dimensions by a quarter of the box length along the third dimension. We make a total of 48,000 periodic matter density maps, split 80-10-10 as training, validation, and test sets. Our real NVP flow is trained on the 128^2 px 2D projections of the simulations with an RTX A4000, using a batch size of 96 with a random rotation and flip given to each map. We minimize the Kullback–Leibler divergence [1] between the flow mapping of a Gaussian noise base distribution and our simulation target distribution with an Adam optimizer of learning rate 10^{-3}, reduced by half on plateauing. The loss converges in $\sim 10^6$ training cycles, in about 10 hours.

After the flow is trained, we make simulated noisy data maps, by adding pixel-wise shot noise to the independent test data, and mask it to mimic a typical survey geometry. On this simulated data, we run an Adam optimizer to find MAP maps with our flow prior by extremizing $\ln P(y|d)$. We found that we obtain the best results by using the flow prior only on small non-Gaussian scales, while optimizing the large linear scales with ordinary Wiener filtering (where it is optimal). Thus we make a Fourier cutoff at about $k = 0.2\ h/\mathrm{Mpc}$, taking the small k modes from Wiener filtering and the large k modes from the flow posterior. An example demonstrating how our flow posterior reconstructs information on a noisy, masked map is in Fig. 1; the noise in this example is a relatively large $1.0\tilde{\sigma}$, where we call $\tilde{\sigma}$ the pixel-wise st. dev. of our training data.

We measure the quality of our posterior maps in several ways: the MSE per pixel, the power spectra, and the reconstruction noise defined by

$$N^{\mathrm{filter}}(k) = \langle (\epsilon^{\mathrm{filter}})^{\dagger} \epsilon^{\mathrm{filter}} \rangle \tag{5}$$

where $\epsilon^{\mathrm{filter}} = y^{\mathrm{filter}} - y^{\mathrm{truth}}$. We also measure the accuracy of the reconstruction with the Fourier mode cross-correlation coefficient:

$$r(k) = \frac{P^{\mathrm{true,filter}}(k)}{\sqrt{P^{\mathrm{true}}(k) P^{\mathrm{filter}}(k)}}. \tag{6}$$

where $P^{\mathrm{true,filter}}(k)$ is the cross power spectrum.

We present results for reconstructing 100 maps in our test set for the $1.0\tilde{\sigma}$ noise and mask setup. The Wiener filtered MSE per pixel (computed in the non-masked region) is $0.130\tilde{\sigma}$, while for the flow posterior it is reduced to $0.101\tilde{\sigma}$. Fig. 2 shows plots of the power spectrum of the posterior maps and $N(k)$ (left), and cross-correlation (center). We find improvement with the flow against Wiener filtering on all scales above the nonlinear scale $k \gtrsim 0.2\ h/\mathrm{Mpc}$, with an improvement of up to a factor of 2 in this setup for large k. Our results are reproducible in Jupyter Notebooks available at `github.com/SubmissionForPapers/DenoisingWithNF`

We also examined how the improvement in the reconstruction depends on the noise in the map (without a mask here). Fig. 2 (right) shows the MSE per pixel calculated as a function of noise level (relative to $\tilde{\sigma}$), comparing the flow posterior with Wiener filtering. We find a lower MSE with the flow at all noise levels. The greatest reduction in the flow's MSE over Wiener filtering is for the

