Robust Neural Network-Enhanced Estimation of Local Primordial Non-Gaussianity

Utkarsh Gir† and Moritz Münchmeyer

University of Wisconsin-Madison, Madison, Wisconsin, USA

Kendrick M. Smith
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, CA

(Dated: May 27, 2022)

When applied to the non-linear matter distribution of the universe, neural networks have been shown to be very statistically sensitive probes of cosmological parameters, such as the linear perturbation amplitude $\sigma_8$. However, when used as a “black box”, neural networks are not robust to baryonic uncertainty. We propose a robust architecture for constraining primordial non-Gaussianity $f_{NL}$, by training a neural network to locally estimate $\sigma_8$, and correlating these local estimates with the large-scale density field. We apply our method to N-body simulations, and show that $\sigma(f_{NL})$ is 3.5 times better than the constraint obtained from a standard halo-based approach. We show that our method has the same robustness property as large-scale halo bias: baryonic physics can change the normalization of the estimated $f_{NL}$, but cannot change whether $f_{NL}$ is detected.

Keywords: non-Gaussianity, neural networks, large-scale structure

I. INTRODUCTION

Observations of cosmological perturbations on quasi-linear scales, with the cosmic microwave background (CMB) and large scale structure (LSS), have resulted in precise measurements of fundamental cosmological parameters. However, quasi-linear scales contain only a small fraction of the theoretically accessible information. On smaller scales, traditional $N$-point correlation function analysis becomes sub-optimal, and inference methods such as forward modelling [1, 2] or machine learning [3, 4] can provide stronger constraints.

Here, we will focus on inference of the important primordial physics parameter $f_{NL}$, which arises in multifield models of inflation [8–11]. In such models, the primordial Bardeen potential $\Phi(x)$ can be parameterized as:

$$\Phi(x) = \Phi_G(x) + f_{NL}(\Phi_G(x)^2 - \langle \Phi_G^2 \rangle)$$ (1)

where $\Phi_G$ is a Gaussian field and $f_{NL}$ quantifies the level of non-Gaussianity. Constraining $f_{NL} < 1$ is a major goal of upcoming galaxy surveys [12].

Several statistical approaches for estimating $f_{NL}$ in LSS have been proposed, including the squeezed bispectrum [13] and the scale-dependent power spectrum approach [14, 15], together with the idea of sample variance cancellation using a variety of probes [16–19].

In this paper, we will propose a neural network (NN) based approach to estimating $f_{NL}$. We will show that our NN-based analysis obtains significantly smaller error bars than an analysis based on large-scale matter and halo fields. This is perhaps unsurprising since a neural network with enough capacity, trained on enough simulations, should give statistically optimal parameter constraints.

However, robustness is a central challenge for neural networks. As is widely appreciated, simulations on small scales suffer from baryonic feedback uncertainties and are not reliable at the sub-percent level accuracy required to tighten parameter bounds. For example, an NN trained to measure $\sigma_8$ on one set of simulations will likely give incorrect results on different simulations or on real data [20, 21].

This problem is not unique to NN-based methods. In particular, the average dark matter halo density $\bar{n}_h$ is statistically a precise probe of $\sigma_8$, but it is not robust because $\bar{n}_h$ is sensitive to uncertain local physics. However, anisotropy in the halo field can be used to place robust constraints on $f_{NL}$. A famous result [14] states that if $f_{NL} \neq 0$, the halo bias $b_h(k)$ contains a characteristic $1/k^2$ term, which cannot be induced by local physics.

The idea of this paper is to replace the halo field by an NN-based local estimate $\pi(x)$ of $\sigma_8$. As in the halo case, $\pi$ is not robust as an absolute measurement of $\sigma_8$, but the bias $b_{\pi}(k)$ contains a term proportional to $f_{NL}/k^2$, which can be used to place robust constraints on $f_{NL}$. We will show that this approach combines the statistical power of neural networks with the robustness of the traditional halo-based analysis.

