

# A Concise Report of Machine Learning's Project

First Milestone : Data

Zahra Farahmand, Reyhaneh Aghaei, Bahar Nikbakht

March 2020

In this report, which is aimed mostly the aspects of the data, we have explained briefly about the whole project, the method we have used for generating data and the general features of Cosmic Microwave Background. Then we have clarified the different types of precision we may face on during the project, and at the end, we have named some of the related articles.

## 1 Introduction

According to inflation theory, the speed of propagation of quantum fluctuations in the early universe is not necessarily equals to the speed of light. This makes it sounds sensible to think of cherenkov effect, in which a particle goes beyond the speed of sound and though the trace of it's path will remain in the surroundings, just like mach cones made by ultrasonic airplanes in the sky. On the other hand, due to a recent claim, massless particles may have intersected the horizon and made the same cones due to the speed of propagation of quantum fluctuations (hereinafter called  $C_s$ ) and the cherenkov effect as a result [1].

On the other hand, Cosmic Microwave Background (CMB) remaining from the era when the universe was young, could be used as a snapshot of the whole universe which involves these traces. However CMB is too dense to look forward special shapes on it without the help of machine learning's methods. On the first step of this project we are going to portrait the general shapes of the traces on the CMB, and find a proper method of simulating them in order to be used in further training.

## 2 Cone-Sphere Intersection

As mentioned above, the traces of the massless particles on CMB, which seems spherical to us, are in different shapes that a cone-sphere intersection may have (Figure 1).

The cone-sphere intersections are various in shape. It made us to look for a general function which enables us to know whether a point is on the intersection or not.

### 2.1 General Solution to Cone-Sphere Intersections

First, consider a sphere at the origin with unit radius, actually the length of the radius doesn't matter, we should just take care that every distance should be normalized by our distance to CMB.

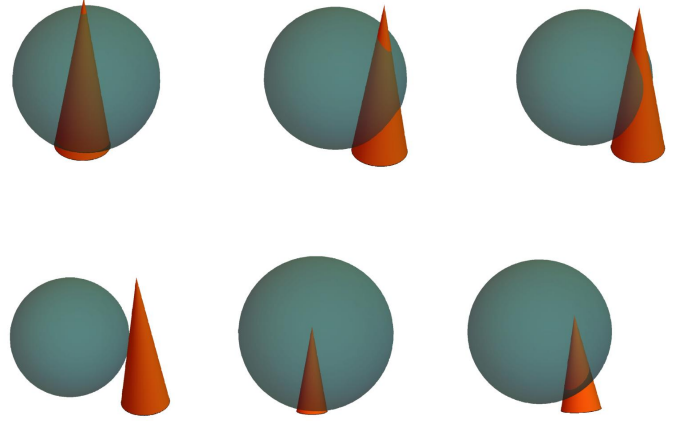


Figure 1: Probable shapes of cone-sphere intersections

Then imagine a cone with angle  $\theta_{C_s}$ , a head positioned at  $\vec{r}_0$  and a direction vector  $\vec{n}$ , which crosses the sphere (Figure 2).

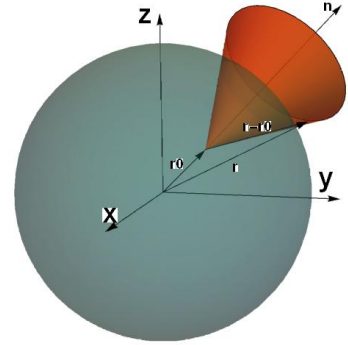


Figure 2: intersection of a sphere and a cone

The parameters of the cone are expressed as:

$$\vec{r}_0 = x_0\hat{x} + y_0\hat{y} + z_0\hat{z}, \quad (1)$$

$$\vec{n} = n_x\hat{x} + n_y\hat{y} + n_z\hat{z}. \quad (2)$$

A cone in general can be defined as a locus of all vectors with the same angle relative to a single vector. here, the single vector which we are looking for is  $\vec{n}$ . So, by using the relation between the magnitude of the product of dot operator and the magnitude of the vectors themselves, the general cone equation can be written as follows:

$$(\vec{r} - \vec{r}_0) \cdot \vec{n} = |\vec{r} - \vec{r}_0| \cos \theta_{C_s}. \quad (3)$$

