Draft version April 3, 2023
Typeset using LaTeX **twocolumn** style in AASTeX631

Using the Metropolis-Hastings Sampling Algorithm and the Markov Chain-Monte Carlo Method to Find Best-Fit Parameters

Akshat S. Chaturvedi[1]

[1] *Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*

ABSTRACT

I created a `python` program that uses the Metropolis-Hastings sampling algorithm (Metropolis et al. (1953)) in conjunction with the Markov Chain-Monte Carlo method to find the best fit parameters for a set of data involving line strengths as a function of frequency.

1. INTRODUCTION

The Metropolis-Hastings sampling algorithm works by sampling from a provided distribution, picking an arbitrary starting value, and then iteratively accepting or rejecting values drawn from a separate distribution, usually one which is easier to sample from. The iterative process involves the Markov Chain-Monte Carlo method. The Metropolis-Hastings algorithm, named after Nicholas Metropolis and W.K. Hastings, who developed it in 1953.

2. METHODS

2.1. *Finding the joint posterior distribution*

The data points are distributed as a normal distribution, given by equation 1 below.

$$P(y_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{(\frac{y_i - m_i}{\sqrt{2}\sigma_i})^2} \tag{1}$$

where m_i is a Gaussian function shown in equation 2

$$m_i = G(x_i, \mu, \alpha_D, A) = \frac{A}{\alpha_D}\sqrt{\frac{\ln 2}{\pi}} e^{-\frac{\ln 2 (x_i - \mu)^2}{\alpha_D^2}} \tag{2}$$

From Bayes' theorem, as shown in equation 3,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

I was able to derive a joint posterior distribution for the data, from which the Metropolis-Hastings algorithm would draw samples. This derivation is shown below:

$$P(\alpha_D, \mu, A|y_i) = \frac{P(y_i|\alpha_D, \mu, A)P(\alpha_D, \mu, A)}{P(y_i)}$$

In this situation, as α_D, μ and A are uniformally distributed, $P(\alpha_D, \mu, A)$ is constant, and as such can be disregarded. Similarly, the assumption that if y_i is known, the paramter values are known as well, can be made and as such, the resultant from Bayes' theorem can be shown as in equation 4.

$$P(\alpha_D, \mu, A|y_i) = P(y_i|\alpha_D, \mu, A) \tag{4}$$

Moreover the likelihood function of this distribution therefore is shown in equation 5

$$\begin{aligned} L(\alpha_D, \mu, A|y_i) &= \prod_{i=1}^{n} P(y_i|\alpha_D, \mu, A) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{(\frac{y_i - m_i}{\sqrt{2}\sigma_i})^2} \end{aligned} \tag{5}$$

The coefficient of the likelihood function can be dropped as σ_i is independent of the parameters and as such becomes a constant. As such, the resultant likelihood function is seen in equation

$$L(\alpha_D, \mu, A|y_i) \propto e^{\Sigma_{i=1}^{n}(\frac{y_i - m_i}{\sqrt{2}\sigma_i})^2} = e^{\psi} \tag{6}$$

From this likelihood, we can determine the posterior probability, as shown in equation 7

$$P(\alpha_D, \mu, A|y_i) \propto e^{\psi} \tag{7}$$

Using this posterior probability, I was able to find the joint posterior distribution as shown below:

$$\begin{aligned} P_J(\alpha_D, \mu, A|y_i) &\propto \prod_{i=1}^{n} e^{\psi} \\ &\propto e^{n\psi} \end{aligned} \tag{8}$$

I then used this joint posterior distribution as the "easier sampling distribution" for the Metropolis-Hastings algorithm.