II. FORMALISM

In an $f_{NL}$ cosmology, the halo bias $b_h(k)$ contains a term proportional to $f_{NL}/k^2$ [14]. In the next few paragraphs, we review the derivation, in a language which will generalize to NN-based observables.

If $f_{NL} \neq 0$, the locally observed amplitude of short-wavelength modes is a function $\sigma_{8,loc}^2(x)$ of position. On large scales $k_L \rightarrow 0$, anisotropy in $\sigma_{8,loc}^2$ is related to the potential $\Phi$ by:

$$\frac{\sigma_{8,loc}^2(k_L)}{\sigma_8^2} = 2f_{NL}\Phi(k_L)$$ (2)
The coupling between large and small scales described by Eq. (2) arises from the term $f_{NL} \Phi^2_{N}$ in Eq. (1), which mixes scales. For a formal derivation of (2), see [11] [22].

On large scales, the local halo abundance $\delta_m$ is sensitive to both $\sigma^8_{\text{loc}}$ and the matter overdensity $\delta_m$:

$$\delta_h(x) = b^G_h \delta_m(x) + \frac{1}{2} b^{NG}_h \log \left( \frac{\sigma^8_{\text{loc}}(x)}{\sigma_8} \right) + \text{(noise)} \quad (3)$$

The first term $b^G_h \delta_m$ is the usual (Gaussian) halo bias, and the second term $b^{NG}_h \log(\sigma^8_{\text{loc}})$ is a non-Gaussian bias term. The coefficient $b^{NG}_h$ is the derivative of the halo density $\bar{n}_h$ with respect to the cosmological parameter $\sigma_8$ [22] [25]:

$$b^{NG}_h = 2 \frac{\partial \log \bar{n}_h}{\partial \log \sigma_8} \quad (4)$$

(Sometimes the approximation $b^{NG}_h \approx 2 \delta_r(b^G_h - 1)$ is used, but we will not use it in this paper.)

On linear scales, the density field $\delta_m$ and potential $\Phi$ are related via the Fourier-space Poisson equation $\delta_m(k) = \alpha(k, z) \Phi(k)$ [26]. Here, $\alpha(k, z)$ is given by

$$\alpha(k, z) \equiv \frac{2k^2 T(k) D(z)}{3 \Omega_m H_0^2} \quad (5)$$

Combining Eqs. (2)–(5), we derive the NG halo bias:

$$\delta_h(k_L) = b_h(k_L) \delta_m(k_L) + \text{(Poisson noise)} \quad (6)$$

where:

$$b_h(k) = b^G_h + b^{NG}_h \frac{f_{NL}}{\alpha(k, z)} \quad (7)$$

In this paper, we generalize the preceding results as follows. First we note that they do not depend on any specific properties of halos, other than the local halo abundance $n_h(x)$ being sensitive to $\sigma^8_{\text{loc}}(x)$. Any field $\pi(x)$ which is derived from the nonlinear density field, in a reasonably local way, should have the same property. We propose constructing a field $\pi(x)$ using a neural network trained to maximize sensitivity to $\sigma^8_{\text{loc}}$.

We assume that our input data consists of the nonlinear density field $\delta_m(x)$ as a 3-d pixelized map at ~2 Mpc resolution without noise. Thus, our constraints on $f_{NL}$ should be interpreted as information content in principle, given complete information at a specified resolution. In future work, we plan to apply our approach to simulated galaxy catalogs, which are more representative of real data.

We define a field $\pi(x)$ by applying a CNN to $\delta_m(x)$ (Figure 1). The CNN has a small receptive field (~18 $h^{-1}$ Mpc), so that the output field $\pi(x)$ is fairly local in the input field $\delta_m(x)$. As described in [11], we train the CNN so that $\pi(x)$ is an estimate of $\sigma^8_{\text{loc}}(x)$ with low statistical noise.