By using the above equation and equations (1) and (2), and considering  $r = X\hat{x} + Y\hat{y} + Z\hat{z}$  we have:

$$(X - x_0)n_x + (Y - y_0)n_y + (Z - z_0)n_z = |\vec{r} - \vec{r}_0| \cos \theta_{C_s}. \quad (4)$$

Intersection points( $\{\theta^*, \phi^*\}$ ) must satisfy both equations of the cone and the sphere, so they can be represented by the following system of equations,

$$\begin{aligned} \vec{r}^*(\theta^*, \phi^*) &= R(\sin \theta^* \cos \phi^* \hat{x} + \sin \theta^* \sin \phi^* \hat{y} + \cos \theta^* \hat{z}), \\ (\vec{r}^* - \vec{r}_0) \cdot \vec{n} &= |\vec{r}^* - \vec{r}_0| \cos \theta_{C_s}. \end{aligned}$$

## 2.2 Cone-Sphere Intersection Data

By solving the system of equations obtained in the previous subsection, the general equation of the intersection points would be:

$$\begin{aligned} f(\theta, \phi) &= (\sin \theta \cos \phi - x_0)n_x + (\sin \theta \sin \phi - y_0)n_y \\ &+ (\cos \theta - z_0)n_z - \cos \theta_{C_s} \left( (\sin \theta \cos \phi - x_0)^2 \right. \\ &\left. + (\sin \theta \sin \phi - y_0)^2 + (\cos \theta - z_0)^2 \right)^{\frac{1}{2}} = 0. \end{aligned} \quad (5)$$

We aim to use simulation for generating intersection curves' data with the same format as the CMB data, in which the space is discretized to pixels and each pixel contains the magnitude of the temperature of that pixel (see section 3.1). To solve the equation(5) numerically, parameters  $\theta$  and  $\phi$  of the space should be discretized.

In order to simulate the two dimensional matrix(intersection matrix) that the nonzero elements' position stand for intersection's  $\{\theta, \phi\}$ , we divide parameter space into  $N_\theta \times N_\phi$  parts(Figure 3).

$$N_\theta = \frac{\pi}{\Delta\theta} \quad N_\phi = \frac{2\pi}{\Delta\phi},$$

$$\text{function matrix} = F = \sum_{i,j}^{N_\theta, N_\phi} F_{i,j} = \sum_{i,j}^{N_\theta, N_\phi} f(\theta_i, \phi_j). \quad (6)$$

Which  $\Delta\theta$  and  $\Delta\phi$  are the CMB's precision of  $\theta$  and  $\phi$ .

