

Baryon density extraction and isotropy analysis of Cosmic Microwave Background using Deep Learning

Amit Mishra^a, Pranath Reddy^a, Rahul Nigam^a

^a*Birla Institute of Technology & Science Pilani - Hyderabad Campus, Hyderabad, India*

Abstract

The discovery of cosmic microwave background (CMB) was a paradigm shift in the study and fundamental understanding of the early universe and also the Big Bang phenomenon. Cosmic microwave background is one of the richest and intriguing sources of information available to cosmologists and one parameter of special interest is baryon density of the universe. Baryon density can be primarily estimated by analyzing CMB data or through the study of big bang nucleosynthesis(BBN). Hence, it is necessary that both of the results found through the two methods are in agreement with each other. Although there are some well established statistical methods for the analysis of CMB to estimate baryon density, here we explore the use of deep learning in this respect. We correlate the baryon density obtained from the power spectrum of simulated CMB temperature maps with the corresponding map image and form the dataset for training the neural network model. We analyze the accuracy with which the model is able to predict the results from a relatively abstract dataset considering the fact that CMB is a Gaussian random field. CMB is anisotropic due to temperature fluctuations at small scales but on a larger scale CMB is considered isotropic, here we analyze the isotropy of CMB by training the model with CMB maps centered at different galactic coordinates and compare the predictions of neural network models.

1. Introduction

CMB is an electromagnetic radiation whose wavelength lies in the microwave region of the spectrum. The CMB maps display the temperature fluctuations of the photons that originate from a very early time when the universe was about one-millionth of its present size. These fluctuations follow an almost random gaussian distribution(Bucher, M., 2015) which extends onto the pixel intensities of the digital images of mollweide projections of these maps. Since there is no spatial coherence between the CMB maps considering they are gaussian random fields, it is rather an intriguing task to check the accuracy with which machine learning models are able to correlate the gaussian maps to cosmological parameters considering the abstract nature of the dataset. Our objective is to build and study the ability of a machine learning model to predict baryon density of a given CMB map and also analyse the isotropy of CMB by comparing the results of models trained with maps centred at different galactic coordinates. Although there is a loss of information while training the model using rendered images of CMB, we have taken only a single quantity (baryon density) as our parameter of interest with a sole purpose of demonstrating the ability of a machine learning model to extract useful information from CMB data.

The variations in temperature of Cosmic Microwave Background (CMB) are similar to the ripples on the cosmic pond and enclose a lot of information about the universe. To collect this information we look at the scales at which these temperature fluctuations occur. The amount

of temperature fluctuations (in micro Kelvin) is plotted against the multipole moment (l). This is the angular power spectrum graph of a CMB temperature map. Such graphs contain a number of peaks which provide with a lot of information and we exploit this for our use.

The first peak is an indication of the geometry of the universe, whether it is flat or curved (Hu, Wayne, et al., 2004). CMB radiation is distorted by the curvature of the universe since the radiation comes from all directions of the visible universe. The fluctuations will appear undistorted if the universe is flat. The fluctuations would appear magnified if the universe is positively curved and de-magnified if it is negatively curved.

The second peak reveals information about the amount of baryon present in the universe. Due to the initial fluctuations in the universe, all matter would tend to gravitationally group towards the higher density fluctuations. However, baryon matter which is interactive with light would heat up as it clumps up, and the resultant pressure would try to push against the grouped matter (Hu, Wayne, et al., 2004). This implies that the second peak will be more damped if there is more matter. Thus, the ratio of the first and second peak gives us the baryon density which we use for each map for training our model.