2.2. *The Metropolis-Hastings Algorithm*

The iterative process/algorithm for Metropolis-Hastings can be described as below:

1. Pick an initial guess for the chosen parameter, θ_0
2. Pick a proposed value for the parameter from a uniform distribution, described as $\theta' \sim Unif(\theta_0 - b, \theta_0 + b)$, where b is a chosen value
3. Find the acceptance probability, where $p = min(1, \frac{P_J(\theta')Q(\theta|\theta')}{P_J(\theta)Q(\theta'|\theta)})$[1]
4. We then either accept or reject the proposed parameter using the following sub-algorithm
   (a) Pick a random number, a between 0 and 1
   (b) If $a < p$, accept the proposed parameter, and set the current iteration value for the parameter equal to θ'
   (c) If $a > p$, reject the proposed parameter, and set the current iteration value for the parameter to θ_0

This algorithm can be repeated for any desired number of iterations.

3. RESULTS AND DISCUSSION

For my program, I used 25000 iterations, and the resultant fits (for three different starting values of the three parameters) are shown in figure 1 The variations of the individual parameters through the iterations are also shown in 2 From figure 2, I determined that the burn in period was $\sim$ 12000 iterations. As such, the parameters as a function of steps after the burn in period is shown in figure 3 The resultant values for each of the parameters is shown in the table below, as well as the values obtained from Levenberg-Marquadt Algorithm:

Parameter	Last Value	Mean Value	LM Value
α_D	15.6061	15.7731	15.0266
μ	43.9852	43.7578	44.9542
A	1.0350	1.0378	1

Parameter	Standard Deviation
α_D	0.1717
μ	0.2624
A	0.0069

Table 1: Caption

The Metropolis-Hastings MCMC algorithm is more efficient than the Levenberg-Marquadt algorithm, being quicker to return values that lead to similar fits.

REFERENCES

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087

[1] In this case, the two Q() distributions can be cancelled out due to the parameters being uniformly chosen from a normal distribution

