Accelerating MCMC algorithms through Bayesian Deep Networks

Hector J. Hortua‡, Dimitri Marinelli*, Riccardo Volpi†, Luigi Malago†

‡UNAL, *FinNet, †RIST

RIST

GRUPO DE GRAVITACION Y COSMOLOGIA

Abstract

MCMC algorithms are commonly used for their versatility in sampling from complicated probability distributions. However, as the dimension of the distribution gets larger, the computational costs for a satisfactory exploration of the sampling space become challenging. In this work we show an alternative way of performing adaptive MCMC, by using the outcome of Bayesian Neural Networks(BNNs) as the initial proposal for MCMC. This combined approach increases the acceptance rate in the Metropolis-Hasting algorithm and accelerate the convergence of the MCMC while reaching the same final accuracy. We demonstrate the main advantages of this approach by constraining the cosmological parameters directly from CMB.

Dataset and Network

Following [3], we use 50.000 images related to the CMB maps projected in $20 \times 20\,\mathrm{deg}^2$ patches in the sky for training the BNNs. These images have a dimensions (256,256,3), where the last channel stands for the Temperature (channel=0) and Polarization (channel=1,2), and each image corresponds to a specific set value of some cosmological parameters. The BNN was implemented in TensorFlow-Probability, and the same version of the VGG architecture along with the presence of Flipout as it was shown in [3] was used in this paper. Finally, we used the calibration method introduced in [2] where α-divergence with $\alpha = 1$ has been included at the top of the BNN.

Method

The method shown here can be seen as a novel way to perform adaptive MCMC in which the output distribution of the BNNs serves as a proposal distribution for the MCMC. The experiments were run in the cobaya software [5], with the likelihood given by [6]

$$-\mathcal{L} \sim \sum_l (2l+1)\Bigg[\ln\left(\frac{C_l^{BB}}{\hat{C}_l^{BB}}\left(\frac{C_l^{TT}C_l^{EE}-(C_l^{TE})^2}{\hat{C}_l^{TT}\hat{C}_l^{EE}-(\hat{C}_l^{TE})^2}\right)\right) + \frac{\hat{C}_l^{BB}}{C_l^{BB}} + \frac{\hat{C}_l^{TT}C_l^{EE}+C_l^{TT}\hat{C}_l^{EE}-2\hat{C}_l^{TE}C_l^{TE}}{C_l^{TT}C_l^{EE}-(C_l^{TE})^2}\Bigg], \tag{1}$$

where $\hat{C}_l$ is the power spectrum of the CMB patches obtained with Lens-Tools [4], and C_l the theoretical model. Cobaya accepts the cosmological parameters as input, compute C_l via CLASS [1] and when the Markov chains have enough points to provide reasonable samples from the posterior distributions, the simulation stops and it return the chains. We run two MCMC experiments taking into account the power spectrum of the CMB maps. In the first MCMC experiments we used the full sky map, while in the second one, we computed the power spectra for CMB patches and used them as an input in cobaya package.

Results

Results of the conditional distributions for the predicted parameters are displayed in Fig. 1 where we compared the MCMC results with the calibrated BNNs (on the CMB patches). We observe that MCMC provides tighter and more accurate constraints. However, the trained Neural Network can generate 8000 samples in approximately ten seconds which it turns out to be 10000 times faster than MCMC for this dataset.

