Data Science – Final project description.

Nicolas Garavito-Camargo

March 28, 2017

In this proposal I will explained my project in detail and in the final I will answer the homework questions.

1 Introduction to the problem:

Aim: Build accurate models of the gravitational potential of the Milky Way in the presence of a Massive Satellite galaxy i.e the Large Magellanic Cloud. The stars in the Milky Way are distributed in two main components a disk and a spherical bulge. However, the main component of the Milky Way is a non-observable halo of Dark Matter, this component is thought to be the biggest and the main source of gravitational potential of the MW.

The Large Magellanic Cloud has roughly 10% of the mass of the Milky way and it is inside the dark Matter halo of the Milky Way, therefore, it is currently interacting with the Milky Way and changing the shape and the gravitational potential of the dark matter halo of the Milky Way.

Problem: Current analytic models of the potential of the Milky Way doesn't account for the time evolution of the Milky Way and for the presence of the large Magellanic cloud.

2 Methods

2.1 N-body simulations:

In order to construct a more realistic potential of the Milky Way and the Large Magellanic Cloud we set-up a N-body simulation (a very common technique used in astronomy) which consist on 40 million particles initialized with observed positions and velocities of the MW disk and bulge. For the Dark Matter halo we assume an initial spherical distribution. Please see the attached movies to see how this simulations looks like. The Large Magellanic Cloud is represented by a spherical distribution and it is placed initially at a position that reproduce the observed orbit of this satellite. This particles are then evolved in time by solving the equations of motion for each particle (this is, solving Newton's second law for each particle). As a result we can study the gravitational potential of the system in a time evolving fashion at any position.

These simulations allow us to study the details of galaxy interactions, however, they require expensive computations. In order to overcome this problem we need a different approach as explained in the next section.

2.2 Characterizing the gravitational potential of the Milky Way galaxy and it's satellite galaxy the Large Magellanic Cloud using N-body simulations:

Towards an analytic and a more realistic description of the Milky Way potential we use a ***Self consistent method*** in which the potential of a given spatial distribution of particles of mass m can be expressed in terms of spherical harmonics and the gegenbauer polynmials.

$$\Phi(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{n=0}^{\infty} Y_{l,m}(\theta)\Phi_{n,l}(r)(S_{nlm} cos m\phi + T_{nlm} sin m\phi) \tag{1}$$

Where $\Phi_{nl}(r)$ encodes how the dependence of the gravitational potential in the radial direction, this is expressed in terms of the ultraspherical polynomials $C_n^{(2l+3/2)}(\xi)$.

$$\Phi_{nl}(r) = -\frac{r^l}{(1+r)^{2l+1}} C_n^{2l+3/2}(\xi)\sqrt{4\pi} \tag{2}$$

$$S_{nlm} = (2-\delta_{m,0})\tilde{A}_{nl} \sum_{k}^{N_{particles}} m_k \Phi_{nl}(r_k) Y_{lm}(\theta_k) cos(m\phi_k) \tag{3}$$

Similarly,

$$T_{nlm} = (2-\delta_{m,0})\tilde{A}_{nl} \sum_{k}^{N_{particles}} m_k \Phi_{nl}(r_k) Y_{lm}(\theta_k) sin(m\phi_k) \tag{4}$$

Where the sum over k is the sum over all the particles in the system. And therefore r_k, θ_k, ϕ_k are the spherical coordinates of the $k-th$ particle.

The coefficient $\tilde{A}_{nl}$ is:

$$\tilde{A}_{nl} = -\frac{2^{8l+6}}{4\pi K_{nl}} \frac{n!(n+2l+3/2)[\Gamma(2l+3/2)]^2}{\Gamma(n+4l+3)} \tag{5}$$

Method to compute the potential of a given distribution of particles of mass m:

1. Use Eq.4 and Eq.5 to compute the coefficients for a given value of n, l, m.
2. Use Eq.2 to compute the potential, here a value of n_{max} and l_{max} have to be chosen, in practice the larger the values the better the representation of the potential. Typical values used in the literature are $n_{max} = 20$ and $l_{max} = 10$, this means that the coefficients are a matrix of dimension $(20, 10, 10)$ meaning that 2000 coefficients have to be computed.