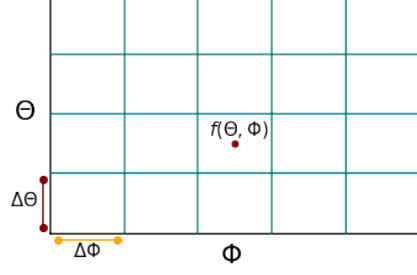


Figure 3: schematic diagram of discrete parameter space

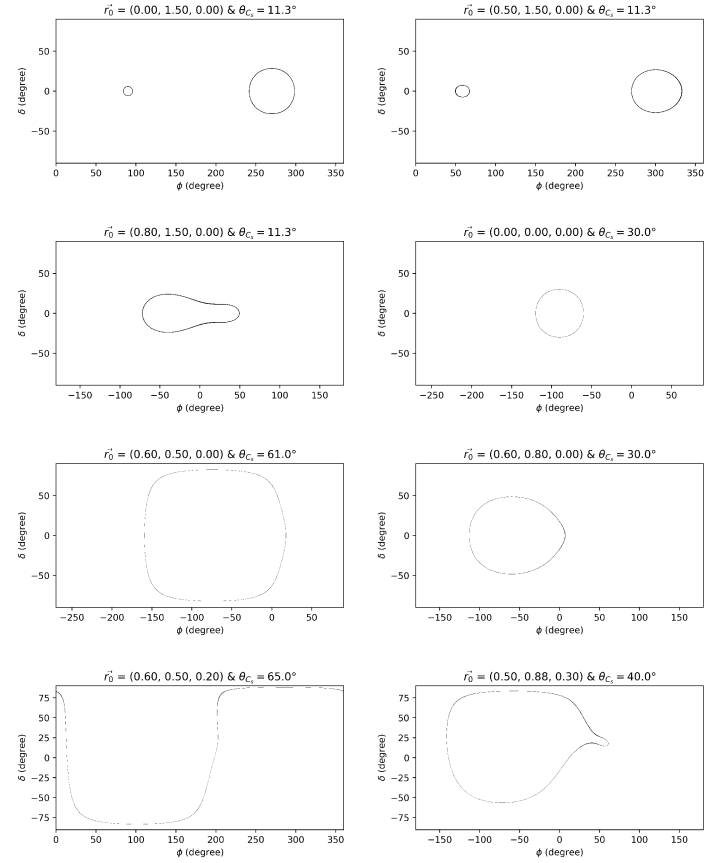


Figure 4: different curves for cones with (0,-1,0) direction vector

Due to discrete parameters considering an error becomes a fundamental issue. We need a threshold for flagging each of function matrix elements as an intersection point and create intersection matrix. Two different ways are tested:

1. Consider an error matrix( $F_{error}$ ) with same dimension as function matrix and compare  $F_{i,j}$  with  $F_{error,i,j}$  for finding corresponding intersection matrix elements.

$$\Delta f(\theta, \phi) = \sqrt{\left(\frac{\partial f}{\partial \theta} \Delta\theta\right)^2 + \left(\frac{\partial f}{\partial \phi} \Delta\phi\right)^2}. \quad (7)$$

2. According to the precision of the physical aspects of the problem, set a constant error to compare function matrix elements with it.

Till now, we have succeeded finding a firm framework of generating data, in which we use  $\vec{n}$ ,  $\vec{r}_0$  and  $\theta_{C_s}$  as inputs, and the output is a matrix of meshed space, as mentioned before, with one element on the intersection line and zero in other parts of the space. However, still some minor issues have remained unsolved. We are not sure about the intervals in which the parameters vary, hence this is concerned with the physical aspects of the problem, and we are not sure about how many data sets would be sufficient for training. We are going to find an accurate answer to these problems meanwhile the next step.

### 3 CMB Data

The European Space Agency's Planck satellite, which was dedicated to studying the early Universe and its subsequent evolution, observing the whole sky for nearly nine years, have equipped us with the precise map of the Cosmic Microwave Background, Known as CMB. Analysis of the Planck 2018 data set indicates that the statistical properties of the cosmic microwave background (CMB) temperature anisotropies are in excellent agreement with previous studies using the 2013 and 2015 data releases, proving that sky is isotropic as was expected before [2].

As mentioned in the previous sections, we are going to look forward to the cone-sphere intersections through temperature map of CMB. In this section, we are going to review one of the common methods of studying CMB, then explaining the method of cleaning the data, and a brief description of data's statistics and accuracy.

#### 3.1 Analysis of Data Distributed on the Sphere

The Hierarchical Equal Area iso-Latitude Pixelization, or HEALPix in short, is a versatile data structure with an associated library of computational algorithms which was first developed to address the data processing and analysis needs of the present generation of cosmic microwave background experiments [3].