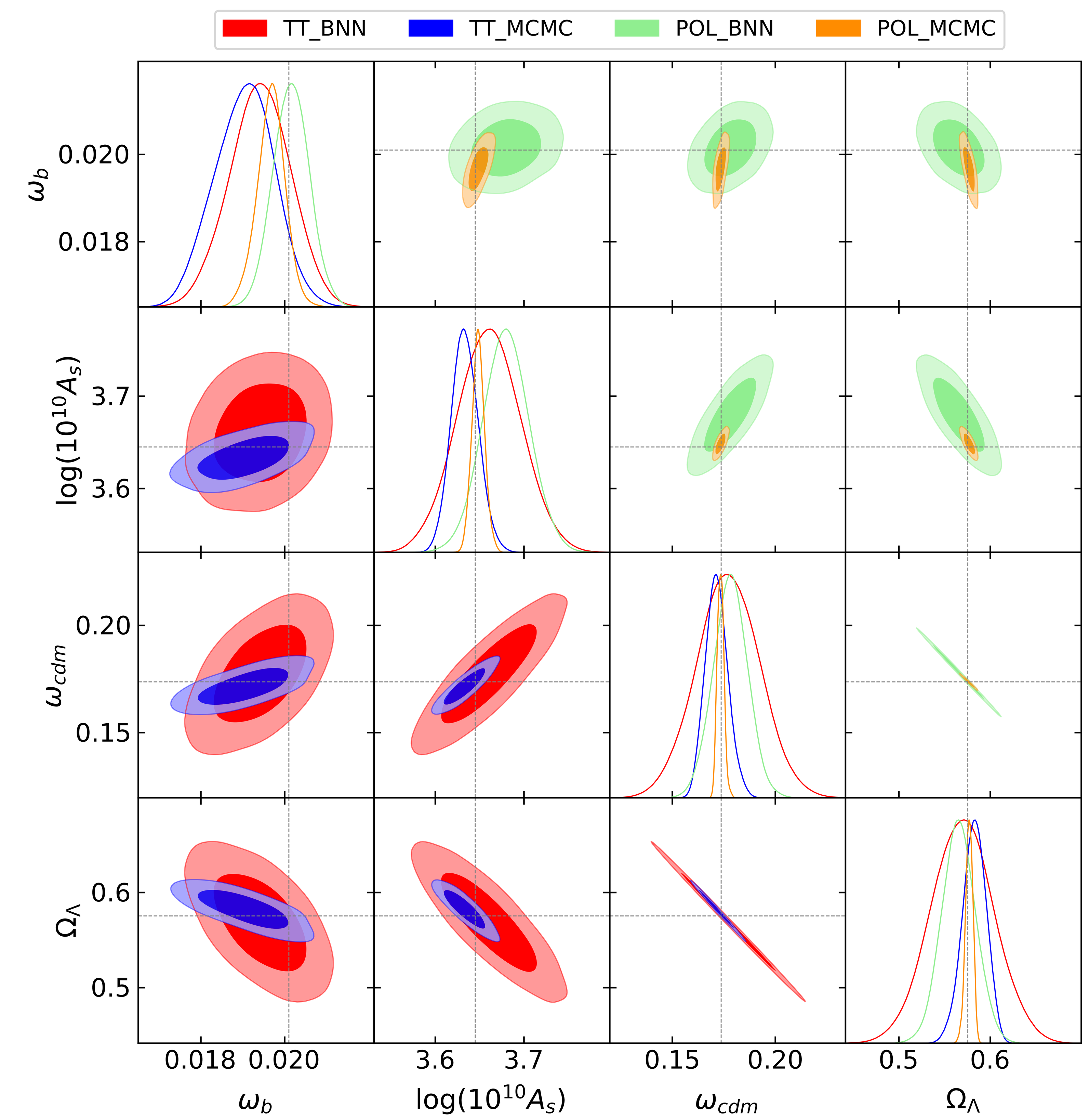

Fig. 1: Marginalized parameter constraints obtained from temperature maps (TT) and combined temperature with polarization (POL) using MCMC and the best BNN model. The black line stand for the real value: $\omega_b = 0.0201$, $\log(10^{10}A_s) = 3.6450$ and $\omega_{cdm} = 0.1736$ taken from the test dataset.

The Statistics and Parameters 95% intervals for the minimal base-ΛCDM model from our synthetic CMB dataset using non-informative prior (MCMC) and a precomputed covariance matrix from VI (covarBNN) are shown in the following table. The last column reports the metrics using the Full CMB map. The bold values in the last column correspond to the implementation of a proposal posterior distribution from VI. Although the full sky gives the smallest credible region, MCMC is 10000 times slower than VI. The real value considered is the same as specified in Fig. 1.