The anisotropies of CMB is determined by two factors, namely, acoustic oscillations and diffusion damping. The pressure of the photons tends to remove anisotropies, whereas the gravitational attraction of matter, makes them

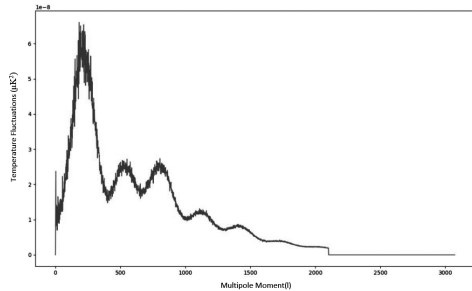


Figure 1: The angular power spectrum of a random CMB temperature sky map

collapse to form over-densities. These two effects conflict and compensate each other to create acoustic oscillations, which gives CMB its characteristic peak structure. The resonant frequency at which photons decouple at a certain mode at its maximum amplitude corresponds to the peaks in the power spectrum of CMB. The temperature anisotropy at any point on the sky (θ, ϕ) can be expressed in spherical harmonics as:

$$\frac{\Delta T}{T}(\theta, \phi) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi) \quad (1)$$

The temperature anisotropies of the CMB are believed to be a result of the inhomogeneous matter distribution at the time of recombination. Since Compton scattering is an isotropic process, any primordial anisotropies should have been erased before decoupling. This provides proof to the results that are derived from the anisotropy that is observed as the result of density perturbations that facilitated the formation of galaxies and clusters. Anisotropy in temperature provides concrete evidence that such inhomogeneity in density existed in the early universe, supposedly in the scalar field of inflation caused by quantum fluctuations or through topological defects resulting from a phase transition.

2. Methodology

We use *CAMB* to generate the CMB temperature angular power spectra data. *CAMB* takes several parameters as input to generate a *FITS* file containing the initial angular power spectrum data of the universe. The Curved correlation function is used as the lensing method and we include reionization. Other physical parameters which are input to *CAMB* include Hubble constant, the temperature of CMB(2.7255 Kelvin), baryon density(0.0226), cold dark matter density(0.112), the effective mass density of dark energy, maximum multipoles data, redshift(11), helium fraction(0.24) and so on(*LAMBDA-Tools*, NASA, 2014). The power spectrum file generated is used by standard cosmological analysis software *healpy* (Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005) to generate random gaussian CMB temperature maps. 2350 such maps are created. Anisotropy from dipole effect due to the movement

of the earth relative to CMB rest frame and galactic contaminants along the equator corresponding to the galactic plane is removed while generating the temperature maps (Bucher, M., 2015)(Planck Collaboration, 2011). The generated full sky maps have the galactic center at the center of the mollweide projection.

Also, baryon density(extracted from the power spectrum) corresponding to each of the generated temperature maps is stored in a separate *CSV* file.

We then snip 64x64 pixel(corresponding to 28x28 deg²) size images from the whole sky map using OpenCV and store them separately. The images are snipped only along the equator such that these portions could be treated as flat. In addition, we also rotate the CMB whole sky map along 4 different axis, moving the center of the maps to different galactic coordinates and repeat the same process in order to verify isotropy at large scales. These images will be treated as our input into the training model. We end up with 141,250 cropped images to train our model.

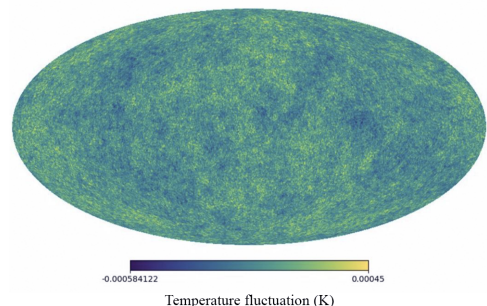


Figure 2: A random full sky CMB temperature map generated using healpy and CAMB

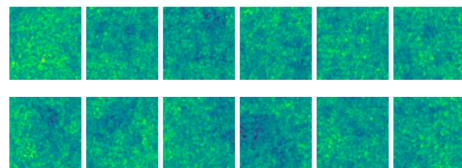


Figure 3: Sample patches cropped along the equator of the full sky CMB temperature map

To remove outliers from our data, we have gone with the box plot approach and found the upper quartile ($Q3$), lower quartile ($Q1$), and also the inter-quartile distance ($IQR = Q3 - Q1$) of all the baryon density values in our dataset and then removed all the data points that are greater than ($Q3 + 1.5 \cdot IQR$) or lower than ($Q1 - 1.5 \cdot IQR$).

Furthermore, to aid our model in better extracting the features, we plot the pixel intensities of the grayscale training images and fit a Gaussian to the plot, thereby deleting the pixels with intensities that fall below a probability value of 0.05.