3 Method implementation:

I have written a code in C and in python in order to compute the coefficients and the potential from the distribution of particle of the N-body simulation. Also a public available code has been developed by a collaborator, the code can be found here The results from both codes are the same.

As a test we start with a spherical distribution of particles given that follows the potential described by:

$$\Phi = -\frac{1}{1+r} \tag{6}$$

Note that this potential correspond to the first term $n = 0, l = 0$ in Eq.3. This is convinient because we expect that when computing the S_{nlm} and T_{nlm} terms the first term is the one contributing more to the potential. Figure 1 shows the particle distribution in which each dot will represent a dark matter particle. Figure 2 shows the values of coefficients of $m = l = 0$ as expected the first coefficient corresponding to the first term to the expansion in Eq. 3 is the larger, however, when looking at the coefficient with $l > 0$ high order terms start to dominate and the trend is to increase for larger l. It is not clear what are these terms describing, this is the main goal of this project: understand what are this terms describing and how to extract relevant information from them.

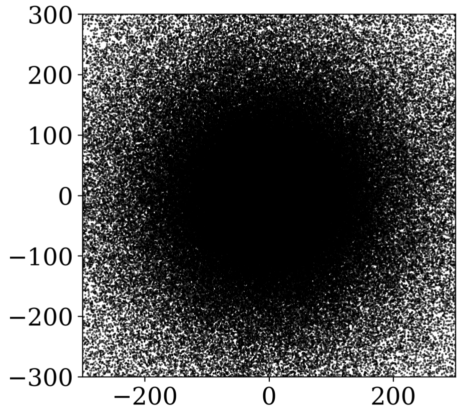

Figure 1: Spherical Dark Matter distribution

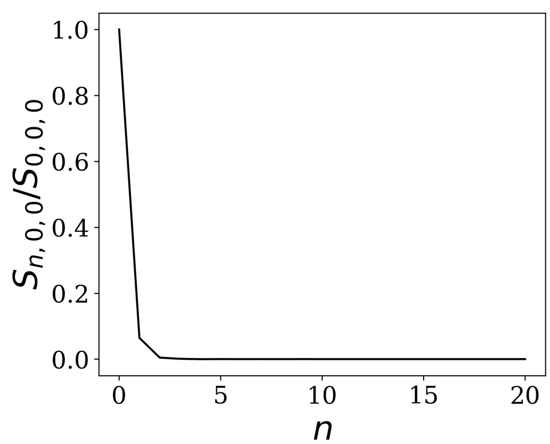

Figure 2: Values of the coefficients for $l = m = 0$

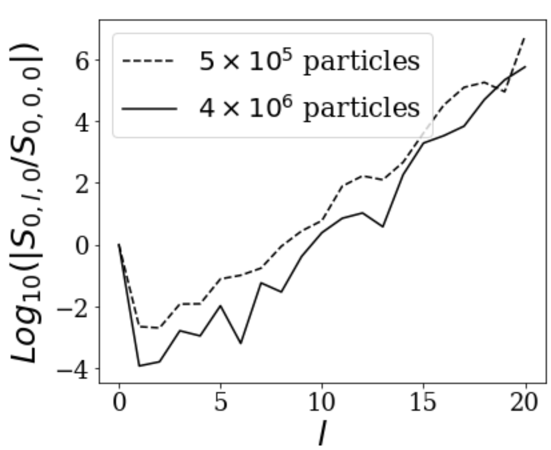

Figure 3: Values of the coefficients for $n = m = 0$

4 Homework questions

What are you going to implement:

At this point it is not entirely clear to me what algorithm I am going to implement to solve this problem because nobody has address this problem before. I guess this might be a classification and optimization problem. This would be the first part of the project to decide and test what algorithm suits better for ths problem.

How it relates to your work:

This is project is part of my current research, as explained in the previous section this is an essential part of the project that I expect to solve using data science.

How will you know that you succeded?:

The final product of this project is to have an algorithm implemented that find the set of coefficients carrying most of the information of the gravitational potential that best reproduce the gravitational potential.

How will you know that the technique is good:

Finding the coefficients that better represent the gravitational potential would be easily tested with the actual gravitational potential computed with the N-body simulation.