Following the logic above for halos, we make the following predictions for the behavior of $\pi(x)$ on large scales.

First, we predict that the matter-$\pi$ and $\pi-$$\pi$ power spectra are given by:

$$P_{m\pi}(k) = b_\pi(k) P_{mm}(k) \quad (8)$$

$$P_{\pi\pi}(k) = b_\pi(k)^2 P_{mm}(k) + N_{\pi\pi} \quad (9)$$

where the linear bias $b_\pi(k)$ is the sum of Gaussian (constant) and non-Gaussian terms:

$$b_\pi(k) = b^G_\pi + b^{NG}_\pi \frac{f_{NL}}{\alpha(k, z)} \quad (10)$$

We also predict that the non-Gaussian bias $b^{NG}_\pi$ is related to the $\sigma_8$ dependence of the mean $\pi$-field:

$$b^{NG}_\pi = 2 \frac{\partial \bar{\pi}}{\partial \log \sigma_8} \quad (11)$$

Finally, we predict that the noise $N_{\pi\pi}$ defined in (9) is constant in $k$. In [11] we will verify these predictions and show that they lead to strong constraints on $f_{NL}$.

### III. NEURAL NETWORK

**Architecture:** Our neural network uses a fully convolutional, sliding-window architecture with a total of 16433 parameters (Figure 1). The network takes the 3D matter density field $\delta_m(x)$ from an $N$-body simulation, and produces an output field $\pi(x)$ with the same resolution as the input. We use small convolution kernels, including several layers with $(1 \times 1 \times 1)$ kernels, so that the total receptive field of the network will be small ($18 \ h^{-1}$ Mpc). The size of the receptive field limits the scales which the neural network can use for estimating $\pi$ and thus enforces locality. We leave systematic exploration of neural network architecture to future work.

**Simulations:** We want to train the network so that its output field $\pi(x)$ is an optimal estimate of $\sigma_8$. To do this, we need a training set of $N$-body simulations with multiple values of $\sigma_8$. We use the s8_p and s8_m datasets from the Quijote simulations [27], with $\sigma_8 = 0.849$ and 0.819 respectively. The remaining cosmological parameters are $\Omega_m = 0.3175$, $\Omega_b = 0.049$, $h = 0.6711$, $n_s = 0.9624$, and $w = -1$. Each dataset contains 400 collisionless simulations. Each simulation has 512$^3$ particles and volume $(1 \ h^{-1}$Gpc)$^3$. For each simulation, we inpaint particles from the $z = 0$ snapshot on a 512$^3$ 3D mesh using the Cloud-in-Cell algorithm implemented in nbodykit. This produces a voxelized 3D matter density field $\delta_m(x)$, which we save to disk for neural network training. We require only two values of $\sigma_8$ in the training data rather than a continuum, because the variance of $\pi(x)$ per receptive field is much larger than the difference between the two $\sigma_8$ values.

**Loss function and optimizer:** For each simulation, let $\pi(x)$ be the CNN output, and let $\sigma_8^{\text{true}}$ be the value of $\sigma_8$
in the simulation. We define the loss function:

\[ L = \left[ \left( \frac{1}{N_{\text{voxels}}} \sum_{\text{voxels}} \pi(x) \right) - \sigma_{\text{true}} \right]^2 \]  

(12)

Intuitively, minimizing \( L \) should produce an output field \( \pi(x) \) which is an optimal estimate of \( \sigma_S \), by minimizing the difference between \( \sigma_{\text{true}} \) and the spatially averaged \( \pi \)-field.

We use the Adam optimizer [28] with a learning rate of \( (5 \times 10^{-5}) \) to minimize the loss function. The learning rate is reduced by a factor of 0.7 whenever the loss fails to register any improvement for 5 successive epochs of training. The architecture is implemented in PyTorch [29] and uses PyTorch-lightning [30] for high level interfacing with mixed precision training [31].