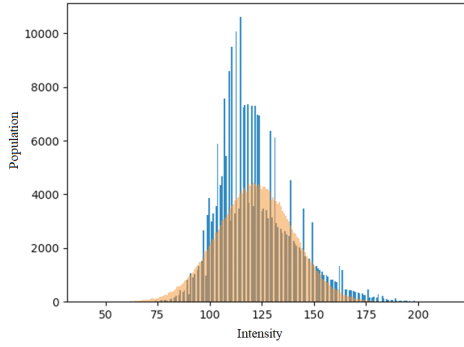
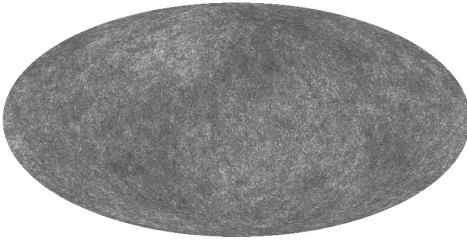
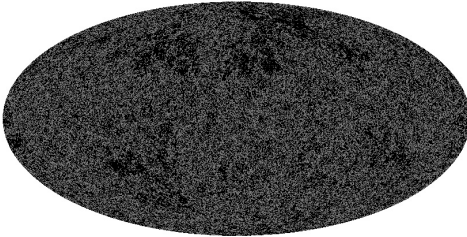


Figure 4: Fitting the pixel intensities with a Gaussian distribution(Population vs Pixel Intensities)



(a)



(b)

Figure 5: (a) Grayscale image of a test CMB temperature map and (b) the same map with pixels with lower than set threshold intensity deleted.

3. Analysis

Convolutional neural networks (CNN) are one of the most famous sets of neural network architectures used for classifying images, CNN takes advantage of local spatial coherence of the input (Rippel, Snoek, Adams, 2015) because we assume that the spatially close images used for training are correlated, but in the case of the CMB dataset, the pixels in the images are random noise following a gaussian distribution, the CNN network will not be able to find any common features in the inputs and thus the training accuracy and test error will be less than favorable. We have tested Resnet-101 and Inception v2 CNN architectures. The training accuracy of Resnet network was very low, which could be due to the high number of classes and a relatively low number of samples per each class in the dataset and the Inception network was suffering from high

variance problem (Liu, Wei, Zhang, Yang, 2017) and was mapping the random noise rather than the distribution of CMB. For this reason, we will be using a fully connected Multi-Layer perceptron network.

A multilayer perceptron is one of the most commonly used architectures of feedforward artificial neural networks. A multilayer perceptron comprises of three classes of layers and nodes, the input layer, hidden layers, and an output layer. Each node in a layer is connected to the nodes of the next layer via a non-linear activation function. Multilayer perceptron makes use of one of the most famous techniques of supervised learning called backpropagation (Goodfellow, Bengio, Courville, 2018) for training the network. A multilayer perceptron can be distinguished from a linear perceptron from its characteristic use of fully connected multilayers. This makes multilayer perceptrons suitable for working with non-linearly separable data (Bullinaria, 2015). Multilayer perceptrons are often informally referred to as vanilla networks (Hastie, Trevor, et al., 2017). A multilayer perceptron can be perceived as a logistic regression classifier. The input is first transfigured with the help of a learned non-linear transformation. The intermediate layers are often mentioned as a hidden layer. These hidden layers make the multilayer perceptron a universal approximator (Sifaoui, Abdelkrim, Benrejeb, 2008).

The weights of the fully connected layers are updated once a batch of data has been passed through the network by measuring the error of the output with the expected result (predetermined labels), this is the essence of learning in neural networks and is carried out with the help of an iterative algorithm called backpropagation. This is an example of supervised learning. Backpropagation uses an iterative optimization algorithm called gradient descent (Goodfellow, Bengio, Courville, 2018) to update the weights of the network.

The configuration of our network is tabulated in Table 1.