Statistics for various MCMC sampling configurations

Metrics	Temperature map		Temperature+Polarization map		
	MCMC	**covarBNN**	**MCMC**	**covarBNN**	**Full-sky**
ω_b	$0.0190^{+0.0013}_{-0.0013}$	$0.0190^{+0.0012}_{-0.0012}$	$0.01967^{+0.00066}_{-0.00066}$	$0.01968^{+0.00064}_{-0.00064}$	$0.02009^{+0.00010}_{-0.00010}$
$\ln(10^{10}A_s)$	$3.633^{+0.031}_{-0.031}$	$3.633^{+0.031}_{-0.030}$	$3.648^{+0.015}_{-0.015}$	$3.648^{+0.015}_{-0.016}$	$3.6449^{+0.0027}_{-0.0027}$
ω_{cdm}	$0.171^{+0.011}_{-0.011}$	$0.170^{+0.011}_{-0.011}$	$0.1734^{+0.0031}_{-0.0032}$	$0.1734^{+0.0031}_{-0.0031}$	$0.1736^{+0.0009}_{-0.0009}$
Ω_Λ	$0.583^{+0.025}_{-0.025}$	$0.583^{+0.024}_{-0.025}$	$0.5769^{+0.0079}_{-0.0080}$	$0.5769^{+0.0079}_{-0.0079}$	$0.5793^{+0.0019}_{-0.0019}$
Runtime	4.02hr	1.56hr	4.40hr	3.14hr	4.52hr//**3.15**hr
Acc. rate	0.19	0.23	0.14	0.25	0.18//**0.23**
$R-1$	0.0093	0.0098	0.0051	0.0084	0.0091//**0.0090**
$(R-1)_{95\%CL}$	0.0827	0.0764	0.0944	0.0642	0.0940//**0.0800**

Convergence in MCMC

In Fig. 2 we report MCMC convergence diagnostic quantities such as $R-1$ and the acceptance rate per iteration. The stopping rule implemented in cobaya ensures that the Gelman-Rubin $R-1$ value and its standard deviation at 95% confidence level interval $(R-1)_{95\%CL}$ computed from different chains (four in our case), satisfy the convergence criterion $R-1 < 0.01$ twice in a row, and $(R-1)_{95\%CL} < 0.2$ respectively to stop the run, [5]. For the Temperature signal alone, the chains achieve a steady state in about 2000 steps working with the covarBNN proposal while it usually takes more than 5000 steps instead with the vanilla MCMC. This behavior can also be explained by observing the acceptance rate in the red and green curves (TTcovarBNN, POLcovarBNN) quickly approaches a considerably high acceptance rate.