We find that the overall normalization \( W \) and additive bias \( b \) of the NN are slow to converge, so as a final training step, we fix all parameters except \((W, b)\), and minimize the loss (12). This minimization can be done exactly in a single epoch, since the loss is a quadratic function of \((W, b)\).

**Validation:** Our NN has been trained so that the spatially averaged \( \pi \)-field \( \bar{\pi} = \frac{1}{V_{\text{box}} \times 3} \int \pi(x) \) is an estimate of \( \sigma_S \) with lowest possible noise. In the top panel of Figure 2, we verify this statement, by evaluating \( \bar{\pi} \) on a test set of 100 + 100 simulations with \( \sigma_S \in \{0.819, 0.849\} \). We see that the network recovers the correct value of \( \sigma_S \), and that the NN obtains better statistical separation between \( \sigma_S \) values than counting halos (bottom panel).

**IV. ESTIMATING \( f_{NL} \)**

In this section, we apply our neural network to simulations with \( f_{NL} \neq 0 \). We use a test set of 10 \( N \)-body simulations with \( \sigma_S = 0.834 \), \( f_{NL} = 250 \), and non-Gaussian initial conditions generated using the Zeldovich approximation.

**Power spectra:** We next verify the predictions in Eqs. [8]–[11] for the large-scale power spectra \( P_{m\pi}, P_{\pi\pi} \) in an \( f_{NL} \) cosmology. First, we note that since \( \bar{\pi} = \sigma_S \) (Fig. 2), our prediction (11) for the non-Gaussian bias \( b_{\pi}^{NG} \) is:

\[ b_{\pi}^{NG} = 2 \frac{\partial \pi}{\partial \log \sigma_S} = 2 \sigma_S \]

(13)

and so our prediction (10) for the total bias \( b_{\pi}(k) \) is:

\[ b_{\pi}(k) = b_{\pi}^G + 2 \sigma_S \left( \frac{f_{NL}}{\alpha(k, z)} \right) \]

(14)

In the the top panel of Fig. 3, we compare the bias model [14] to the empirical bias obtained from cross-correlating \( \pi \) and \( \delta_m \) in \( k \)-bins, and find good agreement. In the bottom panel of Fig. 3 we verify the prediction that the noise power spectrum \( N_{\pi} \) defined in Eq. (9) is constant in \( k \), by plotting the power spectrum of the residual field \( \epsilon(k) = \pi(k) - b_{\pi}(k) \delta_m(k) \).
We emphasize that simulations with $f_{NL} \neq 0$ were never seen during the training process. The predictions in Eqs. [8]–[11] for power spectra in an $f_{NL}$ cosmology (in particular the prediction $b_\pi^{NL} = 2\sigma_8$) are based entirely on the NN response to varying $\sigma_8$, and general considerations of locality. Therefore, the verification of these predictions is a strong test of our formalism.

**MCMC pipeline:** Now that our model for the power spectra $P_{m\pi}, P_{\pi\pi}$ has been verified, we develop an MCMC pipeline that combines large-scale modes of $\pi(k)$ and $\delta_m(k)$ in a joint analysis.

The Gaussian likelihood for our data vector $D = [\delta_m, \pi]$ given model parameters $\Theta = (f_{NL}, b^{NL}_\pi, N_{\pi\pi})$ is:

$$\mathcal{L}(\Theta|D) \propto \prod_k \frac{1}{\sqrt{\text{Det}(C(k))}} \exp\left(-\frac{D(k)^t C(k)^{-1} D(k)}{2}\right)$$

where the $2 \times 2$ covariance matrix $C(k)$ is:

$$C(k) = \begin{bmatrix} P_{mm}(k) & P_{m\pi}(k) \\ P_{m\pi}(k) & P_{\pi\pi}(k) \end{bmatrix} = \begin{bmatrix} b_{\pi}(k) b_{\pi}(k) & b_{\pi}(k) b_{m}(k) + N_{\pi\pi} \\
 b_{\pi}(k) b_{m}(k) & b_{m}(k) b_{m}(k) + N_{\pi\pi} \end{bmatrix}$$ (15)

with $b_{\pi}(k)$ given by Eq. (14). We truncate the likelihood at $k_{\text{max}} = 0.014\ h\text{Mpc}^{-1}$. The posterior is defined using flat priors over a reasonable range for the model parameters $\Theta$. We sample the posterior using affine-invariant sampling implemented in emcee to obtain constraints on $f_{NL}$.