HEALPix may be seen as a division of the sphere into 12 base pixels (Figure 5<sup>1</sup>) which are then subdivided to give the desired resolution. The amount of subdivision is given by the NSIDE number where NSIDE is always a power of 2 hence 1, 2, 4, 8, 16, 32 etc. NSIDE specifies the number of subdivisions in each side of a base pixels, eg. for NSIDE=2, each of the 12 base pixels are subdivided into 4 parts (Figure 6<sup>2</sup>). in the case of CMB Data

which we are going to use, NSIDE=2048 so the total number of pixels is,

$$N_{pix} = 12 \times NSIDE^2 = 12 \times 2048^2 = 50331648. \quad (8)$$

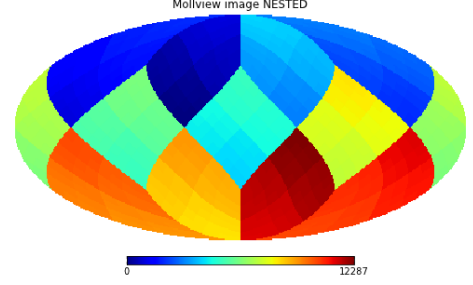


Figure 5: 12 HEALPix base pixels

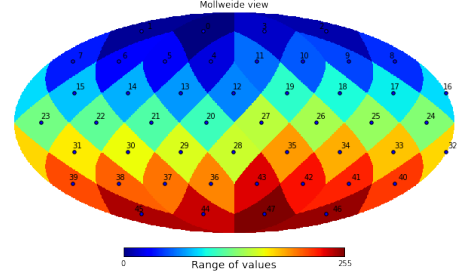


Figure 6: Subdivisions of HEALPix pixels for NSIDE=2

By using HEALPix library, the data can be easily reshape to the form of galactic coordinates and then represent by an array with indices  $\theta$  and  $\phi$  and the temperature in each element, the same as the format we have used for the results of the simulation of the cone-sphere intersections. The Inpainted I intensity map of the 2018 release of Planck data<sup>3</sup> can be shown as Figure 8.

The histogram of the difference of above data from it's mean is shown in Figure 7, showing that the temperature distribution of the universe is gaussian, in agreement with homogeneity and isotropy.

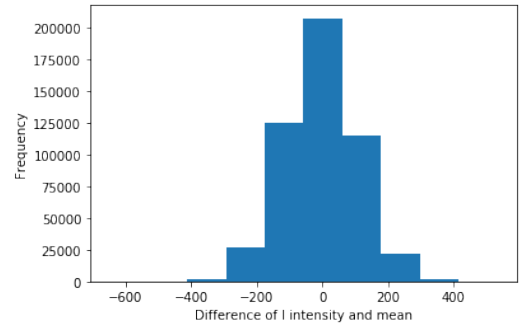


Figure 7: Histogram of the difference of the temperature map of Planck data (2018) and it's mean.

<sup>1</sup>credit:libguides.rhul.ac.uk/referencing/latex

<sup>2</sup>credit:libguides.rhul.ac.uk/referencing/latex

<sup>3</sup>Data are taken from Planck's website : pla.esac.esa.int

## 3.2 Method of Cleaning

Planck detector, like all other CMB detectors that was used before, is a radio detector which counts the number of photons reaching it's CCD. The universe is almost empty making this process precise, however, our position in galactic disk cause the loss of some data, between  $\theta = -18$  deg to  $\theta = +18$  deg. One of the most popular methods <sup>4</sup>that is used in order to clean up the CMB data, is SMICA. SMICA produces a CMB map by linearly combining all Planck input channels (from 30 to 857 GHz) with weights which vary with the multipole. It is a precise method but it's not necessary to keep these data until the end, we can easily throw them away at last, just as the same procedure done by Penrose and Gurzadyan[4].

## 4 Precision Considerations

There are three types of precision which we must take care of:

1. First of all we have to decide whether a pixel in CMB stands for a point or not, in other words, we should define a bound for temperature to divide the points of the shapes from the background.
2. Second is the precision of the CMB itself, which bounded us for getting closer than a certain amount to a line which we expect to see and can be determined by the precision of  $\theta$  and  $\phi$ .
3. Finally, the precision of the traces that specifies the thickness of the lines we are looking for. This precision is straightly connected with the precision of  $\theta_{C_s}$ , that is the angle of the cone.