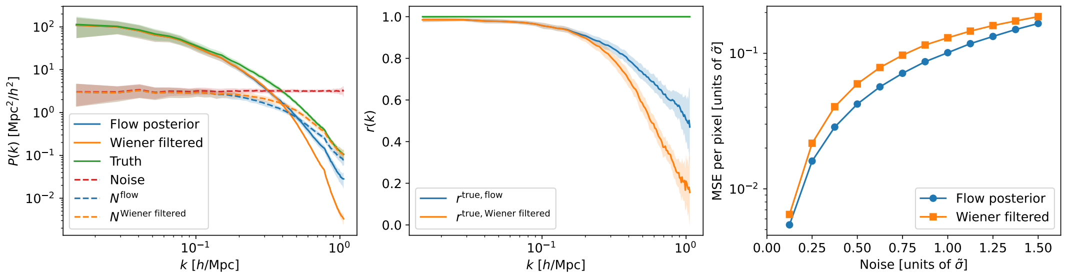


Figure 2: Power spectra (left) and cross-correlation (center) comparing our flow posterior with Wiener filtering for de-noising $1.0\bar{\sigma}$ noise, averaged over 100 maps, along with 1σ confidence intervals. We find improvement with the flow over Wiener filtering on all modes above the nonlinear scale $k \gtrsim 0.2\ h/\mathrm{Mpc}$: the flow power spectrum is closer to the truth, $N^{\rm flow}$ is lower than $N^{\rm Wiener\ filtered}$, and $r^{\rm true,\ flow}$ shows a factor of 2 improvement at larger k modes. (Right) MSE per pixel between posterior and truth maps for a range of noise levels, calculated on 100 maps at each point. The flow improves de-noising at all noise levels relative to Wiener filtering.

noise around half of the signal ($0.5\bar{\sigma}$), giving about a 30% improvement, with lesser improvements at either the low or high noise limit. Intuitively, if the noise is small, the gains by a better prior will be small since the prior matters less. On the other hand, for very large noise, the prior will begin to dominate over the likelihood. A difficulty, as we have found, is that the normalizing flow is trained on IID data, but the posterior optimization enters domains of the configuration space that are OOD and thus the flow may not generalize well to such cases [15]. We found that it is advantageous for generalization to use a flow with relatively few training parameters, and that splitting the k modes as explained above helps at large noise.

4 Patching maps together to reconstruct large maps

Above we discussed that we use an ordinary Wiener Filter on large (linear) scales, and only use the flow on small nonlinear scales. This also allows us to reconstruct very large maps with a patching procedure, without the need to increase the flow dimension. As an example we reconstruct a large map of length $n_{\rm L} = 1024$ px with a flow trained on non-periodic small maps of length $n_{\rm S} = 128$ px. We use the same network architecture as described in Section 1, with zeros for the convolution padding to respect the non-periodicity. Our training data is as described in Section 3, except our 128 px 2D projections are cut out of 384 px 3D simulations to get non-periodicity. The reconstruction and patching procedure is as follows. Divide the large, $n_{\rm L}$ length periodic map into a number of $(2n_{\rm L}/n_{\rm S})^2$ evenly spaced small maps of length $n_{\rm S}$. These small maps have $n_{\rm S}/4$ length overlapping regions with their four neighboring maps. Reconstruct these non-periodic small maps with a trained flow. Apply Wiener filtering to the large map to reconstruct the small k modes. To avoid discontinuities near the edges of the small maps, add only the the large k modes from only the center $n_{\rm S}/2$ length region of the flow reconstructed maps to the Wiener filtered large map.

The critical step that allows the smaller maps to be patched together without discontinuities is the Wiener filtering applied to the entire large map. This 'global' Wiener filtering is well within our computational constraints for very large maps encountered in cosmology, while training a flow on such large maps would be computationally infeasible. High-resolution observed and reconstructed 1024 px maps patched from 256 small 128 px maps are available in the Appendix. There is no visible remnant of a grid where maps were patched together.

5 Conclusion

Flows are a powerful tool to deal with the non-Gaussianity of high-resolution cosmological surveys. The key feature of flows is that their exact likelihoods are tractable and here we have made use of this feature to de-noise simulated cosmological data, gaining up to a factor of 2 in the cross-correlation coefficient. In the future, we will apply this method to various observables, include dependence on cosmological parameters in the prior, and forecast gains for cosmological parameter estimation.

6 Broader impact of our work

This work uses a real NVP normalizing flow to reconstruct noisy and masked cosmological data made from simulations. The real NVP flow can be used to generate and manipulate images, and other similar normalizing flows have recently been used to generate videos (although of low quality). We do not build a new neural network model or improve any pre-existing models, but rather use an already well-known model to learn a cosmological PDF; therefore we do not believe that we contribute to the potential use of normalizing flows in harmful data generation or manipulation. Our de-noising method presented here is specifically designed to run on matter distributions in cosmology, and our results are measured in terms relevant to cosmology and astrophysics. As upcoming surveys of the matter distribution probe more nonlinear scales, we believe that this work benefits the cosmology community by providing a machine learning method of reconstructing future cosmological data.