To compare the neural network to a traditional halo based analysis, we also run our MCMC pipeline using the halo field $\delta_h(k)$ instead of the NN-derived field $\pi(k)$. The only change is that we replace $b^{NG}_\pi = 2\sigma_8$ by the non-Gaussian halo bias $b^{NG}_h$, which we measure in simulations using Eq. (4).

**MCMC results:** We begin by analysing $N$-body simulations with Gaussian initial conditions i.e., $f_{NL} = 0$. We jointly analyze 100 fiducial Quijote simulations by multiplying together their posteriors before sampling. In the right panel of Fig. 4 we show $f_{NL}$ constraints from a joint analysis of the large-scale matter density $\delta_m(k)$ and the NN-derived field $\pi(x)$. In the left panel, we show a similar analysis using $\delta_m(k)$ and the halo field $\delta_h(k)$. In both cases, the result is consistent with $f_{NL} = 0$ as expected. However, the neural network gives an $f_{NL}$ error which is 3.5 times better than the halo based analysis!

In the left panel of Figure 4 we used a single halo field consisting of all halos with $\geq 20$ particles ($M_{\text{min}} = 1.3 \times 10^{13}\ h^{-1} M_\odot$). We checked that if narrow halo mass bins are used with optimal weighting, the Fisher forecasted error $\sigma(f_{NL})$ is only 25% better than the single-bin case.

In Fig. 5, we show $f_{NL}$ constraints from a joint MCMC analysis of 10 simulations with $f_{NL} = 250$. We can see that for these non-Gaussian simulations, the correct value of $f_{NL}$ is recovered, and the NN improvement over halos is just as good as in the $f_{NL} = 0$ case.

**Robustness:** To frame the issue of robustness concretely, imagine that the small-scale astrophysics in the real universe is slightly different from the training set. How will our constraints on $f_{NL}$ be affected?

Since the parameters ($b^{NL}_\pi, b^{NL}_h, N_{\pi\pi}$) are sensitive to small-scale physics, their values will differ slightly from the training set. For $b^{NL}_\pi$ and $N_{\pi\pi}$, this is harmless since we marginalize these parameters in our MCMC anyway.

For $b^{NG}_h$, we note that in the formalism from [11] the parameters $b^{NG}_h$ and $f_{NL}$ only appear in the combination $(b^{NG}_h f_{NL})$. Therefore, a small change in $b^{NG}_h$ is equivalent to a change in the normalization of $f_{NL}$ – it cannot “fake” a detection of nonzero $f_{NL}$. Physically, this is because local physics cannot generate a term in the bias $b_x(k)$ proportional to $1/k^2$. This is qualitatively similar to the familiar case of halo counts.

Our method does depend on having a rough estimate for $b^{NG}_\pi$ based on training data. As a check, we estimated the cosmological parameter dependence of $b^{NG}$ using the Quijote “latin hypercube” simulations, which vary cosmological parameters over wide ranges. We find that $\bar{\pi}$ is well modelled by a quadratic polynomial in $(\Omega_m, \Omega_b, h, \sigma_8)$. Using this quadratic model, we find that if cosmological parameters are varied within the Planck+BAO $2\sigma$ errors [33], the change in $b^{NG}_\pi = 2(\bar{\pi}/\partial \log \sigma_8)$ is $\leq 1\%$. In future work, we hope to extend this analysis to study dependence of $b^{NG}_h$ on subgrid physics.
\[ b_G^2 = 1.408 \pm 0.013 \]

\[ f_{NL} = -3.5 \pm 6.7 \]

\[ N_{hh} = 2591 \pm 68 \]

\[ b_G^2 = 0.8380 \pm 0.0050 \]

\[ f_{NL} = 2.0 \pm 1.8 \]

\[ N_{ππ} = 375.0 \pm 9.8 \]