Table 1: Neural Network Configuration

No of hidden layers	No of Nodes in each hidden layer	Learning Rate
5	3223	0.0001
Batch size	No of epochs	Regularization parameter
512	50000	0.01

The learning rate determines how fast the weights or the coefficients of the network are updated. An epoch can be defined as the number of times the algorithm perceives the entire data-set. Hence, an epoch is completed when all the samples of the data have been perused. An iteration can be defined as the number of times a batch of data has been passed through the algorithm. In the case of a multilayer perceptron, that means the forward pass and backward pass. Hence, an iteration is completed once a batch of data has passed through the network. The batch size is

the number of training examples passed through the network at once (Shen, 2017 & Svozil, Kvasnicka, Pospichal, 1997).

We have used the tensorflow library to implement our MLP model. We have used Adam optimization algorithm (Kingma, Ba, 2017) instead of the traditional stochastic gradient descent for updating the weights of the network (Michelucci, Umberto, 2018).

We have used L2 regularization, also known as ridge regularization to prevent our model from overfitting. In L2 regularization, we add a squared error term as a penalty to the loss function (Goodfellow, Bengio, Courville, 2018).

The training of the network is done in the Google Cloud platform using a Tesla K80 GPU.

4. Results

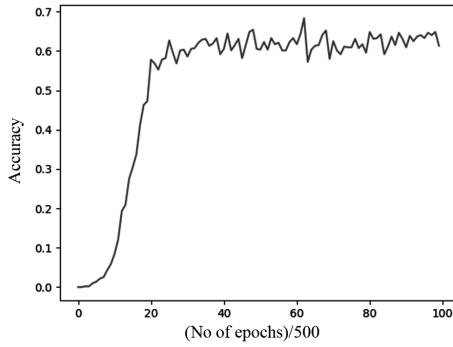


Figure 6: The change in training accuracy of models (vs no of (epochs/500)) trained on data corresponding to a specific galactic latitude and longitude(-30⁰,90⁰) combination at the center of the mollweide projection (input images are cropped along the equator of the projection), shows the general trend of change in accuracy.

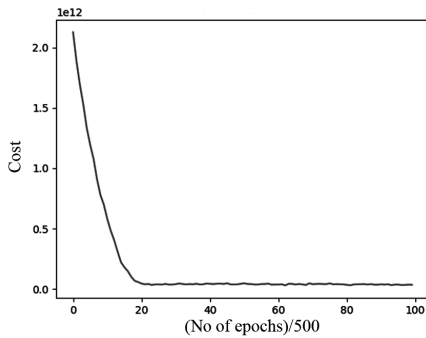
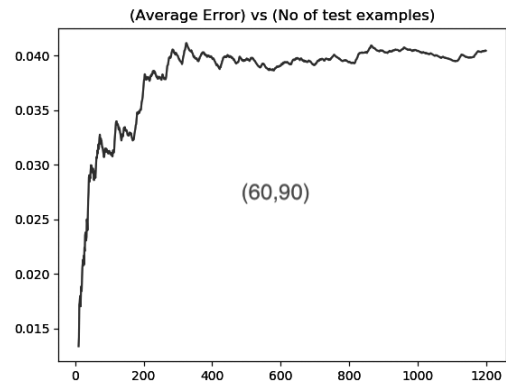
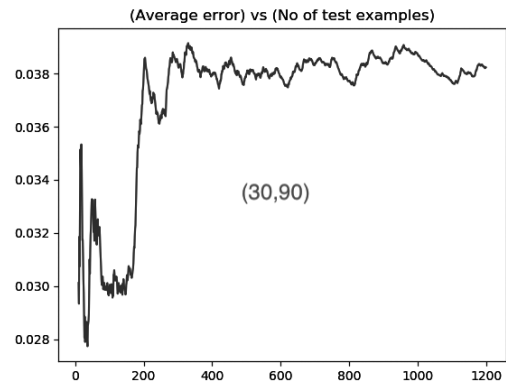
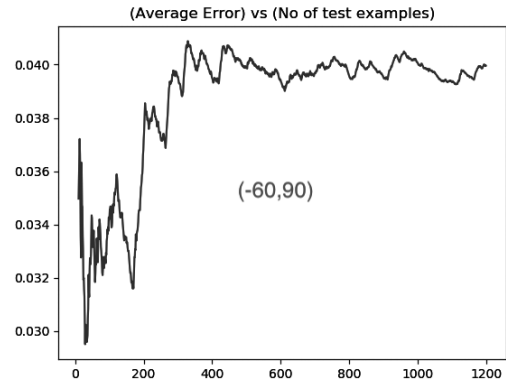
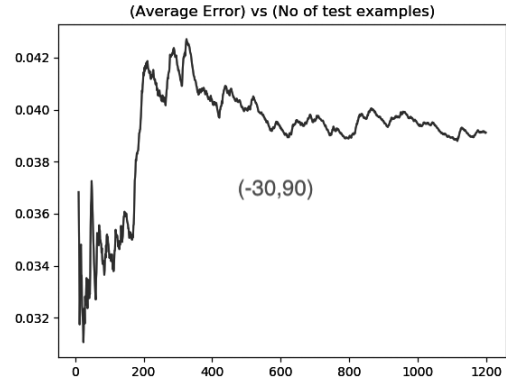