7 Acknowledgements

We thank Kimmy Wu for valuable discussions and initial collaboration. The authors acknowledge support from DOE grant DE-SC0022342. Support for this research was provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

References

[1] George Papamakarios et al. Normalizing Flows for Probabilistic Modeling and Inference. *arXiv e-prints*, page arXiv:1912.02762, December 2019.

[2] Michael S. Albergo et al. Introduction to Normalizing Flows for Lattice Field Theory. *arXiv e-prints*, page arXiv:2101.08176, January 2021.

[3] Adam Rouhiainen, Utkarsh Giri, and Moritz Münchmeyer. Normalizing flows for random fields in cosmology. *arXiv e-prints*, page arXiv:2105.12024, May 2021.

[4] Biwei Dai and Uroš Seljak. Translation and Rotation Equivariant Normalizing Flow (TRENF) for Optimal Cosmological Analysis. *arXiv e-prints*, page arXiv:2202.05282, February 2022.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. *arXiv e-prints*, page arXiv:1605.08803, May 2016.

[6] Benjamin Remy et al. Probabilistic Mapping of Dark Matter by Neural Score Matching. *arXiv e-prints*, page arXiv:2011.08271, November 2020.

[7] Benjamin Remy et al. Probabilistic Mass Mapping with Neural Score Estimation. *arXiv e-prints*, page arXiv:2201.05561, January 2022.

[8] John E. Carlstrom, Gilbert P. Holder, and Erik D. Reese. Cosmology with the Sunyaev-Zel'dovich Effect. *Annual Review of Astronomy and Astrophysics*, 40(1):643–680, September 2002.

[9] Antony Lewis and Anthony Challinor. Weak gravitational lensing of the CMB. *Phys. Rept.*, 429:1–65, June 2006.

[10] Julien Carron and Antony Lewis. Maximum a posteriori CMB lensing reconstruction. *Phys. Rev. D*, 96(6):063510, September 2017.

[11] Moritz Münchmeyer and Kendrick M. Smith. Fast Wiener filtering of CMB maps with Neural Networks. *NeurIPS 2019 workshop Machine Learning for the Physical Sciences*, May 2019.

[12] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, and et al. LSST Science Book, Version 2.0. *arXiv e-prints*, page arXiv:0912.0201, December 2009.

[13] Peter Ade et al. The Simons Observatory: science goals and forecasts. *Journal of Cosmology and Astroparticle Physics*, 2019(02):056–056, February 2019.

[14] Yu Feng, Man-Yat Chu, Uroš Seljak, and Patrick McDonald. FastPM: a new scheme for fast simulations of dark matter and haloes. *Mon. Not. Roy. Astron. Soc.*, 463(3):2273–2286, August 2016.

[15] Polina Kirichenko, Pavel Izmailov, and Andrew Gordon Wilson. Why normalizing flows fail to detect out-of-distribution data. *arXiv e-prints*, page arXiv:2006.08545, June 2020.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Our main limitation is described at the end of Section 3, being difficulties in running experiments on OOD data. We also state in Section 1 that this work is not yet using real data, but simulations.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] We provide a broader impact statement section.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] We read the guidelines at `neurips.cc/public/EthicsGuidelines`.
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] A URL is provided in Section 3.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] At the end of Section 1
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Our power spectrum, reconstruction noise, and cross-correlation plots have confidence intervals for 100 experiments.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] At the beginning of Section 3
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] At the end of Section 1 we cite where our real NVP code was derived from
   (b) Did you mention the license of the assets? [Yes] At the end of Section 1 we mention the license for what real NVP code was derived from
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] A URL is provided in Section 3.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] We mention the license for the real NVP code we used, and we made our own simulation data with `FlowPM`, which is used the MIT License.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] Only cosmological simulations are used as data.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix: High-resolution plots for the large patched map

Here we provide high-resolution plots in Fig. 3 of an observed 1024 px matter distribution with noise and a mask, along with the flow reconstructed map created with the patching scheme described in Section 4.