Figure 4. MCMC posteriors on \( f_{NL} \) and nuisance parameters (either \( (b_G^2, N_{hh}) \) or \( (b_π^2, N_{ππ}) \)) from joint analysis of 100 Quijote simulations with \( f_{NL} = 0 \). Left. Traditional halo based analysis using large-scale modes of the matter field \( \delta_m(k) \) and halo field \( \delta_h(k) \). Right. Neural network based analysis using \( \delta_m(k) \) and the NN output field \( π(k) \). The neural network reduces the error bar on \( f_{NL} \) by a factor \( ∼3.5 \).

Figure 5. MCMC posterior on \( f_{NL} \) from joint analysis of 10 \( N \)-body simulations with \( f_{NL} = 250 \), using either the matter+halo fields (red), or matter+\( π \) fields (black), where \( π(x) \) is the NN output field. The 1-d \( f_{NL} \) likelihoods are marginalized over nuisance parameters (either \( (b_G^2, N_{hh}) \) or \( (b_π^2, N_{ππ}) \)).

V. CONCLUSION AND OUTLOOK

In this letter we have demonstrated that the statistical power of neural networks can be combined with the idea of \( (1/k^2) \) non-Gaussian bias to arrive at a robust measurement of \( f_{NL} \) from the matter distribution. Unlike forward modelling approaches which are difficult at strongly non-linear scales, our approach can use information from very small scales and still remain robust.

Our main next step will be to quantify to what extent the method presented here can improve \( f_{NL} \) constraints from realistic galaxy surveys, rather than the matter field. Machine learning based \( σ_8 \) constraints from simulated galaxy distributions have been examined in [34] (using a halo occupation distribution), [20] (using the hydrodynamic CAMELS simulations) and [35] (using a semi-analytic galaxy formation model). In particular, [20] highlighted the problem that different baryonic subgrid models lead to inconsistent results, which is precisely the issue our method is designed to overcome for \( f_{NL} \). Recently, [36] introduced a non-linear estimator based on the Wavelet Scattering Transform (WST) and even applied it to BOSS data [37] to extract cosmological parameters including \( σ_8 \). The WST behaves similarly to a neural network and thus the claimed improvements in \( σ_8 \) suggest that our \( f_{NL} \) method could also work well for galaxies. We will investigate this question in detail in upcoming work.

A straightforward generalization of our method is to other scale-dependent biases, such as those induced by the trispectrum \( g_{NL} \) parameter [38], neutrino masses [39], and isocurvature perturbations [40]. More generally, our approach of using a neural network as a local probe may generalize to other observables which are large-scale modulations of local non-Gaussian fields or cross-correlations. This is a common setup in cosmology, often exploited for quadratic estimators.
ACKNOWLEDGEMENTS

Part of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. MM acknowledges support from DOE grant DE-SC0022342. KMS was supported by an NSERC Discovery Grant and a CIFAR fellowship. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. Perimeter Institute’s HPC system “Symmetry” was used to perform some of the analysis presented in the letter. We have extensively used several python libraries including numpy, matplotlib, CLASS, getdist, and SciencePlots.
[41] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Nature 585, 357 (2020).

[42] J. D. Hunter, Computing in Science & Engineering 9, 90 (2007).

[43] D. Blas, J. Lesgourgues, and T. Tram, Journal of Cosmology and Astroparticle Physics 2011, 034 (2011).

[44] A. Lewis, (2019), arXiv:1910.13970 [astro-ph.IM].

[45] J. D. Garrett, (2021), 10.5281/zenodo.4106649.