Figure 7: The change in training cost of models (vs no of (epochs/500)) trained on data corresponding to a specific galactic latitude and longitude(-30⁰,90⁰) combination at the center of the mollweide projection (input images are cropped along the central horizontal line of the projection), shows the general trend of change in cost.



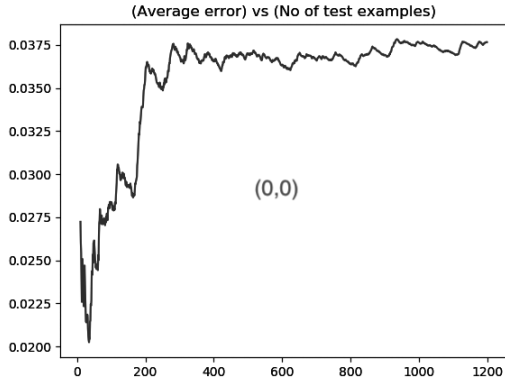


Figure 8: The Average error for predictions of models (vs no of test examples) trained on data corresponding to a specific galactic latitude and longitude combination at the center of the mollweide projection. The fluctuations at the beginning of the plot can be due to high variance of the low sample size of the test data, but the average error is getting saturated as we increase the sample size of our test data giving us a practicable value for the final test set consisting of 1200 maps.

For calculating the test error, we have gone with Mean Magnitude of the Relative Error (MMRE).

$$MMRE = \frac{1}{n} \sum \frac{|y - yp|}{y} \quad (2)$$

where,

n : no of samples

y : actual value

yp : predicted value

The training accuracy percentages indicate the percentage of data samples being correctly classified/predicted during the training process. The model is trained and tested taking five different galactic coordinates at the center of the mollweide projection and cropping images along the central horizontal line. The training accuracy and testing error (average relative error) are tabulated in Table 2. We also calculate the average relative error between the predictions of the models trained with maps centered at different galactic coordinates and are tabulated in Table 3.

Table 2: Training accuracy and testing error for models trained with different galactic coordinates at center

Sl.No.	Galactic coordinates	Training Accuracy	Test Error
1	(0 ⁰ ,0 ⁰)	0.619570	0.035324
2	(30 ⁰ ,90 ⁰)	0.617313	0.038691
3	(-30 ⁰ ,90 ⁰)	0.621992	0.038895
4	(60 ⁰ ,90 ⁰)	0.619843	0.038689
5	(-60 ⁰ ,90 ⁰)	0.617500	0.039340

Table 3: Relative error between different models trained with different galactic coordinates at center

Galactic coordinates	(0 ⁰ ,0 ⁰)	(30 ⁰ ,90 ⁰)	(-30 ⁰ ,90 ⁰)
(0 ⁰ ,0 ⁰)	-	0.029728	0.030671
(30 ⁰ ,90 ⁰)	0.029728	-	0.034365
(-30 ⁰ ,90 ⁰)	0.030671	0.034365	-
(60 ⁰ ,90 ⁰)	0.029723	0.034801	0.034580
(-60 ⁰ ,90 ⁰)	0.030752	0.034343	0.034401