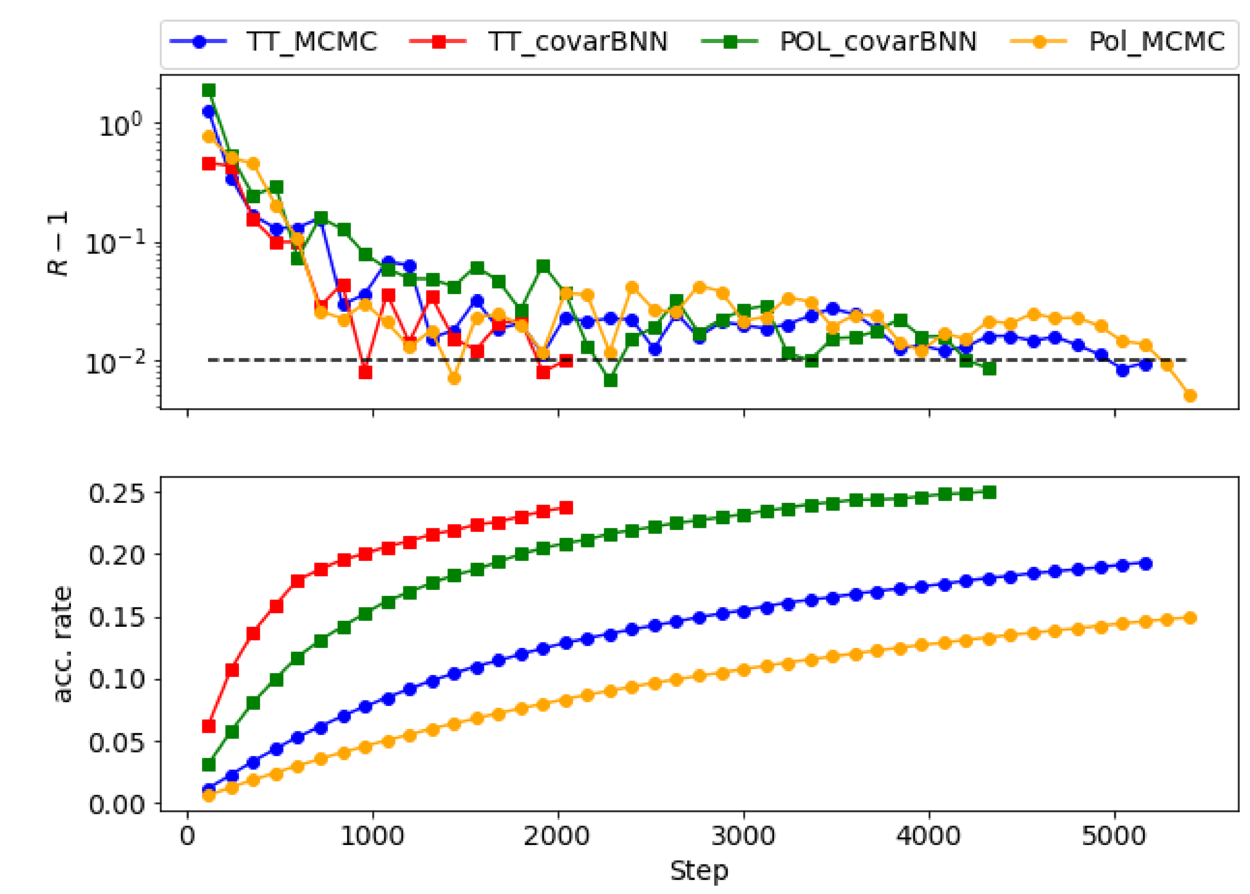

Fig. 2: Graphical representation of convergence in MCMC using non-informative priori (MCMC) and a precomputed covariance matrix from VI (covarBNN). (Top) Gelman-Rubin values with respect to the acceptance step. (Bottom) Acceptance rate with respect to step.

Conclusions

In this work we show that MCMC algorithms excel at quantifying uncertainty with respect to BNNs models, although the latter is about 10000 times faster at inference. Given these properties, we showed an approach in which the covariance matrix efficiently estimated from the BNNs samples, significantly enhance the acceptance rate in MCMC yielding faster convergence.

References

[1] Diego Blas, Julien Lesgourgues, and Thomas Tram. "The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes". In: *JCAP* 2011.07 (July 2011), pp. 034–034.

[2] Héctor J. Hortúa, Luigi Malago, and Riccardo Volpi. "Reliable Uncertainties for Bayesian Neural Networks using Alpha-divergences". In: *Uncertainty and Robustness in Deep Learning Workshop, ICML*. 2020.

[3] Héctor J. Hortúa et al. "Parameter estimation for the cosmic microwave background with Bayesian neural networks". In: *Phys. Rev. D* 102 (10 Nov. 2020), p. 103509.

[4] A. Petri. "Mocking the weak lensing universe: The LensTools Python computing package". In: *Astronomy and Computing* 17 (Oct. 2016), pp. 73–79.

[5] Jesus Torrado and Antony Lewis. *Cobaya: Code for Bayesian Analysis of hierarchical physical models.* 2020. arXiv: `2005.05290 [astro-ph.IM]`.

[6] Licia Verde. *A practical guide to Basic Statistical Techniques for Data Analysis in Cosmology.* 2007. arXiv: `0712.3028 [astro-ph]`.