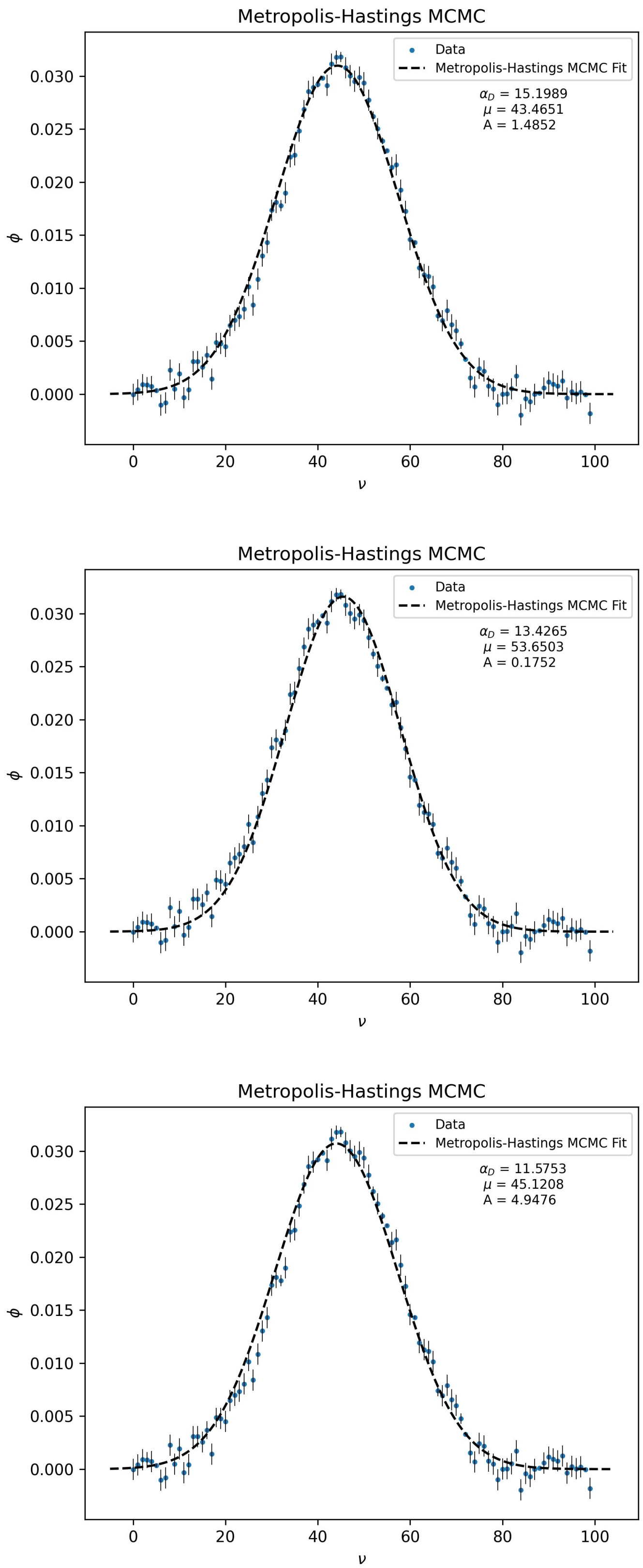

Figure 1: The fit obtained for the provided data using the Metropolis-Hastings-MCMC algorithm