Galactic coordinates	(60 ⁰ ,90 ⁰)	(-60 ⁰ ,90 ⁰)
(0 ⁰ ,0 ⁰)	0.029723	0.030752
(30 ⁰ ,90 ⁰)	0.034801	0.034343
(-30 ⁰ ,90 ⁰)	0.034580	0.034401
(60 ⁰ ,90 ⁰)	-	0.034289
(-60 ⁰ ,90 ⁰)	0.034289	-

5. Conclusion and future plans

The application of deep learning methodologies on the CMB data has steered us to the verification of two well-established results related to the CMB using a completely different approach via deep learning. Firstly, we were able to predict baryon density with a satiable accuracy. The loss of accuracy can be credited to the fact that the CMB temperature maps are of very high resolution and we were bounded by limited computational power but most importantly the relatively random orientation of the pixels considering the fact that CMB is a Gaussian random field was a major complexity for the neural network model to extract any applicable features from the maps. Secondly, when training and predicting with models trained along different galactic latitude and longitude we were able to get very low error between the predictions which reaffirms the well known isotropic nature of CMB at larger scales. Although the training accuracy is low in the domain of precision physics, the test error values are impressive considering the fact that we were limited by the amount of data and also computational power of training our model. We are not trying to compete with traditional well established techniques with the power spectrum, but rather we are proposing a new domain for the study of CMB and subsequently as a demonstration we have chosen baryon density for the same. Although, in the case of extracting baryon density directly from temperature map there is a loss of information but there could be some property of CMB which could be found using deep learning but not from power spectrum. We hope to develop a deep neural network architecture tailored for CMB maps that is capable of correlating random noise to cosmological parameters.

References

- [1] Bucher, M. 2015, International Journal of Modern Physics D, 24, 15300041530303.
- [2] Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, 622, 759.

- [3] Goodfellow, Ian, et al. 2018, Deep Learning, MITP.
- [4] Hastie, Trevor, et al. 2017, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer.
- [5] Liddle, Andrew R 2015, An Introduction to Modern Cosmology, John Wiley and Sons.
- [6] Hu, Wayne & White, Martin 2004, The Cosmic Symphony, Scientific American. 290. 44-53.
10.1038/scientificamerican0204-44.
- [7] Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, The Astrophysical Journal Supplement Series, 208, 20.
- [8] Koberlein, Brian, and David Meisel 2009, Astrophysics through Computation.
- [9] Goddard Space Flight Center, NASA 2005, LAMBDA-Tools: Web interface for CAMB
https://lambda.gsfc.nasa.gov/toolbox/tb_camb_params.cfm
- [10] Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, 536, A1.
- [11] Bennett, C. L., Banday, A. J., Gorski, K. M., et al. 1996, 464, L1.
- [12] Lewis & Challinor. Code for Anisotropies in the Microwave Background(CAMB), <https://camb.info>.
- [13] Rippel, O., Snoek, J., & Adams, R. P. 2015, arXiv e-prints , arXiv:1506.03767.
- [14] Kingma, D. P., & Ba, J. 2014, arXiv e-prints , arXiv:1412.6980.
- [15] Shen, H. 2016, arXiv e-prints , arXiv:1611.05827.
- [16] Michelucci, Umberto. (2018). Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks. 10.1007/978-1-4842-3790-8
- [17] Sifaoui, Amel & Abdelkrim, Afef & Benrejeb, Mohamed. (2008). On the Use of Neural Network as a Universal Approximator. International Journal of Sciences. 2. 386-399.
- [18] Yu Zhang, Qiang Yang, Bo Liu, Ying Wei. Deep neural networks for high dimension, low sample size data. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pages 2287-2293, 2017
- [19] Daniel Svozil, Vladimr Kvasnicka, Jir Pospichal, Introduction to multi-layer feed-forward neural networks, Chemometrics and Intelligent Laboratory Systems, Volume 39, Issue 1, 1997, Pages 43-62, ISSN 0169-7439
- [20] J.A.Bullinaria, Learning in multi-layer perceptrons, back-propagation, Neural Comput. Lect., vol. 7, no. 8, pp. 116, 2015