The above types of precision, are somehow connected to each other. If the CMB resolution is not sufficient, we won't be able to recognize the contrast between a target pixel region and the background.

As mentioned in section 2.1, we have pursuit two different approaches of precision, by calculating the exact amount of precision in each point and by considering a unit amount of precision in all space. The first approach that is much more precise costs the loss of speed. So, depending on the precision we may have to use in next spaces, we will choose what method to use.

## 5 Related Articles

In this section we are going to mention some papers that are relevant to our project, with a small description of their relevance.

1. A. A. Abolhasani and S. Jazayeri, "Cherenkov radiation in the sky: Measuring the sound speed of primordial scalar fluctuations"

This is the main article that we are going to check it's validity. Although it's harder to comprehend by detail, but it may help us with the physics behind the whole story.

2. V. G. Gurzadyan, R. Penrose, "CCC-predicted low-variance circles in CMB sky and LCDM, " :

In this article, Gurzadyan and Penrose have done nearly the same thing. otherwise, they have searched for rings instead of cone-sphere intersections. and have done some statistical analysis that we can keep eye on them.

3. C. L. Bennett, R. S. Hill, G. Hinshaw and others, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Are There Cosmic Microwave Background Anomalies?" :

The above article have examined some of the statistical analysis of finding features and claiming that they are due to the non-standard cosmology. it can help to provide a firm proof of what we find, if there is any!

4. Planck Collaboration, "Planck 2018 results. I. Overview and the cosmological legacy of Planck"

This is the first article of a series of articles written by the Planck Collaboration, explaining the data of the Planck satellite and a complete analysis of it. It can be used to define a proper bound for the first type of precision mentioned above, and it can give us a sight of cleaning the data around the galactic disk.

5. K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, M. Bartelman, "HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere"

The general method of HEALPix is clarified in this article then it allows us to use this library more efficiently.

6. Calabretta, Mark R., Roukema, Boudewijn F., "Mapping on the HEALPix grid "

This paper is just the same as the above article, with some extra explanations about the position of each pixel in HEALPix. So, it can be used in the last part of the project to mark the exact pixels of CMB on the cone-sphere intersections' lines.

7. Amit Mishra, Pranath Reddy, Rahul Nigam, "CMB-GAN: Fast Simulations of Cosmic Microwave background Anisotropy maps using Deep Learning"

If we succeed in finding any cone-sphere intersections, the next step would be to show that this observation is due to non-standard cosmology, so we should run our models on virtual CMB's simulated by the standard model and compare the frequency of features. This article is about simulating CMB.

<sup>4</sup>A complete list of methods can be found here: [wiki.cosmos.esa.int/planckpla/index.php/CMB\\_and\\_astrophysical\\_component\\_maps](http://wiki.cosmos.esa.int/planckpla/index.php/CMB_and_astrophysical_component_maps)

8. Ali Frolof, Douglas Scott, "Pi in the sky":

It's about the same problem discussed in the article number 2, just in shorter form.