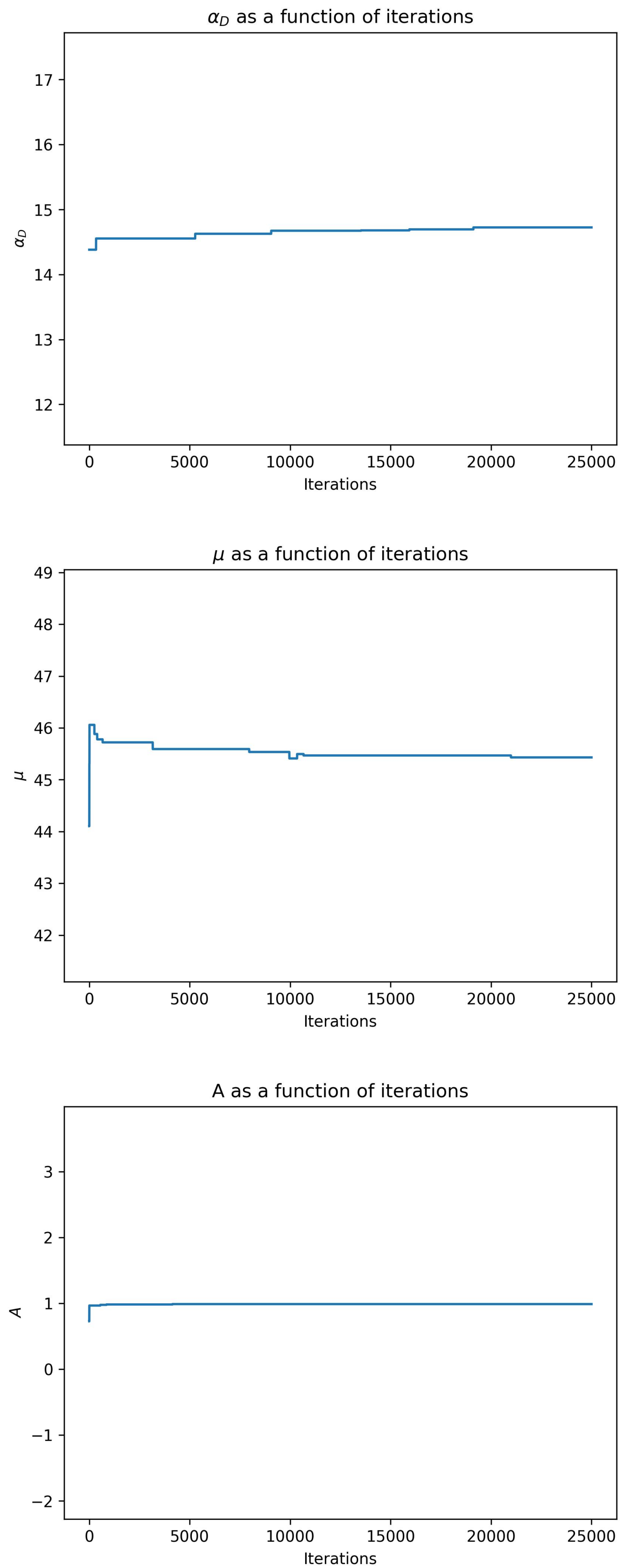

Figure 2: The parameters as a function of iterations

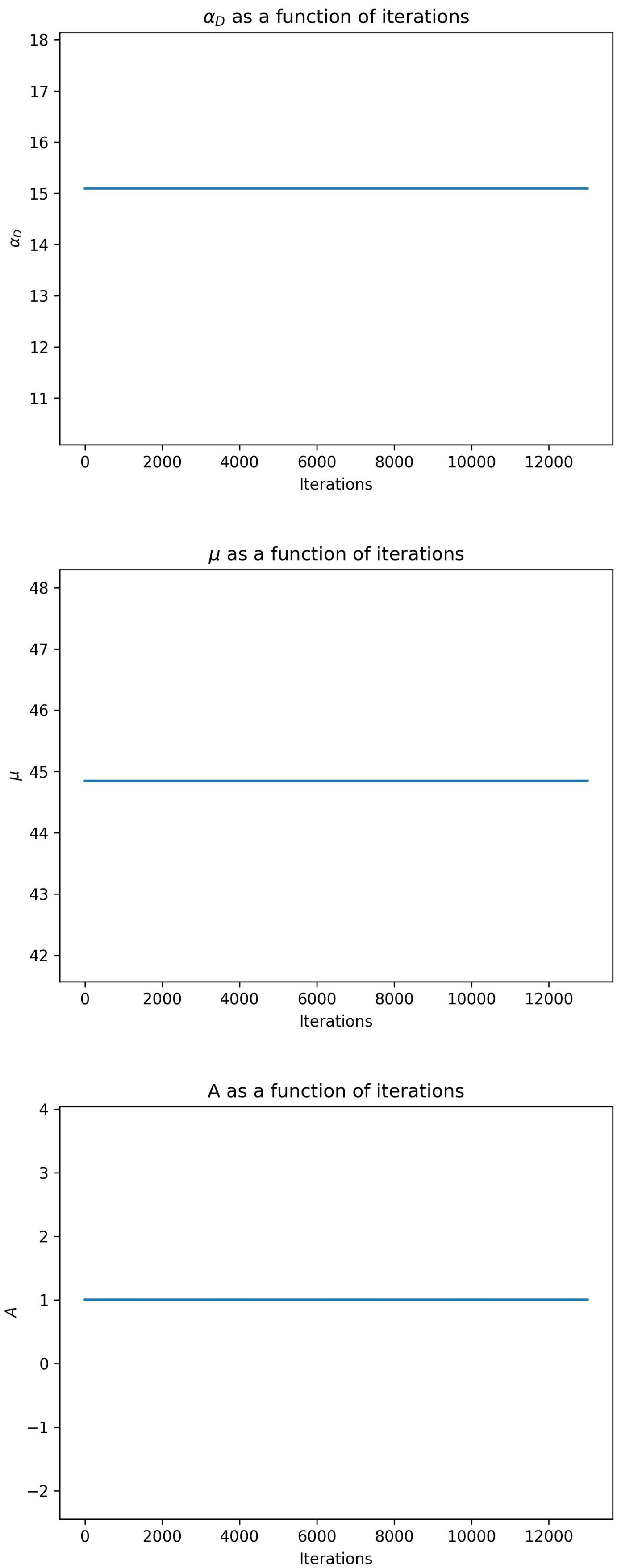

Figure 3: The parameters as a function of iterations after the burn-in period.