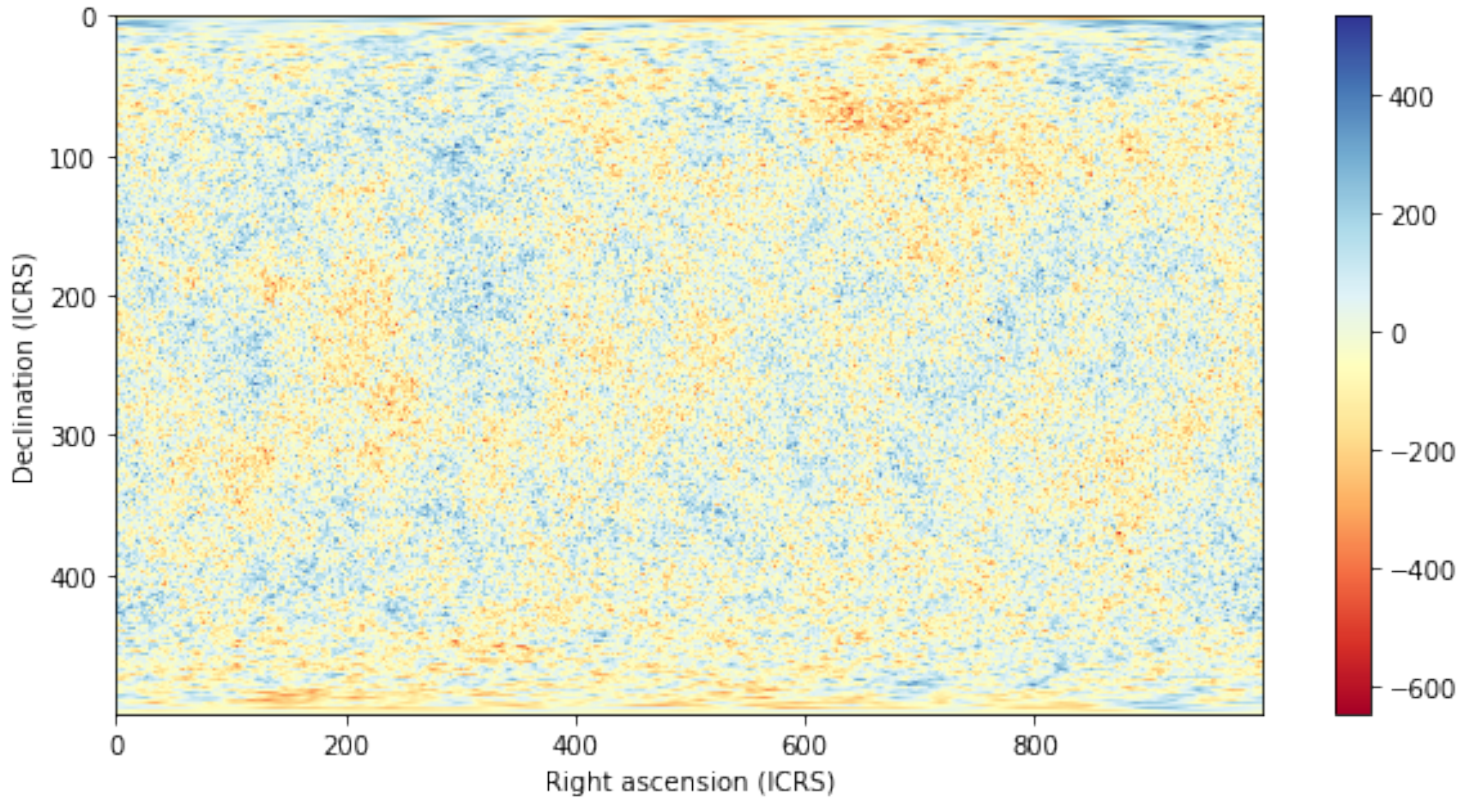


Figure 8: CMB by Planck Data (2018), Temperature map (mK)

## References

- [1] A. A. Abolhasani and S. Jazayeri, "Cherenkov radiation in the sky: Measuring the sound speed of primordial scalar fluctuations," *Phys. Rev.D*100no. 2, (2019) 023520, arXiv:1904.05589[astro-ph.CO].
- [2] Planck Collaboration, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," arXiv:1807.06205 [astro-ph.CO].
- [3] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, M. Bartelman, "HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere" , *Astrophys.J.*622:759-771,2005, arXiv:astro-ph/0409513.
- [4] V. G. Gurzadyan, R. Penrose, "CCC-predicted low-variance circles in CMB sky and LCDM," arXiv:1104.5675 [astro-ph.CO]